Interactions Amongst Polycomb-group Genes Regulating Flowering in *Arabidopsis*

Yindee Chanvivattana

Doctor of Philosophy

The University of Edinburgh

2002

EX BIBL. UNIV. EDINBURGH

Declaration

I declare that this is my own work. Any contributions made by other parties are clearly acknowledged.

Acknowledgements

First and foremost, I would like to thank my PhD supervisor, Justin Goodrich, for his guidance, knowledge, patience and support through the years. Thank you also to past and present members of the Goodrich and Hudson lab especially Christine for her very enjoyable conversations and help, Colin for all his help with the camera and two-hybrid constructs and Carol, Lucia and Tony for their support. I would also like to thank the Thai Government for giving me the opportunity to study in the United Kingdom and for their financial support throughout my stay in the UK and also to the staff at the Thai Government students' office in London for all their help and patience.

Thank you to my current flatmates, Pee Ob and Pee Ekk, for the food, the entertaining conversations and most importantly for putting up with me. Thank you. Also I would like to say thank you to the many Thai students that I have known in Market Harborough, Nottingham and Edinburgh. You know who you are. To Ying, thank you for listening to me and being there when I needed someone to talk to. To Pee Nhoi, thank you for being patient with me on the phone, the office space, the funny conversations and help. Thanks also to my highschool friends; Tan, Jur, Lin, Eung, Milin and Ju for their friendship, help and moral support.

I would also like to take this opportunity to say thank you to my family. My father for believing in me and letting me do what I want to do. My mother for her unconditional love, patience, encouragement, support and for making me wanting to do better in life. I could not have done this without you. This is just the first step I promise you. I love you very much and this is for you. My brother for taking care of my mom through all these years and being such a great chauffeur. We will make it together one day.

Last but not least, I would like to thank Pee Jume, Pee Pom, Lung Boon, Pa Eaeung for their generosity and kindness. You all are so nice to me and I am deeply grateful. Thank you.

Abbreviations

ATP	adenosine 5'-triphosphate
BAC	bacterial artificial chromosome
BASTA	phosphinotricin
BSA	bovine serum albumin
bp	base pair
cDNA	complementary deoxyribonucleic acid
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
DEPC	diethyl pyrocarbonate
dGTP	2'-deoxyguanosine 5'-triphosphate
dH_2O	sterile distilled water
DNA	deoxyribonucleic acid
dNTPs	2'-deoxynucleotide 5'-triphosphates
dTTP	2'-deoxythymidine 5'-triphosphate
dUTP	2'-deoxyuridine 5'-triphosphate
EDTA	ethylene-diaminetetraacetic acid
EMS	ethyl methane sulfonate
EST	expressed sequence tag
g	gravity equivalents
GUS	glucuronidase
IPTG	isopropyl β-D-thiogalactoside
kb	kilobase
M	molar
mM	millimolar
mRNA	messenger RNA
NaCl	sodium chloride
NaOH	sodium hydroxide
ng	nanogram
nm	nanometre
OD	optical density
PCR	polymerase chain reaction

PEG	polyethylene glycol
pmole	picomole
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolution per minute
RT-PCR	reverse transcription polymerase chain reaction
SDS	sodium dodecyl-sulphate
SSC	standard saline citrate
TE	Tris-HCl, ethylenediaminetetraacetic acid
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
Tween	polyethylenesorbitan monolaurate
µg	microgram
µl	microlitre
UTR	untranslated region
UV	ultra violet
X-Gal	4-bromo-4chloro-3-indoyl-β-D-galactopyranoside
X-Gluc	5-bromo-4-chloro-3-indoyl-β-D-glucoronide

Abstract

Cell fate specification is one of the key aspects of development and is relatively well-characterised in plants. However, the mechanisms underlying the process is still rather poorly understood. In animals, Polycomb-group genes (Pc-G) are known to have a prominent role in fate determination by maintaining the silent state of homeotic gene expression through cell division. Recent studies in *Arabidopsis* have indicated that Pc-G functions are also conserved in plants. The first *Arabidopsis* Pc-G member to be isolated, the *CURLY LEAF* (*CLF*) gene, functions to repress transcription of the floral homeotic gene *AGAMOUS* (*AG*) during leaf and flower development. However, unlike the *Drosophila* Pc-G genes which have many homeotic targets, genetic analysis suggested that *CLF* acts largely through *AG*. This and the fact that *clf* mutant exhibits a rather mild phenotype with weak leaf curling and occasional homeotic conversion in flowers suggests some degree of genetic redundancy in this developmental pathway so that other plant Pc-G genes may occur with overlapping function to *CLF*. To address this possibility various genetic and molecular approaches were used to screen for and characterise other *Arabidopsis* Pc-G members.

The recessive *moe leaf* (*moe*) mutation confers a similar phenotype to *clf* mutations. Molecular and genetic studies indicated that like *CLF*, *MOE* activity is required to repress floral homeotic gene expression suggesting that the wild-type *MOE* gene might define an additional Pc-G protein. This was confirmed by the isolation of *MOE*. Thus mapping of *MOE* showed that it is located to the region of a recently isolated *Arabidopsis* Pc-G member, the *EMBRYONIC FLOWER2* (*EMF2*) gene which play an important role in repressing floral genes during vegetative development. Subsequent genetic analysis indicated that *moe* is allelic to *emf2* mutations and cloning of *EMF2* locus from *moe* mutants showed that the *moe* allele harbours a small deletion in the *EMF2* sequence. It appeared that *moe* represents a novel weak allele of *EMF2* and its phenotype was unlike that of other *emf2* alleles described, and informative as to *EMF2* function. In particular, genetic analysis using *moe* and weak *clf* alleles suggested that *EMF2* has a close functional relationship with *CLF*.

Reverse genetic approaches were used to determine the function of an *Arabidopsis CLF* homologue called *CURLY LEAF LIKE1* (*CLK1*). *CLF* and *CLK1* show extensive sequence similarity suggesting that they have a common evolutionary origin. Investigation of mRNA localisation by *in situ* hybridisation indicated that *CLF* and *CLK1* RNA expression patterns are very similar. In addition CLF and CLK1 proteins show common interactions in yeast two-hybrid assay. Together these findings suggested that *CLF* and *CLK1* may show functional redundancy. Characterisation of transgenic lines expressing a dominant negative *CLK1* construct, and of a line carrying a T-DNA insert in the *CLK1* 5'UTR, both provided strong support for this hypothesis. Thus reduced *CLK1* activity had little or no effect in wild-type background, but strongly enhanced the *clf* phenotype. These results implied that null *clf* mutants may not reveal the full extent of *CLF* function due to the redundant activity of *CLK1*. The fact that the *clf clk1* double mutant resembles *emf* mutants suggests that *CLF* and *CLK1* act in the *EMF* regulatory pathway, perhaps as parts of a Pc-G complex.

Contents

Acknowledgements I

Abbreviations II

Abstract IV

Chapter 1. Introduction 1

1.1 General Introduction: Allocation and Control of Cell Fate in Development 2

1.2 Polycomb-group Genes and their Functions 8

1.3 Evolutionary Conservation of Polycomb Function 26

1.4 *Arabidopsis thaliana* as an Experimental System 29

1.5 Plant Polycomb-group Genes and their Roles in Development 31

1.6 Aims of the Project 43

Chapter 2. Materials and Methods 45

2.1 Plant Materials 46

2.2 DNA Analysis 50

2.3 RNA Analysis 61

2.4 *In situ* Hybridisation 65

2.5 Yeast Two-Hybrid Interaction Assay 72

2.6 *In vitro* Mutagenesis 73

2.7 Gene Transfer into *Arabidopsis* 77

2.8 Histochemical Staining of the GUS Reporter Lines 78

Chapter 3. Characterisation of a Novel Leaf Curling Mutant, *moe leaf*, in *Arabidopsis thaliana* 79

3.1 Introduction 80

3.2 *moe leaf* Phenotype 81

3.3 Double and Triple Mutant Analysis 83

3.4 Analysis of Floral Homeotic Gene Expression in *moe* Mutant 90

3.5 Mapping the moe Mutation 95

3.6 *MOE LEAF* Cloning 101

3.7 The Molecular Basis of the *moe* Allele 106

Chapter 4. Characterisation of an *Arabidopsis CURLY LEAF* Homologue: *CURLY LEAF LIKE1* **110**

4.1 Introduction 111

4.2 *CURLY LEAF LIKE1* is an *Arabidopsis* Homologue of *CURLY LEAF* 112

4.3 *CLK1* Expression Pattern 114

4.4 CLK1 Protein Interactions 116

4.5 Isolation of *clk1* Mutant Alleles 119

Chapter 5. Discussion **136**

5.1 Genetic Redundancy between *CLF* and *CLK1* 137

5.2 Interactions between Pc-G Genes, Flowering Time Genes and Floral Homeotic Genes during *Arabidopsis* Development 143

5.3 Role of Plant Pc-G Genes in the Control of Cell Proliferation 149

5.4 Distinct Pc-G Complexes in Plants 154

References **158**

Chapter 1. Introduction

1.1 General Introduction: Allocation and Control of Cell Fate in Development

1.2 Polycomb-group Genes and their Functions

1.3 Evolutionary Conservation of Polycomb function

1.4 *Arabidopsis thaliana* as an Experimental System

1.5 Plant Polycomb-group Genes and their Roles in Development

1.6 Aims of the Project

1. Introduction

1.1 General Introduction: Allocation and control of cell fate in development

The evolution of body forms in multicellular organisms is without doubt one of nature's major triumphs. How it is achieved has provoked a fundamental problem for evolutionary and developmental biologists for many centuries. Through the years, with the help of the increasing wealth of information several aspects of the process have been unravelled. These invaluable little jigsaw pieces have provided scientists with better understandings of the process. Yet there still are many missing links to be found and several questions are left unanswered still. One of the main mysteries that has been of particular concern to developmental biologists is how correct patterns are formed along the body of multicellular organisms. Thus it has been of great interest why organs such as the legs of fruit fly develop in the thorax and not in the head or abdomen, for instance. Molecular genetics has proved to be an extremely useful tool in dissecting the complexities of developmental processes. Genetic data has shown that for a certain cell to acquire its identity and commit itself to undergo a particular developmental pathway, it requires highly co-ordinated regulation of expression patterns of a large number of genes in the cell. It is important that the information stored at the DNA level is correctly interpreted and integrated to ensure that the right set of genes are switched on or off and their expression patterns are precisely regulated both temporary and spatially such that a correct pattern of gene activity is established.

Developmental studies in various model organisms revealed that a number of different strategies are employed in the setting up of the initial pattern of structures along the body axes. Maternal factors deposited in the eggs of *Drosophila* are major determinants of body axes. Differential distributions of maternal mRNAs and proteins provide positional information which set up regional differences along the embryo body. Fine elaboration of the early pattern is achieved through activity of zygotic genes at later stages of development. Similarly the animal-vegetal axis of *Xenopus* is also maternally determined. The gradient of maternal factors provides the egg with a distinct polarity which later becomes a major influence on cleavage pattern. In the nematode, *Caenorhabditis elegans*, asymmetric cell

division at the first cleavage results in the unequal localisation of cytoplasmic factors in the two daughter cells which embark on different developmental pathways giving the organism its initial body plan (reviewed in Wolpert et al., 1998). Similar process also takes place during early development of many plant species including angiosperms to help determining the basic body pattern in embryos. In some cases patterning is not only influenced by intrinsic factors but also by external cues. For examples, polarity of the brown alga, *Fucus*, zygote can be determined by environmental stimuli such as pH gradients and electrical currents in the water (Goodner and Quatrano, 1993).

Once established, pattern is refined using the information from earlier patterning events. These may need to be remembered, for example, if based on short range signaling events which cannot be recreated when the organism is bigger. Thus it is often required that the integrated information is correctly maintained and propagated through cell divisions during later developmental stages. Any slight perturbations or changes could impose detrimental impact on development and in some cases leads to lethality. How this is achieved and regulated has been the subject of intense scrutiny among developmental biologists for many years. Information accumulated from various studies have begun to untangle this mystery. Thus patterning is found to be dependent on cell lineage or cell-cell interactions or both. Cell memory system is widely used in animals to ensure faithful inheritance of the initial states of gene activity through cell divisions. This means that most cell types are irreversibly fated by their history with each newly dividing cell being imprinted with an identity. As a result patterns are maintained cell-autonomously irrespective of the external environments. This determination of cell fate by cell lineage is best studied with regard to the establishment of body segments in the fruit fly, *Drosophila melanogaster*, in which a hierarchical network of genes is a driving force in the pattern-forming processes. Early in embryonic development, the graded concentration of several maternally derived proteins helps establishing the polarity of body axes. Activities of zygotic genes then elaborate on this initial positional information such that each part of the embryo is assigned with an identity. Thus the zygotic gap genes such as *hunchback*, *kruppel*, *giant* and *knirps* are differentially activated by the gradient of the maternally acquired *bicoid* expression. Interactions of these gap gene products result in the sharpening of their expression boundaries. As a consequence the body axis is divided into distinct domains. The combinatorial activities of these proteins in each domain then define the unique expression patterns of the pair-rule genes such as *even-skipped* and *fushi tarazu* which in turns establish the segmentation pattern. Further

elaboration of segment boundaries is achieved through the activation of segment polarity genes. One such gene is *engrailed* which is activated by the expression of the pair-rule genes and plays an essential role in the patterning of body segments. Genetic analysis shows that its activity is required to define the posterior end of each segment. Once the segment pattern is well established, the identity of each segment is conferred by the expression of the so-called homeotic genes lower in the genetic hierarchy. These homeotic genes are found clustered in 2 different complexes; the Antennapedia (ANT-C) and the Bithorax (BX-C) complex. Their initial expression patterns are specified through the interactions of the gap and segmentation genes in a combinatorial manner (Ingham, 1988). The unique combination of homeotic gene activities give each segment its distinct identity and mutations in these genes often result in the inevitable loss of positional identity and subsequent transformations of body segments. For example, the dominant *Antennapedia* allele causes the transformation of the antennae into legs.

Determination of segment identity requires the expression of homeotic genes to persist through later stages of development. The activities of the gap and segmentation genes are only present transiently and disappear only after about 4 hours post-fertilization. This creates a problem of how the correct homeotic gene expression patterns are maintained faithfully during later developmental stages. To overcome this dilemma, two antagonistic mechanisms have evolved. Thus the activity of the Polycomb-group genes (Pc-G) is required to maintain the transcriptional repression of homeotic gene expression while the trithorax-group genes (trx-G) act to maintain the on-state of gene expression (Breene and Harte, 1993; Muller, 1995). The combined activities of these two classes of genes ensure that the determined states of homeotic gene expression is correctly imprinted and maintained in cell lineages, (Paro, 1990).

In addition to cell lineages, cell-cell interaction mechanisms are also employed in animals to specify patterns. For example, positional information and developmental cues have been shown to be essential in cell fate specification in early development of *C. elegans* embryos. Moreover, short range positional signalling between cells and their neighbours are also common in animal development. One such case is the maintenance of the aforementioned segment polarity gene *engrailed* (*en*) by a secreted glycoprotein encoded by the *wingless* (*wg*) gene (Nusse and Varmus, 1992). Genetic mosaic study has shown that correct

maintenance of *en* expression is only achieved when the *wg*-expressing cells lie adjacent to the *en*-expressing cells (Vincent and Lawrence, 1994).

The process of pattern formation provides one fundamental difference between animal and plant development. Thus in contrast to the situation in animals in which patterning takes place almost exclusively during embryogenesis, the establishment of plant body plan happens throughout their life cycles with new organs being formed in a reiterative fashion. Cells that contribute to the body organs such as leaves, stems and roots are derived from proliferating populations of stem cells called meristems. Since plant cells are confined by rigid cell walls and don't migrate, the acquisition of cell fate is, at first glance, unlikely to be dependent on positional information obtained from interactions between neighbouring cells and cell lineage is, therefore, likely to play a prominent role in pattern determination in plant development. Yet there are numerous lines of evidence to suggest that in most cases plant cells do not differentiate according to their lineage but rather rely on their position.

Results from root cell ablation experiment suggest that corpses of the ablated central cells of the meristems in the root cap could be replaced by dividing vascular cells which lose their original fate and express a root cap-specific marker instead suggesting that the fate of these cells can be changed according to their position within the root (van den Berg, 1995). In addition, lineage analysis of the *Arabidopsis* root using GUS-marked sectors showed that cells that contribute to one lineage can also contribute to others (Kidner et al., 2000). Clonal analysis of marked cell lineages during early flower development in the snapdragon, *Antirrhinum majus* also further emphasise the role of developmental plasticity in plant form patterning. Thus by generating marked clones of cells, Vincent et al., found that during early flower development, these cells are not restricted by any boundaries in later development. It is only when individual organ primordia arise that lineage restrictions between adjacent floral whorls start to appear (Vincent et al., 1995). In recent years, a number of plant cell-cell signalling networks have been discovered providing more evidence for interactions between cells of different lineages and hence the acquisition of plant cell identities based on their position and not their ancestry. The most well studied of such networks is the regulation of shoot meristem structure by the *CLAVATA* (*CLV*) genes. The *CLV1* gene encodes a protein with a predicted extracellular leucine rich repeat (LRR) domain and a cytoplasmic serine kinase domain (Clark et al., 1997). The CLV1 protein and the LRR associated CLV2 protein are parts of a complex that initiates a kinase signalling cascade network that monitors the

integrity of the shoot meristem (Clark et al., 1997; Williams et al., 1997; Jeong et al., 1999; Trotochaud et al., 1999). The putative ligand for the CLV1-CLV2 kinase complex is encoded by the *CLV3* gene which is expressed in the two inner layers of the meristem (Fletcher et al., 1999). It is thought that local signalling between the CLV3 ligand in the outermost layer promotes the differentiation of those cells to form organ primordia on the flanks of the meristem. This signalling network is therefore essential for the maintenance of the balance of the meristematic cells in the central zone and the differentiated cells in the peripheral zone of the meristem. In fact the CLV3 protein has recently been shown to bind to the CLV1 complex *in vivo* and its association is vital for the formation of an active CLV complex (Trotochaud et al., 1997; Trotochaud et al., 2000). Mutations in any of the three *CLV* genes result in the overproliferation of the meristem central zone which leads to the disruption in shoot meristem patterning and subsequent abnormalities in organogenesis.

The finding that many plant gene products act non-cell-autonomously and that many developmental patterning processes rely mostly on cell-cell interactions provokes another question of how information is trafficked between cells. One potential mechanism is the use of cytoplasmic continuity created by the plasmodesmata; the plasma membrane-lined channels, across the cell wall that join the cytoplasm of adjoining cells. At least one key transcription factor, the homeo-domain KNOTTED1 protein in maize has been shown to be able to traffick between cells via the plasmodesmata (Lucas et al., 1995).

Most of the evidence gathered so far indicates that complex interactions between cells play a major role in the specification of cell identity during plant development. Determination of cell fate using a memory system is, on the other hand, rarely cited in the context of plant development. The reversible nature of plant cell identity is illustrated by the striking capability of plant cells or tissues to regenerate. Thus plant organs and embryos can be successfully formed from undifferentiated cells in tissue culture. For example it has been demonstrated that single carrot cells can be induced through tissue culture to recapitulate development, so that new carrot plants can be generated from those cells when cultivated in tissue cultures (Stewart, 1958). The remarkable ability of differentiated cells to regenerate can also be illustrated by the fact that many plants can be vegetatively propagated by grafting and bud culture. The totipotency of differentiated plant cells, tissues or organs indicates that in contrast to the strictly irreversible commitment of cells to particular developmental programmes as imposed by cell history, plant development can be regarded as highly plastic.

Maintenance of cell fate determination through cell lineage is thought to be an epigenetic mechanism that involves permanent changes in higher order chromatin structure. The reversible nature of cell identity acquisition in plants and the almost non-reliance on cell lineage in patterning suggest that the chromatin programming mechanism is somewhat less rigid and hence in many cases reprogramming of information is plausible.

Although cell-cell communication has been implicated as a major driving force in the plant body patterning process, there are several emerging lines of evidence to suggest that in some cases cells are irreversibly fated in plant development. The best understood example is the commitment of shoot meristems to form flowers and inflorescences during the vegetative to reproductive phase transition. In fact identity of floral organs is specified by the highly regulated activities of floral homeotic genes reminiscent of the specification of segment identity in *Drosophila*. Further similarity to the *Drosophila* paradigm is extended to the level of gene regulation. Thus in recent years many plant Polycomb-group genes have been identified and characterised. Not only that they share structural similarity with their *Drosophila* or vertebrate counterparts but they also exhibit functional similarity in the maintenance of silent state of target gene expression. This finding indicates that Pc-G mediated gene silencing is widely conserved.

Recently a handful of mutants impaired in epigenetic regulatory processes have been isolated in several plant species. This suggests that epigenetic control of developmental traits in plants might be more common than once thought. To investigate this, the developmental roles of two Pc-G genes in *Arabidopsis thaliana* were characterised in this study. Key features and functions of animal Pc-G genes will be first discussed since much of the current information on Pc-G genes is derived from studies in animals. Later, the evolutionary conservation of Pc-G function, the features of plant Pc-G genes and their role in patterning will be discussed.

1.2 Polycomb-Group genes and their functions

1.2.1 Identification of *Drosophila* Pc-G members and phenotypic characterisation

Our knowledge of Polycomb-mediated gene silencing has increased vastly since the first Pc-G gene, the *Drosophila extra sex combs* (*esc*) was identified more than 60 years ago. In flies, the male has a tuff of bristles on the prothoracic leg, termed the sex combs. In *esc* mutants, however, males unusually show sex combs on the more posterior meso- and metathoracic legs (Figure 1.1). Molecular genetic studies have shown that the ectopic appearance of these sex combs on the second and third legs is a consequence of the misexpression of a homeotic gene *Sex combs reduced* (*Scr*) which is normally expressed only in the first legs, in the affected segments (Kaufman et al., 1990; Patatucci and Kaufman 1991). Over the years many similar mutations that also cause an extra sex combs phenotype have been identified and characterised. The corresponding genes have been collectively classified as Polycomb-Group genes after the second isolated mutation that causes the extra sex combs defect (Jurgen, 1985; Paro, 1990).

At present more than a dozen Pc-G genes have been characterised in *Drosophila* (Simon, 1995; reviewed in Pirrotta, 1997; Lohuizen, 1999). However, the data from genetic studies using screens for enhancers of pre-existing Polycomb genes suggests that there could be as many as 40 Pc-G members in *Drosophila* (Jurgen, 1985; Landecker et al., 1994). Genetic and molecular analysis has shown that not just *Scr* but a spectrum of homeotic genes on the Bithorax (BX-C) and Antennapedia (ANT-C) clusters are misregulated in Pc-G mutants causing the mis-specification of segment identity. For example, homozygous mutant embryos for the gene *Polycomb* (*Pc*) show transformation of all the trunk segments into structures that are characteristic of more posterior segments. This suggests simultaneous misregulation of multiple homeotic genes since the identity of each segment is specified by combinatorial functions of several genes (Lewis, 1978). Detailed analysis of expression patterns suggests that Pc-G functions are not required for the initial set up of homeotic gene expression boundaries but rather are essential for the maintenance of their determined silent state later in embryogenesis (Struhl and Akam, 1985; McKeon and Brock, 1991; Simon et al., 1992). Thus the loss of function of Pc-G members does not affect the initial homeotic gene expression which as previously mentioned is influenced by various gap and pair-rule

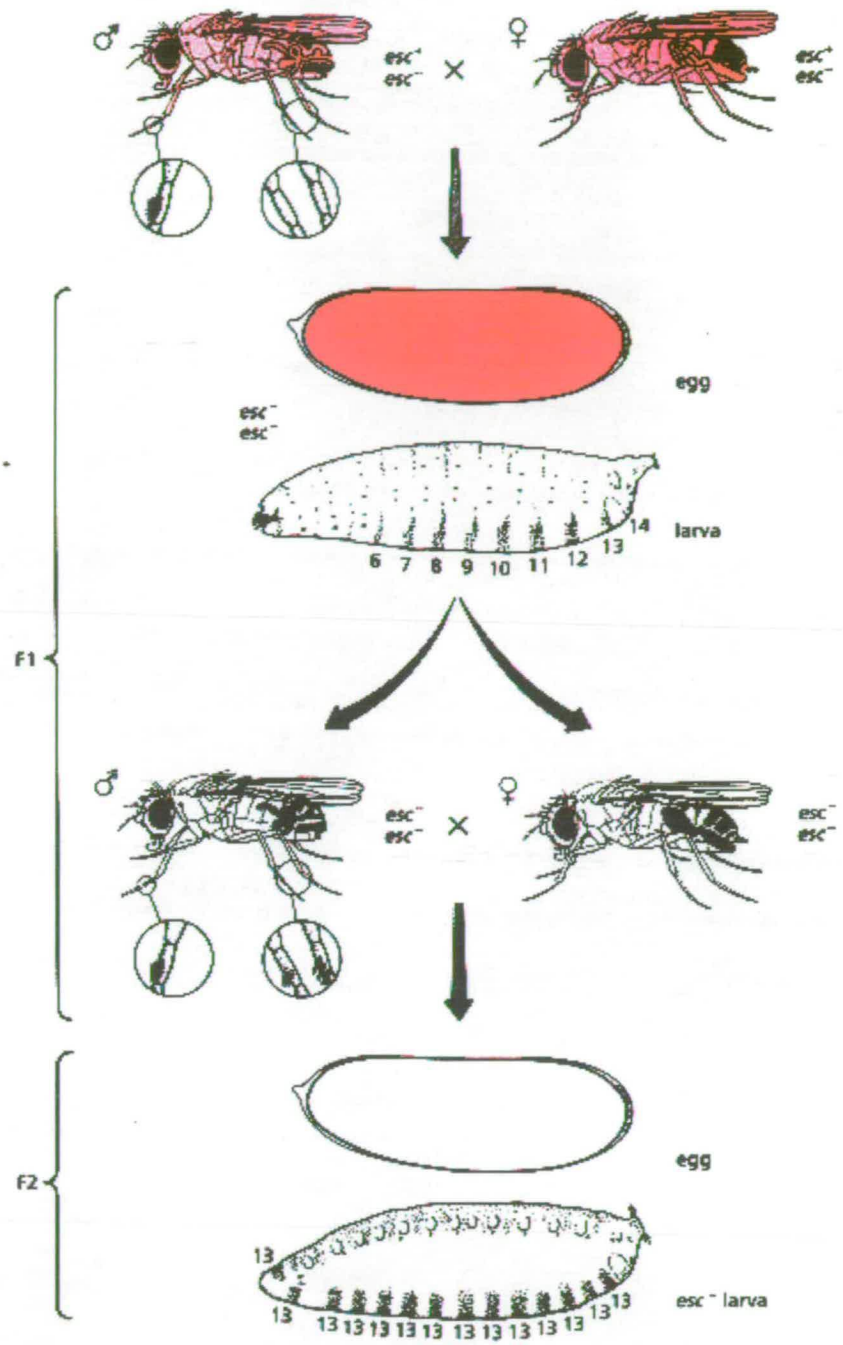


Figure 1.1 The *extra sex combs* (*esc*) phenotype. *esc*⁻ mutant flies from the cross of heterozygous *esc*⁻/*esc*⁺ male and female flies exhibit transformation of the second and third thoracic legs towards the pattern of the first thoracic legs. In males the phenotype is distinct since the sex comb bristles which normally develop on the first legs can be seen on all three pairs of legs. Mating of an *esc*⁻ male to an *esc*- female produce lethal embryos that show complete transformation of all the body segments since there is no maternal *esc*⁺ product deposited in the egg. Taken from Lawrence, 1992.

genes whose activities are present only transiently at the early stages of embryo development and disappear shortly after the blastoderm stage. For example, in embryos that lack product of the Pc-G *Enhancer of zeste* [*E(z)*] gene, the homeotic gene *Ultrabithorax* (*Ubx*) is expressed as in wild-type embryos until about 4 hours into embryogenesis, the time at which gap and pair-rule gene expression start to decay. Subsequently, in mutant embryos regulation breaks down and *Ubx* is expressed ectopically in the anterior body parts (Jones and Gelbart, 1993).

In addition to their function in maintaining homeotic gene expression, many of the *Drosophila* Pc-G genes are involved in other developmental processes as implied by the pleiotropic phenotypes associated with Pc-G mutations. This includes developmental defects in oogenesis, dorsal-ventral patterning and construction of the central nervous system network (Phillips and Shearn, 1990; Adler et al., 1991). For instance, mutations in the *Drosophila Enhancer of Polycomb* [*E(Pc)*] gene not only cause a weak misexpression of the *Abdominal-B* gene in the BX-C complex but also result in strong suppression of position-effect variegation (PEV), a phenomenon which takes place when a euchromatic gene is relocated near or in the transcriptionally inactive heterochromatin by chromosomal rearrangement resulting in its transcription being inactivated. The diverse role of Pc-G genes in *Drosophila* development can be further examplified by the misexpression of several segmentation genes including *engrailed*, *even-skipped* and *knirps* in several Pc-G mutant backgrounds (Moazed and O'Farrell, 1992; McKeon et al., 1994; Pelegri and Lehman, 1994). Given that Pc-G genes have been associated with a number of regulatory events and that they seem to act as general transcriptional regulators with many different classes of target genes, it was of no great surprise that Pc-G proteins were found to be distributed at more than 100 different genomic sites on polytene chromosomes of the larval salivary gland in addition to their homeotic targets (Rastelli et al., 1993; Lonie et al., 1994).

1.2.2 Pc-G complexes and their functions

Cytological analysis has revealed that several of the Pc-G proteins have identical or similar distribution patterns. For example, the protein Polycomb (PC), Polyhomeotic (PH) and Polycomb-like (PCL) are localised at similar positions on the polytene chromosomes. Many of these also coincide with the binding sites of the Posterior sex combs (PSC) protein (Zinc and Paro, 1989; Franke et al., 1992; Rastelli et al., 1993; Lonie et al., 1994). The close

association of these Pc-G proteins invokes the possibility that they can physically interact with each other and form a molecular repression complex. There are several lines of evidence to suggest that this is indeed the case. First, PC and PH together with 10-15 other proteins were found to be co-precipitated in a multimeric protein complex in embryonic nuclear extracts (Franke et al., 1992). Second, PSC and PH proteins disassociated from many genomic sites when the activity of another Pc-G gene *Enhancer of zeste* [*E(z)*] was lost, presumably because this disrupted the protein complex (Rastelli et al., 1993). Third, genetic interaction studies revealed that Pc-G mutations show a synergistic interaction. Thus a combination of two non-allelic Pc-G mutations results in an enhancement in the severity of the phenotype of either one mutant alone. This could be attributed to the progressive impairment of the multimeric protein complex (Jurgen, 1985; Kennison and Tamkun, 1988; Landecker et al., 1994). This result also implicates some degrees of genetic redundancy in homeotic gene repression by Pc-G genes. Furthermore, expression of the fusion construct of the PC protein and the DNA binding domain of the yeast activator GAL4p in a transgenic fly causes transcriptional repression of reporter genes containing the GAL4p binding sites. The silencing effect is dependent on the presence of other Pc-G proteins suggesting that a multimeric protein complex is formed (Muller, 1995). Another compelling piece of evidence for the existence of Pc-G complexes comes from the revelation that Pc-G proteins can physically interact with each other. Thus results from yeast-two-hybrid analysis has indicated direct physical interactions between the protein ESC and E(Z) (Jones et al., 1998). This observation is confirmed by the detection of *in vitro* association of the two proteins using a Glutathione S-transferase (GST)-pulldown assay (Jones et al., 1998). In addition a yeast two-hybrid system and an *in vitro* GST-mediated binding assay also picked up interactions between PSC and PH, PC and PSC, Sex comb on midleg (SCM) and PH and also self-association of PH (Peterson et al., 1997; Kyber and Brock, 1998). Direct association between PC and PH is further suggested by the finding that they co-immunoprecipitate with 10 to 15 other proteins from embryo nuclear extracts (Franke et al., 1992). In addition, PH antibodies also co-immunoprecipitate PSC and vice versa (Strutt and Paro, 1997).

It is, however, unlikely that only a single species of Pc-G complex is all that is required for the wide range of Pc-G-mediated regulatory processes since different Pc-G mutations show subtle differences in phenotypes. The possibility of distinct functional Pc-G complexes gains strong support from several lines of evidence. For example, many of the PSC binding sites lack PC and PH protein and likewise PSC is not found in a number of PC and PH

chromosomal locations (Rastelli et al., 1993). Also distinct mutant phenotypes are observed with the PC and PH proteins although they bind to similar sites on the polytene chromosomes (Franke et al., 1992). Moreover, silencing of the gap genes *knirps* and *giant* expression requires the presence of particular Pc-G proteins such as E(Z) but not the others (Pelegri and Lehmann, 1994). Another line of evidence is provided by the finding that the PC, PH and PSC proteins are found localised at the same sites in the promoter region of the *engrailed* (*en*) gene, PH and PSC presence is not detected at the PC binding sites in the gene *invected* lying just downstream of *en*. On the other hand, PC is missing from the complex bound at the promoter of a gene of unknown function found immediately upstream of *en* (Strutt and Paro, 1997). A likely scenario is that activities of distinct Pc-G complexes, each comprised of a subset of the Pc-G proteins, are required at different target genes. This could account for the variation in the phenotype of different Pc-G mutants and the overlapping but not entirely identical chromosomal binding patterns of different Pc-G members. Distinct multimeric Pc-G complexes with different compositions could enable specificity of silencing to be achieved at various targets in different tissue or organ types. This would generate flexibility and efficiency in repression with only a limited number of Pc-G members being required to regulate a wide range of developmental processes.

The discovery of direct physical interactions between various Pc-G proteins has led to concerted attempts to identify and purify various existing Pc-G partnerships. So far the presence of two distinct classes of Pc-G complexes has been verified. The first complex, namely Polycomb Repressive Complex 1 (PRC1) was isolated from a nuclear extract of transgenic *Drosphila* embryos expressing epitope-tagged alleles of *ph* and *Psc*. Progressive fractionation and purification of the PRC1 complex using antibodies for various Pc-G proteins revealed that this complex contains the proteins PC, PSC, PH, SCM and dRING1 and a number of other proteins (Shao et al., 1999; Saurin et al., 2001). A number of other Pc-G proteins including PCL and E(Z) are, however, not co-purified in the PRC1 complex. The absence of PCL is somewhat interesting since the protein has been shown to completely co-localise with PH and PC on the polytene chromosomes (Lonie et al., 1994). It is possible that association of PCL with the PRC1 complex was lost during the purification process. Alternatively PCL could target the PRC1 complex to its genomic sites.

The second purified Pc-G complexes contains the proteins E(Z) and ESC. This complex is both biochemically distinct and separable from the PRC1 complex (Ng et al., 2000; Poux et

al., 2001; Tie et al., 2001). The close partnership between *esc* and *E(z)* activity was first brought to attention by several genetic observations. Thus mutant *E(z)* and *esc* embryo display near identical patterns of ectopic homeotic gene expression. In addition, the posterior transformation phenotype of the *esc* mutant is very sensitive to the relative level of the E(Z) protein. Both loss and gain of *E(z)* dosage results in a strong exacerbation of the *esc* homeotic defects suggesting that the relative amounts of the ESC and E(Z) proteins are critical for the correct silencing pattern (Campbell et al., 1995). Not surprisingly the posterior transformations of embryos lacking maternal *esc* product produced by *esc* mutant females resemble that seen in embryos produced by females homozygous for temperature sensitive *E(z)* alleles at the restrictive temperature (Struhl and Akam, 1985; Jones and Gelbart, 1990; Phillips and Shearns, 1990; Simon et al., 1992). In addition, it has long been widely recognised that the *esc* gene is somewhat different from the other *Drosophila* Pc-G genes. This is manifested in both its expression pattern and function. Initial expression studies showed that unlike other Pc-G genes which are expressed continuously and whose functions are required throughout development to maintain gene silencing, (Jones and Gelbart, 1993; Martin and Adler, 1993; Paro and Zink, 1993), *esc* expression is largely limited to the first few hours of embryogenesis indicating a primary requirement of *esc* function during early embryogenesis (Frei et al., 1985; Simon et al., 1995). Studies of *esc* temperature sensitive alleles and a heat-inducible transgene revealed that *esc* activity is essential only during a brief period between the first three to six hours of embryogenesis (Struhl and Brower, 1982; Simon et al., 1995). Strikingly, a reduction or loss of *esc* function at earlier or later developmental phases does not result in any phenotypic abnormalities. This finding reinforces the notion that *esc* function is only required transiently early in embryogenesis (Struhl and Brower, 1982; Simon et al., 1995). Another functional disparity between *esc* and other Pc-G members is that *esc* does not seem to play a role in the regulation of other genetic loci such as *en*, *kp* and *giants* but is specific for ANT-C and Bx-C members (Moazed and O'Farrell, 1992; Pelegri and Lehmann, 1994).

Rather interestingly the *esc* expression window coincides with the transition stage after the initial repression has been established by the gap and pair-rule activities and before the take over by the Pc-G proteins. This observation has led to the hypothesis that the early acting ESC protein could recognise the established silent state of homeotic genes and subsequently triggers the repression maintenance process by recruiting its partner E(Z) and other proteins to establish a repression complex. This may in turn initiate the recruitment of constituents of

the PRC1 complex at the target sites. This proposed distinct role of ESC in the establishment of silencing complexes is supported by the finding that ectopic homeotic gene expression is detected at slightly earlier stages in development in the *esc* mutant compared to the other Pc-G mutants (Soto et al., 1995).

As previously mentioned, direct physical interaction between E(Z) and ESC has been confirmed using yeast two-hybrid and *in vitro* binding assays (Jones et al., 1998; Tie et al., 1998). The majority of the ESC protein is composed of seven WD motifs which are thought to fold into a β-propellor structure (Figure 1.2) (Ng et al., 1997). The WD repeats are found in a diverse range of proteins and several lines of evidence have suggested they play an essential role in mediating protein interactions (Koniachi et al., 1994; Neer et al., 1994; Wall et al., 1995; Jimenez et al., 1997). This and the highly conserved nature of the WD motifs makes them prime candidates for E(Z) interacting domains. Indeed yeast two-hybrid studies have implicated them as being indispensable for ESC interaction with E(Z). In accordance with this finding, site-directed mutations that disrupt the WD repeat integrity also perturb its ability to bind E(Z) *in vitro* (Jones et al., 1998; Tie et al., 1998). Results from yeast two-hybrid assays have also mapped the ESC-interacting domain to a 33 amino acid region at the N-terminus of the E(Z) protein (Jones et al., 1998; Tie et al., 1998). This region is predicted to form an extended α-helical structure. Several of the highly conserved residues in this domain have been predicted to provide a surface for the ESC-E(Z) interaction (Tie et al., 1998).

N-terminal to the WD repeat domain of the ESC protein lies a small 70 residue region that is enriched in charged and also serine and threonine residues (Sathe and Harte, 1995). The ESC protein has been found to be phosphorylated *in vivo* and the serine and threonine residues at this N-terminus region are the likely sites for this modification (Ng et al., 2000). Phosphorylation of ESC takes place after the ESC-E(Z) protein complex has been formed. The role of this protein modification is not yet clear but it is thought to contribute to some aspects of *esc* function. Reduced levels of phosphorylated ESC was seen accumulating in the background of site-directed *esc* mutants with impaired *in vivo* function indicating a correlation between ESC phosphorylation and its *in vivo* function (Ng et al., 2000). Although reduction in the phosphorylation levels is observed with mutant esc proteins, those proteins are still entirely capable of forming stable protein complexes with E(Z) suggesting that ESC phosphorylation is not required for complex assembly (Ng et al., 2000).

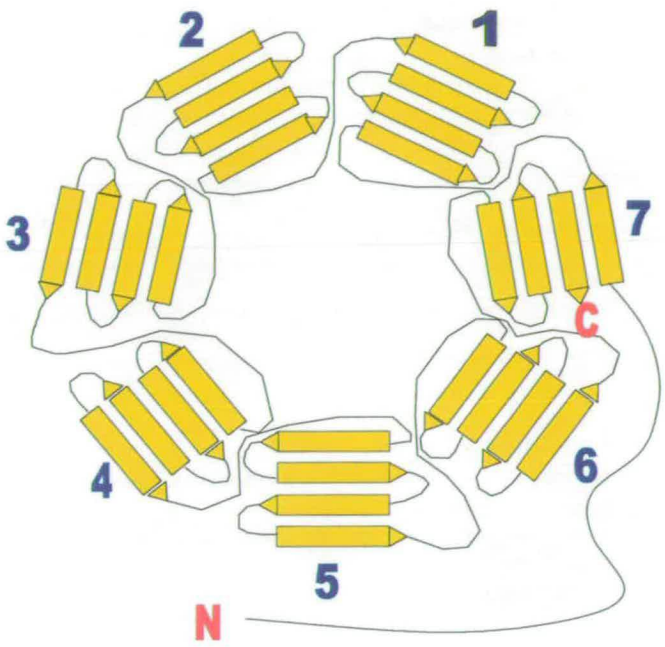


Figure 1.2 The predicted β-propellor structure of the ESC protein. The seven twisted, four stranded b-sheets are numbered 1 through 7. Adapted from Tie et al., 1998.

The E(Z) protein was initially classified as a Pc-G member since a multiple sex combs phenotype was displayed by male flies homozygous for weak *E(z)* alleles (Shearn et al., 1978). Several lines of evidence have since provided support for this notion. Thus it is found that homozygous *E(z)/E(z)* embryos from the cross of heterozygous *E(z)*/+ male and *E(z)*/+ female flies i.e., those that obtain maternal $E(z)^+$ activity but are devoid of any zygotic $E(z)^+$ activity, exhibit transformation of anterior to more posterior segments characteristic of ectopic expression of the genes *abd-A* and *Abd-B* in the ANT-C complex (Jones and Gelbart, 1990; Phillips and Shearns, 1990; Simon et al., 1992). In addition, ectopic expression of the homeotic gene *Ultrabithorax* (*Ubx*) and *Scr* is also detected in the imaginal discs of the embryos hemizygous for a temperature sensitive allele of *E(z)* when reared at non-permissive temperature (Jones and Gelbart, 1990).

At present the exact role of *E(z)* in the establishment of Pc-G silencing complexes is still unclear. However, analysis of *E(z)* temperature sensitive lines shows that binding of several Pc-G proteins such as PSC, SU(Z)2, PH and PC to most loci on the polytene chromosomes is very much reduced when the E(Z) proteins are heat inactivated (Rastelli et al., 1993, Platero et al., 1996). This finding implicates direct involvement of E(Z) in either the recruitment or stabilisation of the repression complexes. As previously stated, it is possible that interaction with ESC recruits E(Z) to its targets and once at the sites E(Z) activity helps assemble other Pc-G members onto the chromosome.

Although *esc* and *E(z)* form an intimate molecular connection as suggested by direct physical interaction in the yeast two-hybrid system and co-immunoprecipitation of the proteins from nuclear extracts, the two genes show some differences in their functions. Thus in contrast to the transient requirement of *esc* during early embryogenesis, *E(z)* activity needs to persist throughout development to maintain repression (Jones and Gelbart, 1990; Phillips and Shearn, 1990). Also unlike *esc* which acts specifically on homeotic loci, *E(z)* is also required in the control of activities of many other target genes suggesting a role of *E(z)* as a general transcriptional regulator (Pelegri and Lehmann, 1993). Presumably E(Z) proteins are recruited to those loci through the interaction with other partners. This non-*esc*-dependent transcriptional silencing function of *E(z)* may explain why *esc* is not found at a number of E(Z) binding sites. The proposed role of *E(z)* as a general regulator is perhaps best illustrated by the fact that it has also been suggested to act as a transcriptional activator as well. Several studies have provided evidence for this putative dual role of *E(z)*. First, mutations in the *E(z)*

gene have been shown to enhance the phenotype of the trx-G genes which as previously mentioned are activators of homeotic gene expression (LaJeunesse and Shearn, 1996). Second, the ability of the Trithorax (TRX) protein to bind the polytene chromosomes is drastically reduced in *E(z)* mutant background and vice versa (Carrington and Jones, 1996). Third, similar enhancement of weak *esc* phenotypes is seen in transgenic flies with increased *E(z)* dosage and those with reduced amount of the E(Z) protein (Campbell et al., 1995). Together these observations suggest that the *E(z)* gene can also act as an activator as well as repressor of gene expression. Interestingly there is some compelling evidence to suggest that the requirement for *E(z)* function as a trx-G or Pc-G gene is tissue-specific and can change during development (LaJeunesse and Shearn, 1996). It is possible that the association of the E(Z) protein with various proteins in different tissues and developmental phases is the main factor that determines the nature of its function.

Given that E(Z) co-localises with many other Pc-G proteins on the polytene chromosome, the strong phenotypic resemblance seen in the mutants of *E(z)*, *esc* and other Pc-G genes, it is likely that even though the ESC-E(Z) and PRC1 complexes are biochemically distinct, their functions at the molecular level are possibly interdependent. Results from a recent study showed that this might indeed be the case. Thus a transient interaction between an ESC-containing complex and the PRC1 complex during the preblastoderm stage early in embryogenesis has recently been demonstrated in co-immunoprecipitation experiments. Moreover, the time window of the interaction appears to coincide with the transition stage in the assembly of Pc-G repression complex (Poux et al., 2001). These data are strongly suggestive of a transient interaction between the two distinct Pc-G complexes which corresponds to the transition stage in the assembly of the Pc-G repression complex and supports the idea that the early acting ESC-E(Z) complex plays a role in recruiting the PRC1 complex to the target sites.

1.2.3 Recruitment of Pc-G proteins to their targets

Many different *cis*-regulatory sequences in the homeotic targets that are able to genetically interact with Pc-G silencing complexes have now been identified within the ANT-C and BX-C complexes. These elements have been named the Polycomb-response elements (PREs) since they are able to confer stable maintenance of homeotic gene expression boundaries. Their activities have been shown to be dependent on the presence of the Pc-G proteins as in Pc-G mutant backgrounds the normally silent segment-specific enhancers are derepressed (Muller and Bienz, 1991; Simon et al., 1993; Chan et al., 1994). The important role of PREs in silencing is further highlighted by the observation that their presence is required throughout to maintain the silent state. Removal of PREs late in development have been shown to cause derepression of the target loci (Busturia et al., 1997).

Genetic analysis has revealed that the PRE elements share a number of common properties. For example, they are able to confer a phenomenon known as 'pairing-sensitive silencing'. For example, when a P element containing a PRE is inserted near the mini-*white* gene, complete repression of the reporter gene was seen in transgenic flies which were homozygous for the transposon. This suggests that the pairing of the PRE-bearing inserts thought to be mediated by PRE interactions somehow facilitates the assembly and the stability of the Pc-G silencing complex. Interestingly, even though PRE elements display the tendency to insert themselves in homologous chromosomal sites or nearby sites containing other PREs, a phenomenon known as 'homing', the pairing effect is also detected with the transposons inserted at different sites or even on different chromosomes (Chan et al., 1994; Pirrotta and Rastelli, 1994; Gindhart and Kaufman, 1995; Hagstrom et al., 1997). The pairing silencing of the mini-*white* reporter gene also indicates that PREs can create novel Pc-G binding sites at sites of transposon insertion which are not regulated by Pc-G complexes under normal circumstances (Chiang et al., 1995; Zink and Paro, 1995).

Of the many PREs that have been identified, the most well characterised ones are those residing in the BX-C gene complex. Genetic analysis reveals that these elements can range from a few hundred base pairs to several kilo base pairs in size and that they can exert the silencing effect on the promoter lying as far apart as 20 kilo base pairs from them (Pirrotta, 1997a). Analysis of the PRE of the *bxd cis*-regulatory domain of the *Ubx* gene shows that each element is composed of several sub-fragments which on their own have weak PRE

affinity (Horard et al., 2000). This finding has led to the speculation that multiple interactions between different recruiter elements, each of which interacts specifically with certain Pc-G proteins could lead to a chain of protein recruitment, and this brings the repressor proteins together to form a functional complex. This model would account for the fact that binding patterns of Pc-G proteins on the polytene chromosomes do not completely overlap and that these binding sites show unequal affinity for different Pc-G proteins. This co-operative assembly model also implies that interactions between weak isolated binding sites, and hence the stabilisation of the Pc-G bindings might be facilitated by the looping of the chromosome (Pirrotta and Rastelli, 1994). This looping model is rather attractive since it would account for the ability of PREs to act over large physical distances and also the pairing sensitive nature of PREs (Figure 1.3B). Alternatively it can be envisaged that formation of Pc-G complexes is a step-wise assembly process. Thus specific DNA binding of certain Pc-G proteins at certain target sites could recruit one or more Pc-G proteins which, in turn, tether a group of other proteins (Figure 1.3A) (Pirrotta, 1997b). In this case the context of DNA target would determine which Pc-G proteins are to bind and the interaction specificities among Pc-G members would then determine the type of complexes that are recruited. This model would also account for the heterogenous nature of Pc-G complexes and the non-overlapping binding patterns exhibited by different Pc-G members. In support of this model, it has been demonstrated that in addition to the GAL4-PC fusion protein repression of a reporter gene containing GAL4 binding sites is dependent on the recruitment of other repressors since it is abolished by Pc-G mutations (Muller, 1995). It is conceivable that several different assembly mechanisms including the step-wise and co-operative interaction mode as used for the formation of Pc-G complexes over the 100 or more predicted target sites and in some cases a combination of different mechanisms might as well be used.

Our knowledge of Pc-G protein recruitment and the establishment of silencing complex has increased substantially over the past few years owing to the detailed characterisation of several new and existing Pc-G proteins and the identification of components of Pc-G complexes. Evidence garnered from genetic studies so far indicates that PREs provide specific binding sites for Pc-G complex binding. Yet most of the cloned Pc-G genes encode proteins that lack DNA binding activity. This raises a question of how Pc-G proteins are brought to their PREs. This enigma is partly solved by the recent cloning of the *Drosophila* Pc-G gene, *pleiohomeotic* (*pho*) that appears to encode a protein containing a known DNA

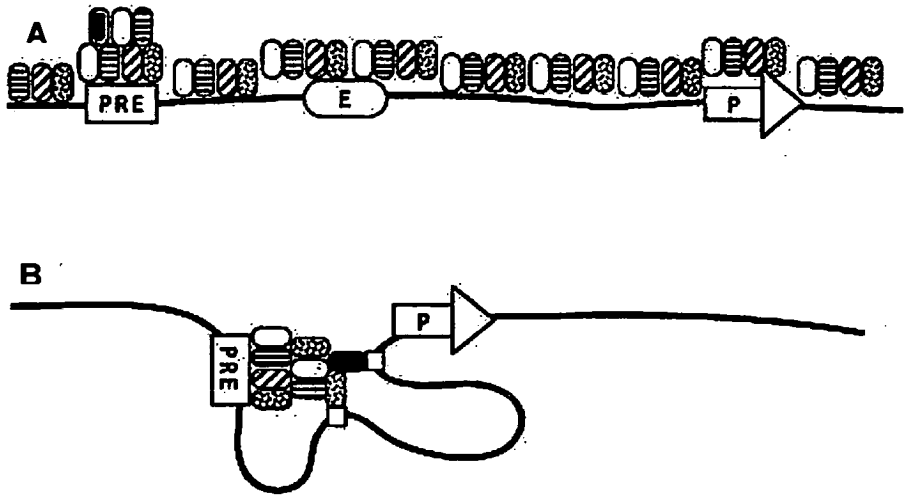


Figure 1.3 Models of the PcG complex. A, In the spreading model, Pc-G complex formation is a step-wise assembly process starting with the binding of sequence-specific Pc-G proteins at the PRE. The bound proteins then recruit more Pc-G proteins to the site and the silencing complex can spread by cooperative interactions with chromatin to involve enhancers (E) and promoter (P). B, In the looping model, the Pc-G complex is stabilised by the cooperative interactions between in the multiple weak Pc-G interaction sites in a PRE. The looping interactions result in the complex being brought to the vicinity of enhancers or promoters. Taken from Pirrotta, 1998.

binding motif. Structural analysis indicates that the PHO protein contains a DNA binding domain consisting of four zinc fingers and is found to be a homologue of the mammalian Yin Yang-1 (YY1) transcription factor (Brown et al., 1998). The ability of PHO to bind DNA was confirmed by the observation that it can interact with the PREs of the gene *engrailed* (*en*). Sequence comparisons of diverse PREs have identified a consensus binding site of YY1. Mutations in this core recognition motif result in the activity of the *en* PRE being abolished (Brown et al., 1998). In addition to that in the *en* PRE, the core element is also found in several other PREs including those in the *Ultrabithorax* and *Antennapedia* gene amongst others (Mihaly et al., 1998). Molecular data from a later study showed that PHO recognises and binds to its putative consensus binding site and site-directed mutagenesis of this core motif totally abolishes PHO binding and hence Pc-G silencing *in vivo* (Fritsch et al., 1999). Although the core consensus is necessary for PHO binding, it is not sufficient for complete PRE activity as proved in the case of the *en* gene (Brown et al., 1998). It is thought that additional sequence motifs are also required for the establishment of a repressing complex. Many other conserved elements are found among various PREs and these could provide recognition sites for other DNA-binding proteins that act in concert with or independently of PHO to recruit Pc-G complexes (Kassis et al., 1989; Karch et al., 1994). This proposal is in agreement with the finding that PHO binding sites are not present in some PREs elements.

In addition to PHO many PREs also contain consensus binding sites for a nucleosomal protein GAGA factor (Tsukiyama et al., 1994). A number of genetic studies have linked the GAGA factor with Pc-G silencing. Thus GAGA has been shown to co-localise with the PC and PSC protein on many targets in the BX-C complex (Strutt et al., 1997; Horard et al., 2000). Moreover GAGA is found to be a component of a number of Pc-G complexes including the aforementioned PRC1 (Shao et al., 1999) and is required for the binding of some Pc-G complexes to their PRE sequences *in vitro* (Horard et al., 2000). Additionally, mutations in the GAGA protein have been shown to affect the pairing-sensitive silencing mediated by the *Fab-7* PRE (Hagstrom et al., 1997). Similar to PHO, it is thought that GAGA is unlikely to be the sole recruiter of Pc-G silencing complexes. This idea is supported by the finding that a number of promoters that appear to contain GAGA binding sites turn out not to be the targets of Pc-G silencing (Poux et al., 2001).

Although a number of genetic data lend support to the Pc-G complex recruiting role of the PHO and GAGA proteins, there are also many lines of evidence that count against it. The fusion construct of the LexA DNA binding domain and either the *GAGA* or *pho* gene does not have the capability to recruit a silencing complex (Poux et al., 2001). This is somewhat not surprising in the case of the GAGA factor since it was originally identified as a transcriptional activator and classified as a member of the trx-G proteins since the loss of its activity results in a concomitant loss of homeotic gene function (Farkas et al., 1994; Biggin and Tjian, 1998). The dual roles of GAGA in the maintenance of both active and repressed states of transcription could reflect a simple role of GAGA in modifying local architecture of the PREs and thereby facilitating subsequent access of Pc-G or trx-G proteins (Tie et al., 1998; Horard et al., 2000). The specific activity of GAGA could be determined by its interaction with different subsets of Pc-G or trx-G proteins in different transcriptional contexts (Tie et al., 1998). GAGA may share this suggested role as a general transcription factor with a number of other proteins including E(Z) whose dual functions as a Pc-G and trx-G protein are well documented (LaJeunesse and Shearn, 1996; Tie et al., 1998). In the case of PHO, the inability of the LexA-PHO protein to recruit Pc-G complexes is rather puzzling given that it is the only known Pc-G member with DNA binding activity. Even more striking and contradictory is the recent finding that PHO is a part of the ESC-E(Z) complex (Poux et al., 2001). A possible explanation for this paradox is that PHO might not directly recruit Pc-G complexes but rather have a primary architectural role and thereby causes conformational changes that facilitates bindings and interactions of other silencing components required to recruit Pc-G complexes (Poux et al., 2001). Interestingly, LexA fusion constructs of any of the four core protein; PC, PSC, PH and SU(Z)2, in the PRC1 complex can silence the reporter gene indicating that each of these Pc-G protein can recruit silencing complexes on its own. The functional significance of this observation is not clear at present since PRC1 is thought to act after the silencing process has been established by the ESC-E(Z) complex. Alternatively the activity of PHO might be required throughout for the maintenance of gene silencing but it is, to an extent, redundant to the function of other, yet unidentified, Pc-G recruiters. Thus loss of *pho* function could be compensated by the activities of those factors. It is also possible that PHO function as a Pc-G recruiter might require activities of additional proteins. This notion would account for the inability of the LexA-PHO protein to establish silencing on its own. Another possible scenario is that PHO itself is initially recruited by another protein. Candidates for this putative recruiter are the gap and pair-rule proteins such as the previously mentioned HB which sets up the initial

repression. There is some evidence that the silencing function of many PREs requires the presence of HB protein and in some cases the addition of HB binding sites can mediate Pc-G silencing (Busturia and Bienz, 1993). However many known PREs which are able to establish silencing appear not to contain HB binding sites and that their activities are not influenced by changes in HB levels (Zhang and Bienz, 1992). This suggests that *hb* function is not directly required for the establishment of Pc-G mediated repression but it does not rule out the possibility of an indirect role of *hb* in the recruitment of Pc-G complex which may be mediated by other proteins' activity. Those proteins might act as a bridge that connects the apparently distinct early and late repression machinery. One such protein which could provide a molecular link between the initial repression by the gap genes and the later maintenance of repression by the Pc-G genes is the *Drosophila* dMi-2 protein. Yeast two-hybrid studies have identified dMi-2 as a HB-interacting protein (Kehle et al., 1998). The interaction domain has been mapped to the D domain of HB which is also required for the repression of the homeotic genes in the BX-C complex. Furthermore, it has been shown that *dMi-2* genetically interacts with *hb*. Strikingly, there is also synergy between *dMi-2* and several Pc-G genes suggesting that *dMi-2* may be involved in Pc-G silencing. However, dMi-2 does not have the ability to bind DNA by itself. This again raises the question of the recruitment of Pc-G complexes at the PREs that lack HB binding sites. Recently a protein ZESTE has been found to co-purify with the PRC1 complex (Saurin et al., 2001). The fact that it co-localises with Pc-G proteins at several silenced loci and that it can bind to both PRE and promoter elements suggest that it might act as yet another Pc-G recruiter by tethering Pc-G proteins to their targets. At present this possibility cannot be discarded and still awaits more investigation.

From the data accumulated so far on Pc-G complex recruitment, it can be envisaged that there are likely to be several distinct Pc-G complex recruiting mechanisms that are employed to cope with the diverse range of Pc-G targets. Recognition of the initial silencing proteins such as HB bound to the target sites by Pc-G proteins could provide one mechanism that triggers a cascade of gene action that leads to the recruitment of Pc-G complexes and subsequently the transition from the initiation to the maintenance of repression (Figure 1.4). Other mechanisms involving different proteins such as ZESTE are likely also used since many PREs function is generally not disrupted by HB depletion or overexpression (Pirrotta, 1997a).

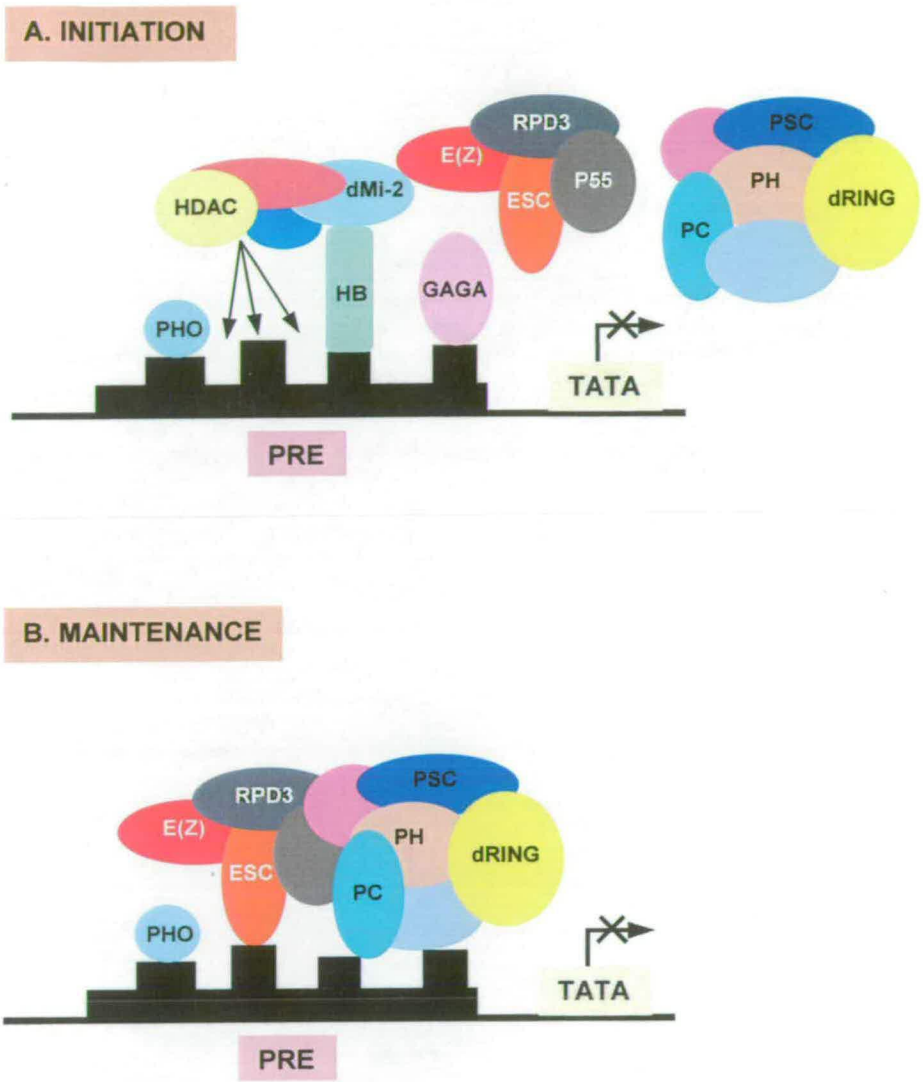


Figure 1.4 Model of Pc-G repression. A, In the initiation stage, the target gene is repressed by the early repressors such as HB. Binding of HB recruits either the histone deacetylase (HDAC)/dMi-2 and the ESC protein complex containing ESC and E(Z) amongst others to the PRE. B, During the maintenance stage, local histone deacetylation and the binding of the ESC complex in turn allow the binding of complexes such as the PRC1 complex containing Pc-G proteins such as PSC, PC and PH which are required for the maintenance of silencing throughout development. Adapted from van Lohuizen, 1999.

1.2.4 Mechanisms of Pc-G action

Although Pc-G members are thought to act on chromatin structures, so far there is still little known as to how this is achieved. Early studies on Pc-G proteins suggested that their activity is related to alteration in higher order chromatin state. In general the chromosome can be subdivided into euchromatic regions which contain coding sequences and are generally transcriptionally active, and heterochromatin which is highly condensed, late replicating and predominantly associated with inactive state of transcription. Heterochromatinization and various forms of chromatin modifications have been associated with a number of gene silencing phenomenon including the previously mentioned PEV. One of the first few clues that suggested the association of Pc-G silencing with structural changes of the chromosome is the highly similar natures of Pc-G mediated gene silencing and PEV. At least two Pc-G proteins have homologies to known modifiers of PEV. Also the PC and the heterochromatin-associated HP1 protein, both of which contain a highly conserved chromodomain, have been found to be a dominant suppressor of PEV (Paro and Hogness, 1991). The precise role of HP1 in heterochromatinization is not clear but there is some speculation that it might act as a general organiser of higher order chromatin structure in the nucleus. This view is supported by evidence from a number of recent studies. Thus direct interactions between several chromatin factors such as the chromatin assembly factor 1 (CAF-1) and the HP1 protein have been reported (Murzina et al., 1999). Furthermore, it has been shown that HP1 protein interacts via the conserved chromodomain with the histone protein H3 *in vitro* (Lade-Nielsen et al., 2001). Histone proteins are components of nucleosomes; the basic repeating unit of chromatin. A nucleosome contains an octamer of core histone proteins consisting of two histone H2A-H2B dimers and a H3-H4 tetramer (Luger et al., 1997). Structural organisation of higher order chromatin has been linked to post-translational covalent modifications such as acetylation, phosphorylation and methylation of the hydrophilic N-terminal of histone tails. Distinct combinations of histone modifications is thought to provide a 'histone code' that confers information to recruit chromatin-associated proteins which modify higher order chromatin structure and hence the transcription state of gene expression (Strahl and Allis, 2000). Recent studies show that HP1 has high affinity for histone H3 protein methylated at lysine 9 residue (Bannister et al., 2001; Lachner et al., 2001). Comparative studies in yeast and mammalian systems have revealed that specific methylation of lysine 9 of histone H3 is carried out by protein members of the suppressor of variegation 3-9 [Su(var)3-9] histone methyltransferase family (Rea et al., 2000). These results suggest that the establishment of

'histone mark' is achieved by stepwise modifications of histone tail that is recognized by chromatin associated proteins including HP1 whose activity somehow mediates the assembly of heterochromatin. The chromodomain homology between the PC and HP1 proteins suggests that PC might also exert its repression effect on transcription of target genes by packaging the DNA into a transcriptionally inactive heterochromatin-like structure.

Interestingly the methylase activity of the histone methyltransferase protein Su(Var)3-9 is catalysed by the evolutionary conserved SET domain initially identified in the *Drosophila* protein SU(VAR)3-9, E(Z) and TRITHORAX (Tschiersch et al., 1994). This methylation activity of the SU(VAR)3-9 protein also requires the presence of two cysteine-rich stretches of amino acids lying on either side the SET domain (Rea et al., 2000). Deletion of either of these cysteine-rich region results in the inability of the protein to methylate *in vitro*. There is a growing amount of evidence that indicates a role of the SET domain in transcriptional silencing mechanisms in a diverse range of organisms including the fission yeast and mammals (Nislow et al., 1997; Ivanova et al., 1998; Firestein et al., 2000; Peters et al., 2001). The presence of the SET domain at the C-terminus of the E(Z) protein suggests that there might be a molecular link between Pc-G silencing and histone methylation. One problem, however, exists. Thus E(Z) is devoid of histone methyltransferase activity *in vitro*, since it lacks one of the cysteine-rich regions required for the catalytic function of the SET domain, suggesting that it might not be able to methylate histone H3. This was shown to be the case in *in vitro* assays using E(Z) protein (Rea et al., 2000). Perhaps E(Z) SET domain has another yet unidentified histone modification activity which is also recognised by heterochromatin factors. In addition to E(Z) functional links with histone modification and more generally, alteration of chromatin structures have also been suggested with a number of Pc-G silencing components. For example the YY1 protein, the mammalian PHO homologue, can interact directly with histone deacetylases and their interaction is thought to contribute directly to the gene silencing activity of the protein (Yang et al., 1996). By analogy PHO binding could result in the conformational change of chromatin structure at the PRE and thereby facilitates the co-operative binding of other Pc-G silencing components. However, the YY1 histone deacetylase interacting domain is not conserved in PHO and therefore another motif is likely used by PHO for such interactions (Poux et al., 2001). Further Pc-G connection with higher order chromatin is suggested by the recent purification of the ESC-E(Z) complex in *Drosophila*. The purification process has uncovered the histone deacetylase RPD3 and the histone binding factor p55 which is also part of the CAF1 protein complex as

its constituents (Tie et al., 2001). It has been postulated that binding of the p55 protein helps targeting the RPD3 protein to its histone substrate for deacetylation. Genetic analysis revealed that RPD3 is also required for PRE-mediated silencing *in vivo* and that E(Z) and RPD3 co-localise at the *Ubx* PRE *in vivo* (Tie et al., 2001). A direct molecular link between Pc-G silencing and chromatin structure is further highlighted by the revelation that both RPD3 and p55 proteins are also co-purified with the PRC1 complex (Saurin et al., 2001). Strikingly the previously mentioned HB-interacting protein dMi-2 is also a component of the PRC1 complex (Saurin et al., 2001). The mammalian homologue of dMi-2 has recently been identified as a subunit of the repressive nucleosome remodelling and deacetylation (NURD) complex (Wade et al., 1998; Zhang et al., 1998). This complex can deacetylate histone H3 and H4 amino-terminal tails when stimulated by ATP hydrolysis. This chromatin remodelling activity is thought to result in a repressive chromatin configuration which is recognised by Pc-G complexes (Struhl, 1998).

Modifications of higher order chromatin structures by the binding of Pc-G complexes are thought to result in the packaging of chromatin in a more compact state by increasing the order of nucleosome array (Pirrotta and Rastelli, 1994). The intimate relationship between structural changes of nucleosomes and Pc-G function has gained support from a number of recent genetic and molecular studies. At least one Pc-G member, PC is able to interact with the histone H3 protein in the nucleosome core via its conserved C-terminus region. Interaction of PC and nucleosome is seen accompanied by an associated binding of PSC suggesting that the whole Pc-G complex might also bind directly to nucleosomes (Brieling et al., 1999). It is striking that PC interaction with nucleosomes does not require the N-terminal histone tail which has been implicated as targets of various forms of histone modifications required for the formation of many repressive complexes. In accordance with this finding, the repressive activity of the PRC1 complex, of which PC is a component, is not affected when only tailless histones are present (Shao et al., 1999). Interactions between Pc-G complexes and higher order chromatin could hinder the access of activators or the basal transcription machinery to the DNA. Alternatively, nucleosomes might be modified such that they are locked in a certain configuration which makes them difficult to be displaced and hence subsequent remodelling of chromatin is prevented. This anti-remodelling mechanism is substantiated by the finding that pre-incubation of nucleosome array with the PRC1 complex blocked subsequent remodelling activity of the antagonistic SWI-SNF complex (Shao et al., 1999). Another possible scenario is that Pc-G induced change in chromatin

structure causes sequestration of the target genes into transcriptionally inactive nuclear compartments.

It is important to note that although there are many intriguing lines of evidence to support the heterochromatin model of Pc-G action, there is also an overwhelming piece of evidence to suggest that Pc-G proteins do not act by heterochromatin formation. Thus unlike HP1, so far none of the Pc-G proteins has been shown to localise to heterochromatin suggesting that they do not play a structural role in heterochromatin.

1.2.5 Propagation of Pc-G repression

In order to ensure that correct spatial homeotic gene expression patterns are retained throughout development, the silent state of gene expression established by Pc-G activity must be faithfully maintained through multiple rounds of cell division. This means that it must survive chromatin alterations involved in DNA replication and the chromosome condensation events that occur during mitosis. Thus the altered chromatin structure induced by Pc-G activity needs to be either stable through cell cycle or reassembled at each cell cycle. It has been demonstrated that *in vivo* a number of Pc-G proteins such as PC, PSC and PH dissociate themselves from the chromosome and remain in the cytoplasm during mitosis (Buchenau et al., 1998). However recent studies have shown that some residual amount of Pc-G proteins remain associated with mitotic chromosomes (Dietzel at al., 1999; Voncken et al., 1999). It is possible that these remaining proteins somehow mark the target loci and directly recruit other Pc-G proteins to reassemble the silencing complexes after the decondensation of daughter cell chromosomes. Alternatively, other marking systems such as the state of acetylation of nucleosomes might be required. It is thought that during DNA replication structurally altered nucleosomes are equally distributed to both daughter strands. If some proteins such as PC remain associated to the nucleosome while replication takes place they would be co-segregated with the nucleosome cores. Both daughter strands would then have the capability to initiate reassembly of Pc-G complexes and therefore reform the repressive chromatin structures (Breiling et al., 1999). Maintenance of the integrity of the silenced state throughout rounds of DNA replication can also be accounted for by the looping model of Pc-G function. Local Pc-G complexes would be dissociated from their targets as the replication fork arrives. Consequently, certain loops are removed from the main complex without its integrity disturbed. The displaced complex could reassemble after

the replication wave has passed by interacting with the other repressed areas which have not yet been replicated (Pirrotta, 1997b).

Although Pc-G silencing is generally stably propagated through many cell generations, reversions to active state are sometimes observed. This loss of cellular memory has been demonstrated by many genetic studies. Initial study showed that overexpression of an activator whose binding sites were inserted next to a PRE can reverse silencing mediated by the PRE (Zink and Paro, 1995). In a subsequent, more elaborate experiment, strong pulses of activator during early embryogenesis can impose a stable chromosome reprogramming from a silent to active state. However, if an activator is supplied later during larval development the derepression is only transient and silencing can be re-established (Cavalli and Paro, 1998). This suggests that in larval development Pc-G complexes might be stable and the silencing mark is well established such that Pc-G reassembly can take place more readily compared to the embryo state. This in turn implies that the silencing mark can be retained for several cell divisions even after derepression has occurred. An emerging new line of evidence suggests that a number of Pc-G proteins are involved in the establishment and propagation of those silencing marks and that some of them might act as the marks themselves. Thus depletion of some of the PC-G protein such as PSC, PH and SU(Z)2 causes a rapid and complete reversal of silencing of the homeotic targets. Repression, however, can be re-established provided that wild-type gene function is re-supplied within 48-72 hours after the initial removal of the activity. This implies that the mark which could be residual PSC, PH or SU(Z)2 or other proteins associated with the targets could be propagated for at least a few cell divisions. It also indicates that a prolonged absence of PSC-PH-SU(Z)2 activity and hence an increasing amount transcriptional activity causes a gradual decrease in the efficiency of the propagation of the mark and eventually the irreversibility of the active state (Beuchle et al., 2001). On the other hand, removal of another set of Pc-G function including PC, SCM, and PCL only results in ectopic expression of the targets after a long period of time. Presumably this group of proteins are not required for the maintenance and propagation of the silencing marks but rather are required for the repression of the target loci that bear these marks (Beuchle et al., 2001).

experiments on *C. elegans* embryo extract revealed that MES-2 and MES-6 and also another MES protein, MES-3 are associated in a complex (Xu et al., 2001). Estimation of the size of the complex indicates that other factors apart from the MES proteins are also parts of this complex. The association of these yet unidentified co-factors and the MES-3 protein might contribute to the germline specific role of the MES proteins. This finding clearly demonstrates that the ESC-E(Z) partnership is ancient and that although the two proteins have evolved and adapted a range of functions in evolutionary diverged organisms, their biochemical role at the chromatin level seems to have been retained. Astonishingly, the unique combination of the ESC-E(Z) protein pair that sets them apart from the other Pc-G genes appears to be also conserved across kingdoms. Thus while there is no recognisable homologues of Pc-G apart from *E(Pc)* has been found outside the animal kingdom, both *esc* and *E(z)* homologues have been identified in plants.

1.3 Evolutionary conservation of Polycomb function

Given that the Pc-G genes play a prominent role in the process of pattern formation in *Drosophila* development, it is unsurprising that they are highly conserved in evolution and that their homologues have been identified in a diverse range of organisms. In fact, nearly all the *Drosophila* Pc-G proteins characterised have their representatives in vertebrates. Characterisation of the vertebrate Pc-G homologues revealed that they also function to repress homeotic gene expression. This finding is not surprising given that the vertebrate homeotic targets, the hox genes, are homologues of the *Drosophila* homoetic genes (Reviewed in Simon, 1995; Brock and van Lohuizen, 2001). In addition to the homeotic targets, vertebrate Pc-G genes also have targets that are involved in cell proliferation. For example, in mammals the role of Pc-G genes has extended to the regulation of hematopoietic cell proliferation. Targeted replacement of the mammalian *Psc* homologue *Bmi-1*, *Mel-18* and also the *Pc* homologue *M33* all result in the defects in the proliferation of hematopoietic cells (van der Lugt et al, 1994; Akasaka et al., 1996; Core et al., 1997). The observed cellular proliferation phenotypes suggest that Pc-G proteins are somehow involved in the control of cell-cycle progression. Further evidence to support this view comes from the finding that mouse Pc-G genes also act upon tumorigenic pathways. Thus an additional target of *Bmi-1*, the *Ink4* gene which encodes the tumour-suppressors and cell-cycle inhibitors $p16^{ink4a}$ and $p19^{ARF}$ has recently been identified. *Ink4* mutations partially suppress the proliferative defects in *bmi-1* mutant mice. Conversely, both $p16^{ink4a}$ and $p19^{ARF}$ are downregulated in *Bmi-1* overexpressing transgenic lines (Jacobs et al., 1999).

High level of conservation of Pc-G genes at both structural and functional level and similarities in their mutant phenotypes suggest that, similar to their *Drosophila* counterparts, vertebrate Pc-G members may function together in big protein complexes. This notion is supported by the initial observation that mutations in mammalian Pc-G genes show synergy and thus enhance each other phenotypes indicating a progressive impairment of Pc-G complex stability (Bell et al., 1998). Indeed results from yeast two-hybrid and *in vitro* immunoprecipitation analyses have confirmed that mammalian Pc-G proteins are associated in big complexes. Not surprisingly, the interaction patterns between those Pc-G proteins seem to have been conserved in evolution as well. Thus mammalian Pc-G proteins are found engaged in two distinct multimeric complexes. One of which contains homologues of PC,PSC, dRING and PH whereas the other complex includes the ESC homologue EED/eed

(embryonic ectoderm development) and the E(Z) counterparts EZH2/Enx1 and EZH1/Enx2 (Alkema et al., 1997; Gunster et al., 1997; Satijn et al., 1997; Denisenko et al., 1998; Sewalt et al., 1998; van Lohuizen et al., 1998; Satijn and Otte, 1999).

Strikingly, it appears that in addition to the existence of their mammalian counterparts both E(Z) and ESC sequences are extensively conserved in evolution with their homologues being found in the relatively primitive genome of the nematode *C.elegans* (Holderman et al., 1998; Korf et al., 1998). So far no homologues of other Pc-G members, apart from E(PC), have not been identified from database searches.

The *C. elegans* ESC and E(Z) homologues are encoded by the *maternal effect sterile (MES) mes-6 and mes-2* gene respectively. Mutations in the *mes* loci result in identical phenotypic defects. Thus germ line proliferation is limited and germ cells often degenerate during hermaphrodite larval development. Consequently, no mature gametes are formed (Garvin et al., 1998; Lapowski et al., 1999). However, in contrast to the *Drosophila* and mammalian paradigms in which the ESC and E(Z) protein function in the regulation of homeotic gene expression in the soma, *mes-2* and *mes-6* functions in *C. elegans* are restricted to the germline with their mutation having very little effects on soma development.

Although the developmental roles of *mes-2* and *mes-6* seem to be distinct from their *Drosophila* counterparts, there is an increasing amount of evidence to suggest that at least the basic biochemical relationship between them has been conserved. For example, expression of transgenes present in extrachromosomal arrays is repressed in the germline of wild-type worm while in the germline of the either *mes* mutants the repetitive arrays are de-silenced (Kelly and Fire, 1998). In parallel to the *Drosophila* paradigm, silencing of the transgenes is thought to be achieved by the *mes* gene-mediated packaging of the repetitive chromosomal arrays into the repressed chromatin state in the germline. These observations indicate that the MES proteins function as transcriptional repressors and that they impose a repressive effect on chromatin structure. The intimate relationship between the MES-2 and MES-6 proteins is also implied from the observation that their spatial and temporal distribution are identical and that the nuclear localisation of one protein requires the wild-type activity of the other protein (Holdermann et al., 1998; Korf et al., 1998). These results are consistent with the notion that MES-2 and MES-6 protein physically interact with each other and likely are parts of a functional complex. In fact, recent immunoprecipitation

1.4 *Arabidopsis thaliana* as an experimental system

Within the past 50 years or so the inconspicuous weed *Arabidopsis thaliana* (Figure 1.5) has been adopted and used as a model system in plant genetic analysis. There are several good reasons to explain why *Arabidopsis* has been an object of intense investigation by plant geneticists in recent time. Its life cycle and various aspects of genomic organisation offer a number of advantages that make it suitable for most genetic studies. Its short generation time of 5 to 6 weeks allows rapid construction and analysis of strains of interest. In addition its small size and the ability to grow on soil or defined media in a greenhouse or under laboratory environments allow easy and inexpensive maintenance of large populations of plants. Furthermore large scale mutagenesis can be carried out with relative ease in *Arabidopsis*. Recent concerted effort to saturate the *Arabidopsis* genome using large mutagenesis result in a vast improvement of functional genomics, so that many knockouts are available for many genes of previously unknown function. Indeed a number of these insertional collections are now freely available to the research community making it very convenient for plant biologists to analyse gene families, for instance. Also the ease of *Arabidopsis* transformation using *Agrobacterium* allows systemic and highly efficient T-DNA tagging of the genome which in turn provide geneticists with efficient genetic cloning tools.

Since *Arabidopsis* is a flowering plant and typically propagates itself by self-fertilisation with large number of seeds produced from a single plant, there is virtually no limit in the maintenance of stocks of the strain carrying the phenotypic trait of interest. The diploid nature of the *Arabidopsis* genome also allows many lethal mutations to be maintained in heterozygous state.

One of the greatest assets of *Arabidopsis* is that genetic manipulations can be readily carried out since its genome size is small consisting of just 5 chromosomes and 125 Mb of DNA sequence. The recent completion of sequencing of its entire genome has provided plant geneticists with golden opportunities to apply reverse genetic approaches to determine the functions of every gene in the genome. Dense genetic and molecular maps have been constructed and this greatly facilitates the mapping, positional cloning and comparative analysis of phenotypic traits of interest. A more comprehensive picture of plant structure and function can now be established. This will provide biologists with tools to address several



Figure 1.5 An *Arabidopsis thaliana* plant at the flowering stage. Taken from Walbot, 2000.

fundamental problems such as the evolution of pattern formation and common developmental networks among multicellular organisms.

1.5 Plant Polycomb-group genes and their roles in development

1.5.1 *Arabidopsis* Pc-G gene *CURLY LEAF* regulates floral homeotic gene activity

The development of plant flowers can be envisaged as being analogous to the anteroposterior patterning of animal body axis. Floral organs in plants consist of serially repeated structural units which are comparable to the segments of insects. In the model plant *Arabidopsis thaliana*, a wild-type flower is composed of four concentric whorls of organs; four sepals, four petals, six stamens, and two fused carpels, arranged from the outermost to the innermost ring of meristem tissue (Figure 1.6).

The formation of such a complex structure that is a flower from a mass of undifferentiated cells raises the question of how those cells acquire their identity and maintain it through many rounds of cell division. Extensive studies of floral mutants with perturbed homeostasis in recent years have provided parts of the answer to this question. As in *Drosophila*, mutations that cause homeotic transformations, which in this case result in changes in the identity of floral organs such that one organ type is replaced by another, have been identified. Genetic and morphological studies of these floral homeotic mutants have put them into three distinct genetic groups each of which alters the identities of organs in two adjacent whorls. In the first class of mutants, whorl 1 and 2 are affected resulting in the transformation of sepals to carpels and petals to stamens. The second class of mutations affect whorl 2 and 3 giving rise to sepals in place of petals and carpels instead of stamens. In the final mutant class stamens are converted to petals in whorl 3 and sepals are formed instead of carpels in whorl 4. Thus, according to this classification, a floral meristem can be divided into 3 distinct regions reflecting three classes of organ identity functions. This has led to the proposal of the 'ABC model' for floral organ identity (Coen and Meyerowitz, 1991; Weigel and Meyerowitz, 1994). In this model the floral homeotic genes are required for 3 classes of activities namely A, B and C and it is the unique combination of one or two classes of these activities that specifies the identity of each floral organ. Hence class A genes act in whorl 1 to specify sepals and the combination of A and B class gene activity directs petal development in whorl 2 while that of B and C genes gives rise to stamens. Lastly, C activity alone is required to specify carpel development in the innermost whorl. In addition, the ABC model also suggests that activities of floral homeotic genes are confined to spatially

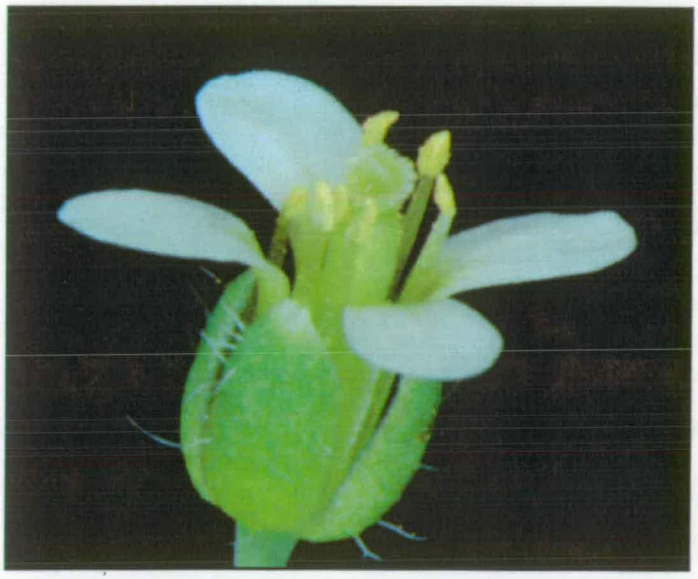


Figure 1.6 An *Arabidopsis thaliana* flower consisting of four concentric whorls of organs with four sepals, four petals, six stamens and two fused carpels arranged from the outermost to the innermost floral whorls.

restricted domains i.e., the two adjacent whorls of organs whose identities are altered when the corresponding sets of genes are mutated. Cloning and characterisation of plant homeotic genes show that this is indeed mostly the case. Thus, with some exceptions such as the *Arabidopsis APETALA2* (*AP2*) gene, which encodes a transcription factor with a novel AP2 domain and is expressed ubiquitously (Jofuku et al., 1994), most of the floral homeotic genes encode putative transcription factors which share a highly conserved motif called the MADS-box and whose expression are largely restricted to the regions in which they function to specify organ identity (Reviewed in Riechmann and Meyerowitz, 1997; Theissen et al., 2000). In *Arabidopsis*, genes that encode class A, B and C functions are well known and characterised. Thus the genes *APETALA1* (*AP1*), *LEUNIG* (*LUG*) and *AP2* contribute A function. B function requires the activity of the genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) while the C activity required the function of *AGAMOUS* (*AG*) and the recently identified *HUA1* and *HUA2* genes (Chen and Meyerowitz, 1999; Li et al., 2001)

Over the years the ABC model has been widely accepted since it correctly predicted the phenotypes of organs of double and triple mutant combinations and the effects of ectopic expression of homeotic genes in flowers of transgenic plants. However, there are several lines of evidence to suggest that although the ABC functions are necessary, they are not sufficient for the specification of floral organ identity. Most notable is the observation that although triple mutants that carry combined alleles of A, B and C functions exhibit a conversion of floral structures into leaf-like organs thought to represent a developmental ground state, simultaneous ectopic expression of A, B and C activities fail to convert vegetative leaves into floral organs (Mizukami and Ma, 1992; Krizek and Meyerowitz, 1996). This indicates that additional factors are also required to specify floral organ identity. Recently these missing factors have been identified and characterised in detail. The genes, namely *SEPALLATA* (*SEP*) 1, 2 and 3 represent a new functional class of floral organ identity MADs-box genes (Pelaz et al., 2000; Pelaz et al., 2001). These three genes encode functionally redundant activities which are required with the combined functions of class B and C genes to specify petal, stamen and carpel identity. Triple mutant that lacks all 3 *SEP* functions produce flowers that consists of sepals only (Pelaz et al., 2000). Conversely, combinatorial ectopic expression of the SEP genes and various ABC genes is sufficient to convert leaves into floral organs (Honma and Goto, 2001; Pelaz et al., 2001). Further characterisation of the *SEP* genes has provided another striking twist to the classic ABC model. Thus, it appears that the SEP proteins can directly interact and form complexes with

the ABC proteins and the multimeric complexes formed are thought to provide the molecular basis of the ABC model (Honam and Goto, 2001). Therefore, in the revised ABC model, petal identity is specified by the combination of class A, B and *SEP* activities while the combined class B, C and *SEP* functions are essential for stamen development and the presence of the AG/SEP complex in whorl 4 is required for the specification of carpel fate (Figure 1.7). Since the *sep123* triple mutant does not display transformations of the first whorl organs it is thought that the wild-type genes are not required for proper first whorl organ development. Instead sepal identity might be specified through the complex of the class A proteins and other yet unidentified factors (Goto et al., 2001).

An important tenet of the ABC model is that the A and C functions are mutually antagonistic such that the A products inhibit the activity of the C genes in the outer two floral whorls and conversely the C function represses the A activity in the inner two whorls. This basic principle would account for the homeotic defects associated with the A and C mutations. Thus in the absence of the C activity it is predicted that class A function expands into whorl 3 and 4, and likewise, the C function becomes active in all 4 floral whorls when the A genes are mutated. By far the most well understood case of the A-C negative interactions is that of the *Arabidopsis AP2* and *AG* gene.

Transcription of the class C gene *AG* is initiated in the centre of a wild-type flower at stage 3 of flower development (Yanofsky et al., 1990; Drews et al., 1991). Expression analyses indicated that *AG* expression is primarily and mainly induced by the activity of the *LEAFY* (*LFY*) gene which is also required previously in the establishment of floral meristem identity (Schultz and Haughn, 1991; Weigel et al., 1992; Weigel and Meyerowitz, 1993). It has been shown that the onset of *AG* expression is delayed in the *lfy* mutant background and when expressed *AG* RNA is initially confined to a noticeably smaller domain than in wild-type (Weigel and Meyerowitz, 1993). Subsequent studies revealed that *AG* is a direct target of *LFY*. Thus in transgenic lines harbouring a gain-of-function *LFY* allele; *LFY*::VP16, in which *LFY* is fused to the strong activation domain from the viral transcription factor VP16, *AG* expression is both ectopic and precocious (Parcy et al., 1998). Recent findings have located a *LFY* response element in the second intron of *AG* which is both necessary and sufficient for LFY binding and the response to *LFY in vivo* (Sieburth and Meyerowitz, 1997; Busch et al., 1999).

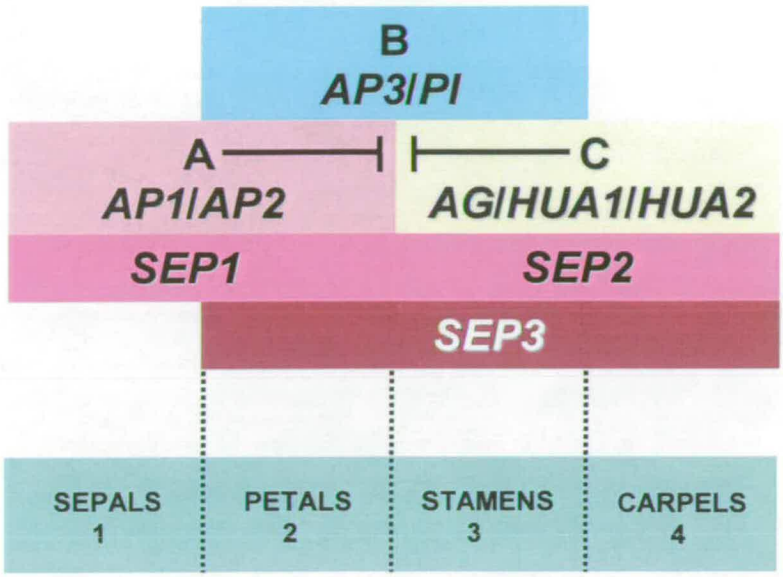


Figure 1.7 The ABC model of floral homeotic genes. The identity of organs in each floral whorl is specified through a combination of the activities of 3 classes of floral homeotic genes and the *SEP* activities. The class A function genes *AP1* and *AP2* specify sepals in the first whorl. In whorl 2, petals are specified by the A, B (*AP3* and *PI*) and *SEP* activities. Stamens in whorl 3 are specified by the B, C and *SEP* activities while the C (*AG*, *HUA1*, *HUA2*) and *SEP* activities are required in the specification of carpel identity in whorl 4. *AP1* (*APETALA1*), *AP2* (*APETALA2*), *AP3* (*APETALA3*), *PI* (*PISTILLATA*), *AG* (*AGAMOUS*), *SEP* (*SEPALLATA*). Adapted from Jack, 2001.

After the initiation stage, regulation of *AG* expression takes place at the RNA level such that *AG* RNA is being confined to the inner two floral whorls which are affected by *ag* mutations. The establishment of a spatially limited domain of *AG* expression is mainly brought about by the antagonistic activity of the class A gene *AP2*. In strong *ap2* mutants, *AG* RNA expands into the two outer floral whorls from late stage 2 of flower development onwards causing the homeotic transformation of sepals to carpels and the failure of petals to develop (Drews et al., 1991). Similar phenotypic defects are observed in 35S::*AG* transgenic flower in which *AG* expression is placed under the control of the strong 35S promoter from the cauliflower mosaic virus (Mizukami and Ma, 1992). These results indicate that the wild-type *AP2* gene is required to repress *AG* transcription in whorl 1 and 2 early in flower development. Rather unexpectedly, however, *AP2* is not only expressed in whorl 1 and 2 but also in whorl 3 and 4 where *AG* normally functions (Jofuku et al., 1994). This raises a question of how the transcriptional repression of *AG* by *AP2* is specifically limited to the outer two floral whorls. One possibility is that *AG* expression is differentially induced to a greater extent in the two inner whorls by the interaction of LFY with specific factors such that the *AP2*-mediated repression of *AG* can be overcome in the centre of the flower (Parcy et al., 1998). This idea is supported by the finding that rather then being uniformly transcribed, *AG* expression is to some extent weaker in the outermost whorl as compared to the level in the inner whorls in *ap2* flowers (Drews et al., 1991). Alternatively, the ability of *AP2* to negatively regulate *AG* expression in the outer floral whorls requires either post-translational modification or specific interactions with other *AG* regulators. A good candidate for such an AP2-inteacting factor is the product of the gene *LUG*. As in the *ap2* mutants, mutation in *LUG* causes precocious and ectopic AG expression in all 4 floral whorls (Liu and Meyerowitz, 1995). An intimate relationship between the two genes is also suggested by the fact that *ap2* and *lug* mutations show synergistic enhancement of defects in *AG* repression (Liu and Meyerowitz, 1995), and that both AP2 and LUG act upon redundant enhancer sequences in the second intron of *AG* (Sieburth and Meyerowitz, 1997). Cloning of *LUG*, however, revealed that it is ubiquitously expressed in all 4 floral whorls and that LUG does not co-immunoprecipitate with AP2 (Connor and Liu, 2000). This indicates that the interaction between LUG and AP2 is indirect and may require additional factors and these or other whorl-specific factors are also needed to mediate *LUG* and *AP2* repression activity in spatially restricted domains.

Study of temperature sensitive *ap2* allele suggests that although *AP2* is expressed continuously throughout flower development, its repression function is not required beyond

stage 4 of development (Bowman et al., 1989). This creates a problem of the maintenance of the repressed state of *AG* expression during later stages of flower development. This paradox is likely overcome by the activities of additional genes that take over the repressor role from the early acting genes like *AP2*. One such activity is provided by the recently identified *CURLY LEAF* (*CLF*) gene.

As the name suggests, mutations in the *CLF* gene result in a prominent leaf-curling phenotype. In addition, *clf* plants are early flowering and exhibit occasional weak homeotic transformation of sepals into carpels and petal into stamens in whorl 1 and 2 of flowers respectively (Goodrich et al, 1997). These defects resemble the observed phenotypes of transgenic plants with constitutive *AG* expression (Mizukami and Ma, 1992; Mizukami and Ma, 1997), suggesting that *AG* was ectopically expressed in the *clf* mutant background. This notion was verified by both molecular and genetic evidence. Expression analysis by RNA blotting showed that *AG* RNA is ectopically expressed in leaves, stems and outer floral whorls of *clf* plants. In addition the *ag* mutation is almost completely epistatic to *clf* mutation. Thus *clf ag* double mutant has normal leaves, sepals and petals and near wild-type flowering time suggesting that ectopic *AG* activity is largely responsible for the *clf* phenotype (Goodrich et al., 1997). RNA blot analysis also indicates that, in addition to *AG*, the homeotic gene *AP3* is also ectopically expressed, albeit at a much lower level in *clf* mutant leaves. Apart from *AG* and *AP3*, misexpression of other floral homeotic genes has so far not been detected in *clf* leaves. Detailed expression studies using *in situ* hybridisation showed that in flower development the initial expression of *AG* is normal but ectopic accumulation of *AG* transcripts become apparent in later stages of development with weak expression detected in the petals from stage 9 onwards (Goodrich et al., 1997). This finding indicates that the wild-type *CLF* gene is not required in the establishment of correct AG expression pattern but rather has a significant role in the maintenance of the repressed state of AG expression throughout development. This is reminiscent of the maintenance of homeotic gene expression by Pc-G genes in *Drosophila*. This parallel between the CLF and the *Drosophila* Pc-G function was extended by the finding that the *CLF* gene encodes a protein which is a homologue of the D*rosophila* Pc-G member *E(z)* (Goodrich et al., 1997). This finding is striking since the homeotic targets in *Arabidopsis* and *Drosophila* encode structurally unrelated sets of proteins with the MADS-box factors found in *Arabidopsis* and the homeobox proteins in *Drosophila*. These differences imply that the repression activity of Pc-G proteins might have evolved independently in plants to create a cellular memory

system that ensures proper propagation of the initially established correct patterns of gene expression.

Protein alignment reveals extensive homology between the CLF and E(Z) proteins with several highly conserved domains found along the length of the two proteins. Most notable is the SET domain found at the C-terminus of the CLF and E(Z) proteins. In addition to the SET domain and the adjacent cysteine-rich CXC region, the E(Z) and CLF protein also share a highly conserved SANT (SWI3, ADA2, N-COR and TFIIIB" DNA-binding domain) which lies just upstream of the CXC domain. The SANT domain is found in many transcriptional regulators including a number of chromatin remodelling proteins such as those in the SWI/SNF complexes. Evidence from biochemical analysis suggests that this domain is involved in non-specific DNA binding (Aasland et al., 1996). At present the precise role of this domain and its possible co-ordination with the SET domain in Pc-G silencing still remain to be investigated. Furthermore, two other conserved motifs specific to the E(Z) family are found at the N-terminus. The first of these is a highly acidic region of about 75 amino acids, while the second adjacent domain is rich in cysteine residues. Currently, the function of either domain is unknown. Finally, the CLF protein also contains a short basic region which is thought to be a putative bipartite nuclear localisation signal (NLS) lying between the SANT and the CXC domain. A recent study using CLF::GFP fusion constructs indicated that in accordance with its putative role as a chromatin regulator the CLF proteins are indeed localised in the nucleus (L Primavesi pers. comm.).

In situ analysis showed that *CLF* RNA is expressed in all four floral whorls (Goodrich et al., 1997). Thus, similar to the *AP2* scenario, expression of *CLF*, at least at the RNA level, is insufficient to repress *AG* since their expression patterns overlap in the inner two floral whorls. This again invokes the question of how the activity of *CLF* is modulated in a whorl specific manner. It is conceivable that specificity of *CLF* action is acquired by post-translational modifications or selective degradation of CLF proteins in whorl 3 and 4. Alternatively, *CLF* function may be modulated by its interactions with whorl specific proteins.

1.5.2 The *Arabidopsis* ESC and E(Z) homologues control seed development

Since the landmark discovery of the first plant Pc-G member; the *Arabidopsis CLF* gene, the involvement of Pc-G function in plant development has garnered much interest from plant developmental biologists in recent years. Many of the studies carried out have focused on the identification and characterisation of additional plant Pc-G members. Indeed, a number of novel plant Pc-G homologues have recently been discovered and unlike *CLF* which regulates homeotic targets during vegetative and floral development, these Pc-G genes act in the female gametophyte during early seed development.

As is typical of angiosperms, *Arabidopsis* seed is a product of double fertilisation. Fusion of the haploid egg cell and the diploid central cell of the female gametophyte with two sperm cells from the male parent produces a diploid embryo and a triploid endosperm respectively. In *Arabidopsis*, embryo development follows a well-defined series of distinct morphological stages from globular to heart, torpedo, walking stick, early maturation and maturation (Holdberg et al., 1994). Early endosperm development is marked by a series of synchronous nuclear divisions which are not accompanied by cell division resulting in the production of a syncytium of nuclei that surround the embryo. Cellularisation eventually takes place at the heart stage of embryo development with these nuclei and the cytoplasm being partitioned into cells (Brown et al., 1999). Although the exact nature of their relationship is still not known, available data indicates that endosperm functions to support the growth and development of the embryo. Thus, endosperm cells produce large amounts of storage proteins, starch and lipids which are absorbed by the embryo during seed development (Lopes and Larkins, 1993).

The *Arabidopsis FIS* (*FERTILISATION-INDEPENDENT-SEED*) genes were identified in a genetic screen devised to isolate mutants that exhibit seed development without fertilisation (Chaudhury et al, 1997). In wild-type, fertilisation is needed for seed development and elongation of carpel to form silique. In the *fis* mutants, however, carpel elongation and some aspects of seed development, including autonomous endosperm proliferation, occur in the absence of pollination. Three *fis* mutants were identified and these mapped to three separate loci on different chromosomes; *FIS1*, *FIS2* and *FIS3*. Separate genetic screens have identified additional mutants namely *fertilisation-independent endosperm development* (*fie*)

and *medea* (*mea*) that also display similar abnormalities in seed development (Ohad et al., 1996; Ohad et al., 1999; Grossniklaus et al., 1998). Molecular characterisation of *mea* and *fie* mutations indicated that they are allelic to the *fis1* and *fis3* mutation respectively. Cloning and sequencing of the genes showed that the *FIS3/FIE* gene encodes a protein with structural homology to the *Drosophila* ESC protein (Ohad et al., 1999). The *MEA* gene, on the other hand, was found to be another homologue of the *Drosophila E(z)*, in addition to the *CLF* gene. The corresponding MEA protein contains a SET domain and a CXC cysteine-rich domain which share 55% and 43% amino acid identity with the corresponding counterparts in the E(Z) protein respectively (Grossniklaus et al., 1998). Cloning of the *FIS2* gene suggested that it encodes a protein with predicted zinc finger and nuclear localisation signal and thus is likely to be involved in transcriptional regulation (Luo et al., 1999). Detailed analysis of the three *fis* mutants revealed that in *mea* and *fis2* mutant ovules autonomous endosperm development progresses up to the multicellular stage and occasionally embryo-like structures are formed (Ohad et al., 1996; Chaudhury et al., 1997). In a *fis3/fie* mutant, however, endosperm develops only up to the syncytium stage (Chaudhury et al., 1997). These phenotypes of the *fis* mutants suggest that the wild-type *FIS* genes play an important role in preventing proliferation of the central cell before fertilisation occurs (Kiyosue et al., 1999; Luo et al., 1999; Ohad et al., 1999).

In addition to the similarities in fertilisation-independent seed development, all three *fis* mutants also exhibit defects in seed development even if fertilised (Ohad et al., 1996; Chaudhury et al., 1997). Genetic evidence indicates that proper seed development requires the maternal but not the paternal copy of the 3 *FIS* genes. Pollination of a female plant heterozygous for any one of these 3 mutations with wild-type pollen resulted in developmental abnormalities in those that inherited a maternal mutant allele. The *fis* seed aborts late in development. Characterisation shows that the endosperm overproliferates and the embryo arrests at the heart stage. Conversely, no seed abortion was observed when the reciprocal cross was carried out. These results indicate that the *FIS* genes behave as classic gametophytic maternal-effect genes (Chaudhury et al., 1997; Grossniklaus et al., 1998; Kiyosue et al., 1999). A number of possible explanations might account for the observed parent-of-origin specific effects of the *fis* mutants. First, the *FIS* gene products could be required in a dose dependent fashion. In this case a single copy of a given *FIS* gene inherited through the pollen might not be sufficient to support embryogenesis due to the failure to compensate for the loss of two maternal copies in the triploid endosperm. Evidence from

recent investigations has, however, argued strongly against this endosperm haplo-deficiency hypothesis. Thus, it has been confirmed that the mutant female gametophytes cannot be rescued by the presence of additional paternal copies of the wild-type *FIS* genes (Grossniklaus et al., 1998; Kiyosue et al., 1999; Ohad et al., 1999; Vielle-Calzada et al., 1999). A second explanation is that the maternal activity of the *FIS* genes in the female gametophyte is required before fertilisation to ensure proper embryo and endosperm development. For instance, loading of the maternal *FIS* RNA or protein into the egg or central cell before the introduction of the paternal allele by fertilisation might be essential for early seed development post-fertilisation. In support of this model, the *MEA*, *FIS2* and *FIE* genes are all expressed in the female gametophyte prior to fertilisation (Luo et al., 2000). However, unlike early animal development in which the requirement of stored maternal information in the unfertilised egg is well established, it is thought that plant embryo development requires little stored maternal material. This stems from two observations. First, plant embryos can often be readily formed from cultured somatic cells. Second, there are only few maternal effect mutations on embryo development isolated in plants. A third model is that only the maternal allele is active during early seed development. In support of this model, reporter gene studies using *FIS* promoters have revealed that, the paternally inherited allele is not activated until the later stages of seed development (Kinoshita et al., 1999; Luo et al., 2000). The preferential expression of the maternal allele of the *FIS* genes suggests that they are imprinted during seed development. Genomic imprinting is widely recognised as an epigenetic phenomenon caused by differential expression of the genes inherited from maternal and paternal gametes. Thus unlike most genes whereby the maternally and paternally inherited copy are identically expressed, only one copy of an imprinted gene is transcriptionally active. It is thought that imprinting has evolved as a consequence of the competition between the maternal and paternal genomes over resource allocation from the mother to the offspring (Haig and Westoby, 1989). According to the parental conflict theory, in polygamous species of plants and mammals, a mother's interest is to allocate resources equally among her offspring since she has equal investment in each embryo. Meanwhile, a father must ensure that his offspring has the best fitness, thus giving a better chance of his genes being passed on through later generations. Therefore, for the father's best benefit his embryos must be able to extract the maximum amount of nutritional resources from the maternal tissues that is the endosperm. These conflicts led to the proposal that genes that promote zygotic growth would be preferentially expressed by the paternal derived allele while the maternal allele is silenced. Conversely, the maternal copy of the growth restricting

genes are preferentially transcribed and the paternally inherited allele is silenced. The main support of this parental conflict model comes from the analysis of genome dosage effects in interploidy crosses (Haig and Westoby, 1991; Scott et al., 1998). In general, increases in paternal genetic load result in endosperm overproliferation and hence larger seeds. On the contrary, endosperm reduction, hence the production of smaller seeds, are associated with excess maternal genetic dosage. By analogy, the phenotypes of the *fis* mutants can be interpreted as diagnostic of paternal excess. This and the fact that the mutants exhibit maternal effect lethality strongly indicate that the wild-type *FIS* genes are regulated by imprinting mechanisms. In fact there are now several lines of evidence to suggest that the parent-of-origin effects associated with the *MEA* locus is caused, at least in part, by genomic imprinting. Detailed characterisation of *MEA* mRNA distribution revealed high levels of *MEA* transcripts in the developing seeds suggesting that *MEA* is zygotically transcribed (Vielle-Calzada et al., 1999). Moreover, analysis of nascent transcripts showed that *MEA* transcripts are readily detected in the central cell nucleus as two visible transcription focuses which presumably correspond to the products of the two copies of the two maternal copies of the *MEA* gene. Post-fertilisation analysis revealed no additional dots in the derived endosperm. This implies that the paternal copy of the *MEA* gene is transcriptionally silent in the endosperm (Vielle-Calzada et al., 1999). In accordance with these observations, RT-PCR analysis has shown that only the maternal *MEA* allele is active in the developing endosperm (Kinoshita et al., 1999; Vielle-Calzada et al., 1999). These results indicate that *MEA* is at least imprinted in the endosperm. At present the nature of parent-of-origin effect of *mea* mutations on embryo development is still under dispute since the available data are contradictory. However, the fact that occasionally *mea* mutations can be transmitted through the female lines resulting in viable and phenotypically normal homozygous plants suggests that the female gametophytic function of the *MEA* gene is not absolutely critical for embryo development (Grossniklaus et al., 1998). Therefore, the defects seen in *mea* embryos are likely indirect consequences of the main defects in the surrounding endosperm. In addition to *MEA*, the other two *FIS* genes, *FIS2* and *FIE*, are also likely regulated partly by imprinting at least in the very early stages of seed development. Analysis of the expression patterns of *FIE::GFP* transgenes showed that only the maternal *FIE::GFP* allele is transcribed in the central cell of the female gametophyte and during early stages of endosperm development while the paternal copy remains silenced during this time interval (Yadegari et al., 2000). Similarly, expression studies of the FIE and FIS2::GUS fusion constructs indicated that the paternally inherited constructs are transcriptionally inactive during early seed development

(Luo et al., 2000). Expression of the FIE::GUS fusion protein was later detected in the heart stage embryos and the cellularised endosperm suggesting that imprinting effects are somewhat relieved at later stages of development (Luo et al., 2000).

At present the targets of the *FIS* genes have not yet been determined. However, given the autonomous endosperm proliferation defects and the parent-of-origin effects on seed development associated with the three *fis* mutations, it is rather likely that their targets are those genes that promote growth (via the control of cell proliferation) during seed development. Those genes must also be normally repressed in the maternal genome and derepression is only allowed to take place after fertilisation.

As previously mentioned, the ESC/E(Z) subset of the Pc-G proteins are widely conserved in evolution and are thought to act together as part of a distinct class of multimeric Pc-G protein complexes. The striking structural similarities of the three *Arabidopsis* FIS proteins to the *Drosophila* ESC/E(Z) proteins suggest that the *FIS* genes might also act together in a complex. This notion is supported by the fact that similar phenotypic abnormalities are associated with the three *fis* mutants. In fact it has recently been shown in several independent studies that not only the E(Z) and ESC pairs are structurally conserved in plants but their biochemical partnership appears to be also retained in evolution. Thus, results from yeast two-hybrid studies and *in vitro* binding assays have revealed direct physical interaction between the MEA and FIE proteins (Luo et al., 2000, Spillane et al., 2000; Yadegari et al., 2000). Rather strikingly, the domains of interaction are also conserved between plants and animals. Thus the amino-terminal regions of E(Z) and MEA are both necessary and sufficient for the interactions with their respective partner (Tie et al., 1998; Jones et al., 1998; Luo et al., 2000; Spillane et al., 2000; Yadegari et al., 2000). Even more striking are the findings that MEA can also weakly interact with ESC despite such sequence divergence between MEA and E(Z) at their N-terminus region (Spillane et al., 2000). The observed conservation of molecular partnership of WD and SET domain Polycomb proteins in *Arabidopsis* provides a strong piece of evidence for the existence of a functional FIS multiprotein complex which functions during seed development. In support of this view, expression analysis indicates that *MEA* and *FIE* share overlapping expression domains in the female gametophyte and in early embryo and endosperm and that they are co-localised in the nucleus (Luo et al., 2000; Spillane et al., 2000; Vielle-Calzada et al., 1999).

As mentioned previously in *Drosophila*, assembly of the Polycomb complex is thought to be initiated by recruitment of the transiently acting ESC protein by the zinc finger transcription factor HB (Kehle et al., 1998). ESC in turn functions to recruit other Pc-G proteins including its direct partner E(Z) to form a stable repression complex at their target sites (Sewalt et al., 1998; van der Vlag and Otte, 1999; Ng et al., 2000;). The observed conservation at both the structural and biochemical levels suggest that a similar repression complex might exist in plants. Thus, the putative zinc finger protein FIS2 might recruit the WD protein FIE which then recruits other Pc-G proteins including MEA to establish a stable repression complex that might modify higher order chromatin structure into a transcriptionally inactive state, mediated perhaps by the associated histone deacetylase activity, resulting in a long term stable repression of target gene expression (Figure 1.8). It would be of particular interest to find out if the targets are also conserved in plants, since although the biochemical properties of the WD and SET domain protein pair are highly conserved, the developmental processes that they regulate appear not to be so. Thus, whilst ESC and E(Z) are involved in homeotic gene regulation in *Drosophila*, the MES-2 and MES-6 partnership functions in the regulation of germ line development and on the other hand the FIS/MEA pair is required in the control of cell proliferation during seed development.

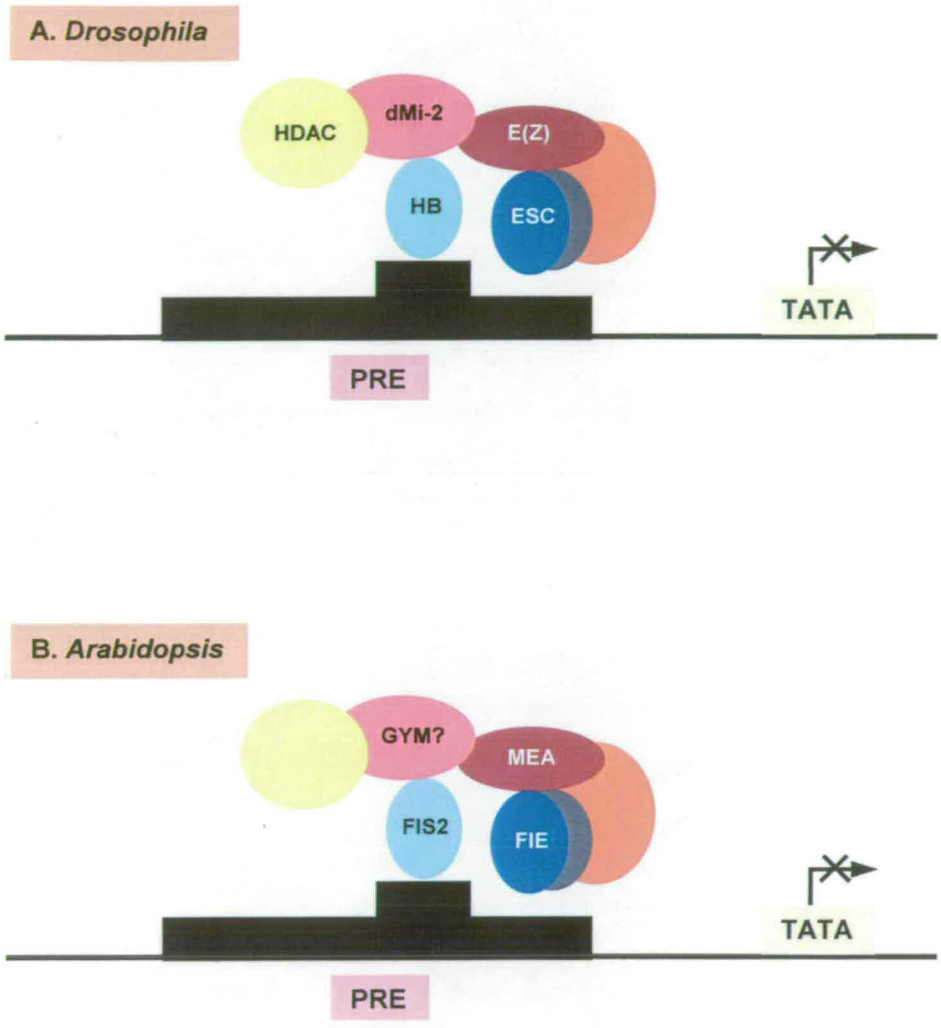


Figure 1.8 Comparison of Pc-G mediated repression in *Drosophila* (A) and *Arabidopsis* (B) system. DNA-binding proteins such as HB and FIS2 recruits dMi-2/histone deacetylase (HDAC) complexes and the ESC-E(Z) (*Drosophila*)/ MEA-FIE (*Arabidopsis*) protein complex. Local chromatin changes induced by the histone deacetylase and Pc-G activities result in the conformational changes in the chromatin structures into the transcriptionally silent state. Adapted from Grossniklaus et al., 2001.

1.6 Aims of the project

The recent cloning of several plant Pc-G genes revealed not only that their structures are highly conserved but their molecular functions may also be faithfully maintained. This suggests that plants have also evolved a Pc-G mediated silencing mechanism as part of an epigenetic network that helps maintain the stable state of gene expression through cell division in diverse developmental processes.

In the well-characterised *Drosophila* system, Pc-G genes are found to localise at several target sites on the polytene chromosomes. Genetic studies have shown that Pc-G proteins interact together both *in vitro* and *in vivo* and that these interactions are essential for their function as transcriptional repressors. It has also been shown that in *Drosophila* at least two functionally distinct Pc-G complexes exist in Pc-G mediated silencing of homeotic genes.

On the contrary, relatively little is known about the Pc-G mediated repression mechanism in plants. Evidence gathered so far has suggested some fundamental differences between plant and *Drosophila* Pc-G paradigms. For instance, unlike the promiscuous nature of the *Drosophila* Pc-G genes, study of the *Arabidopsis CLF* gene so far has only assigned *AG* and *AP3* as its targets. Up until now, *CLF* has not yet been shown to form a functional protein complex with other proteins, although physical interactions with other *Arabidopsis* Pc-G proteins have been detected *in vitro*. Also the loss of *CLF* function does not result in lethality and the phenotypic abnormalities seen in the *clf* mutant are rather mild compared to the detrimental developmental defects conferred by strong alleles of its *Drosophila* counterpart, *E(z)*, and other *Drosophila* Pc-G members. One possible explanation that could account for these discrepancies is that in the *clf* background the effect on the expression of other target genes is minimal due to a considerable degree of functional redundancy in the regulatory pathways that *CLF* is involved in. Thus, the activities of other genes, possibly other Pc-G members, that act in the same complex or developmental processes as *CLF* could to some extent compensate for the reduction or loss of CLF activities.

The main aim of this study is, therefore, to use various genetic and molecular approaches to identify and characterise genes in *Arabidopsis* that interact with the *CLF* gene. Such genes could also encode additional members of the *Arabidopsis* Pc-G proteins. Analysis of their functions and possible interactions with *CLF* during development would give invaluable

insights into the regulation of gene silencing mediated by *CLF* and more generally plant Pc-G genes. This would in turn provide a more comprehensive understanding of the epigenetic regulation that takes place during plant development.

The project is divided into two parts. The first of these concerns the use of forward genetics to characterise a novel mutant which bears striking similarities to *clf* plants. Techniques such as reporter gene assay and double mutant analysis was carried out to establish possible genetic interactions with *CLF* and floral homeotic genes. In addition, map-based cloning was carried out in order to isolate and characterise the corresponding gene. This showed that the mutant was an unusual allele of the gene *EMF2*, recently found to be another *Arabidopsis* Pc-G member.

The latter half of the project employed various reverse genetic tools to characterise a homologue of the *CLF* gene identified from searches of the sequence databases. Several mutagenesis approaches were applied to attempt to isolate mutant alleles of this gene. Phenotypic analysis of the mutants was then carried out to help elucidating the function of the wild-type gene. Possible interactions with *CLF* and other known *Arabidopsis* Pc-G genes were also investigated.

Chapter 2. Materials & Methods

2.1 Plant Materials

2.2 DNA Analysis

2.3 RNA Analysis

2.4 *In situ* Hybridisation

2.5 Yeast Two-Hybrid Interaction Assay

2.6 *In vitro* Mutagenesis

2.7 Gene Transfer into *Arabidopsis*

2.8 Histochemical Staining of the GUS Reporter Lines

2 Materials and Methods

2.1 Plant materials

2.1.1 *Arabidopsis thaliana* stock lines

***clf-2* line**

The strong *clf-2* mutant was originally derived from a transposon mutagenesis experiment (Long et al., 1993) in the Landsberg *erecta* (L*er*) background. Detailed characterisation revealed an insertion of the *Ds* element at the first exon-intron junction and a subsequent 8 nucleotide duplication of the sequence. As a consequence a cryptic splice site is used and the mRNA produced carries a 5 nucleotide insertion. The homozygous *clf-2* line used in this study was provided by Justin Goodrich.

***clf-50* line**

The *clf-50* line confers a severe *clf* phenotype. This allele was identified in a T-DNA mutagenesis experiment in Wassilewskija (Ws) background and was kindly donated by Dr. E Huala. Molecular analysis indicated that there was no T-DNA insertion associated with the mutation. However, Southern and PCR analysis revealed that the *CLF* locus (and possibly adjacent sequences) is deleted in this background. The allele, therefore, represents a null allele (J Goodrich, pers. comm.).

***clf-9* line**

The weak *clf-9* allele was isolated in a screen for revertants in progeny of *clf-2* parents that carried an *Ac* transposase source. Molecular analysis revealed that *clf-9* results from imprecise excision of the *Ds* insertion from the *clf-2* allele. The excision results in the 5 bp duplication of the 5' exon/intron junctions and impairs splicing of the *CLF* message. *clf-9* homozygotes exhibit near wild-type phenotype with only very mild leaf curling. The homozygous *clf-9* line used in this study was provided by Dr. J Goodrich.

***ag-3/+* line**

The strong *ag-3* allele was generated by EMS-induced mutagenesis in the L*er* background (Bowman et al., 1991). Seeds for the heterozygous *ag-3/+* line used in this experiment were provided by Martin Yanofsky.

***lfy-5* line**

The weak lfy-5 allele was derived from an EMS mutagenesis experiment in the L*er* background. Characterisation of this allele revealed a single amino acid change from proline to leucine at position 240 in the LFY protein sequence (Weigel et al., 1992). The homozygous *lfy-5* line used in this study was provided by Detlef Weigel.

***ap1-1* line**

The homozygous *ap1-1* line was obtained from Martin Yanofsky. The strong *ap1-1* mutation was first isolated in an EMS mutagenesis experiment in the L*er* ecotype (Koornneef et al., 1982). Subsequent cloning of the *AP1* gene showed that the mutation in this allele occurs at the splice acceptor site of the third intron and is predicted to truncated most of the C-terminus of the protein. (Mandel et al., 1992).

***ap2-1* line**

The temperature sensitive *ap2-1* mutation was originally isolated by Maarten Koornneef (Koornneef et al., 1980) and characterised by John Bowman (Bowman et al., 1991). DNA sequence analysis indicated that the *ap2-1* allele carries a serine to glycine substitution at residue 251 in the conserved AP2 domain (Jofuku et al., 1994). The homozygous *ap2-1* line used in this study was obtained from the Nottingham *Arabidopsis* Stock Centre (NASC).

***ap3-3* line**

The strong *ap3-3* allele (L*er* background) was generated by EMS mutagenesis. The mutation is a single nucleotide change that causes a conversion of a glutamine residue at amino acid 18 in the protein sequence to a stop codon (Jack et al., 1992). The *ap3-3/+* heterozygous line used in this study was obtained from Thomas Jack.

pi-1 line

The strong recessive *pi-1* mutant allele was characterised in details by John Bowman (Bowman et al., 1989). It carries a single nucleotide substitution in the first exon that results in a change from tryptophan residue at position 80 to a stop codon (Goto and Meyerowitz, 1994). The *pi-1* line used was obtained from NASC.

The *AGAMOUS* reporter lines

The pAG-I::*GUS* reporter gene construct was obtained from Leslie Sieburth. This fusion construct contains a 9.8 kb *AG* sequence consisting of ~6 kb upstream and 3.8 kb genomic sequence of the *AG* gene fused to the *GUS* coding sequence at nucleotide 331 of *AG* cDNA near the 5' end of the exon 3. Genetic analysis has shown that this construct contains *AG* intragenic sequences necessary for response to *CLF* (Sieburth and Meyerowitz, 1997).

The pI::*GUS* fusion line was provided by Detlef Weigel. This line contains a 3 kb *Hin*dIII restriction fragment of the *AG* gene which largely coincides with the large second intron fused to a minimal promoter in a GUS reporter construct (Busch et al., 1999).

The *AP3* and *PI* reporter lines

The pAP3::*GUS* fusion construct was developed by Thomas Jack. This line contains 3.7 kb of the *AP3* promoter sequence and a 1.5 kb fragment of 3' *AP3* sequence fused upstream and downstream of the *GUS* coding sequence respectively (Jack et al., 1994). The pPI::*GUS* line was obtained from Jose Luis Riechman. The fusion construct was created by joining a 480 bp promoter sequences of the *PI* gene with the *GUS* coding sequence.

The pAG-I::*GUS* and pI::*GUS* constructs were crossed to the *clf-2* and *moe*/+ background by Justin Goodrich and Lucia Primavesi. All the other constructs used were introduced into the *moe*/+ background by Justin Goodrich. The F1 progeny were allowed to self and kanamycin resistance was used to select F2 plants that carried the reporter construct.

Steroid inducible *clf-2/clf-2* line

The dexamethasone inducible p(CLF)::CLF-GR fusion construct was developed by Justin Goodrich. The rat glucocorticoid receptor steroid binding domain was joined in frame to the 3' end of the *CLF* coding sequence and cloned into the binary pGR0029 vector. Wild-type *clf-2*/+ heterozygotes were transformed by floral dip. Primary transformants were identified

from the T1 plants by selecting for the acquired resistance to the herbicide BASTA conferred by the *BAR* gene in the T-DNA sequence. In subsequent generations, lines that were homozygous for the *clf-2* allele and for a single locus transgene insert were selected. These all showed steroid dependent rescue of the *clf* mutant phenotype

2.1.2 Plant Growth Conditions and Pollination

i. Seed sterilisation

Chlorine vapour treatment was used to sterilize seeds. Three ml of hydrochloric acid was added to 100 ml of sodium hypochlorite solution to generate chlorine gas in a sealed desiccator. Treatment was carried out for 6 to 16 hours. In some cases seeds were sterilized by treating with 70% ethanol solution for 2 minutes and then 5% (v/v) sodium hypochlorite: 0.5% (v/v) SDS solution for 15 minutes followed by several rinses in sterile water. After sowing imbibed seeds were stratified by keeping at 4°C for 1-3 days to synchronize germination.

ii. Growing *Arabidopsis* on soil

Plants were grown on a 3:1 mixture of Levingston's compost:grit in greenhouse or long-day growth room.

iii. Growing *Arabidopsis* on plates

Sterilised seeds were sown on petri dishes containing GM agar, pH 5.7 (1 litre contains 4.604g MS salt, 10g glucose, 100 mg inositol and 0.5g MES). In cases of transgenic lines, selection were made by adding appropriate antibiotics (50 µg/ml kanamycin or hygromycin) to the media. Plates were put in a growth cabinet at 22°c under constant light.

iv. Cross-pollination

Female parents were emasculated by removing immature anthers from floral buds. Pollination was performed on open flowers by smearing pollens from anthers of the male parents onto the stigma of the emasculated female flowers. Seeds were collected from mature siliques and allowed to ripen at room temperature for 2 weeks before sowing. When possible, transgenic plants were used as male parents, so that F1 seeds could be distinguished from self-pollinated contaminants using antibiotic or herbicide resistance.

2.2 DNA Analysis

2.2.1 Isolation of Plant DNA

i. Alkali lysis method for PCR analysis

A small piece (3 mm^2) of rosette leaf or a young cauline leave was put in an eppendorf tube containing 50 µl 0.25M sodium hydroxide. The tube was boiled for 30 seconds and 50 µl of 0.25M hydrochloric acid and 25 µl of 0.5M Tris-HCl pH8.0/0.25% (v/v) IGEPAL CA-630 were then added and the tube boiled for another 2 minutes. The treated leaf tissue was picked with a sterile pipette tip and used directly in PCR analysis.

ii. A small-scale DNA extraction for PCR analysis

One small leaf or half a large leaf was put in an eppendorf tube containing with fine glass beads and 500 µl DNA extraction buffer (100mM Tris pH8.5, 50mM EDTA pH8.0, 500mM NaCl, 10mM β-Mercaptoethanol) and 35 µl 20% SDS. The leaf was mashed thoroughly with plastic pestle and quickly frozen in liquid nitrogen. When all the samples were collected, the tubes were incubated at 65° C for 5 minutes then 130 µl 5M potassium acetate was added and the samples incubated on ice for 5 minutes and then spun for 10 minutes at 13,000 rpm on a bench top mini centrifuge. The supernatant was transferred to a fresh tube and 640 µl isopropanol and 60 µl 3M sodium acetate were added. The samples were then incubated at -20°C for 2 hours and then spun for 15 minutes. The DNA pellet was washed in 70% ethanol and resuspended in 50 µl dH_20. In a typical PCR reaction 2 to 5 µl of DNA solution were used as template.

iii. High Throughput DNA Isolation from Plant Leaf Tissue

Small pieces of plant leaf was collected and put in a rack of 12x8 microtubes each containing a tungsten carbide bead. Once all the samples had been collected the rack was rapidly cool in liquid nitrogen and the samples disrupted in the MM300 Mixer Mill (Retsch GMbtt & Co.) by sandwiching the rack between 2 adapter plates and fixing them in mixer mill clamps before shaking vigorously for 1 minute. Extraction solution (400 µl buffer AP1, 1µl 100 mg/ml RNaseA and 1 µl Reagent DX; Qiagen) was then added to each collection microtube and the rack tightly sealed, shaken vigorously for 15 seconds and centrifuged for 10 seconds

at 1500 x *g* in centrifuge 4-15C (Qiagen). To precipitate the DNA, 130 µl of Buffer AP2 was added to each microtube and the rack manually shaken for 15 seconds, spun for 10 seconds at 1500 x *g*, incubated at –20°C for 10 minutes and then spun at 5600 x *g* for 5 minutes. The supernatants were transferred to a new rack of collection tubes and 600 µl Buffer AP3/E (Qiagen) was added to each supernatant. The rack was then shaken vigorously for 15 seconds and centrifuged for 10 seconds at 1500 x *g*. A 1ml aliquot of each sample was then added to a DNeasy 96 plate placed on top of a square-well block and centrifuged at 5600 x *g* for 4 minutes. To wash the DNA, 800 µl Buffer AW (Qiagen) was added to each well of the DNeasy plate and the samples centrifuged for 15 minutes at 5600 x *g* to dry the DNeasy membrane. Finally, to elute the DNA the DNeasy plate was placed on a new rack of collection microtubes and 100 µl Buffer AE (Qiagen) added to each well. After 1 minute incubation at room temperature, the samples were centrifuged for 2 minutes at 5600 x *g*. The elution step was repeated with another 100 µl Buffer AE. DNA samples were frozen immediately and 2 µl were used in the PCR-based mapping analysis.

iv. Genomic DNA Extraction for Southern Analysis

Several leaves were ground to a fine powder with a pre-cooled pestle and mortar in liquid nitrogen. The tissue was then quickly transferred to a chilled eppendorf tube containing 500 µl extraction buffer (50mM EDTA, 0.1M NaCl, 0.1M Tris-HCl pH8, 1% SDS). An equal volume of phenol/chloroform was added and the sample vortexed, left on the bench for 5 minutes, vortexed again and centrifuged for 5 minutes. The aqueous phase was taken to a new tube and the phenol/chloroform extraction step was repeated. To precipitate the DNA, 50 µl 3M sodium acetate pH5.5 and 350 µl isopropanol were added and the sample centrifuged for 5 minutes. The DNA pellet was rinsed with 70% ethanol, air-dried and resuspended in 50 µl TE solution. One third of the DNA extracted was used for Southern analysis.

v. Plasmid DNA Extraction

Single colonies were inoculated in independent 5 ml LB medium (1% tryptone, 0.5% yeast extract, 1% NaCl) containing an appropriate antibiotic. Cultures were grown overnight at 37°C with constant shaking. The next day bacterial cells were harvested from 1.5 ml of culture by centrifugation at 13,000 rpm in a microcentrifuge for 1 minute. Cells were resuspended in 100 µl GTE solution (50mM glucose, 25mM Tris-HCl pH7.5, 10mM EDTA)

EX BIBL. UNIV. EDIN.

then lysed in 200 µl 0.2M NaOH/1%SDS solution for 2 minutes. Once the cells were lysed 150µl 3M potassium acetate, pH5.2 was added and the tube inverted several times then centrifuged at 13000 rpm in a microcentrifuge for 5 minutes. The clear supernatant was carefully transferred to a fresh microtube and 1 ml of ethanol then added and the solution vortexed briefly before being spun immediately for 5 minutes. The DNA pellet was washed in 70% ethanol, air-dried and resuspended in 50 µl TE.

2.2.2 Restriction Analysis of DNA

Plasmid, phage and DNA from PCR reactions was digested in 20 µl reactions containing 1 x of the appropriate restriction buffer and 3.5 units/µg template DNA of the corresponding enzyme as supplied from the manufacturers. The samples were incubated for 60-90 minutes at 37°C. If large amounts of DNA were to be digested, the reactions described were scaled up to appropriate volumes accordingly. For Southern analysis, 5 µg of genomic DNA was digested in 1x restriction buffer containing 1mM spermidine for at least 6 hours at 37°C. When double digests were performed the buffer was chosen such that optimized restriction efficiency for both enzymes was obtained.

2.2.3 DNA gel electrophoresis

Gel electrophoresis was performed in submerged, horizontal gels of size 4-8 mm thick to determine the size and the presence of DNA and to separate DNA fragments from restriction digests and PCR reactions. Gels used were 0.7-4% agarose (Bioline) or 2-4% Metaphor (Fisher Scientific) in TBE (2.5mM EDTA, 90mM boric acid and 90mM Tris-HCl pH8.0). Before running the gel 1 x loading buffer (0.25% Ficoll, 0.0025% xylene cyanol and 0.0025% bromophenol blue) was added to each sample. Electrophoresis was carried out at 8 volts/cm for 1 to 2 hours in 0.5 x TBE buffer. DNA from genomic digests for Southern analysis was run in 0.7% agarose in a larger tank at 1-2 volts/cm for 20-72 hours. Gels were stained in ethidium bromide (10 µg/ml) for 20-30 minutes and the DNA was visualised under UV light.

2.2.4 Purification of DNA from agarose gel

Ethidium bromide-stained DNA was visualised using a low energy UV translilluminator, and the desired DNA fragments were excised from the gel. DNA was extracted and purified with the Qiagen's QIAquick Gel Extraction kit. In order to elute the DNA 30 µl of the Elution buffer (10mM Tris-HCl pH8.5) was added to the QIAquick column. After 1 minute standing at room temperature, the column was centrifuged for 1 minute at 15000 rpm in a microcentrifuge and the DNA was collected in an eppendorf tube.

2.2.5 Southern Blotting

Ethidium bromide-stained gels were soaked in 0.25M HCl with gentle agitation for 15 minutes to depurinate large DNA fragments and aid DNA transfer during blotting. If the desired fragments were less than 10 kb big this stage was bypassed. This was followed by a 30 minute wash in denaturation solution (0.5M NaOH, 1.5M NaCl) and two 15 minute washes in neutralization solution (0.5M Tris-HCl pH7.2, 1.5M NaCl, 1mM EDTA). As described by Maniatis et al., 1982, DNA was allowed to transfer to a Hybond-N (Amersham) Nylon membrane on a wick platform with 20 x SSC (1M NaCl, 100mM sodium citrate) as the transfer buffer. After 15-20 hours of transfer, the filters were washed briefly in 2 x SSC solution, air-dried and baked under vacuum at 80°C for 1 hour to fix the DNA onto the membrane. Before hybridisation, the filters were incubated at 65°C with 100 ml of hybridisation buffer (4 x SSC, 1% SDS and 0.5% (w/v) dry milk) for at least 1 hour. The solution was then replaced by 30 ml of fresh hybridisation solution and the appropriate denatured probe was added. Hybridisation was allowed to proceed in a 65°C incubator with gentle shaking. After overnight incubation the probe was collected and stored at -20°C for future use. The filter was washed at 65°C with 2 x SSC, 0.1 % SDS for 15 minutes, then 1 x SSC, 0.1% SDS for 25 minutes and lastly 0.1 x SSC, 1%SDS for another 15 minutes. The filter was then wrapped in clingfilm and put in X-ray cassettes. Exposure with X-ray films (Cronex, Dupont) with intensifying screens were carried out at -70°C for 1-14 days.

2.2.6 Radiolabelling of DNA probe

The DNA probe fragment (500 ng) was added to 60 ng random hexamer primer, boiled for 2 minutes spun briefly and put on ice. A labeling mix was added containing 5mM Tris-HCl pH6.9, 1mM $MgSO_4$, 10 µM dithioreitol, 6 µM dATP, 6 µM dTTP, 6 µM dGTP, 50 µCi α32-P dCTP and 5 units of DNA polymerase I (Klenow fragment, Roche Scientific) in a 20 µl reaction. The reaction was incubated for at least 1 hour at room temperature. To fragment the probe, the reaction mix was applied onto a 1ml Sephadex G50 column followed by the addition of 180 µl TE pH8.0 and 20 µl 10 x NT dye (1% Dextran Blue, 0.1% Orange G). The blue fraction containing the longer probe fragments was collected and heated at 100°C prior to hybrisation to denature the probe or stored at –20°C otherwise.

2.2.7 Preparation of DNA samples for sequencing

DNA sample (40 ng/1kb) was mixed with dH_2O to a volume of 4.4 µl in a PCR tube. To each tube 1.6 µl of 1 µM sequencing primer and 4 µl of dRhodamine terminator mix (PE Applied Biosystems) were added. PCR was performed using the following cycles,

96°C for 30 seconds
96°C for 30 seconds ⎫
50°C for 15 seconds ⎬ 25 cycles
60°C for 4 minutes ⎭

The extension products were transferred to a microtube and DNA was precipitated by adding 1 µl 3M sodium acetate pH5.5 and 25 µl ethanol. The tube was incubated on ice for 10 minutes then centrifuged for 15 minutes at full speed. The DNA was washed in 70% ethanol and air-dried. Prior to sequencing the DNA sample was mixed with 6-9 µl loading buffer (5:1 ratio of deionized formamide to 25mM EDTA, pH8.0 / 50 mg/ml Blue Dextran), heated for 2 minutes to denature the DNA and then loaded onto an ABI Prism 377 sequencing gel (Perkin Elmer). Analysis of sequences was performed using the GCG10 program (University of Wisconsin's sequence analysis package, Devereux et al., 1984).

2.2.8 DNA Cloning

i.Ligation of DNA fragments to plasmid vectors

A 10 µl reaction was set up with 1 x ligation buffer (Roche Scientific) and 1 unit of T4 DNA ligase (Roche Scientific) and a vector to DNA fragment molar ratio of 1:1. Reactions were incubated in a 14°C water bath overnight.

ii.Transformation with electrocompetent cells

Cell Preparation

A 1 litre culture of *E.coli* DH5α strain was grown in LB medium from a 10 ml overnight culture at 18°C. When an OD_{600} of 0.5-0.7 was reached, cells were incubated on an ice waterbath for 30 minutes, then centrifuged at 4500 rpm in a chilled Sorval GSA rotor for 15 minutes. The cell pellet was resuspended in 1 litre ice cold 10% glycerol and centrifuged at 4500 rpm for 15 minutes in a chilled Sorval GSA rotor. This resuspension/centrifugation step was repeated with 500 ml, 20 ml glycerol volume etc. until the cells were resuspended in a final volume of 2-3 ml. Cell aliquots of 100 µl were prepared, quick frozen in liquid nitrogen and stored at -70°C.

Electro-transformation

Frozen competent cells were placed on ice and after complete thawing 40 µl of cells were mixed with 1 µl ligation reaction in an ice cold microfuge tube and then transformed into a chilled transformation cuvette. Cells were shocked with a 18 kV electric pulse in a Gene Pulser machine (Biorad) set to 200 Ohms and 25 µF and immediately resuspended in 1 ml SOC media (2% bactotryptone, 0.5% bacto yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$ and 20mM glucose). This was followed by incubation at 37°C with constant shaking. After 1 hour incubation, cells were spun at 10000 rpm in a microcentrifuge for 1 minute, resuspended in 100 µl LB medium and spread out on ampicillin (100 µg/ml) or kanamycin (50 µg/ml) LB plates containing X-Gal (40 µg/ml) and IPTG (200 µg/ml). Plates were incubated overnight at 37°C.

iii. PCR Cloning using TOPO TA Cloning Package (Invitrogen)

A 5 µl reaction was set up with 2 µl PCR product, 1 µl pCR®II TOPO® vector, 1 µl salt solution (1.2M NaCl, 0.06M $MgCl_2$). After a 5 minute incubation at room temperature, 2 µl of the ligation mix were added to 50 µl of thawed chemically competent TOPP10F' cells (Invitrogen) or Epicurian coli XL2-Blue MRF' ultracompetent cells (Stratagene) together with 2 µl β-mercaptoethanol. Cells were incubated on ice for 30 minutes, heat-shocked for 30 seconds in a 42°C water bath, immediately chilled on ice for 2 minutes and incubated with 250 µl SOC media at 37°C. After 1 hour the cells were spread on LB plates containing 100 µg/ml ampicillin, X-Gal (40 µg/ml) and IPTG (12.5 µg/ml) and the plates incubated overnight at 37°C. Colonies were selected the next day and plasmid DNA was prepared from overnight cultures. Restriction digest with the enzyme *Eco*RI was carried out to identify the presence of DNA insert.

2.2.9 Polymerase Chain Reaction (PCR)

i. Oligonucleotide preparation

Oligonucleotides were obtained from several manufacturers. The dried samples were dissolved in 200 µl sterile water. The absorbance at 260nm of a 5 µl sample was determined for each oligonucleotide and the concentration then calculated using the following formula;

$$\text{Concentration (M)} = \text{Optical Density}_{260} / \text{Extinction Coefficient } (\varepsilon)$$

where ε = the number of nucleotide 'A' x (15.4×10^3) + 'G' x (13.7×10^3) + 'C' x (8.7×10^3) + 'T' x (7.4×10^3)

The oligonucleotides used in PCR and sequencing experiments are listed as shown in table 2.1;

Table 2.1 Oligonucleotide sequences

NAME	SEQUENCE (5' TO 3')	DIRECTION	GENE
YC5	CCT TCC ATG GAT CTT GGT GC	Forward	*PISTILLATA*
YC6	CCA CGA TAG TAT AGA TAG GAG	Reverse	*PISTILLATA*
YC7	CCA CAA CGA AGG AGA TCG	Forward	*APETALA-3*
YC8	CAA TGA TGG CAC CAG CAA ACC	Reverse	*APETALA-3*
YC9	CGG AGC TAG GAG GAG ATT CCT CTC	Forward	*AGAMOUS*
YC10	GCT TGC TCA ACC CAA TTC TGG	Reverse	*AGAMOUS*
YC11	CGG CGA GCA CGC TTG TGG GTA TG	Forward	*LEAFY*
YC12	CTG ATG ACA CCT TCA CAG TTC	Reverse	*LEAFY*
YC13	CAT ACC AAA GTC TGA GCT C	Forward	*APETALA1*
YC14	GGC TGG AAA GAG CCT AGC CAG	Reverse	*APETALA1*
YC15	TTG GGA AAG TCT GAT AAT GCT GGA TGG GGT GCT	Forward	*CURLY LEAF LIKE1*
YC16	AGA AAG CAC CAA TCG GAT CCA TGA TTA CTG GCT	Forward	*CURLY LEAF LIKE1*
YC17	TGA GAG ATG ATA TTT GAA TCC TGA TAA CTT AAG CTT	Forward	*CURLY LEAF LIKE1*
YC18	AGA CGC GGA GAA GGG CAA TGC	Forward	*CURLY LEAF-LIKE1*
YC19	CAG GAG ATC ACA GGG TCG GGA	Forward	*CURLY LEAF LIKE1*
YC22	CAG AAG AGG AAT GCA GCA AGG C	Forward	*CURLY LEAF LIKE1*
YC23	CTT CCA AAG ATT TCG ACT CC	Reverse	*CURLY LEAF LIKE1*
YC24	TGC CTG ATA TCA AAC GTT TGT C	Forward	*CURLY LEAF*
YC25	ACA GGA ACG TCA TCC TTC ACAG	Reverse	*CURLY LEAF*
YC26	TCC ATG AGT ACT GTA GCA GGG TCA TTT GG	Reverse	RAT GR-LBD
YC28	AGA AGC GGA AAG CTC GAG GAG TC	Forward	*CURLY LEAF LIKE1*
YC29	CAC CTA TAC AAA GTC CAA ACT CGA TGC	Reverse	*CURLY LEAF LIKE1*
YC32	AAG CTC GAG TCG ACA ATG GTG ACG GAC	Forward	*CURLY LAEF LIKE1*
YC33	ATC ACC TAT ACA GCG GCC GCA CTC GAT GTC AA	Reverse	*CURLY LEAF LIKE1*
YC40	TCT CTC AGC AAG CGG CCG CTT GAC ACT TGC	Reverse	*CURLY LEAF LIKE1*
YC44	CAA GTG TCA AGC TTC CTA TTG CTG	Forward	*CURLY LEAF LIKE1*
YC45	GGT GCA GCT CTA GAT TCT TTT G	Forward	*CURLY LEAF LIKE1*
YC46	GTC TGT CTG GAA AAG AAT AGC TGG	Forward	*CURLY LEAF LIKE1*
YC47r	GAC CGG GCA AAT TCG TCC TTC	Reverse	*CURLY LEAF LIKE1*

YC48r	GCC ACC AGC TAT TCT TTT CCA G	Reverse	*CURLY LEAF LIKE1*
YC49r	GTC AGT TGT TCC ACC TTG ATC AC	Reverse	*CURLY LEAF LIKE1*
YC50r	GAA TCT AGA GCT GCA CCA AGG C	Reverse	*CURLY LEAF LIKE1*
YC51r	CAG CAA TAG GAA GCT TGC CAC	Reverse	*CURLY LEAF LIKE1*
JL202	CAT TTT ATA ATA ACG CTG CGG ACA TCT AC	Forward	T-DNA
JL270	TTT CTC CAT ATT GAC CAT CAT ACT CAT TG	Forward	T-DNA
CLK-F	GGA CGA ATC GTA CTC AGA TTT ACC TTT CA	Forward	*CURLY LEAF LIKE1*
CLK-R	TGG TTT AGT GTG ATT GGT TCG TTC ACT TA	Reverse	*CURLY LEAF LIKE1*
CLK-U	CAA ACG TAA CTG CTA TAA CCT GTG TGG CT	Forward	350 bp 5' OF *CLK1*
JG15	CGT ACT CAG ATT TAC CTT TCA ATC TCC TC	Forward	*CURLY LEAF LIKE1*
JG16	CAA TGA GAT TGG TGC TTT CTG GCT C	Reverse	*CURLY LEAF LIKE1*
JG17	CTC TAA TTC CAT GCT CTT GTT CTA CAC C	Reverse	*CURLY LEAF LIKE1*
JG18	CGT GAG GTC CCA ATT CTA AAA GAC TC	Forward	*CURLY LEAF LIKE1*
CLK-FN	CGT GAA ATC CGT TTG AGT TCA TCA AAT TA	Forward	*CURLY LEAF LIKE1*
JG17N	GCA GCA GCA TGG GGA ACA TCA GAG GAG	Reverse	*CURLY LEAF LIKE1*
M13f	GTA AAA CGA CGG CCA GT		
M13r	GGA AAC AGC TAT GAC CAT G		
Sp6	GAT TTA GGT GAC ACT ATA G		Sp6 PROMOTER
TUB-F	GTT CTT GAT AAC GAG GCC TT	Forward	*β-TUBULIN*
TUB-R	ACC TTC TTC CTC ATC CTC G	Reverse	*β-TUBULIN*
KNAT1-F	CAT AGA TGA GTC GTC TAG TCG	Forward	*KNAT1*
KNAT1-R	CAT GTC ACA GTA TGC TTC CAT G	Reverse	*KNAT1*
EMF2.7	GAG TTA ACT GTT CCA TCT CCA T	Forward	*EMF2*
EMF2.1	CTG TGA TGG TGT TAT GCA CTG	Forward	*EMF2*
EMF2.2	ACG CCG TGC TAT TAG GAA TG	Forward	*EMF2*
EMF2.3	TAC AGA GGG AAT AAT TTT TG	Reverse	*EMF2*
EMF2.4	AAA AGT AAA CCA ATA CAT G	Reverse	*EMF2*
EMF2-P1	AGA TGC TGG GGT ACA AAC TCA AAA A	Forward	*EMF2*
EMF2-P3	GGA GCA AAA TAC CTC TGC AAT C	Forward	*EMF2*
EMF2-P4	GCA GAA GAA GTT GGG TCA ACG G	Forward	*EMF2*
EMF2-P5	ATG TTG TAG GGC CTG AGA TAT C	Forward	*EMF2*
EMF2-P2	TAC AAC TAA GAC AGG CTC TCA TCA C	Reverse	*EMF2*

EMF2-P6	CAT AAA TCG AAC AAA CAA GAT GTT C	Reverse	*EMF2*
EMF2-P7	GCA GAC TAA ACT GAA GCT TAC C	Reverse	*EMF2*
EMF2-P8	CTC TTA TAA TTG GAG TCA AAG	Reverse	*EMF2*
EMF2-F1	TGT GTG AGA CCA TAA AAA TTC TTC	Forward	*EMF2*
EMF2-F2	CGG ACC GGG ATA GTG AAG ATG AAG	Forward	*EMF2*
EMF2-F3	CAA TCA TGT GGG GAT CCA GAT TAT G	Forward	*EMF2*
EMF2-F4	TTC AGT ACA TGT CAC GTG GAG TTT TCC	Forward	*EMF2*
EMF2-F5	CCC AAG AAA TCA TTA ATT TAA G	Forward	*EMF2*
EMF2-R1	ATT GTA CTT CCA GCT ATC GAC	Reverse	*EMF2*
EMF2-R2	ACC CAA CTA CTG AGC TTT ATT CTA C	Reverse	*EMF2*
EMF2-R3	AGT CAG AGA ACA TGG ATG CGT ATG	Reverse	*EMF2*
EMF2-R4	CAA ATT CAT GCA TAA ATA TGC TC	Reverse	*EMF2*

ii. PCR reaction

The PTC-200 DNA engine (MJ Research) was used to carry out the PCR reactions. A typical 50 µl reaction contained approximately 50 ng of template DNA mixed with 10 x PCR buffer (10mM Tris-HCl, 15 mM $MgCl_2$, 50mM KCl, pH8.3), 0.4 µM of each primer primer, 200 µM of each deoxynucleotide 5' triphosphate and 0.2 units of Taq DNA polymerase (5 units/µl). Reactions were performed with a 'HOT' start. Thus the reactions were preheated to 94°C before the Taq polymerase was added. Typical cycle reactions consisted of the initial incubation step at 94°C for 2 minutes. This was followed by repeated cycles; 2 minutes denaturation at 94°C, 30 seconds primer annealing step and 3 minutes polymerase extension at 72°C. The annealing temperature used was between 55 and 65°C depending on the primer Tm and the reaction stringency required. The extension time required was approximately 1 minute per 1 Kb of DNA. The cycle was repeated 25-30 times for RT-PCR in which PCR-induced mutations need to be minimized and 40-50 times for all other PCRs including those used for the mapping purpose. Finally, the PCR products were completely extended by incubation at 72°C for 10 minutes and then promptly run on agarose gels or stored at 4°C for subsequent analysis.

iii. Wisconsin Knockout PCR Screen

PCR screen for T-DNA insertions in the *CURLY LEAF LIKE1* gene was carried out at the *Arabidopsis* Knockout Facility, The University of Wisconsin, USA. Each PCR reaction contained 0.4 ng of DNA template, 12 pmole of each primer, 200 μM of each deoxy nucleotide 5' triphosphate, 1 x Takara's Ex-Taq Polymerase buffer (Panvera). The reaction was made up to 40 μl with sterile water and 40 μl of mineral oil was added on top to prevent evaporation of PCR products. A 'Hot' start was carried out. The sample was heated to 96°C and 10 μl of enzyme mix containing 1 x Ex-Taq polymerase buffer, 2.5 units Ex-Taq polymerase (Panvera) were added to each tube. The PCR was carried out using 36 cycles of 15 seconds at 94°C, 30 seconds at 65°C and 2 minutes at 72°C followed by a 4 minute extension time.

2.3 RNA Analysis

2.3.1 RNA Isolation

i. Isolation using TRIzol Reagent

Plant tissue was harvested and homogenized in TRIzol reagent (GIBco BRL) with 1 ml of the reagent per 50-100 mg of tissue. The homogenized sample was incubated for 5 minutes at room temperature. Chloroform was added (0.2 ml per 1 ml of TRIzol originally used) and the sample shaken vigorously for 15 seconds, then left at room temperature for 3 minutes. The sample was then centrifuged for 15 minutes at 4°C and the aqueous layer transferred to a fresh sterile eppendorf tube. RNA was precipitated by mixing the sample with isopropanol (0.5 ml per 1ml of TRIzol originally used). After a 10 minute incubation at room temperature the tube was spun for 10 minutes at 4°C. The RNA pellet was washed once with 70% ethanol (1ml ethanol per 1 ml TRIzoL reagent used) and air-dried briefly. Finally the pellet was dissolved in RNase-free water and store at -70°C. The concentration of RNA was determined by spectrophotometry using the estimation that 1 unit of absorbance at 260 nm corresponds to approximately 40 μg/ml.

ii. RNA Isolation using QIAGEN RNeasy Plant Mini Kit

Approximately 100 mg of plant tissue was completely ground into fine powder using liquid nitrogen with a pre-cooled pestle and mortar. Tissue powder was transferred to an eppendorf tube and 450 μl buffer RLT was added to the sample and the tube shaken vigorously. The resulting lysate was applied to the QIAshredder spin column and centrifuged for 2 minutes. The flow-through fraction was then transferred to a fresh tube. A half volume of ethanol was added and mixed well with the lysate. The sample was then applied onto an RNeasy mini spin column and centrifuged for 15 seconds at 10000 rpm. To wash the RNA, 700 μl buffer RN1 was added and the sample centrifuged for another 15 seconds at 10000 rpm in a microcentrifuge. The column was transferred onto a new collection tube and washed twice with 500 μl buffer RPE. To elute the RNA, 50 μl RNase-free water was added to the RNeasy membrane and the column spun for 1 minute at 10000 rpm in a microcentrifuge.

2.3.2 3'-End cDNA Amplification

i. First strand cDNA synthesis

For each sample, 5 µg of total RNA was denatured in 13 µl of RNase-free water by incubating at 80°C for 3 minutes. After a short incubation on ice the sample was spun briefly and then added to the reverse transcription components containing 4 µl of 5 x M-MLV reverse transcription buffer (250mM Tris-HCl pH8.3, 375mM KCl, 15mM $MgCl_2$ and 50mM DTT), 1 µl of 10mM dNTPs, 2 µl 0.1M DTT, 0.5 µl 8µM Q_T primer and 10 units of RNase inhibitor. Subsequently 200 units of M-MLV reverse transcriptase were added and the reaction incubated for 5 minutes at room temperature, 1 hour at 42°C and 10 minutes at 50°C. After a 15 minute incubation at 70°C to inactivate the reverse transcriptase, the reaction mixture was diluted to 500 µl with TE pH8.0 and stored at -20°C.

ii. cDNA Amplification

An RT-PCR reaction was set up using an appropriate primer pair. Typically a 50 µl reaction was performed using 2 µl of the first strand cDNA template. PCR amplification was carried out on a PTC-200 DNA engine (MJ Research) using the following amplification cycles;

94°C for 3 minutes
94°C for 30 seconds
58°C for 30 seconds } 25-35 cycles
72°C for 3 minutes
72°C for 5 minutes

5 µl of the reaction products were run on 1% agarose gel and visualised under UV light after being stained with ethidium bromide.

2.3.3 5'-End cDNA amplification

i. Preparation of poly A^+ mRNA using Qiagen's Oligotex mRNA kits

A 250 µl solution containing 200 µg of total RNA in RNase-free water was mixed with 1 volume of 2 x binding buffer (20mM Tris-HCl pH7.5, 1M NaCl, 2mM EDTA and 0.2%SDS) and 15 µl Oligotex suspension (10% w/v suspension of Oligotex particles in 10m Tris-HCl pH7.5, 500mM NaCl, 1mM EDTA, 0.1% SDS and 0.1% NaN_3) prewarmed to 37°C in a RNase-free microfuge tube. The RNA mix was incubated for 3 minutes at 65°C, 10 minutes at room temperature then spun for 2 minutes at 13,000 rpm in a microcentrifuge. The

oligotex pellet was resuspended in 1 ml of wash buffer OW2 (10mM Tris-HCl pH7.5, 150mM NaCl and 1mM EDTA), vortexed and centrifuged for 2 minutes. This washing step was repeated and the mRNA was then eluted twice with 100 µl of preheated elution buffer (5mM Tris-HCl pH7.5) and stored at -70°C.

ii. First-strand cDNA synthesis using CLONTECH SMART™ RACE cDNA Amplification kit

A 5 µl reaction was set up with 1 µg poly A^+ RNA, 1ul 10µM 5'-CDS primer [5'-$(T)_{25}$N-1N-3' (N= A, C, G or T, N-1 = A, G or C)], 1µl 10µM SMART II oligo (5'AAGCAGTGGTAACAACGCAGAGTACGCGGG-3') and sterile water in an RNase-free microfuge tube. The tube was spun briefly, incubated at 70°C for 2 minutes, put on ice for 2 minutes and spun again briefly. A 5 µl cocktail mix containing 2 µl 5 x first-strand buffer (250mM Tris-HCl pH8.3, 375mM KCl and 30mM $MgCl_2$), 1 µl 20mM DTT, 10mM dNTP mix and 1 µl M-MLV reverse transcriptase (200 units/µl) was added to the reaction tube and the contents were mixed by gentle pipetting. The tube was centrifuged briefly and then incubated at 42°C for one hour and a half. The first-strand reaction product obtained was diluted with 250 µl TE pH8.0 and heated at 72°C for 7 minutes.

iii. Rapid Amplification of cDNA ends

A PCR master mix containing 5 µl 10 x PCR buffer (100mM Tris-HCl, 15mM $MgCl_2$, 500mM KCl pH8.3), 1µl 10mM dNTP mix, 1µl Taq DNA polymerase (5 units/µl) and 34.5 µl dH_2O was set up and added to PCR reaction tubes containing 2.5 µl 5'-RACE-ready cDNA, 5 µl 10 x universal primer mix (5'-CTAATACGACTCACTATAGGGCAAGC AGTGGTAACAACGCAGAGT-3') and 1µl 10µM primer JG17. The contents were mixed by gentle pipetting and PCR was carried out using the following cycles;

94°C for 5 seconds
65°C for 10 seconds } 25 cycles
72°C for 3 minutes

The primary PCR product was diluted in TE buffer, pH8.0 (5 µl PCR product in 245 µl TE). Nested PCR reaction was performed using 5 µl of the diluted primary PCR product, 1 µl 10 µM nested universal primer (5'-AAGCAGTGGTAACAACGCAGAGT-3') and 1 µl 10

μM nested gene specific primer JG17N in a cocktail of reaction mix and same reaction cycles as described above.

2.4 *In situ* Hybridisation

The expression pattern of the *CURLY LEAF LIKE1* gene was analysed by *in situ* hybridisation with labelled RNA probe. *In vitro* transcribed RNA probe was non-radioactively labelled with digoxygenin-11-UTP and hybridised to sections of *Arabidopsis* inflorescence and vegetative tissue. Probe detection was carried out by an immunochemical assay with an anti-digoxygenin antibody conjugated to an alkaline phosphatase. The anti-digoxygenin antibody binds to digoxygenin and the phosphatase enzyme catalyses the colour reaction with Western Blue substrate (Promega) such that the signal can be detected as a blue/purple precipitate.

2.4.1 Tissue Fixation and Embedding

Fresh inflorescence and vegetative meristem tissues were harvested and fixed in 4% (v/w) paraformaldehyde. Paraformaldehyde was dissolved in 1 x phosphate buffered saline (PBS, 130mM NaCl, 7mM Na_2HPO_4, 3mM NaH_2PO_4) which had been warmed to 60°C and the pH adjusted to 11 with NaOH pellets. The solution was then slowly cooled down to 4°C and its pH readjusted to 7 using H_2SO_4. In order to improve the penetration of the fixative into the tissues, 0.1% (w/v) Triton-X-100 and 0.1% (w/v) Tween 20 were added. The plant materials were put in Histoprep plastic grills (Fisher Scientific) and immersed in the fixative. Vacuum was then applied for 15 minutes in order to aid the penetration of the fixative. The fixative was later replaced with fresh solution and fixation proceeded overnight at 4°C with gentle shaking.

After fixation, the tissue was washed twice in 1 x PBS with gentle shaking to remove the paraformaldehyde. Subsequently the tissue was dehydrated through an ethanol series (Table 2.2). All the steps involving paraformaldehyde were carried out in a fume hood. The tissue was then removed from the plastic grills, placed in universal bottles and cleared with histoclear (CellPath). Wax was introduced at the second 100% Histoclear treatment by adding ¼ volume of paraplast chips (BDH-Merke) and leaving overnight in a 42°C heating block. After overnight incubation, a further ¼ volume of melted chips were added and the wax/histoclear mixture was replaced with fresh molten wax after 2 hour treatment at 60°C. The tissue was incubated in wax overnight at 60°C. The molten wax was replaced twice in the morning and evening of the following day. This was repeated for the next two days. The

details of the treatments are illustrated in table 2.2 Finally the tissue was placed in plastic moulds containing molten wax. The position of the tissue was carefully adjusted for subsequent sectioning. The mould was floated on water to solidify the wax and later stored at 4°C until sectioning.

2.4.2 Sectioning of embedded tissues

The wax block containing fixed tissue was carefully cut using a razor blade into a trapezoid, mounted onto a small metal disc and kept at 4°C for at least 15 minutes. The disc was secured onto a microtome (Reichet-Jung) and its position adjusted such that the face of the trapezoid was aligned next to and parallel to the blade. Sections were cut into ribbons of 8 μM think and floated on sterile water on a Probe On Plus coated slide (Fisher Scientific). The slide was placed in a 42°C hot plate for about 15 minutes until the ribbons flattened. Water was removed by suction with a sterile Plasteur pipette and blotting with a lens cleansing tissue (Whatman). Finally the slides were left in the 42°C hot plate overnight to allow the ribbons to stick onto the slide surface. The slides were stored at 4°C until use.

Table 2.2 Tissue treatment for embedding

TREATMENTS	TIME	TEMPERATURE
30% Ethanol	1 hour	4°C
40% Ethanol	1 hour	4°C
50% Ethanol	1 hour	4°C
60% Ethanol	1 hour	4°C
70% Ethanol	1 hour	4°C
85% Ethanol	1 hour	4°C
95% Ethanol and 0.1% eosin	Overnight	4°C
100% Ethanol and 0.1% eosin	30 minutes	Room temperature
100% Ethanol and 0.1% eosin	30 minutes	Room temperature
100% Ethanol and 0.1% eosin	1 hour	Room temperature
100% Ethanol and 0.1% eosin	1 hour	Room temperature
25% Histoclear, 75% ethanol	1 hour	Room temperature

50% Histoclear, 50% ethanol	30 minutes	Room temperature
75% Histoclear, 25% ethanol	30 minutes	Room temperature
100% Histoclear	1 hour	Room temperature
100% Histoclear	1 hour	Room temperature

2.4.3 *In vitro* transcription of RNA probes

The desired *CLK1* DNA fragment to be transcribed was subcloned into the polylinker of pBluescript KS$^-$ (Stratagene) next to the T7 polymerase promoter (for details see section 4.2). Plasmid DNA was extracted and purified. The inserts were then PCR amplified. The resultant PCR product was purified from agarose using the QIAGEN's QIAquick PCR purification kit, dried in a vacuum concentrator (Savant) and 0.5-1 µg was resuspended in 18 µl RNase-free water. The diluted DNA was added to a labeling reaction mix containing 2.5 µl 10 x RNA polymerase buffer (Roche Scientific), 1 µl RNase inhibitor (20 units/µl), 2.5 µl of 5mM ATP, GTP and CTP, 2.5 µl of 1mM DIG-UTP and 1 µl of RNA polymerase (20 units/µl). The reaction was incubated at 37°C for 2 hours and then stopped by adding 75 µl RNase-free water, 1µl of 100 mg/ml tRNA and 3 µl of DNase I (10 units/µl). After 15 minute incubation at 37°C to allow DNaseI to digest template DNA, the RNA was precipitated using 1 volume of 3.5M sodium acetate and 2 volumes of ice-cold ethanol. The RNA pellet was washed in DEPC-treated 70% ethanol, air-dried and resuspended in a mixture of 100 µl RNase-free water and 100 µl of 2 x carbonate buffer. The probe was then hydrolysed at 60°C. The time required for hydrolysis was calculated according to the formula:

$$T = \frac{Li - Lf}{k.Li.Lf}$$

where Li = initial probe length (700 bp)
Lf = final probe length (150 bp)
K = rate constant (0.11 Kb/min)

The hydrolysis was stopped with the addition of 10 µl 10% acetic acid, 1/10 volume of 3M sodium acetate pH5.2 and 2 volumes of 100% ethanol. The RNA was precipitated for 2 hours at -20°C, pelleted, washed in 70% DEPC-treated water, air-dried, resuspended in 50% deionized formamide and stored at -70°C. The incorporation success of the Dig-UTP labelling was estimated by spotting 1 µl of the newly labelled probe and 1/100 dilution of the DIG-labelled control RNA (Roche Scientific) onto a nylon filter. The RNAs were UV crosslinked at 0.4 Jcm^{-2}. The detection steps were then carried out. Thus, the filter was wetted with buffer 1 (100mM Tris-HCl pH7.5, 150mM NaCl), incubated with gentle shaking in buffer 2 (1x Blocking agent [Roche Scientific] in buffer 1) for 30 minutes, washed briefly with buffer 1, and then incubated in buffer 1 with 1 µl of anti-Dig antibody added. The filter was washed twice in buffer 1 for 15 minutes to remove unbound antibody and then briefly washed in buffer 5 (100mM Tris-HCl pH9.5, 100mM NaCl, 50mM $MgCl_2$) to raise the pH. Finally the filter was immersed in the Western Blue substrate (Promega) and put on gentle shaking on a rocking platform for 15 to 30 minutes until blue/purple spots became visible. The relative intensities of the RNA probe spot and the control spots gave an estimate of the success of the transcription reaction.

2.4.4 Hybridisation

i. Section pre-treatment

Twenty four slides containing sectioned tissues were loaded into a baked stainless steel rack. Slides were treated with a series of solutions as shown in table 2.3. All the steps were carried out in RNase-free tubs and 300 ml of solution were used in each step (600 ml were required for the acetic anhydride treatment). The histoclear and the first ethanol steps were performed in a glass container. The rest of the procedure was carried out in plastic tubs. Due to the toxicity and irritant nature of the chemicals involved the formaldehyde, acetic anhydride and histoclear steps were performed in a fume hood. Since acetic anhydride is unstable in water, fresh acetic anhydride was added to the solution with vigorous mixing using a magnetic stirrer with the slide rack elevated above the stirrer on an empty rack. Apart from the Proteinase K digestion which was carried out at 37°C, all the treatments were performed at room temperature.

ii. Hybridisation

Prior to hybridisation, the hybridisation mix was made up. For 5 pairs of slides, the final mix comprised of 100 µl 10x *in situ* salt (3M NaCl, 100mM Tris-HCl pH8, 100mM $NaPO_4$ pH6.8, 50mM EDTA), 400 µl deionized formamide, 200 µl 50% dextran sulfate, 20 µl 50 x Denhardts solution (1% polyvinylpyrrolidone, 1% Ficoll, 1% BSA fraction V), 10 µl tRNA (100 mg/ml) and 70 µl of DEPC-treated water. An excess amount of the hybridisation mix was made up due to its viscous nature and subsequent loss during pipetting. The prepared mix was kept in a 70°C heating block to reduce its viscosity. Pre-treated slides were dried completely of ethanol on clean paper towels. At the same time the digoxygenin-labelled RNA probe was thawed on ice and 2 µl of the probe was then diluted in 19 µl of DEPC-treated water. The diluted probe was added to 20 µl of formamide, denatured for 2 minutes at 80°C and then added to the hybridisation mix. The probe mix was then applied onto the slides. Two slides were sandwiched together carefully to allow the probe mix to spread evenly throughout the slide surface. The slide pairs were placed on top of glass rods in a plastic tray saturated with 50% formamide. The tray was sealed carefully and incubated at 55°C over night with no shaking.

Table 2.3 Pre-hybridisation treatment

	TREATMENT	TIME
1	Histoclear	10 minutes
2	Histoclear	10 minutes
3	100% Ethanol	2 minutes
4	100% Ethanol	2 minutes
5	95% Ethanol	2 minutes
6	90% Ethanol	2 minutes
7	80% Ethanol	2 minutes
8	60% Ethanol	2 minutes
9	30% Ethanol	2 minutes
10	DEPC-treated water	2 minutes
11	2x SSC	20 minutes
12	Proteinase K (1ug/ml in 100mM Tris-Hcl pH8.0, 50mM EDTA)	30 minutes
13	Glycine (2mg/ml in PBS)	2 minutes

14	PBS	2 minutes
15	PBS	2 minutes
16	Acetic anhydride (0.5% v/v in 0.1M triethanolamine, pH8.0)	10 minutes
17	PBS	2 minutes
18	PBS	2minutes
19	Sections rehydrated through the ethanol series (repeat step 9 to 3)	30 seconds each

iii. Post Hybridisation washes

After overnight hybridisation, the slide pairs were separated by rinsing in prewarmed 0.2 x SSC Solution then placed in a clean rack. The rack was put through a series of washes with gentle agitation as shown in table 2.4 below

Table 2.4 Post-hybridisation treatment

	TREATMENT	TEMPERATURE	TIME
1	0.2 x SSC	Room temperature	1 hour
2	0.2 x SSC	Room temperature	1 hour
3	NTE	37°C	5 minutes
4	NTE	37°C	5 minutes
5	RNase (20 ug/ml in NTE)	37°C	30 minutes
6	NTE	37°C	5 minutes
7	NTE	37°C	5 minutes
8	0.2 x SSC	55°C	1 hour
9	PBS	Room temperature	45 minutes

iv. Detection of bound digoxygenin labelled probe

The slides were washed in 1% Roche Block (in buffer 1). After 45 minutes the block solution was replaced by buffer 3 (1% BSA in 100mM Tris-HCl pH7.5, 150mM NaCl, 0.3% Triton-X-100) and the slides washed for another 45 minutes. Following the washes, the anti-

digoxygenin –AP antibody (Roche Scientific) was diluted 1:1250 in buffer 3. The slides were sandwiched together again and the antibody solution was drawn up by capillary action. The solution was drained off and fresh solution was reapplied. The slide pairs were then placed on glass rods in a plastic tray and incubated at room temperature for 2 hours. Later on the antibody was drained off the slides and the slides separated, returned to the rack and washed 4 times for 15 minutes each in buffer 1 with gentle shaking. This was followed by a 10 minute wash with buffer 5. Each slide was then rinsed briefly in this solution to remove all trace of the detergent. Slides were re-sandwiched and capillary action was used to draw Western Blue substrate up. The first application was drained on a clean paper towel and the solution reapplied. The slide pairs were elevated on glass rods in a sealed container and placed in total darkness for 2-3 days. Subsequently the slides were separated, placed back in the rack and immersed in TE buffer pH8.0 for 15 minutes to stop the colour reaction. Slides were then incubated in 0.1% (w/v) Calcafluor White (Sigma) for 10 minutes to counter stain the cell walls and rinsed briefly in water and air-dried overnight in a fume hood. The next day, each slide was mounted with Entellan (BDH-Merck). The calcafluor-stained sections were visualised and photographed using a Polyvar microscope (Riechet-Jung) under UV illumination.

2.5 Yeast Two-Hybrid Interaction Assay

2.5.1 Yeast Transformation

A 10 ml overnight culture was grown in YPD media (1% w/v yeast extract, 2% w/v bacto peptone, 2% w/v glucose) at 30°C. Cells were counted the next day and added to 100 ml fresh YPD media such that a cell concentration of 2 x 10^7 cells/ml was obtained. On day three, cells were spun for 5 minutes at 3000 x *g* , washed in 25 ml dH_2O and recentrifuged. Cells were then resuspended in 1 ml 100mM lithium acetate, aliquoted into eppendorf tubes and spun again briefly. After lithium acetate was removed, cells were resuspended in fresh 100mM lithium acetate so that the final cell density of 2 x 10^9 cells/ml was obtained and 50 µl cell aliquots were made in eppendorf tubes. To transform the yeast cells, 50 µl cell aliquots were centrifuged at top speed for 15 seconds and lithium acetate removed. This was followed by the addition of 240 µl 50% PEG, 36 µl 1M lithium acetate, 25 µl 2 mg/ml boiled salmon sperm DNA and 1 µg of DNA. The sample was vortexed well for 1 minute, incubated for 30 minutes at 30°C, 25 minutes at 42°C and spun for 15 seconds at 8000 rpm in a microcentrifuge. Cells were resuspended in 200 µl dH_2O then plated on freshly made YMM-L-W plates [2% w/v glucose, 0.67% yeast nitrogen base without amino acids, 2% w/v agar (Difco), 39.84 mg/l adenine sulphate, 19.92 mg/l L-arginine-HCl, 99.60 mg/l L-aspartic acid, 99.60 mg/l L-glutamic acid, 19.92 mg/l L-histidine-HCl, 29.88 mg/l L-isoleucine, 29.88 mg/l L-lysine-HCl, 19.92 mg/l L-methionine, 49.80 mg/l L-phenylalanine, 373.50 mg/l L-serine, 199.20 ms/l L-threonine, 30 mg/l L-tyrosine, 19.92 mg/l uracil and 149.40 mg/l L-valine] (all from Sigma) and incubated at 30°C for 3 days.

2.5.2 Yeast X-Gal Colony Filter Assay for β-galactosidase Activity

Six transformed yeast colonies were individually picked and streaked onto a fresh YMM-L-W plate and grown at 30°C. After 3 days colonies were replicated onto Nylon-backed nitrocellulose filters (Amersham). After complete transfer, the filter was submerged with the colony side down for 10 seconds in liquid nitrogen then placed with the colony side up in a petridish containing Whatman filter paper disc saturated with 1.8 ml Z-buffer (60mM Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgSO_4$, 1mg/ml X-Gal and β-mercapthoethanol) and the reaction incubated at 37°C for 1 to 6 hours.

2.6 *In vitro* Mutagenesis

Site-directed mutagenesis was performed according to the method as described in Kunkel, et al., 1987.

2.6.1 DNA preparation from BAC clone T10M13

BAC clones from agar stab stock were streaked out on LB plate containing chloramphenicol (12.5 μg/ml) and grown overnight at 37°C. A single colony was inoculated in 100 ml of L-Broth containing chloramphenicol and grown for 16 hours at 37°C. The culture was spun for 15 minutes and the cell pellet was resuspended in 0.2 ml resuspension buffer (50mM glucose, 10mM EDTA, 25mM Tris-HCl pH8.0, 5 mg/ml lysozyme). The sample was put on ice immediately and after 5 minute incubation, 0.4 ml 0.2M NaOH/1% SDS solution was added to the tube and the content gently mixed and incubated on ice for 5 minutes. Subsequently 0.5 ml 3M potassium acetate was added to the tube and the contents gently mixed then frozen at -80°C for 15minutes and spun at full speed for 15 minutes. About 0.75 ml of the supernatant was removed and the sample was transferred to a fresh microtube and 0.45 ml of isopropanol was added. The tube was then frozen at -80°C for 15 minutes then spun for 5 minutes. The DNA pellet was rinsed with ice-cold 70% ethanol, air-dried for 15 minutes and dissolved in 200 μl TE pH8.0 at 4°C overnight. The DNA concentration was determined by spectrophotometry, and 10 μg DNA was digested with the enzyme *Eco*RI. The reaction was run on an agarose gel and the 8.7 kb DNA fragment containing the *CLK1* gene was excised from the gel using a clean razor blade. The gel fragment was mixed with 300 μl phenol and frozen at -20°C for 2 hours. The sample was then spun for 10 minutes and the resulting supernatant was mixed with 1 volume of phenol/chloroform mix. The content was vortexed vigorously, spun for 5 minutes and the supernatant transferred to a new tube, mixed with 1 volume chloroform, vortexed vigorously and spun again for 5 minutes. DNA was precipitated by adding 0.6 volume isopropanol, 0.1 volume of 3M sodium acetate pH5.2 and 1 μl 10 mg/ml glycogen. The tube was left at -20°C for 1 hours then spun for 20 minutes. The DNA pellet was rinsed with 70% ethanol, air-dried and let to dissolve in 50 μl TE pH8.0.

2.6.2 Ligation Reaction

A 10 µl ligation reaction was set up containing 6 µl of the gel purified 8.7 kb *Eco*RI fragment of the BAC clone T10M13, 1 µl phosphatase-treated *Eco*RI cut pBluescript SK$^-$ (Stratagene), 1 µl 10 x ligation buffer, 1µl ligase (Roche Scientific) and 1 µl dH_2O. The reaction was incubated at 14°C overnight. The next day the electrocompetent cells of *E.coli* strain DH5α were transformed with 2 µl of the ligation reaction. Transformants were selected on LB + ampicillin plates. Plasmid DNA was extracted and in order to identify the orientation of the *CLK1* insert with respect to the m13 replication origin in the pBluescript SK$^-$ vector, a double digest was carried out on plasmid DNA using the enzyme *Eco*RI and *Sal*I which cuts internally in the 8.7 kb fragment.

2.6.3 Preparation of single-stranded DNA for mutagenesis

The bacteria *E.coli* strain CJ263 (*dut*$^-$ *ung*$^-$ F') was transformed with plasmid DNA to be mutagenised. The *dut*$^-$ mutation leads to a deficiency in dUTPase and hence the intracellular pool of dUTP is vastly increased resulting in dUTP being incorporated in to some of the DNA sites instead of thymine. These uracil residues cannot be removed since the cells carry the *ung*$^-$ mutation which results in the inability to synthesise the enzyme uracil-*N*-glycosylase which removes incorporated uracil residues. When a bacteriophage M13 recombinant is grown in the CJ263 strain, uracil-containing single-stranded DNA is produced and subsequently used as the template in the mutagenesis procedure. This results in the generation of heteroduplex molecule with uracil in the template strand and thymine in the newly synthesised strand. Subsequent transformation of an *ung*$^+$ strain with this DNA results in the destruction of the uracil-filled template strand. The DNA of the majority of the progeny bacteriophages therefore is uracil-free and would carry the desired mutation primed by the mutagenic primer. Cells were spread onto LB plates containing ampicillin and chloramphenicol to selects for plasmid DNA and the F' episome which is required for infection with the M13 phage. Plates were incubated at 37°C overnight and double resistant colonies were picked the next day. Plasmid DNA was extracted and digested with *Eco*RI to reconfirm the presence of the insert DNA. Single colonies were picked and individually inoculated into 20 ml 2x YT medium (1 litre contains 16g bactotryptone, 10g yeast extract and 5g NaCl) containing ampicillin in a 250 ml flask. The cells were incubated at 37°c for 30 minutes. Subsequently 10 µl of VCMSM13 phage stock (2 x 10^{12} pfu/ml) was added. After 2

hour incubation kanamycin was added to select for infected cells and the culture was grown overnight. The next day the culture was transferred into a 30 ml Corex tube and spun in a HB4 swingout rotor at 8k for 10 minutes. 10 ml of the supernatant was taken and transferred to a fresh Corex tube and 2.5 ml 20% PEG/2.5 M NaCl were added to precipitate the phage particles and the reaction incubated for 15 minutes at room temperature. The tube was then spun for 5 minutes and the supernatant removed and the pellet resuspended in 0.4 ml 2 x YT medium. The solution was transferred to a fresh eppendorf and 0.1 ml 20% PEG/2.5 M NaCl was added to the tube and the content mixed thoroughly by gentle inversion. The tube was the spun for 5 minutes, the supernatant removed and the tube respun briefly to get rid of the remaining supernatant. The phage pellet was resuspended in 500 ml 0.3 M sodium acetate/1mM EDTA solution. The DNA was then extracted with 500 ml phenol and chloroform mixture. 400 ml of the supernatant was transferred into a fresh tube and 1 ml of ethanol was added and the tube incubated at -20°C. After 1 hour incubation the tube was spun for 10 minutes, the supernatant removed and the DNA pellet washed with 70% ethanol, air-dried and resuspended in 50 ml TE.

2.6.4 Kinasing the mutagenic primers

The kinase enzyme adds the 5'-phosphate group to the mutagenic primers such that they can be ligated to the 3' terminus of the newly synthesised DNA strand. A reaction was set up in an eppendorf tube containing 10 µl 10µM mutagenic primer, 2 µl 10 x kinase buffer (500 mM Tris-Hcl, pH8.0, 100 mM $MgCl_2$, 1 µl 100 mM DTT, 4 µl 5 mM ATP, 2 µl dH_2O, 1 µl T4 polynucelotide kinase [5 units/µl]). The tube was incubated at 37°C for 30 minutes and then stored at -20°C for future use.

2.6.5 Annealing reaction

A 20 µl annealing reaction was set up containing 5 µl of the kinased mutagenic primer, 0.5 µg single-stranded DNA and 2.5 µl Fritz standard KCl/Tris annealing buffer. The tube was incubated at 75°C for 5 minutes then 55°C for 30 minutes, 37°C for 15 minutes and finally at room temperature for 15 minutes.

2.6.6 DNA synthesis

After the annealing step, 2.5 µl 10 x T4 synthesis buffer, 1 µl 10^{-1} dilution of T4 ligase, 1 µl T4 DNA polymease and 0.5 µl T4 gene 32 protein (all from Roche Scientific) were added to the tube. The reaction was put on ice for 5 minutes then placed at room temperature for 5 minutes and finally incubated at 37°C overnight. The competent *E.coli* cells of the dut^+ ung^+ DH5α strain were transformed using 2 µl from the DNA synthesis reaction and the cells plated onto LB+ampicillin plates. After an overnight incubation at 37°C, positive colonies were picked from the plates and the plasmid DNA extracted. Clones that contained the desired mutation were identified by sequencing using YC18 primer, and YC18 and YC19 primer for the inducible construct.

2.7 Gene Transfer into *Arabidopsis*

2.7.1 *Agrobacterium* transformation

i. Cell Preparation

Agrobacterium tumefaciens strain GV3101 mp90 was grown in LB medium with appropriate antibiotic at 28°C until an OD_{680} of 0.8 was reached. The culture was chilled on ice and centrifuged briefly at 1500 rpm. The cell suspension was spun at 4000 x *g* for 6 minutes at 4°C. Cells were resuspended in ice-cold 20mM $CaCl_2$ (1 ml per 100 ml of starting culture) and 200 µl aliquots were dispensed into chilled sterile eppendorf tubes.

ii. Transformation

About 5 µg of plasmid DNA was added to the cells in an eppendorf tube. The cells were then snap-frozen in liquid nitrogen and subsequently thawed in a 37°C water bath for 5 minutes. To allow the bacteria to express the antibiotic resistance genes, 1 ml LB medium was added to the tube and the content mixed by gentle pipetting followed by inversion. The sample was incubated for 2-4 hours at 30°C, spun for 30 seconds at 13000 rpm in a bench top centrifuge and the cells resuspended in 100 ml LB medium. Finally the cells were spread onto selection plates containing rifampicin (100 µg/ml), gentamycin (50 µg/ml) and kanamycin (50 µg/ml). The plates were incubated at 30°C for 3 days. To identify clones containing the desired recombinant plasmid, colonies were picked individually and inoculated in 5 ml LB medium containing appropriate antibiotic. Plasmid DNA was extracted as previously described and digested with the appropriate restriction enzymes to detect the insert.

2.7.2 *Arabidopsis* Transformation

i. Plant growth

Four to five plants in Ws or *clf-50*/+ background were grown on soil in small pots. After the plants had bolted, the primary bolts were clipped from each plant to encourage the growth of secondary bolts. Infiltration was carried out approximately 5 days after clipping.

ii. Transformation by vacuum infiltration

Two days prior to infiltration, a 25 ml overnight culture was grown in LB medium containing rifampicin, gentamycin and kanamycin and 5 ml of this culture was used to inoculate a 400

ml culture the next day. The culture was grown to an OD_{260} of 1.2. Cells were harvested and resuspended in infiltration medium (0.5 x MS salts, 1 x B5 vitamins, 5% sucrose, 0.44 μM benzylamino purine and 0.03% Silwet L-77) to an OD_{260} of 0.8. Plant pots were inverted, carefully immersed into the infiltration mix. Vacuum infiltration was then carried out in a dessicator jar for 20 minutes at 10-15 in^3Hg. After the vacuum was released, pots were laid on their side in a mouse tray and then covered with Saran wrap. The cover was removed the next day and pots were put in an upright position and the plants were grown in a growth room with constant light. Seeds were collected when siliques were ripe and sown on GM + kanamycin plates to select for transformants.

iii. Transformation by floral dipping

A 5 ml overnight *Agrobacterium* culture grown in LB medium containing rifampicin, gentamycin and kanamycin was used to inoculate a 500 ml LB + gentamycin and kanamycin medium. After overnight incubation at 30°C with constant shaking the cells were spun and resuspended to the OD_{260} of 0.8 in 5% sucrose solution containing 0.05% Silwet L-77. Plants were dipped in the *Agrobacterium* solution with the above ground parts totally immersed for 3 seconds with gentle swirling. Dipped plants were covered for 24 hours and left in a growth cabinet. After 5 days the plants were re-dipped. Seeds were collected when dried and transformants were selected on GM + kanamycin plates.

2.8 Histochemical staining of the GUS reporter lines

Plant materials were fixed in 90% ice-cold acetone. After 20 minutes of incubation, acetone was removed and tissues were washed briefly with the rinse solution (50mM $NaPO_4$ pH7.2, 0.5mM $K_3Fe(CN)_6$, 0.5mM $K_4Fe(CN)_6$). The staining solution (50mM $NaPO_4$ pH7.2, 0.5mM $K_3Fe(CN)_6$, 0.5mM $K_4Fe(CN)_6$, 2mM X-Gluc in dimethylformamide, 0.05% Triton-X-100) was then added. The tissues were vacuum-infiltrated once and then transferred to a 37°C oven for an overnight incubation. The next day chlorophyll and other pigments were removed by washing the tissues several times with 70% ethanol.

Chapter 3. Characterisation of a Novel Leaf Curling Mutant, *moe leaf*, in *Arabidopsis thaliana*

3.1 Introduction

3.2 *moe leaf* Phenotype

3.3 Double and Triple Mutant Analysis

3.4 Analysis of Floral Homeotic gene Expression in *moe* Mutant

3.5 Mapping the *moe* Mutation

3.6 *MOE LEAF* Cloning

3.7 The Molecular Basis of the *moe* Allele

3. Characterisation of a novel leaf curling mutant, *moe leaf*, in *Arabidopsis thaliana*

3.1 Introduction

The availability of complete *Arabidopsis* genome sequence and the ever growing amount of sequence data from chromosomal DNA and expressed sequence tags have contributed enormously to a wealth of information on plant genetics. Recently an exhaustive functional genomic effort has been made in order to apply reverse genetic tools to manipulate the genome sequence of *Arabidopsis* with the main priority to assign functions to a large number of newly isolated genes to which no phenotype has been associated. As a consequence, the traditional forward genetics which aim to identify genes responsible for mutant phenotypes of interest has become somewhat less popular with number of geneticists. This, however, is rather unfortunate since forward genetics could benefit as much as reverse genetics from the recent avalanche of sequence information. The availability of the complete sequence of *Arabidopsis* means that many novel genetic and molecular markers have been identified and this results in a vast improvement in the efficiency of linkage analysis and map-based cloning. The advance in mapping technology allows systematic gene function search to be carried out with great speed and precision. For example, several developmental networks could be dissected and reconstructed using second site suppressor/enhancer screens. The information obtained would help providing us with better understanding of the organisation and regulation of the genetic hierarchies in those networks of gene functions.

In this study attempts were made to identify and characterise mutants that shows phenotypic similarities to *clf* plants. It has usually been the case that genes that confer similar mutant phenotypes often act in the same genetic pathway. Thus those genes are more than often closely related in function and either encode proteins that share conserved motifs that are essential for their function or proteins that, although not structurally related, perform similar functions.

By searching for mutants that resemble *clf* but define new loci and characterising them using genetic and molecular approaches such as epistasis analysis of various double and triple mutant combinations with potential partners and targets including the *CLF* and floral

homeotic genes, more insights into the nature and function of those genes and the proteins that they encode. These could potentially be additional members of plant Pc-G genes or those that act closely or in the same regulatory network as *CLF* could be obtained. This in turn would lead to a better understanding of the function of *CLF* and other genes in the genetic hierarchy that helps maintaining homeotic gene expression status during plant development.

3.2 *moe leaf* phenotype

A novel mutant that confers a phenotype similar to that of *clf* plants was identified by Mark Running during a T-DNA mutagenesis experiment. Genetic analysis showed that this mutation, *moe leaf*, is recessive and not allelic to the *clf* mutation and hence defines a second locus that is distinct to *CLF* (J Goodrich pers. comm.).

As in the *clf* mutants, *moe* plants confer pleiotropic developmental defects. Firstly, leaf morphology is severely affected. Detailed phenotypic analysis revealed that the leaves were small and curled upwards along the longitudinal axis of the leaf blade. The degree of curling was much more pronounced than that seen in *clf* leaves and increased acropetally. When compared with wild-type and *clf* plants, both the rosette and cauline leaves of *moe* mutants were much smaller (Figure 3.1A-D). In addition to the leaf morphology defects, flowering time was also altered in the *moe* background. Thus, the *moe* plants bolted at least 1 week earlier than the wild-type under long-day conditions (Figure 3.1E). The inflorescences produced were noticeably smaller than those in wild-type plants as a result of shorter stem internode elongation.

The *moe* mutation also affected the development of flowers in a similar manner to the *clf* alleles. Thus flowers were generally smaller and open prematurely due to the failure of the small petals to enclose the floral buds (Figure3.1G). Other organs were also affected with petals being smaller and narrower while the stamen filaments are thinner with small anthers which produced lower amount of pollen than wild-type flowers, resulting in the *moe* plant being semi-sterile. The carpels were occasionally bent but otherwise the morphology was generally not affected. The number of floral organs was also affected in *moe* plants. Thus although the average number of sepals and carpels appeared to be normal, *moe* plants

showed a general reduction in the number of petals and in particular stamens in the second and third whorls respectively.

Several *moe* flowers also exhibited homeotic conversions in the organs of the two outer whorls as seen in *clf* flowers. In those flowers, first whorl sepals showed several carpelloid features such as the appearance of stigmatic papillae at the tips and ovules along the margins (Figure 3.1H-I). Petals, on the other hand, had staminoid features with patches of yellow tissues characteristic of anthers. In addition, the shape of some of the petals appeared to be intermediate between that of normal petals and stamen filaments (Figure 3.1I). In extreme cases sepal and petal numbers were reduced and in some cases petals were totally absent. Closer investigation showed that the severity of the phenotypes seen increased from the base to the apex of the inflorescence (Figure 3.1J). Generally, no homeotic transformations were seen in the third and fourth whorl organs.

The occasional carpelloidy and staminoidy seen in *moe* flowers are suggestive of ectopic *AG* expression in the first and second whorls of *moe* flowers. This in turn suggested that in wild-type flowers, *MOE* activity is required to regulate *AG* expression in the two outermost whorls. However, this aspect of *MOE* function is likely to be redundant with the activities of other proteins since most of the time the *moe* flowers appeared relatively normal with respect to organ identity and the transformations of organs seen were rather mild. Only flowers at the more apical position and those on plants grown in adverse environments showed the homeotic conversions described above. This view is supported by the fact that the *AG* regulatory network is highly redundant consisting of proteins with overlapping functions including *AP2*, *LUG* and *CLF* amongst others. Thus, the effect of the loss of *MOE* function on the specification of organ identity in the two outermost whorls may be somewhat slight since any potential disturbance to the network could be counteracted by the activities of the other *AG* repressors.

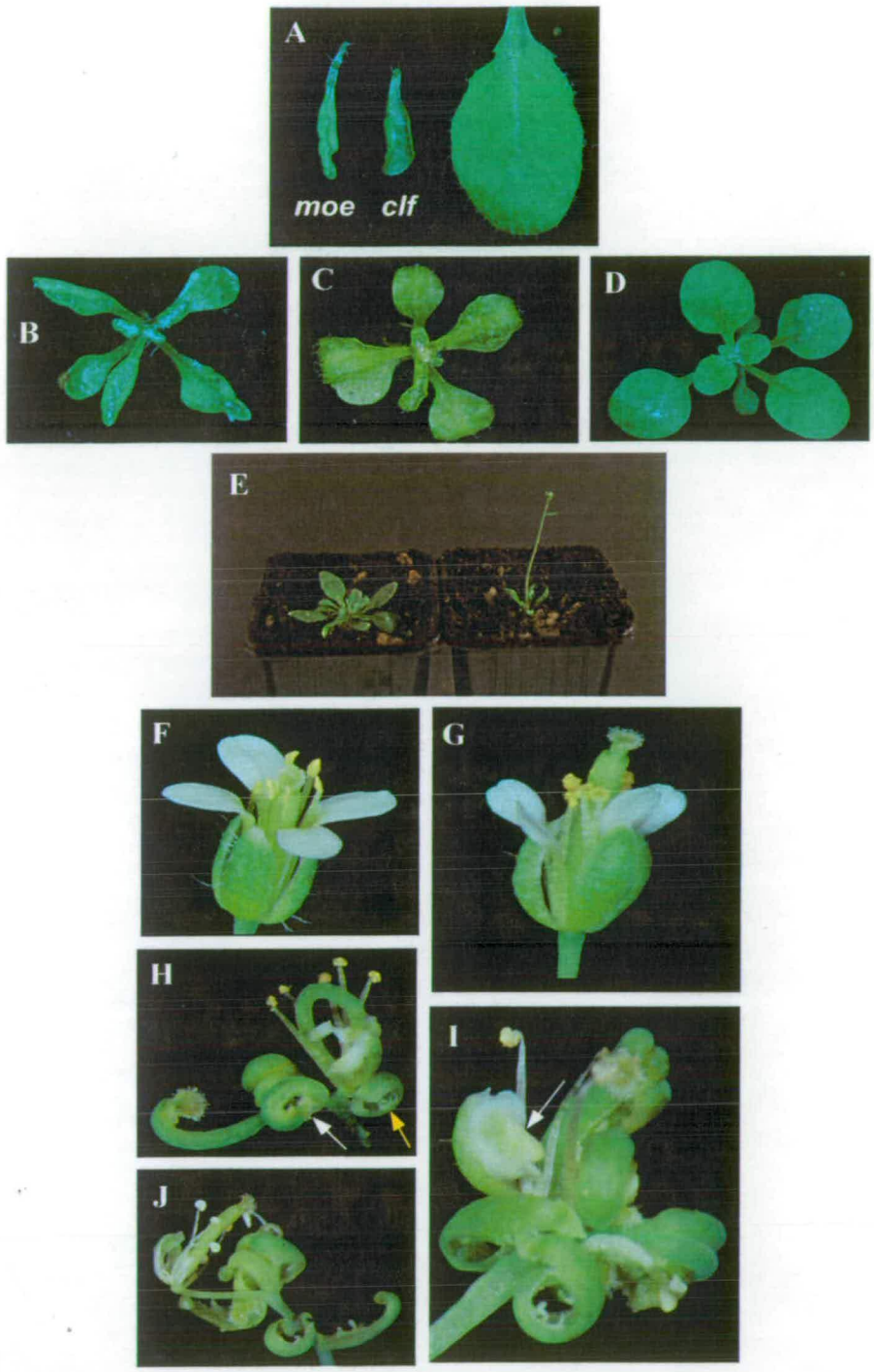


Figure 3.1 *moe* phenotypic defects. A, B, C and D, *moe leaf* mutant shows severe leaf curling (A left, B) compared to *curly leaf* (A middle, C) and wild-type plants (A right, D). E, *moe* plant (right) exhibits an early flowering phenotype. *moe* plant bolts and form inflorescence earlier than the wild-type sibling (left). F, A wild-type flower showing a normal arrangement of floral organs with four sepals, four petals, six stamens and two fused carpels from whorl 1 to 4 respectively. G, A prematurely opened *moe* flower with small and narrower sepals and petals in whorl 1 and 2 respectively. H, A *moe* flower showing strong homeotic conversion in the first whorls with carpelloid organs formed in the first whorl. The white arrow indicates stigmatic papillae while the yellow arrow indicates ovule structure characteristic of carpels. I, A *moe* flower showing staminoidy in the second whorl with staminoid petals formed (arrow). J, The frequency of *moe* floral homeotic transformations increased acropetally with the apex of the inflorescences often most highly affected.

3.3 Double and triple mutant analysis

Analysis of *moe leaf* mutant revealed striking similarities with *clf* plants at the phenotypic level indicating that the wild-type *MOE* gene might act as a repressor of floral homeotic genes. In order to gain a better view on possible genetic interactions between *MOE* and the floral homeotic genes, various double mutant combinations were produced and their phenotypes analysed and compared with the defects seen in the single mutant parents.

3.3.1 *moe agamous*

The leaf curling phenotype in *clf* mutants has been shown to be a consequence of the misexpression of the gene *AG* in the leaves. The alleviation of leaf curling in the *clf ag* double mutant suggests that *CLF* acts as a negative regulator of *AG* in the leaves. In order to determine if the leaf curling defect in *moe* plants shared similar molecular basis with that of *clf* mutants, heterozygous *moe* plants were crossed to female *ag-3* heterozygotes. The F1 progeny were allowed to self and families that segregated both *moe* and *ag* mutations identified. In the F2, the putative *moe ag* double mutant was observed at a 1/16 frequency. To confirm that these were double mutants, seed was collected from the F2 *moe* plants. When sown out, 2/3 of the F3 progeny segregated the *moe ag* double mutant at 1/4 frequency, and the double mutants were recovered and analysed in detail in the F2 generation.

As expected the *moe ag* double mutants showed suppression of the *moe* leaf phenotype. Thus the plants were more wild-type looking with much less leaf curling and the flowering time was intermediate between that of the *moe* and wild-type plants (Figure 3.2A-C). This is consistent with the notion that, like *CLF*, the wild-type *MOE* gene acts as a negative regulator of *AG* function. The fact that the phenotypic defects of *moe* plants can be partially rescued when *AG* function is eliminated indicates that the ectopic *AG* expression is the main contributor to the leaf curling defect in *moe* plants. However, the fact that the suppression of the leaf curling phenotype was incomplete in the *ag* background suggested that other targets of *MOE* also influencing leaf development and flowering time were still mis-regulated. Complete epistasis was, nevertheless, observed with respect to flower development. Thus *moe ag* double mutant produced *ag*-like flowers with flowers developed within flowers with a repeated pattern of floral organs of sepals-petals-petals from the outermost to the innermost

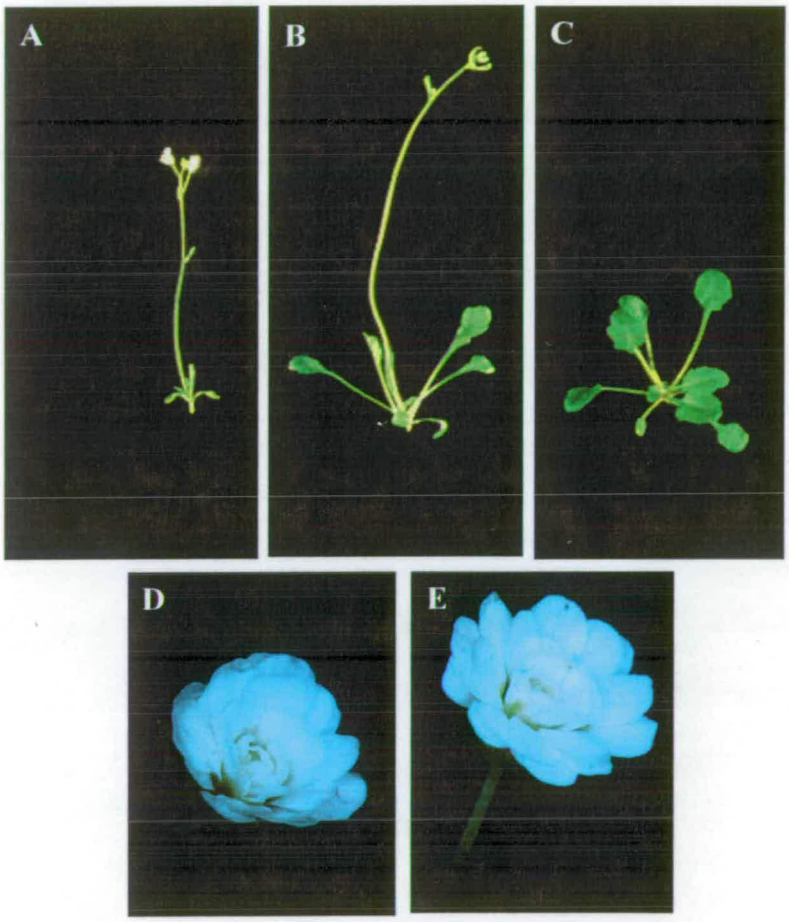


Figure 3.2 Alleviation of *moe leaf* phenotype in *agamous* background. A double *moe ag-3* mutant (B) shows alleviation of the *moe* phenotype (A) with mild leaf curling and flowering time that is intermediate between that of *moe* and wild-type (C) plants. The *ag-3* mutation is completely epistatic to the *moe* mutation in the flowers with the *moe ag-3* flower (D) showing a flower-within-flowers phenotype identical to that seen in the *ag-3* flower (E).

floral whorls (Figure 3.2D-E). This indeterminate pattern of organogenesis reflected the role of wild-type *AG* activity in floral meristem termination which is genetically separable from its function in organ identity specification (Sieburth et al., 1995). The observation that the strong *ag-3* mutation was completely epistatic to the *moe* allele in the flower indicated that *AG* function is essential for the homeotic transformations of organs in the two outer whorls of the *moe* flowers. It also suggested that the wild-type *MOE* function is required only for *AG* repression and not for the specification of sepal and petal identity.

3.3.2 *moe lfy-5*

Mutations in the *LFY* gene cause a partial transformation of flowers into inflorescence shoots. In strong *lfy* mutants such as *lfy-6* and *lfy-7* plants, flowers are partially converted into inflorescence structures. Early arising flowers often develop into inflorescence shoots while the late arising ones show partial inflorescence characteristics such as the arrangement of floral organs that is intermediate between whorled and spiral pattern and homeotic conversions of the outermost whorl floral organs into leaf-like structures and the inner floral whorls into sepaloid and carpelloid structures. In addition, secondary inflorescence shoots are often subtended by leaf-like bracts which occasionally show some carpelloidy (Schultz & Haughn, 1991; Huala & Sussex, 1992; Weigel et al., 1992; Schultz & Haughn, 1993). The transformations of floral structures into inflorescence shoots is much less pronounced in plants homozygous for the weak *lfy-5* allele. The leaf-like bracts subtending the secondary shoots are less frequently observed and the development of floral organs is less perturbed and hence the flowers formed are relatively normal and only infrequently show weak homeotic conversions in the second and third whorl organs (Figure 3.3A-1). The severity of the phenotype is dependent upon the position of the flowers in the inflorescence. The first arising flowers are usually more abnormal than the late arising ones with fewer or no organs in the second and third whorls. Later in development, as seen with all other *lfy* alleles, a mass of carpelloid tissue composing of solitary carpelloid bracts is usually found at the apex of the inflorescence of the *lfy-5* plant (Figure 3.3A-2), (Weigel et al., 1992).

To generate the *moe lfy* double mutant, plants that were heterozygous for the *moe* mutation were used as males and crossed to *lfy-5* female lines. The F2 lines carrying both the *moe* and *lfy* mutations were identified in the same way as in the case of the *moe ag* double. Phenotypic analysis revealed that the defects of the *lfy-5* mutant were dramatically enhanced

in the *moe* background. In general, double mutant plants exhibited more severe transformation of flowers into inflorescence shoots resembling the strong *lfy* mutant phenotype. Floral bracts showing carpelloidy such as the ectopic ovules along the margins were often found subtending secondary inflorescence shoots with the flowers in the axils of the bracts frequently absent (Figure 3.3A-3 & 4). Flowers formed were severely transformed. Petals and stamens were largely absent and first whorl organs were mosaic between leaf-like structures and carpelloid organs. Carpel development was also affected with the number of carpels increased presumably at the expense of the third whorl organs and the carpels formed often did not fuse properly (Figure 3.3A-3).

The drastic phenotypic enhancement seen in the *moe lfy* double mutant implied that the wild-type *MOE* and *LFY* gene interact with each other and that they likely act in the same genetic pathway controlling flower development. It was also suggestive of a possible role of *LFY* as an *AG* repressor.

3.3.3 *moe ap1-1*

The wild-type *AP1* gene has a dual role in the determination of floral meristem identity and also the specification of organ identity in the outer floral whorls. Mutations in the *AP1* gene, therefore, affect both aspects of *Arabidopsis* reproductive development. The strong *ap1-1* allele confers a partial conversion of floral meristem into inflorescence meristem. Thus first whorl floral organs develop as bract-like organs and secondary floral buds often form in their axils (Figure 3.3B-1). This developmental pattern is sometime reiterated in the axillary secondary flowers resulting in the formation of tertiary floral buds in the axils of each of the secondary leaf-like organs (Figure 3.3B-2). In addition to this, second whorl organs often fail to develop in the *ap1-1* background. In contrast, the development of the third and fourth whorl organs are unaffected. Thus the stamen and carpel morphology and position appear normal in *ap1-1* flowers (Irish & Sussex, 1990).

When the *ap1-1* allele was combined with the *moe* mutation synergistic enhancement of the floral defects was observed. First whorl organs were carpelloid with ovules formed along the margins and stigmatic papillae at the tips. Second whorl organs were absent in the *moe ap1-1* flowers. Stamens in the third whorl developed relatively normally, although in some cases

they were absent. Carpels could also exhibit some developmental abnormalities such as an increase in number, shape irregularity and improper fusion (Figure 3.3B3-4).
The phenotypic enhancement seen in the *moe ap1-1* flowers as compared to the *ap1-1* mutant suggested that the wild-type *MOE* and *AP1* gene interact synergistically to regulate flower development. The development of ectopic ovules and stigmatic papillae in the first whorl organs indicated that *MOE* activity is required to repress class C function, *AG*, in the first whorl of *ap1* flowers. It also suggested that *AP2* and *LUG* activity alone are not sufficient to prevent precocious *AG* expression in the outermost whorl and that there is, to some extent, non-overlapping requirements of *MOE* and *AP2* as a repressor of *AG*. Also *AG* was possibly misexpressed in the second whorl but the effect on the phenotype was masked since *AP1* appears to be essential for second whorl development.

3.3.4 *moe ap2-1*

As previously mentioned, the A class gene *AP2* plays an essential role both in the specification of organ identity in the two outermost whorls of *Arabidopsis* flowers and in repressing the C function in whorl 1 and 2. In general the organs in *ap2* mutant flowers develop as carpels-stamens-stamens-carpels from the outermost to the innermost whorls. In strong *ap2* alleles such as *ap2-2*, first whorl organs develop as carpels or carpelloid organs. Development of second and third whorl organs is often abolished although in some cases fertile stamens could be found. Carpels in the fourth whorl are often irregular in shape and fail to fuse properly (Bowman et al., 1989; Bowman et al., 1991).

Plants that are homozygous for the weak temperature sensitive *ap2-1* allele used in this study display homeotic conversions in the outer two floral whorls. Thus first whorl organs develop as leaf-like structures with simple sepal trichomes being replaced with stellate trichomes which are characteristic of leaves or in some cases carpelloid leaves. Second whorl organs are usually present and contain both petal and stamen characteristics (Figure 3.3C-1), (Bowman et al., 1989).

The *moe ap2-1* double mutant was generated by crossing female *moe* heterozygotes with male *ap2-1/ap2-1* plants. The F2 lines carrying both mutations were identified in the same way as with the *moe lfy* double mutants. Detailed analysis of the *moe ap2-1* double mutant revealed that the *moe* mutation enhanced the phenotype of the *ap2-1* allele in the flowers.

Detailed investigation showed that stamen and carpel development in the third and fourth floral whorls was relatively normal. However, first whorl organs were transformed into carpelloid structures with ectopic ovules formed along the margins and stigmatic papillae found at the tip of the organs (Figure 3.3C-2). Second whorl organs were largely absent and when present developed as stamens or stamen-like structures (Figure 3.3C-2). The increase in phenotypic severity of the *ap2-1* allele in the *moe* background indicated that the two wild-type genes interact synergistically and possibly act on a common pathway to repress *AG* activity during flower development.

3.3.5 *moe ap3-3*

The strong *ap3-3* allele affects the identity of floral organs in the second and third whorl. Thus organs in the second whorl acquire sepal identity but develop on the time scale which is characteristic of second whorl organs in wild-type i.e., they form more slowly than first whorl sepals, and are therefore often found enclosed by sepals in the first whorl. Floral organs in the third whorl show variable phenotype. In some cases they are totally absent while in other flowers they could develop into carpels or filamentous structures. The third whorl carpels have stigmatic papillae at their tips as in normal carpels and they either partially or completely fuse with each other and the central gynoecium in the fourth whorl resulting in a large mass of carpelloid organs in the centre of the flower. The occasional filamentous structures seen in the third whorl were similar to stamen filaments and were either free-standing or partially or totally fused to the fourth whorl carpels along their length (Figure 3.3D1-2), (Jack et al., 1992).

The double mutant of *moe* and *ap3-3* allele was created by crossing male *moe/+* plants to *ap3-3/ap3-3* females. Analysis of the floral phenotype revealed that the *moe/ap3-3* double mutant showed additive phenotype of the *moe* and *ap3-3* mutants. Thus petals were converted to small sepals and stamen to carpels or filamentous structures in whorl 2 and 3 respectively (Figure 3.3D-3), while whorl 1 organs either developed as sepals with carpelloid features or carpelloid structures (data not shown).

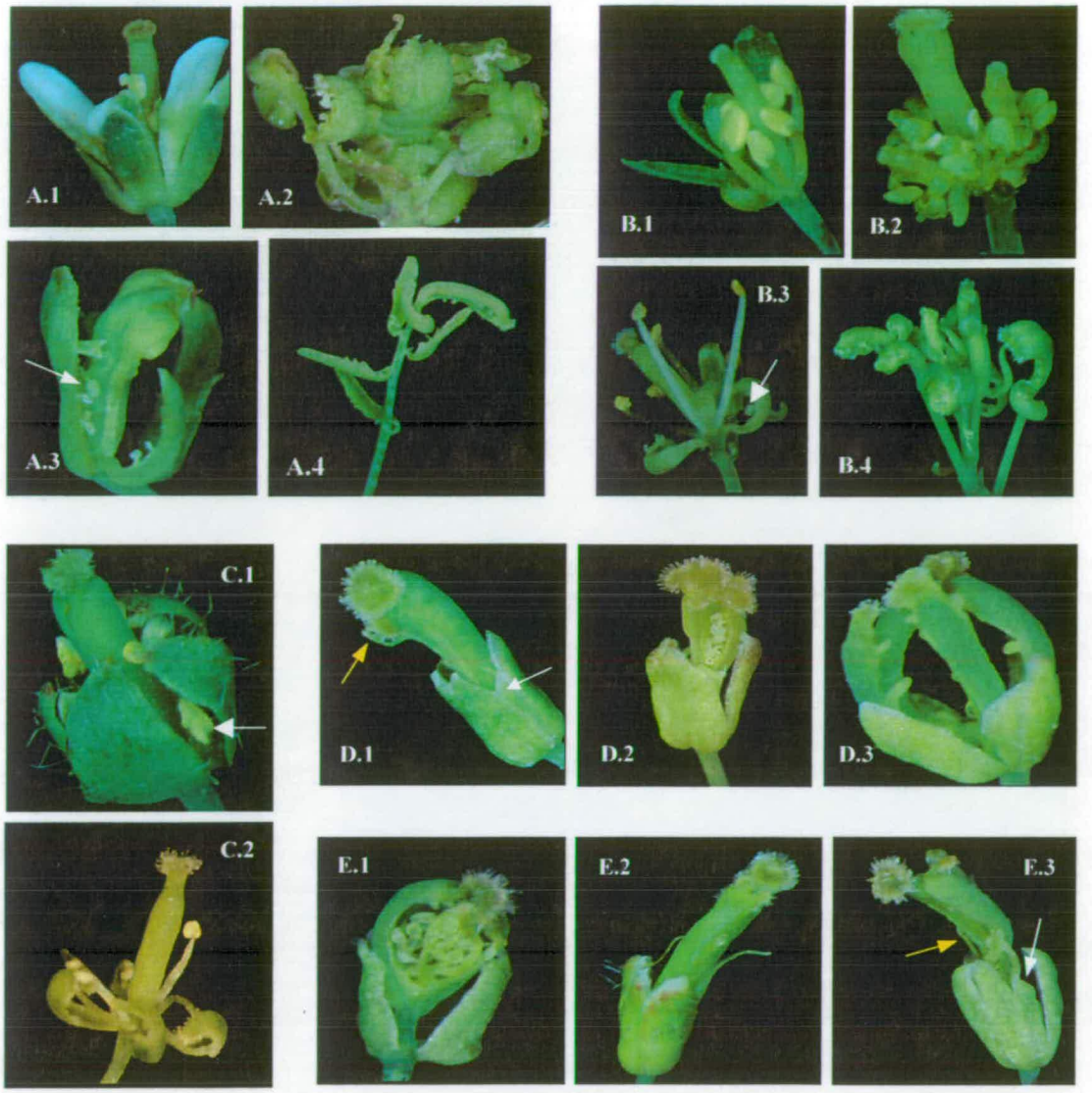


Figure 3.3 Double mutant phenotypes of various *moe* and floral homeotic gene combinations. A.1, A weak *lfy-5* flower comprises of a correct set of floral organ identity in all four whorls. A.2, A mass of carpelloid structures forms at the apex of the inflorescence of *lfy-5* mutant. A.3, A *moe lfy-5* flower showing a strong enhancement of the weak *lfy-5* phenotype with first whorls organs being mosaic between leaf-like structures and carpelloid organs. An arrow indicates ectopic ovules forming along the margin of the organ. A.4, An inflorescence of *moe lfy-5* plant showing floral bracts with ectopic ovules along the margins. B.1, An *ap1-1* flower showing bract-like structures in the first whorls. Petals are absent in the second whorl while stamens and carpels form as normal in the third and fourth whorl respectively. B.2, An *ap1-1* flower showing inflorescence characteristics with a reiterative pattern of floral buds forming in the axils of the leaf-like organs in whorl 1. B.3, A *moe ap1-1* flower showing transformation of first whorl organs into more carpelloid structures. An arrow indicates ectopic ovules forming at the margin of the carpelloid organ. B.4, An apex of *moe ap1-1* inflorescence showing severe floral transformations into carpelloid structures. C.1, An *ap2-1* flower showing leaf-like structures in the first whorl and staminoid petals in the second whorl (arrow). C.2, First whorl organs are transformed into carpelloid structures and petals fail to form in the second whorl of a *moe ap2-1* flower. D.1 and D.2, *ap3-3* flowers showing homeotic transformations in the second and third floral whorls. Second whorl organs develop as small sepals (white arrow, D.1) which are enclosed by first whorl sepals while third whorl organs develop as either carpels, which can be completely (D.1) or partially (D.2) fused to the fourth whorl carpels, or filamentous structures (yellow arrow, D.1). D.3, A *moe ap3-3* flower with transformation of second and third whorl organs into small sepals and carpels respectively. E.1, A *pi1-1* flower showing enlarged carpels at the expense of third whorl stamens. E.2, A *pi-1* flower with some third whorl organs develop as carpels and the others as filamentous structures. E.3, A *moe pi-1* flower with small sepals (white arrow) form in the second whorl instead of petals and filamentous organs (yellow arrow) or carpels in the third whorl instead of stamens.

3.3.6 *moe pi-1*

Plants lacking the class B *PI* function were phenotypically similar to *ap3* mutants with respect to the specification of floral organ identity. Hence, flowers of the strong *pi-1* homozygotes have small sepals developed in places where petals are normally formed and the third whorl organs are transformed into carpels, filaments or carpelloid organs which fuse either partially or completely with the central gynoecium (Figure 3.3E1-2) (Bowman et al., 1989).

The *moe pi-1* double mutant were generated in the same way as previously described for the *moe ap3-3* double mutant. Phenotypically the *moe pi-1* double mutant was indistinguishable from the *moe ap3-3* double. Thus the petal and stamen fate was lost in the second and third whorl respectively. The organs developed in the second whorl were small sepals enclosed by the first whorl sepals. On the other hand, third whorl organs were either carpels which often found incorporated into the central carpels resulting in an enlarged gynoecium or filamentous structures which resembled stamen filaments (Figure 3.3E-3).

The additive effect of the *moe* and the *ap3-3* and *pi-1* mutation suggested that the wild-type *MOE* gene functions independently of the *AP3* and *PI* activity in the regulatory pathway controlling flower development.

3.3.7 *moe clf-2*

Plants carrying the transposon-tagged *clf*-2 allele show strong leaf curling and occasional transformation of organ identity in the outer two whorls (Goodrich et al., 1997). Double mutants for *moe* and the *clf*-2 mutation were obtained in the F2 generation from the cross of male *moe*/+ to female *clf*-2/*clf*-2 plants. The combination of the *moe* and *clf*-2 mutation resulted in plants with very severe phenotypes. Thus the *moe clf*2 double mutant showed much more severe vegetative phenotype than either *clf* or *moe* single mutation. The early developmental pattern of the *moe clf*-2 plants did not show any deviation from the pattern seen in either parent. Therefore, roots, hypocotyl and cotyledons were relatively normal, albeit noticeably smaller than usual. Later on, however, it became apparent that the vegetative growth of the double mutant was severely affected with only a few very small sessile leaves formed and the overall size of the plants was very much smaller than the *moe*

or *clf*-2 parent (Figure 3.4A1-3, B1-3). The double mutants flowered at 10-14 days after sowing which was about 1 week earlier than their parents. Inflorescence shoots formed were very short and cauline leaves were absent. A number of flowers were produced at the top of the inflorescence structures but these were considerably smaller than normal. However, the identity of floral organs in all 4 whorls was largely unaffected although some carpelloidy and staminoidy was occasionally seen in the first and second whorl respectively (data not shown).

3.3.8 *moe clf-9*

Plants homozygous for the weak *clf*-9 allele were phenotypically almost indistinguishable from the wild-type plants. Thus leaf development of the *clf*-9 mutants is normal, although very weak leaf curling was occasionally seen particularly if plants were stressed, for example, by crowding (Figure 3.4C1-3). Reproductive development of the *clf*-9 plant is also unaffected with the flowers produced not showing any homeotic transformation of floral organs (data not shown).

The *moe clf*-9 double mutant was generated in the same way as the *moe clf*-2 plants. Detailed analysis of the double mutant plants revealed pleiotropic effects on the phenotypes like that of the *moe clf*-2 plants. Thus the plants were extremely small and produced only a few sessile leaves before entering reproductive development. The defects were, however, not as severe as those of the *moe clf*-2 double. Thus more leaves were produced in the *moe clf*-9 double mutants and their size was larger compared with those in the *moe clf*-2 mutants. This implied that vegetative growth was allowed for a longer period of time in the *moe clf*-9 background. The discrepancies seen between the two classes of *moe clf* double mutants might reflect the nature of the two *clf* alleles used in this study with *clf*-9 being a much weaker allele than *clf*-2.

3.3.9 *moe clf-2 ag-3*

In order to obtain a further information on the nature of the relationship between the *MOE* and *CLF* gene analysis of the triple mutant combination of *moe*, *clf* and *ag* was carried out. To generate the *moe clf ag* triple mutant female plants homozygous for the *clf*-2 allele and heterozygous for the *ag-3* mutation were pollinated with pollen from *moe*/+ heterozygous

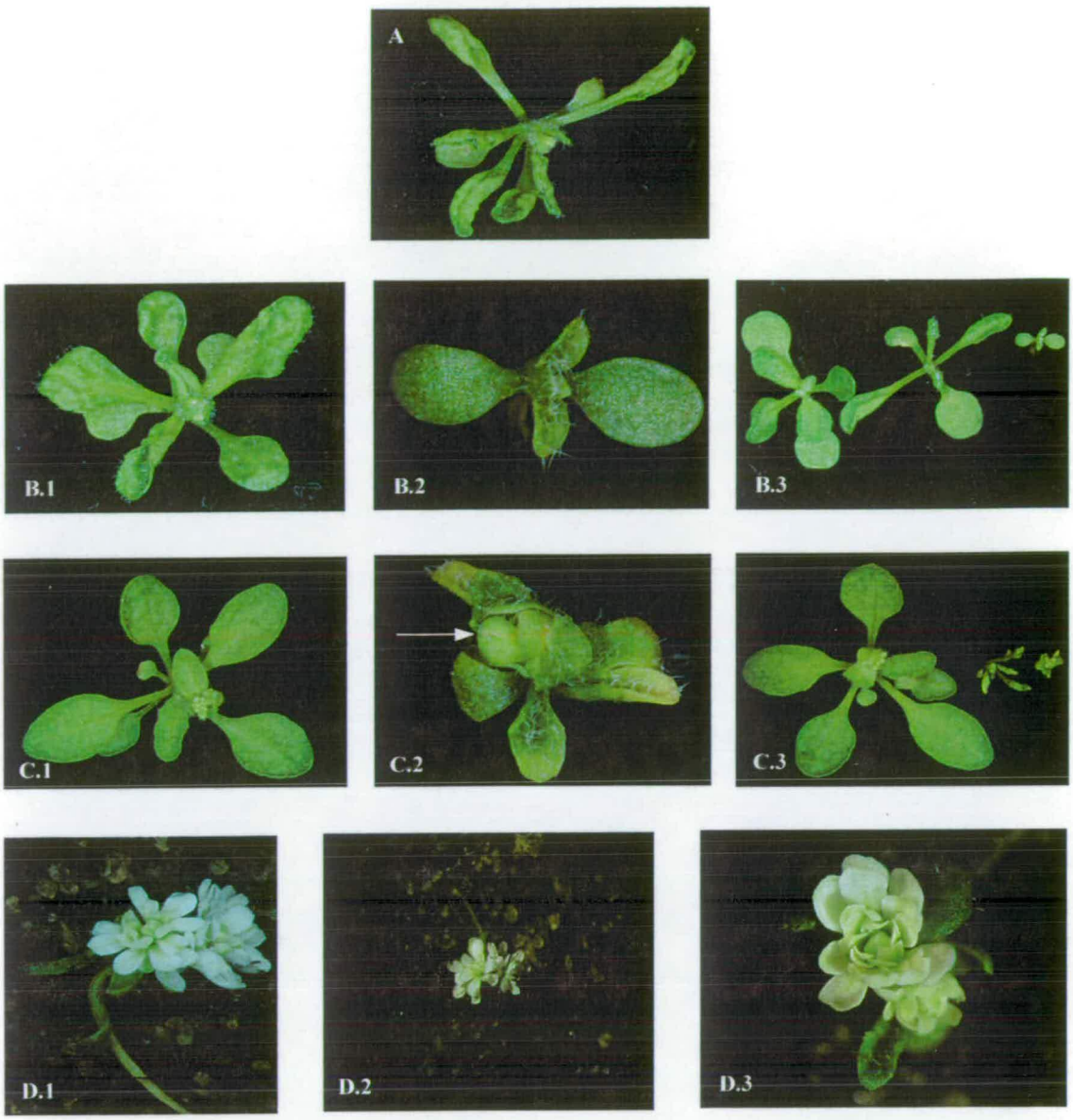


Figure 3.4 Enhancement of *moe* phenotype in *clf* and *clf ag* backgrounds. A, A *moe* mutant plant showing strong leaf curling. B.1, A *clf-2* plant with leaf curling phenotype. B.2, A *moe clf-2* double mutant showing severe developmental defects. The plant is extremely small in size with only a few sessile leaves formed. B.3, Comparison of the *moe clf-2* mutant (right) with its *moe* (middle) and *clf-2* (left) siblings. C.1, A plant homozygous for the weak *clf-9* allele showing near wild-type appearance with very mild leaf curling. C.2, A *moe clf-9* plant exhibiting a strong enhancement of the weak *clf-9* phenotype (C.1). The plant is tiny with only a few sessile leaves formed before entering the reproductive phase. An arrow indicates a floral bud. C.3, A comparison with its *moe* (middle) and *clf-9* (left) sibling highlights the small size of the *moe clf-9* plant (right). D.1 and D.2, A *moe clf-2 ag-3* triple mutant showing severe growth defects with only two or three leaves, some of which are curled, formed and no stem structure is seen indicating a global developmental defect with several target genes misregulated. The small size of the triple mutant is highlighted in D.2. D.3, The flowers of the *moe clf-2 ag-3* mutant are typical of *ag* mutant flowers with the flower-within-a-flower phenotype.

plants. The F1 plants were allowed to self-fertilise and in the F2 generation families that segregated *clf*, *moe* and *ag* alleles were identified. In these families, about 1/64 of plants had the severe *clf moe* phenotype but with *ag* flowers suggesting that they were the triple mutants.

The *moe clf*-2 *ag-3* triple mutants showed extreme developmental defects. Thus the vegetative developmental phase was drastically shortened with stem development severely affected and only 3 or 4 sessile leaves, some of which were curled, formed. No inflorescence structure was observed and flowers formed at the apex of the plant. Flower organs seen were typical of *AG* mutant flowers with reiterated whorls of petals formed (Figure 3.4D1-3).

The striking phenotype of the *moe clf*-2 *ag-3* triple mutant provided evidence for the diverse role of the *CLF* and *MOE* gene in *Arabidopsis* development. The failure of *ag* mutation to suppress the mutant phenotype of the *clf moe* plant indicated that the extreme *clf moe* phenotype was not simply a result of ectopic *AG* activity. Presumably, in the *clf moe* doubles, many other targets are misregulated and contribute to the severity of the phenotype.

3.4 Analysis of floral homeotic gene expression in *moe* mutant

The double mutant studies suggested that like *CLF*, *MOE* also regulates floral homeotic gene activities in leaves and flowers. In order to obtain more insights into the relationship between the *MOE* and various floral homeotic genes at the molecular level, extensive analysis of homeotic gene expression patterns in leaves and floral tissues of *moe* mutants were investigated.

3.4.1 RT-PCR Experiment

To verify that ectopic *AG* expression was a cause of the leaf curling seen in the *moe* mutant as suggested by the phenotype of the *moe ag* double mutant, the level of *AG* transcripts in the *moe* background was monitored and compared to the wild-type level using RT-PCR analysis. Possible changes of other floral homeotic gene expression in the *moe* mutant were also investigated.

Total RNA was extracted from young leaves of *moe* plants. Reverse transcription was performed and PCR analysis was carried out on first strand cDNA using specific primers for the *AG* (YC9 & YC10), *PI* (YC5 &YC6), *AP3* (YC7 &YC8), *LFY* (YC11 & YC12) and *AP1* (YC13 &YC14) genes. The *AP2* gene was not included in this study since it is ubiquitously expressed and its transcripts are found in leaves in wild-type (Jofuku et al., 1994).

PCR products were run on 1% agarose gel and Southern blotting was performed and the filters hybridised with the cDNA specific probe from each gene (gift from J. Goodrich). The results obtained revealed that *AG*, *AP3* and *PI* were ectopically expressed in the *moe* mutant leaves. Thus transcripts of *AG*, *AP3* and *PI* were present in *moe* leaves and their cDNA were thus amplified by PCR (Figure 3.5). No expression of *AG*, *AP3* and *PI* was observed with the wild-type control. This agreed with results from previous studies that *AG*, *AP3* and *PI* are not expressed in the leaves in wild-type (Yanofsky et al., 1990; Jack et al., 1992; Goto and Meyerowitz, 1994). No misexpression of *LFY* and *AP1* was detected in the *moe* background in this experiment. However, possible ectopic expression of these two genes cannot be ruled out since the RT-PCR analysis might not have been sensitive enough to pick up the transcripts of *LFY* and *AP1* which could have been present at very low abundance. It could also be possible that those transcripts were present only transiently at a particular period of leaf development.

Of the three genes, ectopic *AG* expression was particularly strong with the cDNA band being visible on the gel (data not shown). This finding was in accordance with previous double mutant analysis that *AG* was the main target of *MOE* activity. Weak *AP3* expression was also observed in the leaves of *clf* plants but is unlikely to contribute to the leaf curling defect or other aspects of the *clf* phenotype since constitutive *AP3* expression under the strong 35S cauliflower mosaic virus promoter does not have any effect on leaf development (Jack et al., 1994). However this could be misleading since it has been shown that AP3 protein is not stable unless PI protein is also present (Jack et al., 1994). Therefore although *AP3* is misexpressed in *clf* mutant, the protein might not be able to exert its function as its partner, PI is absent in the leaves. The situation is, however, slightly different in the *moe* background with the *PI* gene also being ectopically expressed in the leaves. Thus the AP3 protein may be able to form active heterodimers with the PI protein and their activity might contribute to the stronger leaf curling defect seen in the *moe* mutants. Alternatively the function of *AP3* in the

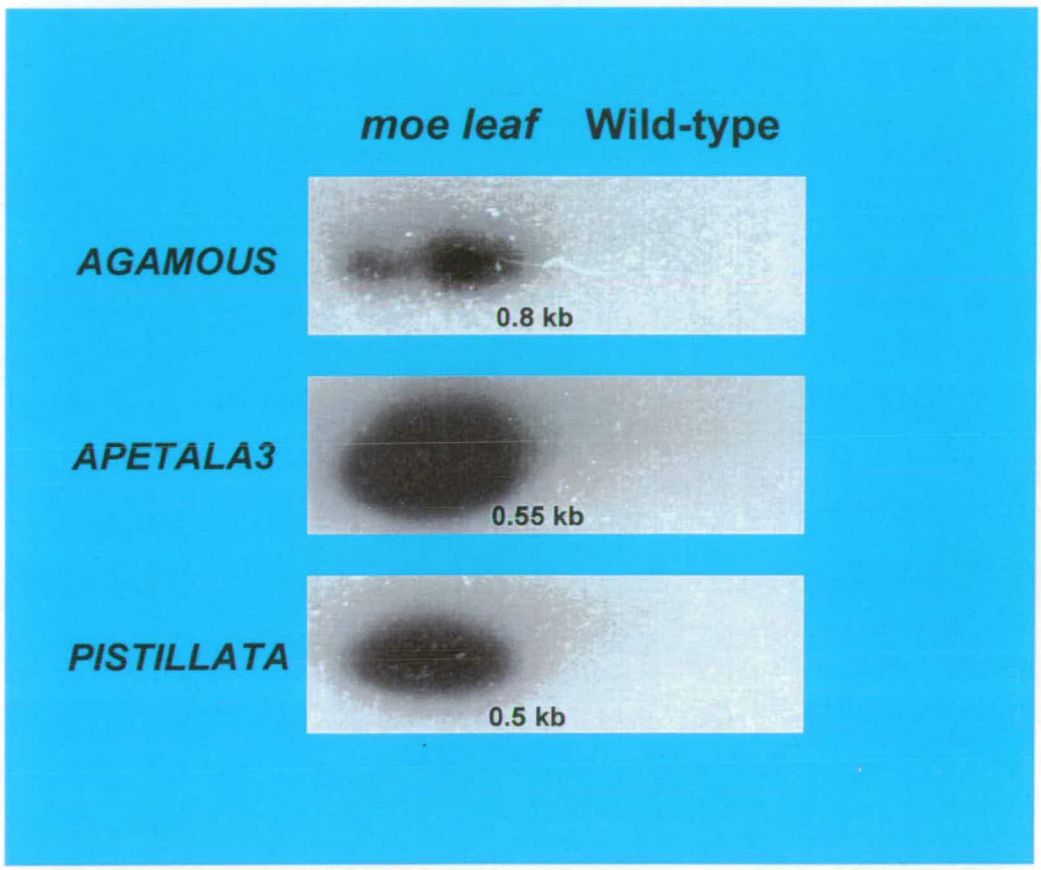


Figure 3.5 RT-PCR analysis of floral homeotic gene expression in *moe* background. Southern analysis shows that transcripts of the floral homeotic genes *AGAMOUS*, *APETALA3* and *PISTILLATA* are amplified using the primer pair YC9 & YC10, YC7 & YC8 and YC5 & YC6 respectively from the total RNA sample from *moe* but not wild-type leaf tissues indicating that the three genes are misexpressed in the leaves of *moe* mutants. Note that the exposure time of the *AGAMOUS* Southern blot was much shorter than that of *APETALA3* and *PISTILLATA*.

leaves might not be dependent on *PI* activity and therefore ectopic expression of *AP3* might not have any phenotypic effect on *moe* leaves.

3.4.2 GUS-staining analysis

To confirm the conclusion from RT-PCR analysis that not only *AG* but also *AP3* and *PI* transcripts were present in the leaf tissues from the *moe* mutant, it was decided to carry out a histochemical staining experiment to visualise the expression pattern of GUS reporter gene constructs for the aforementioned 3 floral homeotic genes in the *moe* background. In each case, *moe* heterozygotes were crossed as a male to female transgenic lines homozygous for the homeotic gene GUS reporter construct. The F1 plants carrying the reporter gene was identified on the basis of kanamycin resistance, the selectable marker gene present on the T-DNA construct harbouring the GUS reporter. To identify lines that inherited the *moe* allele, seeds were collected from kanamycin resistant F1 plants and sown out on soil. F2 seeds from lines that carried both the *moe* mutation and the GUS reporter construct were sown out on soil and used in the histochemical staining analysis.

3.4.2.1 pAG-I::*GUS* & pI::*GUS*

The pAG-I::*GUS* fusion construct contains about 6 kb of *AG* upstream sequences plus 3.8 kb of intragenic sequences which extend to the 5' end of the third exon of *AG* (Sieburth & Meyerowitz, 1997). Results from previous studies have shown that the pAG-I::*GUS* construct faithfully mimics the expression pattern of the endogenous *AG* gene and contains the *cis*-acting sequences necessary for response to *trans*-acting regulators such as *AP2* and *CLF*.

In this study, no GUS staining was detected in the leaves of the wild-type plants carrying the pAG-I::*GUS* transgene (Figure 3.6A-2). Normal *AG* expression pattern was detected in the flowers with blue staining in the stamens (both the anthers and the filaments) and the fourth whorl carpels with particularly strong staining in the stigmatic papillae. The staining pattern was very different in the *moe* background. The GUS activity was detected from the early stages of development. Thus deep blue stain could be seen in the cotyledons, young leaves, and also in the vein of the hypocotyl (Figure 3.6A-1). Ectopic *AG* expression in the leaves

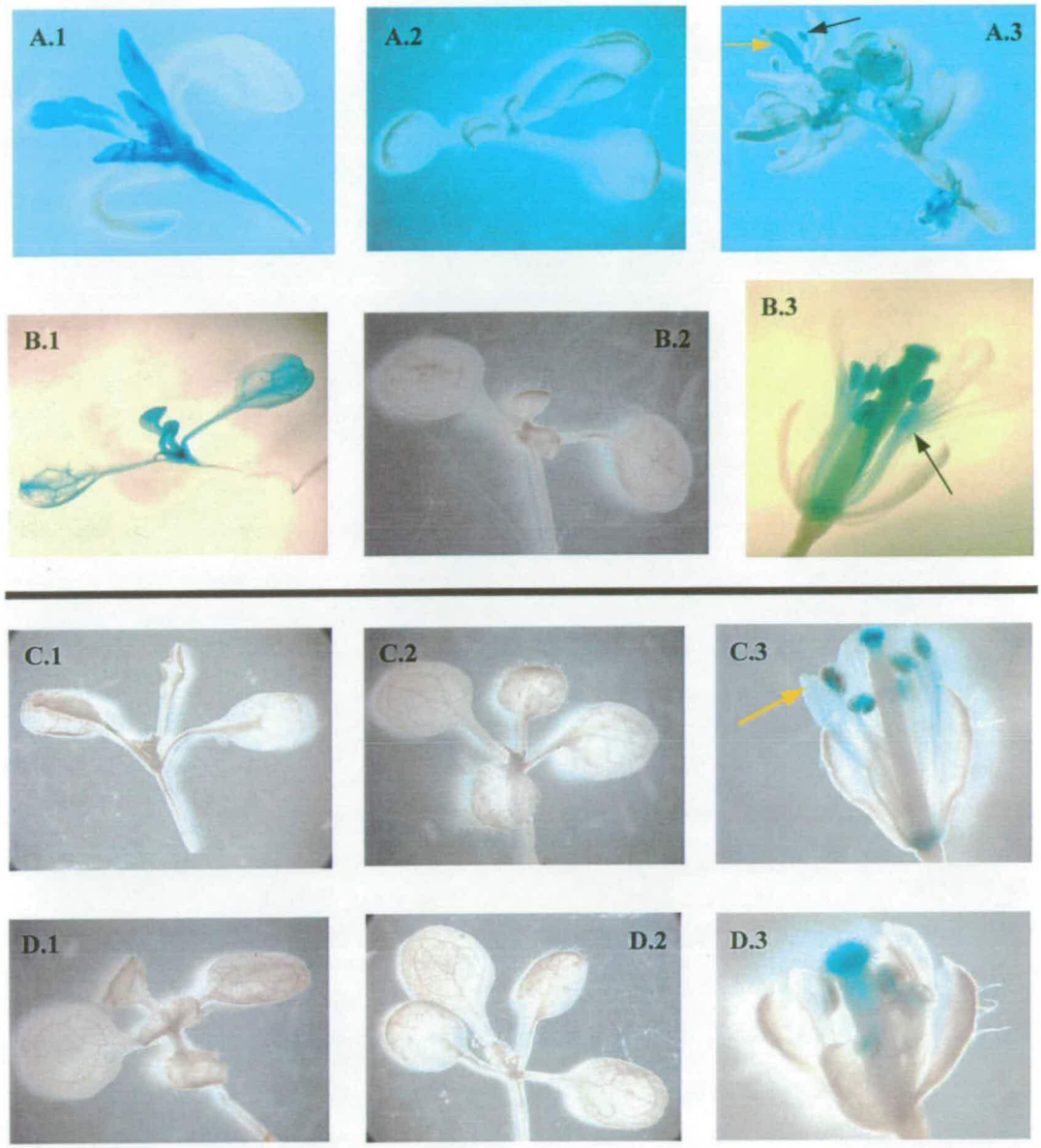


Figure 3.6 *AG* expression pattern in *moe* and *clf* background. Plants in row A and B contain the pAG-I::*GUS* construct while plants in row C and D contain the pI::*GUS* construct with just the second intron sequence fused to the GUS reporter gene. A.1, *AG* is expressed both ectopically and precociously throughout young *moe* seedling. Strong blue staining is detected in the hypocotyl vein and young leaves and weak staining is seen in the cotyledons. A.2, A young wild-type seedling showing no *AG* expression in the leaves. A.3, *AG* is expressed in stamens (black arrow) and carpels (yellow arrow) the inner two whorls of *moe* flowers. B.1 and B.2, A young *clf* seedling showing *AG* misexpression in the cotyledons and young leaves (B.1) while no *AG* expression is seen in the leaves of young wild-type sibling (B.2). B.3, Weak *AG* staining is seen in the petals of *clf* flower (black arrow) in addition to the normal staining pattern in stamens and carpels. C and D, No ectopic *AG* expression is seen in the *moe* and *clf* seedlings harbouring the pI::*GUS* construct indicating that the intron 2 alone does not give ectopic *AG* expression in *moe* or *clf* mutant leaves (C.1 and D.1 respectively). Wild-type controls also show no *AG* staining (C.2 and D.2). As with the pAG-I::*GUS* construct, similar *AG* staining patterns were detected in the flowers of *moe* (C.3) and *clf* (D.3) plants with occasional blue staining detected in the second whorl petals (yellow arrow in C.3) suggesting that *MOE* and *CLF* act similarly upon response elements in the intron 2 of *AG* to regulate its expression in the flowers.

seemed to persist into later stages of development after the plants had produced inflorescences and flowers. The pattern of *AG* expression seen correlated with the high abundance of *AG* transcripts detected in the RT-PCR experiment. Staining pattern in the *moe* flowers was rather similar to the wild-type plants (Figure 3.6A-3). However, occasionally weak staining was seen in the morphologically normal petals in the second whorls (data not shown). Staining was not detected in the sepals but this could not be interpreted as conclusive since *AG* might have been expressed at a very low level in the first whorl organs as suggested by the occasional carpelloidy in the sepals but the GUS assay might not have been sensitive enough to allow visualisation of the *AG* transcripts in the sepals.

The GUS staining patterns seen with the pAG-I::*GUS* construct in the *moe* mutants were very similar to the patterns in the *clf* mutants (Figure 3.6B1-3). In order to compare further similarities between *CLF* and *MOE* with respect to the regulation of *AG* expression, AG-GUS staining patterns in the lines carrying a modified AG::*GUS* construct was investigated. The pI::*GUS* construct used harbours the large *AG* second intron fused upstream of a minimal promoter-*GUS* transgene. This construct has been demonstrated to be both necessary and sufficient to confer the *AG* expression pattern (Sieburth and Meyerowitz, 1997; Busch et al., 1999). Previously, study using the *clf-2* mutants revealed that unlike pAG-I::*GUS*, the pI::*GUS* construct was not expressed in leaves of *clf or CLF*+ plants (J Goodrich pers. comm.) (Figure 3.6D1-3). However, the staining pattern in the flowers of plants carrying the pI::*GUS* construct was the same as seen with those from plants carrying the pAG-I::*GUS* construct with blue staining visible in the stamens and carpels and occasionally in the second whorl petals (white arrow in figure 3.6D3).

The AG-GUS staining patterns in the *moe* mutants carrying the pI::*GUS* construct were identical to those seen in the *clf* plants. Thus GUS activity, and hence the deep blue stain was lost in the leaves but still retained in the two innermost whorls of flowers (Figure 3.6C1-3). This finding further highlighted the similarity between *CLF* and *MOE* mediated repression mechanisms.

3.4.2.2 pAP3::*GUS*

The *moe* mutant lines harbouring the pAP3::*GUS* reporter construct were developed in the same way as with the pAG-I::*GUS* construct. The *AP3* fragment fused to the *GUS* gene contains a 3.7 kb of 5' *AP3* promoter sequences required for the wild-type *AP3* expression pattern.

In the wild-type siblings carrying the pAP3::*GUS* construct, no GUS staining was detected in the leaves (Figure 3.7A-4). Similarly, the staining pattern in the flowers was as expected with the GUS activity detected in the petals and stamens in the second and third whorl respectively (Figure 3.7A-5). Analysis of the *moe* mutant showed no difference in the staining pattern in the flowers compared to the wild-type pattern (Figure 3.7A-3). This suggested that *MOE* plays no part in the regulation of *AP3* expression during flower development. Alternatively, the *MOE* repression activity might be redundant with other factors and as a consequence the effect of the loss of *MOE* function on *AP3* expression was minimal. In contrast to the flower staining pattern, AP3-GUS staining in the leaves was markedly different between the *moe* and wild-type background. Blue staining, and hence, GUS activity was detected from the very young seedling stage. Although not as intense as the *AG* pattern the staining was seen in leaves and sometimes in the cotyledons and the hypocotyl (Figure 3.7A-1). However, compared to that of *AG*, the *AP3* staining patterns were rather different. Thus the blue patterns were not uniformed with the staining lost in some areas of the leaves and intense blue spots seen at some areas of the leaf blades (Figure 3.7A-2). It could be that the repression effects are somewhat stochastic, as with PEV, so that there are some differences in staining pattern from plant to plant.

3.4.2.3 pPI::*GUS*

The pPI-*GUS* reporter was introduced into the moe mutant background using the same strategy as for the other reporters. Histochemical staining showed that the expression pattern of the gene *PI* in the *moe* background was similar to that of the pAP3::*GUS* line. Thus as in the wild-type, petals and stamens gave strong staining in the flowers. Furthermore, the *PI* staining in the leaves of the *moe* mutants was uneven and found at specific areas in the leaves (Figure 3.7B1-4).

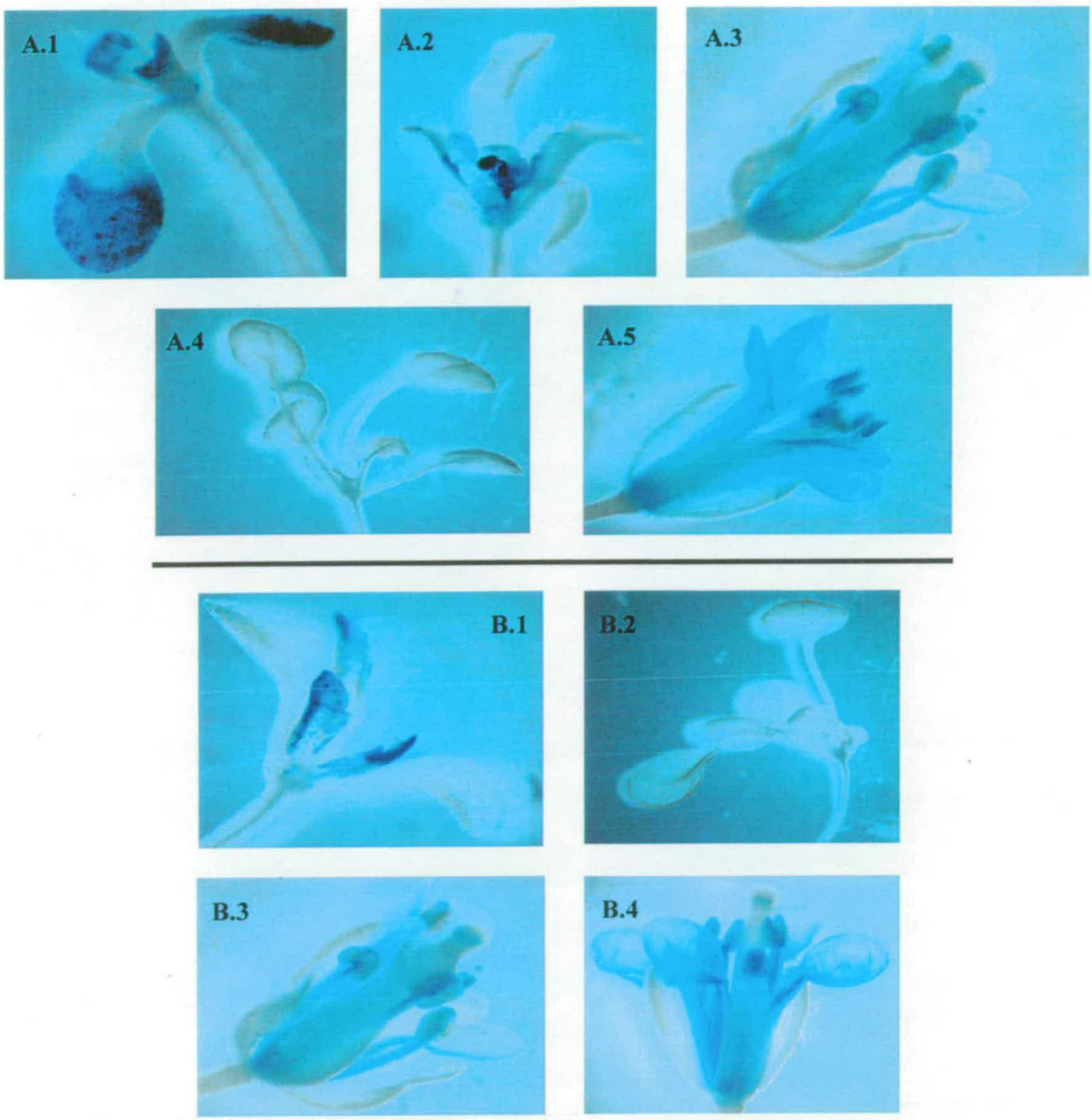


Figure 3.7 Expression of *AP3* and *PI* in *moe* mutant background. *AP3* is ectopically expressed in the cotyledons and young leaves of *moe* seedling (A.1). Later on expression pattern of *AP3* becomes stochastic with blue staining seen in certain areas of the leaves (A.2). Staining pattern in the *moe* flower shows no misexpression of the *AP3* gene with signal detected in the petals and stamens (A.3). No GUS staining is observed in the leaves of wild-type seedling and normal staining pattern was seen in the petals and stamens of wild-type flower (A.4 and A.5 respectively). *PI* shows similar expression pattern as *AP3* in the *moe* background. Expression is seen as clonal i.e., being confined to particular areas of the adult leaves (B.1). No staining is detected in the wild-type control (B.2). Like *AP3*, *PI* staining pattern in *moe* flower (B.3) is identical to that seen in wild-type flower (B.4).

3.5 Mapping the *moe* mutation

Data accumulated from various genetic analyses described previously indicated striking similarities between the *moe* and *clf* mutation at both the phenotypic and functional level. This in turn suggested that the *MOE* and *CLF* genes may act very closely to regulate the expression patterns of homeotic gene and possibly other yet unidentified target gene. It is, therefore, highly possible that the *MOE* and *CLF* genes code for proteins with similar structure and functional motifs. Hence the wild-type *MOE* gene is a very strong candidate for another Pc-G member in *Arabidopsis*. Identification of the *MOE* gene would inevitably prove if this prediction was right.

Although the *moe* mutation was isolated from a T-DNA mutagenesis experiment, genetic and molecular analysis suggested that it was not tagged by T-DNA insertion (C MacDougall and J Goodrich, pers. comm.). Therefore in order to clone the *MOE* gene, conventional mapping strategies were taken with the first step of using various genetic and molecular markers to pinpoint the location of the *MOE* gene on the *Arabidopsis* genome.

3.5.1 Morphological markers

Linkage analysis using visible morphological markers such as mutants with scorable phenotype often provide a good indication of the rough position of the gene of interest. In practice, the mutant in question is crossed to other well-defined mutants with known phenotypes. The F1 double heterozygotes are allowed to self and the segregation pattern of the two phenotypes is analysed in the F2 generation. If the mutation of interest is linked with the marker used, the segregation ratio in the F2 population is usually distorted i.e., it shows deviation from the ratio predicted by the Mendel law as the close physical distance of the two mutations reduces the amount of recombination that takes place between them.

As the *moe* mutation was crossed to various floral homeotic mutants with clear visible phenotypes in the attempt to study genetic interactions, the segregation data obtained from the F2 populations was analysed to test for linkage between the *moe* mutation and any of those floral homeotic traits. In those crosses, according to the Mendel law of segregation, the frequency of the F2 double mutants from the selfing of the F1 double heterozygotes was predicted to be 1/16 if the 2 mutations were unlinked and did not show epistasis or lethality

in double mutant combinations. However if the two loci were linked the double mutants would be much rarer than expected as the mutant alleles would seldom be present on the same chromosome.

Analysis of the data from the crosses showed that the *moe* mutation did not co-segregate with the *ag*, *ap1*, *ap2* and *ap3* mutation but showed significant linkage with the *pi* and *lfy* mutations, both of which are located on chromosome 5, with the *pi* mutation mapped onto the upper arm and the *lfy* mutation onto the lower arm of the chromosome. The frequency of the *moe pi* double mutants was 0.096 while that of the *moe lfy* double was 0.039 compared to the *moe* single mutants. These numbers showed big departure from the expected frequency of 1 in 4 (0.25) had the genes and the *moe* mutation segregated in a Mendelian fashion. Mathematical analysis indicated that both values were statistically significant (*moe pi* χ^2=15, $p<0.01$, *moe lfy* $\chi 2$=30.56, $p<0.01$). The fact that the *moe lfy* double mutants were about half as rare as the *moe pi* plants suggested that the *moe* allele was more tightly linked to the *LFY* locus on the lower arm of chromosome 5.

3.5.2 Molecular markers

Although gene mapping using morphological markers often gives a rough map position for the mutant of interest, the resolution is rather low because the number of markers available are limited. To overcome these problems molecular markers are used instead of the morphological markers, to allow fine mapping of the mutation to within a small chromosomal region. Molecular markers offer several advantages over the traditional morphological markers. Firstly, their number is virtually unlimited due to the very high polymorphic nature at the DNA level between distinct ecotypes. For example it has been estimated that the Columbia (Col) and Landsberg *erecta* (L*er*) ecotypes differ by about 0.5-1% at the DNA sequence level. The difference can be attributed to random point mutations, small deletions or insertions in the genome of one ecotype but not the other. This means that for a high resolution mapping effort a single cross is all that is needed for the detection of the molecular polymorphisms between two different ecotypes. Furthermore, the genotype of each plant used with respect to a particular marker is easy to score with only simple molecular analysis required for the detection of differences at the DNA level. These features of molecular markers allow high resolution mapping to be carried out with high efficiency and reliability in a short period of time.

Historically the well defined restriction length polymorphism (RFLP) markers were widely used. However, mapping a new mutation onto a pre-existing RFLP map can also be a time consuming process involving the extraction of large amounts of DNA from plants in the F2 or F3 generation and several DNA blot analysis. Therefore, considering the time and work involved, successful mapping of the *moe* mutation using RFLP markers within the given time span would be highly unlikely and PCR based molecular markers were employed instead in this study. These marker are highly abundant and found randomly in the *Arabidopsis* genome. Scoring is also relatively easy and requires only a small amount of DNA with no DNA blotting and hybridisation steps needed.

Two types of PCR-based markers were used in this study. The cleaved amplified polymorphic sequences (CAPS) are co-dominant markers that give ecotype-specific restriction pattern when digested with an appropriate restriction endonuclease. On the other hand, the simple sequence length polymorphisms (SSLPs) are small microsatellite repeat sequences that are polymorphic in size between different ecotypes. Like CAPS, SSLPs are also co-dominant. Mapping using these markers offer a number of advantages. Their co-dominant nature means that homozygosity and heterozygosity for the parental alleles can be readily distinguished and hence a good map position can be obtained by scoring only a relatively small number of plants. In addition, only a small amount of DNA is required from each plant since these markers utilise PCR for the detection of ecotype-specific differences.

Mapping of the *moe* mutation was carried out on genomic DNA samples from the *moe* mutant F2 plants from an outcross of *moe* heterozygous males in the Ws background to Col and L*er* wild-type females. Two wild-type backgrounds were used in order to cover as many markers as possible since some markers are polymorphic between Ws and Col while the others show polymorphism between Ws and L*er*. DNA samples were subjected to PCR reactions with primers for a marker of interest. In the case of SSLPs the PCR product sizes were determined by running the products on agarose gels. PCR products amplified from CAPS markers were digested with an appropriate restriction enzyme and the digest products then run on the gels and the genotype scored.

If a given marker is linked to the *moe* mutation, most of the F2 *moe* plants should then be Ws homozygous with respect to the size (in the case of SSLPs) or restriction pattern (CAPS) of that marker since most of the time the marker would co-segregate with the *moe* allele and

thus recombination is unlikely to take place between them. If, however, that marker is not linked or located on a different chromosome to the *moe* mutation, then the restriction patterns or the sizes seen from the F2 plants studied should segregate randomly with respect to the *moe* mutation. For each marker the recombination frequency was worked out using the following formulae,

$$\text{Recombination frequency (\%)} = \frac{\text{the number of recombinant chromosomes (non-WS chromosomes)}}{\text{total number of chromosomes}} \times 100$$

The value of the frequency was then converted to an estimate genetic map distance between the marker and the *moe* allele. A rough estimation is that one centimorgan (cM) of genetic distance equals to 1% recombination frequency. By assuming that the lower the recombination frequency was, the nearer the corresponding marker was to the *MOE* gene, it was therefore possible to place *MOE* to an interval of two nearest markers on both sides of the gene.

The data from mapping using the morphogical markers indicated that *MOE* is tightly linked to *LFY* on the lower arm of chromosome 5. In order to obtain a more precise location of the *moe* allele on chromosome 5, finer mapping using a series of published CAPS and SSLPs markers on the lower arm of chromosome 5 and others kindly suggested by Dr. R Buell (TIGR) who identified them in the course of mapping in similar region was carried out. The annealing temperature used was 55°C and PCR was carried out for 50 cycles. The mapping result is shown in Figure 3.8.

The data presented in Figure 3.8 suggested that there was significant linkage between the *moe* mutation and the markers lying between and inclusive of DFR and nga129. In order to narrow down the position of the *MOE* gene to the nearest 2 markers, larger scale mapping was performed. A new set of plant DNA was extracted using a Qiagen high throughput DNA isolation kit. PCR mapping was carried out on those DNAs using a number of chosen markers that lie within the distance between DFR and nga129. These are DFR, RPS4-NT and nga129. A recently published new CAPS marker; AtTED2, lying just below nga129 was also included in the analysis. Its details are shown as follows;

Marker	Map Position (cM from top of chromosome 5)	Marker type	Polymorphism	Primers	Recombination Frequency (%)
AthSO262	65.20	SSLP	Ws = 151 bp *Ler* = 159 bp	5'-TTGCTTTTTGGTTATATTCGGA-3' 5'-ATCATCTGCCCATGGTTTTT-3'	25
AthPHYC	74.30	SSLP	Ws = 226 bp *Ler* = 211 bp	5'-CTCAGAGAATTCCCAGAAAAATCT-3' 5'-AAACTCGAGAGTTTTGTCTAGATC-3'	21.95
AthSO191	Between AthPHYC and RBSC-B	SSLP	Ws = 162 bp *Ler* = 156 bp	5'-TGATGTTGATGGAGATGGTCA-3' 5'-CTCCACCAATCATGCAAATG-3'	20
RBCS-Ba	78.97	CAPS	*Ssp*I Digest Ws = 0.866/0.286 kb *Ler* = 0.609/0.286/0.257 kb	5'-TATCGACAATTTATTGTGG-3' 5'-TGAATGGAGCGACCATAGTGGCCTGAGCCAA-3'	16.7
DFR	89.5	CAPS	*Bsa*AI Digest Ws = 0.609/0.534 kb *Ler* = 0.609/0.318/0.216 kb	5'-AGATCCTGAGGTGAGTTTTTC-3' 5'-TGTTACATGGCTTCATACCA-3'	4
RPS4-NT	Between DFR and PAP548 marker at 94.18 cM	CAPS	*Xho*I Digest Ws = 0.21/0.21 kb Col = 0.42 kb	5'-TCATCAATTTCCGTGGGGCA-3' 5'-GGTACCTTTTCTTGTCAATG-3'	3.5
RB55/RB56	Just 5' of PAP548 At 94.18 cM	CAPS	*Eco*RV Digest Ws = 330/140 bp *Ler* = 140/160 bp	5'-GGTCGTCAGTTTCATGCACA-3' 5'-AAGCAGACCCGATAACAACG-3'	3.3
nga129	105.41	SSLP	Ws =165 bp Col= 177 bp	5'-TCAGGAGGAACTAAAGTGAGGG-3' 5'-CACACTGAAGATGGTCTTGAGG-3'	3
LFYa	116.88	CAPS	*Rsa*I Digest Ws = 855/236/ 126/78/35 bp *Ler* = 708/236/ 147/126/78/ 35 bp	5'-TAACTTATCGGGCTTCTGC-3' 5'-GACGGCGTCTAGAAGATTC-3'	13.9

Figure 3.8 PCR-based mapping of *moe leaf* mutation. Various CAPS and SSLPs markers were used to map the *moe* mutation. DNA samples were prepared from F2 plants from an outcross of *moe*/+ plants (Ws background) to wild-type plants in either L*er* and Col background. A number of markers at the lower arm of chromosome 5 were chosen. For each marker, PCR was performed on the DNA samples using the appropriate primer set. PCR products were run on agarose gels and the genotype of each sample was scored by direct visualisation of differences in PCR product sizes for SSLPs markers or differences in restriction endonuclease digest patterns in the case of CAPS marker. In each case recombination frequency was calculated as the percentage of the number recombinant chromosomes over the total number of chromosomes.

Marker	Position (cM from top)	Polymorphism	Primers
AtTED2	115.3	*Eco*RV digest Ws=0.43+0.23 kb L*er*= 0.35+0.23+0.08 kb	5'-CGTAGACAAGGT ACTGTCAACC-3' 5'-GATAATCTCGTC TCCAAGTGTCC-3'

Figure 3.9 A table showing the details of the CAPS marker AtTED2.

The genotype of each plant was scored for each of the 4 markers. Detailed analysis of the data from about 200 plants used would give invaluable information on the position of the *moe* allele relative to each marker. The recombination breakpoint in each plant, if any, could also be deduced by comparing the genotype of the plant with respect to each marker. Thus as shown in figure 3.10, in the case of event 1, recombination takes place between marker A & B the plant will be heterozygous for marker A and homozygous Ws for marker B and also the rest of the markers on the right side of marker B. In event 2 the recombination breakpoint lies between marker B & C then the plant will be heterozygous for both marker A & B and homozygous Ws for the markers lying on the right hand side of A and B. If the *moe* allele lies between marker C & D which are close together it is very unlikely that recombination will occur between them and therefore most of the *moe* plants would likely be homozygous Ws for both markers and that those that are heterozygous for one of the markers would not be heterozygous for the other marker i.e., any given plant would not be heterozygous for both markers. The exception is when the rare double recombination event takes place on either side of the *MOE* gene resulting in the plant being heterozygous for all the markers.

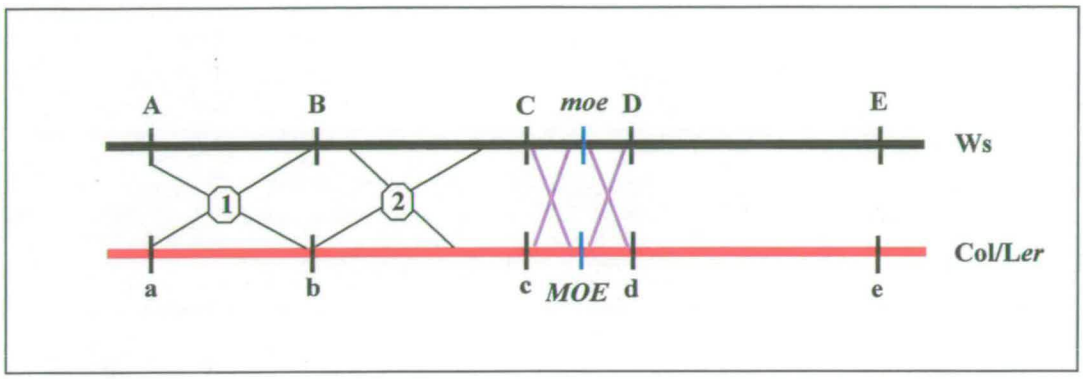


Figure 3.10 Schematic diagram showing possible recombination events that can occur between *MOE* and the markers lying close by. In the event 1, the progeny *moe* plant will be Ws heterozygous for marker A and homozygous for the other 4 markers. In the event 2, the *moe* plant will be heterozygous Ws for marker A and B and homozygous for the rest of the markers. Finally, if double recombination takes place (purple crosses) near and on both sides of the *MOE* gene the *moe* plant will be Ws heterozygous for all the markers on both sides of the gene.

The data from finer mapping of *MOE* using the 4 markers chosen was pooled together and part of it is summarised in figure 3.11.

Marker / Plant	DFR (89.5 cM)	RPS4-NT (~97 cM)	nga129 (104.5 cM)	AtTED2 (115.3 cM)
1	Ws	Ws	Ws	Het
2	Ws	Ws	Ws	Het
3	Het	Ws	Ws	Ws
4	Het	Het	Het	Ws
5	Het	Het	Het	Ws
6	Ws	Ws	Ws	Ws
7	Het	Ws	Ws	Ws
8	Ws	Ws	Ws	Ws
9	Het	Het	Ws	Ws
10	Ws	Ws	Ws	Het
11	Ws	Ws	Ws	Ws
12	Ws	Ws	Ws	Ws
13	Het	Het	Ws	Ws
14	Ws	Ws	Ws	Ws
15	Het	Het	Het	Ws
16	Ws	Ws	Ws	Ws
17	Ws	Ws	Ws	Het
18	Ws	Ws	Ws	Ws
19	Het	Ws	Ws	Ws
20	Het	Het	Ws	Ws
21	Ws	Ws	Ws	Ws
22	Het	Ws	Ws	Ws
23	Ws	Ws	Ws	Ws
24	Ws	Ws	Ws	Het
25	Ws	Ws	Ws	Ws

Figure 3.11 Finer mapping of *moe* mutation. PCR mapping was performed on *moe leaf* DNA samples. Genotype of each sample was scored for each of the four markers used between the DFR-AtTED2 interval. The table shows that no plants are simultaneously heterozygous for the markers nga129 and AtTED2 suggesting that recombination is extremely rare between these markers and therefore *moe* is likely lying between them.

3.6 *MOE LEAF* Cloning

It was apparent from the figure 3.11 that none of the plants used was double heterozygous for the markers nga129 and AtTED2. This implies that the *MOE* gene is likely to lie between these 2 markers. However, the distance between nga129 and AtTED2 is about 10 cM which corresponds to approximately 2 Mb of DNA sequence. According to the estimate gene density of one gene per 4.5 kb of DNA in the *Arabidopsis* genome there could be as many as 450 genes located in the 10 cM interval between nga129 and AtTED2. It was, therefore, necessary to carry out finer mapping and try to create more molecular markers within this region and screen larger size *moe* population to narrow down the position of *MOE* to a nearest 2 markers interval. However, given that this would involve a lot of laborious analysis and that the time span left was rather short, it was decided to undertake another approach. Thus the candidate gene approach was taken. Since this region was completely sequenced and well-annotated in the databases, it was therefore possible to search the databases for the candidate genes such as Polycomb homologues or genes that code for the SET domain.

Of the few hundred genes pulled out from the databases, one in particular stood out. Thus the recently identified Pc-G gene *EMBRYONIC FLOWER 2* (*EMF2*), although mainly known as a major chaperone of phase transition from vegetative to reproductive growth, has also been shown to play a role in the control of floral meristem identity and homeotic gene expression (Yang et al., 1995; Chen et al., 1997; Haung & Yang, 1998; Chou et al., 2001).

Two classes of *emf* mutants; *emf1* and *emf2*, were isolated from EMS and gamma-ray-induced mutagenesis experiments (Sung et al., 1992; Yang et al., 1995). These mutants exhibit similar phenotypic abnormalities. Both vegetative and reproductive development of *emf* plants are adversely affected. In the weak *emf2-3* mutant the vegetative phase is drastically shortened compared to the wild-type plant (Figure 3.12A). In addition several aspects of growth appear to be affected. Thus seedlings are found with short hypocotyls and petiole-less, oval-shaped cotyledons. Later on a few sessile leaves are formed (Figure 3.12B). The majority of these leaves appear to be cauline rather than rosette. Shortly afterwards, the plants enter the reproductive phase and produce inflorescence shoots of varying sizes depending on the extent of internode elongation (Figure 3.12C). Two or three solitary flowers are formed on the inflorescence. These flowers appear to have prominent pistils and usually lack petals (Figure 3.12D). Stamen and carpel identity and the organ

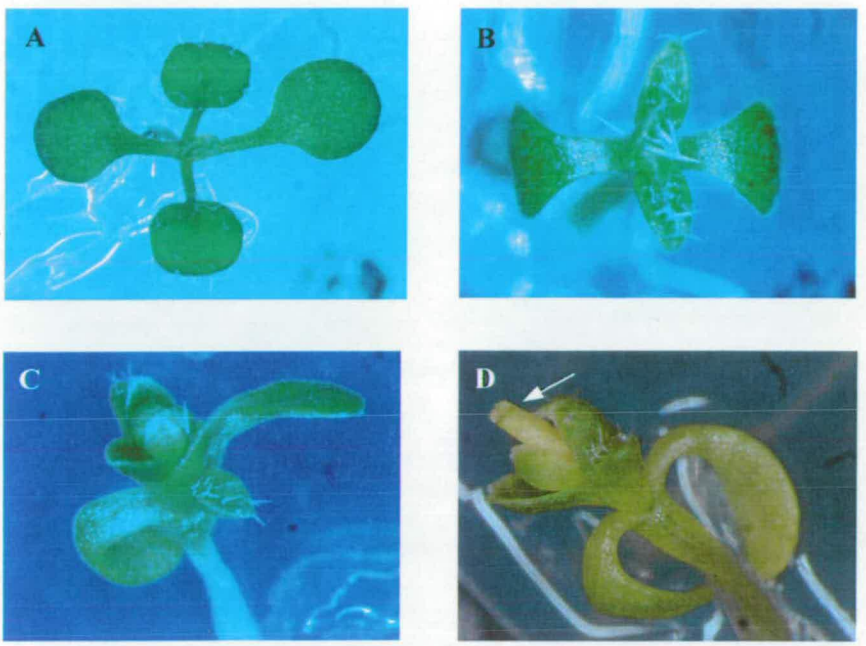


Figure 3.12 *emf2-3* mutant phenotype. A, A typical wild-type plant grown on GM agar media. B, An *emf2-3* mutant seedling is very small in size and forms only a few sessile leaves. C, An *emf2-3* mutant has a very short vegetative growth and enters reproductive growth phase soon after germination. The inflorescence shoot formed is extremely small due to short internode elongation. D, An *emf2-3* plant produces one or two flowers with the prominent pistil characteristic (white arrow) on the inflorescence.

number seem largely unaffected. However, in very strong *emf2* alleles such as *emf2-2*, stamens are often absent and only carpelloid structures are seen (Bai et al., 1995; Yang et al., 1995; Chen et al., 1997).

These phenotypic abnormalities indicate that the *EMF* genes play important roles in repressing the vegetative to reproductive phase transition in *Arabidopsis*. In addition, *EMF* activities are also likely required to ensure proper development of flowers. Strikingly, a number of the reproductive defects exhibited by *emf* mutants are similar to those seen in *terminal flower 1* (*tfl1*) mutants. By analogy it is likely that like the *TFL1* gene which acts to maintain inflorescence development by negatively regulating the activity of floral meristem identity genes during inflorescence development (Alvarez et al., 1992), the *EMF* genes might act as floral repressors. It has been proposed that reproductive development is a default state and high activities of genes such as *TFL1*, *EMF1* and *EMF2* during early stages of growth suppress the development of reproductive structures (Chen et al., 1997). Documented data from genetic studies suggest that *EMF* genes might exert this repressor function by prohibiting the expression of floral meristem identity genes such as *LFY* (Chou et al., 2001). It has also been proposed that the vegetative to reproductive growth transition in wild-type plants is accompanied by a gradual decline in *EMF* activities. Thus floral meristem activities are continually derepressed and the formation of reproductive structures is then ensured. In addition, it is thought that the increase in *LFY* activity in turn suppresses the activities of the *EMF* genes suggesting a reciprocal negative regulatory relationship between the *EMF* and floral meristem identity genes during reproductive development (Chen et al., 1997). Expression analysis has also indicated a role of the *EMF* genes as negative regulators of the floral homeotic genes. Thus *in situ* hybridisation analysis revealed that *AG* is ectopically expressed in the carpelloid structures of the *emf* mutants (Chen et al., 1997).

So far data obtained from various studies indicate that the *EMF* functions are required in many aspects of *Arabidopsis* development especially the control of reproductive switch and the regulation of many floral genes. Both *EMF* genes have recently been cloned and in depth analysis at the molecular level has provided valuable information on the role of the two genes during reproductive development. It appeared that the *EMF1* gene encodes a novel protein with no significant homology with any protein of known function. However, a homologue in rice exists and both proteins share some common motifs including nuclear localisation domains and ATP/GTP binding motifs suggesting that both may act as

transcriptional regulators during shoot development (Aubert et al., 2001). Expression analysis revealed that the *EMF1* gene is ubiquitously expressed and that an alteration of the *EMF1* level by overexpression using the 35S::sense *EMF1* construct or suppression using the antisense construct had dramatic impacts on flowering time. For example, it was shown that *EMF1* antisense transgenic plants exhibited an early flowering phenotype and also produced terminal flowers on the inflorescences similar to the *tfl1* flowers suggesting a shift from indeterminate to determinate growth. Thus when the wild-type *EMF1* activity was lost or reduced, several of the floral promotion genes which were normally repressed by the *EMF1* function became ectopically expressed early in development resulting in a premature shift to reproductive growth (Aubert et al., 2001).

Cloning of the *EMF2* gene revealed that it codes for a C_2H_2 type zinc finger protein which is highly conserved in plants and animals (Figure 3.13, 3.14A). It turned out that *EMF2* and its *Arabidopsis* homologues *FIS2* and *VRN2* represent a small novel class of plant Pc-G genes. As previously mentioned the *FIS2* gene has been implicated to function in Polycomb-mediated epigenetic control of seed development (Chaudhury et al, 1997; Luo et al., 1999). The gene *VERNALIZATION2* (*VRN2*), on the other hand, acts to promote floral transition through vernalization response pathway. Detailed analysis showed that *VRN2* functions to maintain the stable repression of a floral repressor; the *FLOWERING LOCUS C* (*FLC*) gene after cold treatment of seeds. Mutations in the *VRN2* gene results in a concomitant loss in cellular memory of vernalization and hence delayed flowering caused by elevated *FLC* expression (Gendall et al., 2001).

The *Drosophila* counterpart of the *EMF2* gene, the recently cloned *Suppressor of zeste 12* [*Su(z)12*], is required for the maintenance of the repressed state of homeotic gene expression. A spectrum of *Drosophila* homeotic genes are misexpressed in *su(z)12* mutants which exhibit strong homeotic phenotypes comparable to those of the most severe classes of Pc-G mutants. Furthermore, *su(z)12* mutations has also been indicated as strong suppressors of position-effect-variegation (PEV). This suggests a role of the wild-type *Su(z)12* gene in heterochromatin-mediated repression (Birve et al., 2001).

In addition to the zinc finger motif, two other well conserved domains are also present in EMF2 and its homologues. Thus, the N-terminal region of the VRN2, EMF2 and SU(Z)12 proteins contains a conserved region of approximately 60 amino acid which includes a

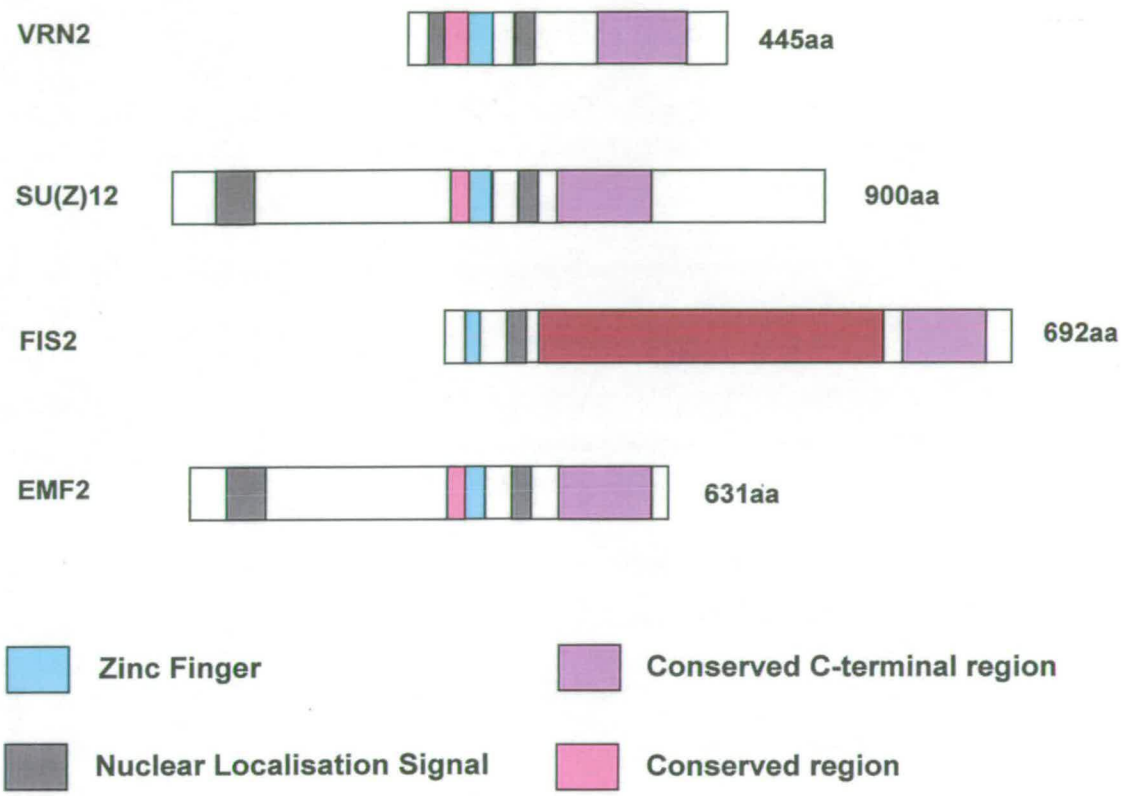


Figure 3.13 Alignment of VRN2, SU(Z)12, FIS2 and EMF2 proteins. The EMF2 protein shares a number of motifs with the VRN2, SU(Z)12 and FIS2 proteins. The numbers refer to the total number of amino acid residues. The conserved C_2H_2 type zinc finger is shown in light blue, the highly conserved acidic C-terminal domain is highlighted in purple, the conserved N-terminal region is indicated in pink and the putative nuclear localisation signal is indicated in grey (Modified from Gendall et al., 2001).

putative nuclear localisation signal. In addition, the C-terminal part of all 4 proteins is characterised by a large cluster of acidic amino acids followed by 2 repeats of leucine-tryptophan-asparagine rich sequence (Figure 3.14B), (Luo et al., 1999; Birve et al., 2001; Gendall et al., 2001; Yoshida et al., 2001). The presence of the zinc fingers suggests that these proteins might act as transcriptional regulators. The conserved C-terminal half is also likely indispensable for their function since mutations in this region have been shown to cause severe functional defects (Luo et al., 1999; Gendall et al., 2001).

Expression analysis using *in situ* hybridisation has demonstrated that, like *EMF1*, the *EMF2* gene is also ubiquitously expressed throughout *Arabidopsis* life cycle. Thus *EMF2* RNA was detected in the shoot apical meristems, young leaves and later in the inflorescence and floral meristems. In addition, expression was also detected in the developing embryos and endosperm (Yoshida et al., 2001). Overlapping expression patterns of *EMF1* and *EMF2* suggest that the two genes might act redundantly to delay the onset of reproductive development and later to repress floral homeotic genes. Since *EMF2* encodes a Pc-G protein it is tempting to speculate that the protein exert its repressive function in a similar manner to other Pc-G proteins by modifying higher order chromatin structure.

The fact that *EMF2* encodes a Pc-G protein which functions as a floral repressor fitted in very well with the striking phenotypic similarities between the *moe* mutant and the mutant plants for the *CLF* gene which also codes for a Pc-G protein. This in turn provided a tantalising thought that *MOE* and *EMF2* are the same locus. In support of this view, the strong leaf curling and early flowering phenotypes exhibited by transgenic lines harbouring the *EMF2* antisense construct were remarkably similar to the defects seen in *moe* plants (Yoshida et al., 2001). This suggested that *moe* might be an allele of the *EMF2* gene. To test this an allelism analysis was carried out to find out if the *emf2* and *moe* mutation complemented one another. In the test *moe* heterozygous plants were crossed as females with pollen from plants that were heterozygous for the weak recessive *emf2-3* allele. If *moe* was allelic to the *emf2-3* mutation then 1 in 4 plants in the F1 generation should exhibit an *emf2-3* or *moe*-like phenotype. If, on the other hand, the 2 mutations were in separate genes, all the F1 progeny would have been wild-type in appearance.

The result from the crosses proved that *moe* was indeed allelic to the *emf2-3* mutation (Figure 3.15). Thus a quarter of F1 progeny from the cross of *moe*/+ to *emf2-3*/+ plants

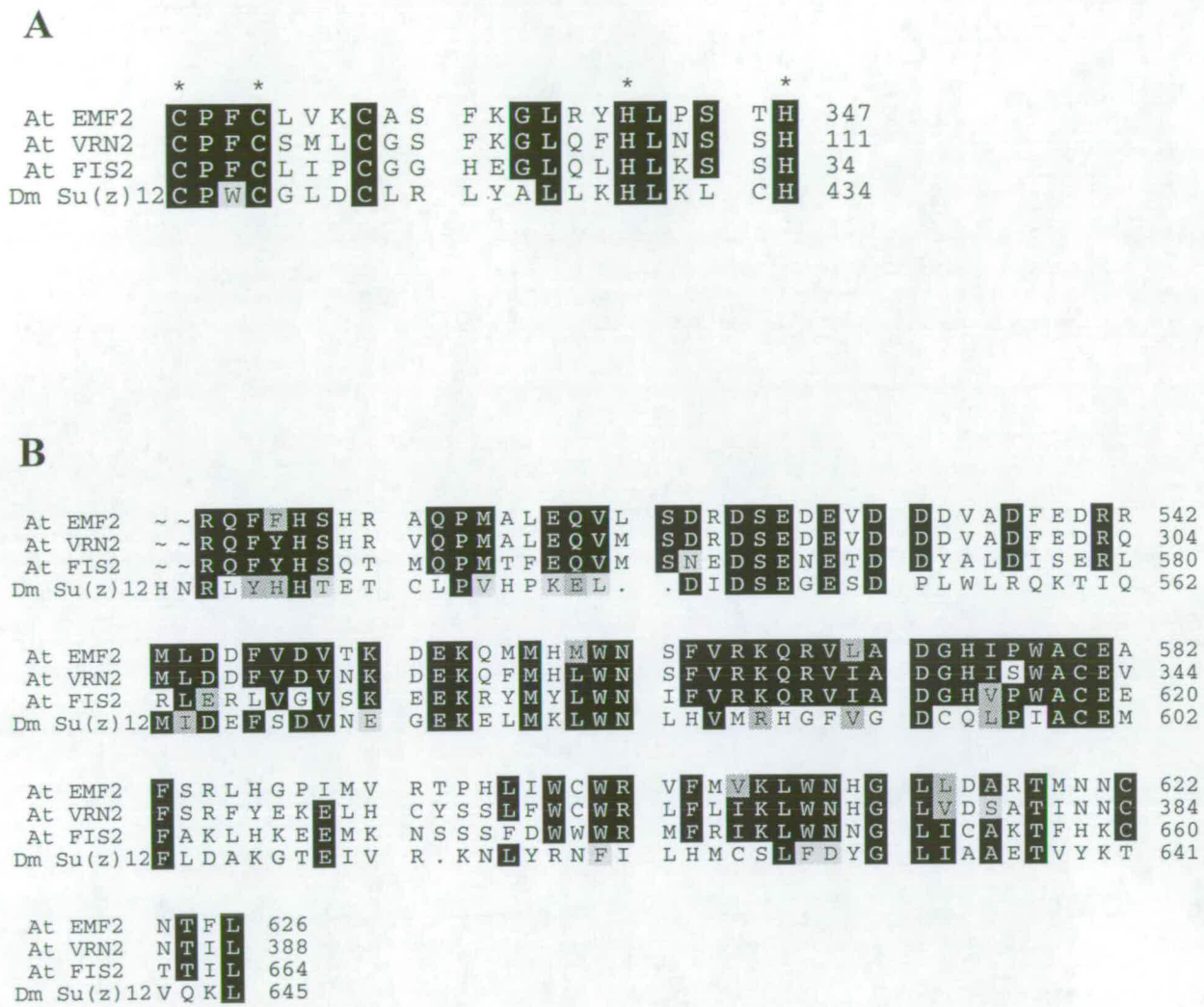


Figure 3.14 Structural similarity in the EMF2 protein family. A, Sequence alignment of the conserved C_2H_2 zinc finger motifs in the *Arabidopsis* EMF2, VRN2 and FIS2 protein and the *Drosophila* SU(Z)12 protein. Identical residues are boxed in black. B, Sequence alignment of the highly conserved C-terminal acidic domains. Identical amino acids are boxed in black, conserved amino acids are boxed in dark grey and similar residues are boxed in light grey.

showed *moe* or *emf2-3* like phenotype. Control crosses were made whereby *moe*/+ females were crossed to wild-type male plants. As expected none of the F1 progeny from those crosses were *moe* mutant.

Female parent	Male parent	F1 phenotype	
		moe	Wild-type
MOE/MOE	*emf2-3/EMF2-3*	0	37
moe/MOE	*emf2-3/EMF2-3*	16	32
moe/MOE	*emf2-3/EMF2-3*	19	54
moe/MOE	*EMF2-3/EMF2-3*	0	30
moe/MOE	*EMF2-3/EMF2-3*	0	39
MOE/MOE	*EMF2-3/EMF2-3*	0	35

Figure 3.15 *moe leaf* is allelic to *emf2-3* mutation. Allelism test showed that the *moe* mutation co-segregated with the *emf2-3* mutation. When heterozygous *moe*/+ plants were crossed to heterozygous *emf2-3*/+ plants, a quarter of the F1 progeny obtained showed a *moe*- or an *emf2-3* like phenotype suggesting that the *moe* and *emf2-3* phenotypes are caused by mutations in the same gene.

3.7 The molecular basis of the *moe* allele

The allelic nature of the *moe* and the *emf2-3* mutation indicated that the *moe* phenotype was caused by a mutation in the *EMF2* gene. Since the *moe* phenotype was considerably weaker than the *emf2-3* phenotype which is the weakest documented *emf2* allele, it is unlikely that the *moe* mutation was a big deletion, large chromosomal rearrangement or a mutation in one of the conserved functional motif that would result in the protein produced being truncated or non-functional. A small deletion or single mis-sense mutation would probably be a more likely cause. Alternatively the *moe* allele could be a consequence of a defect in the intron-exon splicing process. The possibility that *moe* is a epigenetic allele could not be discarded either. Thus hypermethylation of the *EMF2* gene could cause the defects seen in *moe* plants. Ectopic cytosine methylation has generally been associated with a silent state of gene expression and several heritable genetic alleles have been identified in a number of plant genes including the *Arabidopsis SUP* and *FWA* genes (Jacobsen and Meyerowitz, 1997; Soppe et al., 2000). However ectopic hypermethylation was rather unlikely to be the cause of the *moe* phenotype since occasional somatic reversion which is one of the features of epigenetic alleles was never seen on the *moe* plants.

In order to find out if a point mutation or a small deletion was the basis of the *moe* allele, direct sequencing of the *EMF2* gene was carried out on the *moe* and Ws genomic DNA. To cover the whole gene three sets of primers; EMF2.7/EMF2.3, EMF2-F1/EMF2-R4 and EMF2-P1/EMF2-P2 were used to amplify genomic DNA from *moe* and Ws plants. Direct sequencing reactions were then performed on both strands of DNA. For the EMF2.7/EMF2.3 fragment, forward sequencing primers were EMF2.7, EMF2.1 and EMF2.2 while the reverse primers were EMF2.4 and EMF2.3. In the case of the EMF2-P1 and EMF2-P2 clone, the forward primers used were EMF2-P1, EMF2-P3 and EMF2-P5 and the reverse set consisted of EMF2-P2, EMF2-P6, EMF2-P7 and EMF2-P8. Finally sequencing of the EMF2-F1/EMF2-R4 fragment employed the forward primers EMF2-F1, EMF2-F2, EMF2-F3, EMF2-F4 and EMF2-F5 and the reverse primers EMF2-R1, EMF2-R2, EMF2-R3 and EMF2-R4.

Comparison of the sequences from the *moe* and Ws wild-type genome revealed no change in the fragment EMF2-P1/EMF2-P2 and EMF2-F1/EMF2-R4. However, the EMF2.7/EMF2.3 PCR fragment from the *moe* DNA appeared smaller on gels suggesting that it might harbour

a deletion relative to wild-type. This was confirmed by sequence analysis which showed that the *moe* sequence did contain a small deletion when compared to the Ws sequence. Thus a 17 nucleotide bp region was deleted: this extended from the 3' end of the coding exon 2 into the 5' end of intron 2 with 9 bp of the deletion being in the exon sequence and the other 8 bp in the intron sequence. This was accompanied by a mutation from cytosine to guanine at the following residue (Figure 3.17). This deletion and the single nucleotide change could be the consequences of spontaneous mutations or alternatively it could be induced by the abortive T-DNA insertion.

As the splice donor site at the exon2-intron2 junction would have been eliminated by the deletion, it was highly likely that splicing at this particular junction would have been disrupted or abolished and alternative splice sites might have been used instead. As a consequence, the *moe* cDNA could differ from the wild-type sequence due to an incorporation of extra intron sequences or a loss of small coding exon sequences. In order to gain more insights into the possibility that *moe* is a splicing mutant, splicing of the *EMF2* transcript in the *moe* background was investigated and compared to the wild-type. Thus sequencing of *EMF2* cDNA in the *moe* and Ws background was carried out. Total RNA was prepared from young leaf tissue of *moe* and Ws plants, and RT-PCR was subsequently performed using the primer EMF2.7 and EMF2.3 pair and also the EMF2.1 and EMF2.3 pair which cover the 5' region of the *EMF2* containing the *moe* lesion.

The PCR products were run on an agarose gel and as expected several different splicing products were seen in the *moe* cDNA sample compared to wild-type (Figure 3.16). These splicing products were cloned into the Invitrogen pCR®II TOPO® vector and sequencing was then performed using the primer m13f which has a binding site in the vector.

Sequence analysis of the splicing products indicated that they could be put into 5 different classes, most of which resulted from the use of an alternative splice site (Figure 3.17). The most common group, which accounted for more than 50% of the transcripts analysed, used an alternative gt splice acceptor site which lies at 34 nucleotides upstream of the ATG start site (No.1 in figure 3.17). This would have led to the majority of the exon 2 including the translation initiation codon being spliced out. It is therefore likely that the translation machinery would use an alternative translation initiation codon. Analysis of the downstream sequence revealed the presence of an alternative in-frame ATG translational start site at the

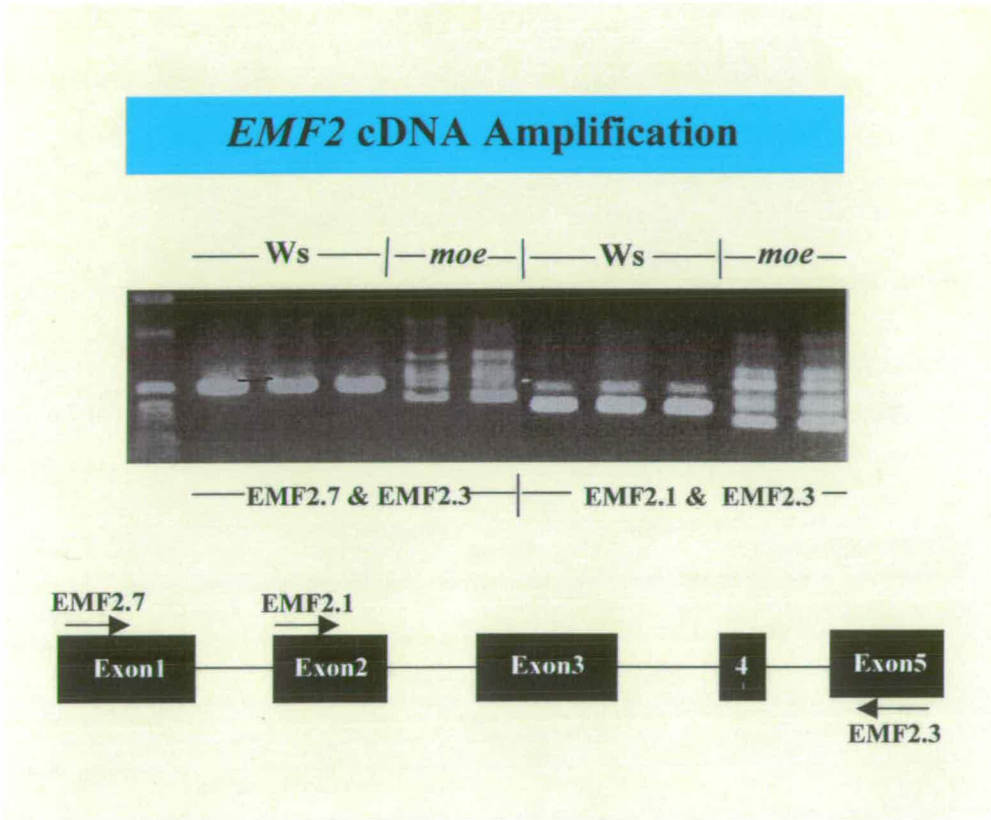


Figure 3.16 RT-PCR analysis of *EMF2* transcripts in *moe leaf* and wild-type background. Total RNA was extracted from *moe* and wild-type seedlings and reverse transcribed. *EMF2* cDNAs were amplified using two sets of primers; EMF2.7/EMF2.3 and EMF2.1/EMF2.3. With both primer sets, several different transcripts were amplified from the *moe* sample compared with one from the wild-type tissue suggesting that *moe* is a splicing allele. The upper fainter bands in the Ws lanes represent amplification of genomic DNA which presumably only accounts for one of the band in the *moe* lanes.

5' end of exon 3 (highlighted in black in figure 3.17). If this ATG was indeed used instead of the spliced ATG in exon 2, the protein produced would be 17 amino acid shorter at the N-terminus. The shortened EMF2 protein produced could be nearly as functionally efficient as the wild-type protein since no conserved or functional motif is found in this short stretch of amino acid residues that was missing from the protein.

In the second class of *EMF2* transcripts, the splicing machinery chose a splice site lying just downstream of the ATG start site (No.2 in figure 3.17). This would have resulted in another 11 nucleotides of DNA being spliced out. Thus together with the original 9 bp deletion, 20 nucleotides of DNA at the end of exon 2 would have been eliminated in the splicing process. This inevitably led to a translational frameshift and the peptide produced is predicted to have contained numerous translational stop codons at the N-terminal region. Consequently, the corresponding protein would have been truncated and therefore non-functional.

Similar scenarios happened with the third and fourth classes of the *moe* EMF cDNA. In these 2 classes, 2 different alternative splice sites in the intron 2 were used instead (No. 3 and 4 in figure 3.17), both of which resulted in extra intron sequences being incorporated into the transcript. Thus although a 9 bp sequence was deleted in the exon 2 of the genomic DNA, another 26 and 66 nucleotides would have been gained in the third and fourth class cDNA respectively. For the third class, this again would have led to the inevitable translational frameshift and hence protein truncation. In the fourth class of cDNA even though the gain of DNA sequence was in frame, the protein produced would have also been truncated and hence non-functional since an in-frame 'TAA' stop codon was incorporated into the sequence. Splicing of intron 2 was skipped in the last class of the *EMF2* transcript in the *moe* background. This means that an extra 163 nucleotide stretch of the intron 2 DNA sequence would have been incorporated into the cDNA sequence. This would have also led to translational frameshifts with premature stop codons being incorporated resulting in protein truncation.

Although the nature of the different EMF2 proteins produced in the *moe* background is not yet known, the weak *moe* phenotype could be explained in part by the varying amounts of the several different peptides produced. Thus as previously mentioned, the slightly shorter EMF2 peptide as coded for by the most common class of cDNA might be as efficient as the wild-type protein. As this alternative version of EMF2 cDNA accounted for 30-50% of the

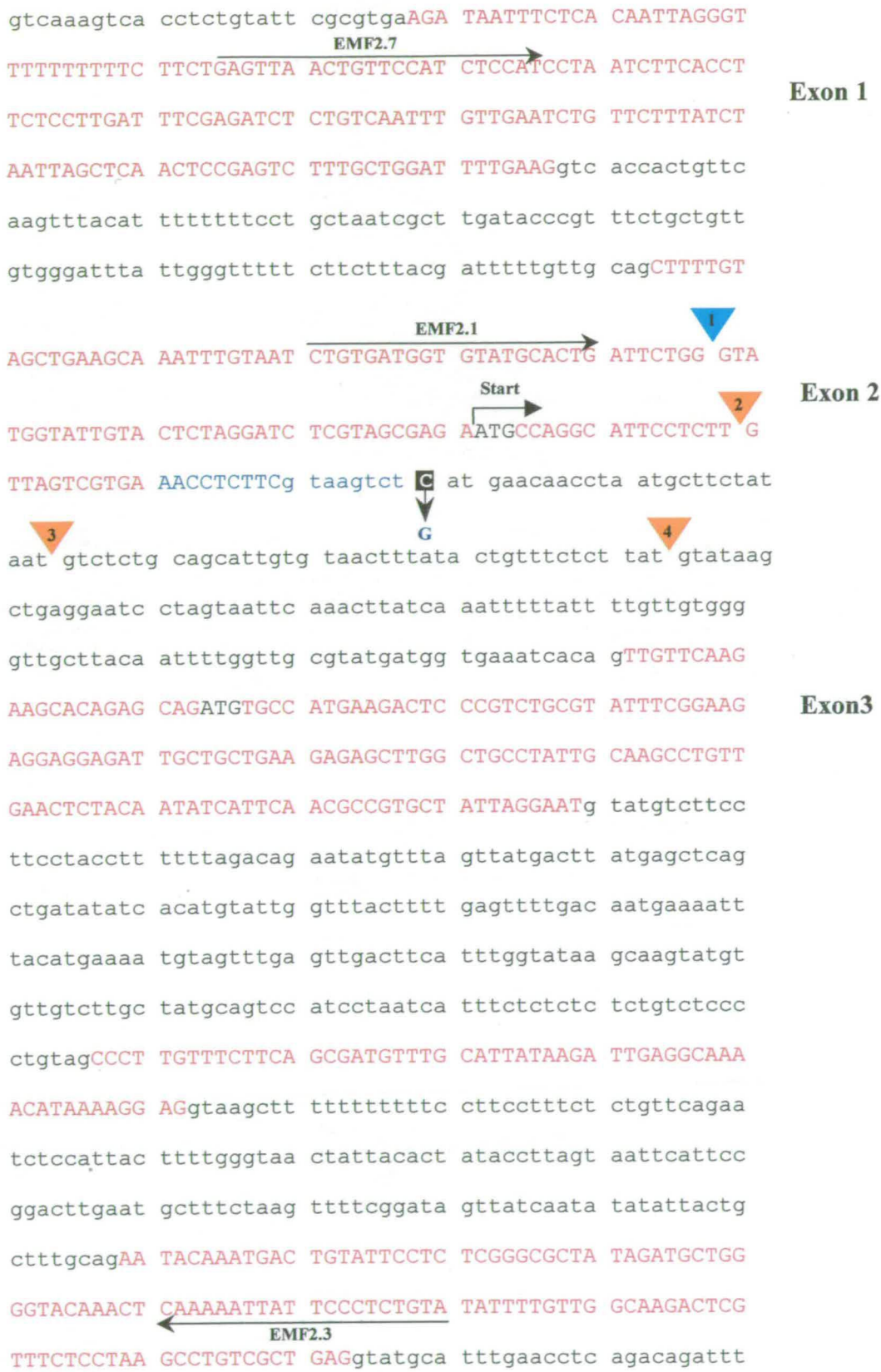


Figure 3.17 Molecular basis of *moe leaf* allele. Sequencing of *moe* genomic DNA revealed a 17 bp lesion at the exon2-intron2 splice junction followed by a nucleotide substitution from cytosine to guanine residue in the *EMF2* gene. Analysis of the *moe EMF2* cDNA transcripts indicated that *moe* is a splicing mutant with 5 different splicing products detected. In the most common transcript, an alternative splice junction in the untranslated region of exon 2 is used (upside-down green triangle) causing the splicing of the ATG start site. It is likely that a downstream ATG start codon in exon 3 is used instead (highlighted in black) resulting in a slightly truncated EMF2 protein. In three other classes of transcripts alternative splice junctions are used (orange upside-down triangles) all resulting in premature protein truncation. In the final class of *EMF2* transcript in the *moe* background, no splicing takes place and the whole intron two is incorporated into the peptide sequence resulting also in protein truncation.

transcript produced judging by the relative numbers of clones containing the corresponding transcript and the other transcripts, most of the aspects of *EMF2* function would have been mildly affected, if at all. This is in accordance with the fact that the *moe* phenotype is considerably weaker than the weakest published *emf2-3* allele of *emf2* mutants. It is possible that there is a threshold level of the amount of functional EMF2 proteins required for proper development and when this is disturbed as in many of the *emf2* alleles, several of the developmental pathways that *EMF2* regulate would be simultaneously disrupted resulting in severe developmental abnormalities. In the *moe* mutant the amount of the functional protein might hover at about or just slightly below this threshold value and the overall effect on the phenotype is, therefore, much milder.

Chapter 4. Characterisation of an *Arabidopsis CURLY LEAF* Homologue: *CURLY LEAF LIKE1*

4.1 Introduction

4.2 *CURLY LEAF LIKE1* is an *Arabidopsis* Homologue of *CURLY LEAF*

4.3 *CLK1* Expression Pattern

4.4 CLK1 Protein Interactions

4.5 Isolation of *clk1* Mutant Alleles

4. Characterisation of an *Arabidopsis CURLY LEAF* homologue: *CURLY LEAF LIKE1*

4.1 Introduction

The *Arabidopsis* genome sequencing project was recently completed. Since then a large amount of effort has been made to analyse the integrated sequence databases. The information obtained will provide an invaluable contribution to the ultimate goal of comprehensive plant functional genomics. The ongoing characterisation of the published sequence has revealed one striking, if not unexpected, fact. Thus despite its small size, a vast majority of the *Arabidopsis* genome (more than 60%) is covered by large-scale duplications of chromosomal fragments and tandem gene duplications (Blanc et al., 1999). As a result of the high number of genome duplication events and subsequent gene divergence, many genes in *Arabidopsis* exist in big gene families such as the KNAT homeodomain and the MADS-box group. In many cases wholesale gene duplication and the subsequent sequence divergence provided a mechanism for the evolution of new functions and hence a new level of complexity. On the other hand, a large number of duplicated genes have been found to display complete or partial redundant genetic function. As a consequence, the loss-of-function of several regulatory genes can be partly or completely compensated for by the activities of closely related genes. For instance, redundant activities of the MADS-box gene *APETALA1*, *CAULIFLOWER* and *FRUITFULL* regulate the inflorescence architecture during *Arabidopsis* reproductive development. A mutation in *CAL* has no phenotypic consequence on its own but results in a dramatic enhancement of the flower-to-shoot transformation of the *ap1* mutant. More strikingly, flowering is completely abolished when *AP1*, *CAL* and *FUL* activities are knocked out simultaneously (Ferrándiz et al., 2000). The fact that numerous developmental pathways are potentially regulated by redundant gene functions suggests that several aspects of the function of a given gene might not be observed in a single mutant background and will only become apparent when several or all other genes in the same family are knocked out as well.

Since the complete *Arabidopsis* genome sequence has been made accessible to the public, it is therefore possible to search against the databases for homologues of a gene of interest. Reverse genetic approaches can then be employed to generate loss-of-function alleles of all the genes in the family. Subsequent manipulations such as double or triple mutant crosses

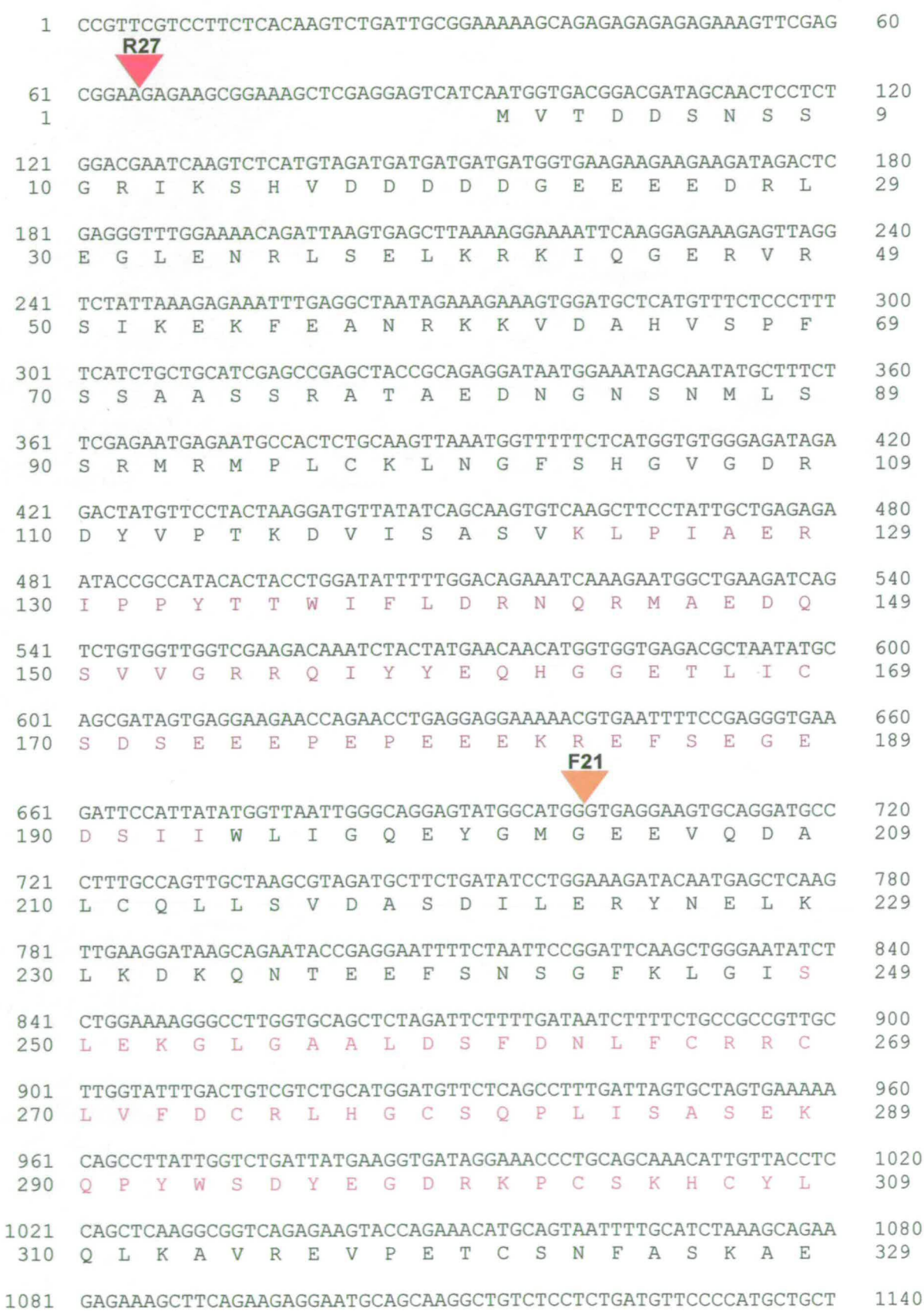
would enable phenotypic assessment of various mutant combinations. Not only may this help untangle the genetic hierarchy in the regulatory pathway that those genes are involved in, but it might also reveal some novel functional properties of the genes that are not previously evident.

In this study several reverse genetic approaches were employed in an attempt to characterise a *CLF* homologue. High sequence homology between the two genes and the proteins that they encode suggests that this gene is likely to act closely and perhaps redundantly with *CLF* during *Arabidopsis* development. The data obtained from the genetic and molecular analysis of the gene could therefore not only provide information on the function of this gene but it might also give invaluable insights into some aspects of the *CLF* function which cannot be elucidated by the analysis of *clf* single mutants.

4.2 *CURLY LEAF LIKE1* is an *Arabidopsis* homologue of *CURLY LEAF*

Database searches against the *Arabidopsis* genome sequence identified a homologue of the *CLF* gene. This gene, namely, *CURLY LEAF LIKE1* (*CLK1*), also known as *EZA1*, is located on the BAC clone T10M13 on the upper arm of chromosome 4 and contains the RFLP marker MI390 which has been mapped to 11.89 cM on the Lister & Dean RI map.

DNA sequence alignment indicated a high degree of similarity between the *CLK1* and *CLF* gene with 45% of the nucleotide sequences of the genes being identical. Because results from gene prediction programmes are often inaccurate and there was no *CLK1* EST sequence in the database, it was necessary to isolate the *CLK1* cDNA. Thus an RT-PCR experiment was performed on total RNA extracted from leaf tissue using the primers YC28 & YC29 which bind to the sequence nearby the predicted translation start codon and the stop codon respectively. Both strands of the cDNA were then sequenced. Finally the 5' UTR of the *CLK1* transcript was amplified in a 5' RACE experiment using the primer JG17 and its nested primer JG17N. It appeared that the *CLK1* cDNA sequence contained a long open reading frame and a short 5' UTR with the translational start codon ATG at position +93 (Figure 4.1). Strikingly, comparison of the *CLF* and *CLK1* cDNA and genomic sequences revealed near identical intron/exon structures. Not surprisingly, therefore, the corresponding CLF and CLK1 proteins exhibited extensive sequence homology throughout their entire lengths (51% identity and 59% similarity) (Figure 4.2). Thus protein alignment showed that


R27
F21

```
 330  E  K  A  S  E  E  E  C  S  K  A  V  S  S  D  V  P  H  A  A    349

1141  GCTAGTGGTGTCAGTCTGCAAGTTGAGAAGACTGATATTGGTATCAAGAATGTAGATTCA  1200
 350  A  S  G  V  S  L  Q  V  E  K  T  D  I  G  I  K  N  V  D  S    369

1201  TCCTCTGGTGTAGAACAAGAGCATGGAATTAGAGGAAAGCGTGAGGTCCCAATTCTAAAA  1260
 370  S  S  G  V  E  Q  E  H  G  I  R  G  K  R  E  V  P  I  L  K    389

1261  GACTCCAATGATCTGCCTAATTTATCGAACAAGAAACAGAAGACCGCAGCCTCAGATACA  1320
 390  D  S  N  D  L  P  N  L  S  N  K  K  Q  K  T  A  A  S  D  T    409

1321  AAAATGTCATTTGTTAATTCTGTCCCTAGCTTAGATCAGGCATTGGATAGCACAAAGGGT  1380
 410  K  M  S  F  V  N  S  V  P  S  L  D  Q  A  L  D  S  T  K  G    429

1381  GATCAAGGTGGAACAACTGACAATAAAGTAAACAGAGACTCAGAAGCTGATGCAAAAGAA  1440
 430  D  Q  G  G  T  T  D  N  K  V  N  R  D  S  E  A  D  A  K  E    449

1441  GTAGGTGAGCCTATTCCAGACAATTCGGTCCATGATGGTGGTTCCTCAATTTGTCAGCCA  1500
 450  V  G  E  P  I  P  D  N  S  V  H  D  G  G  S  S  I  C  Q  P    469

1501  CACCATGGTAGTGGAAACGGAGCAATAATCATTGCAGAAATGTCTGAGACAAGTCGACCA  1560
 470  H  H  G  S  G  N  G  A  I  I  I  A  E  M  S  E  T  S  R  P    489

1561  TCTACAGAGTGGAATCCTATCGAGAAGGATCTTTACTTGAAGGGAGTCGAAATCTTTGGA  1620
 490  S  T  E  W  N  P  I  E  K  D  L  Y  L  K  G  V  E  I  F  G    509

1621  AGAAACAGCTGTCTTATTGCAAGAAACCTGCTTTCTGGCTTGAAGACATGCCTAGATGTG  1680
 510  R  N  S  C  L  I  A  R  N  L  L  S  G  L  K  T  C  L  D  V    529

1681  TCCAATTACATGCGTGAAAACGAAGTTTCAGTTTTTCGAAGATCTAGTACCCCAAATTTG  1740
 530  S  N  Y  M  R  E  N  E  V  S  V  F  R  R  S  S  T  P  N  L    549

1741  CTGTTGGATGATGGCAGGACTGACCCAGGGAATGATAATGATGAGGTGCCTCCAAGGACA  1800
 550  L  L  D  D  G  R  T  D  P  G  N  D  N  D  E  V  P  P  R  T    569

1801  AGATTGTTCCGTAGAAAAGGCAAAACCCGGAAGCTAAAATACTCTACAAAGTCTGCTGGT  1860
 570  R  L  F  R  R  K  G  K  T  R  K  L  K  Y  S  T  K  S  A  G    589

1861  CATCCGTCTGTCTGGAAAAGAATAGCTGGTGGCAAAAACCAGTCCTGTAAACAATACACG  1920
 590  H  P  S  V  W  K  R  I  A  G  G  K  N  Q  S  C  K  Q  Y  T    609

1921  CCGTGTGGATGCCTGTCAATGTGCGGAAAGGATTGCCCTTGTCTAACTAATGAAACTTGC  1980
 610  P  C  G  C  L  S  M  C  G  K  D  C  P  C  L  T  N  E  T  C    629

1981  TGCGAGAAATATTGCGGGTGCTCAAAAAGCTGTAAAAATCGTTTCCGAGGATGTCATTGT  2040
 630  C  E  K  Y  C  G  C  S  K  S  C  K  N  R  F  R  G  C  H  C    649

2041  GCAAAGAGTCAATGCAGAAGTAGGCAGTGTCCCTGCTTTGCTGCTGGCAGAGAATGTGAT  2100
 650  A  K  S  Q  C  R  S  R  Q  C  P  C  F  A  A  G  R  E  C  D    669

2101  CCAGATGTTTGCAGAAATTGCTGGGTTAGTTGTGGAGATGGTTCTCTCGGTGAAGCACCA  2160
 670  P  D  V  C  R  N  C  W  V  S  C  G  D  G  S  L  G  E  A  P    689

2161  AGACGCGGAGAAGGGCAATGCGGAAACATGAGACTTCTCCTGAGGCAACAACAGAGGATC  2220
 690  R  R  G  E  G  Q  C  G  N  M  R  L  L  L  R  Q  Q  Q  R  I    709
```

```
2221  CTATTGGGAAAGTCTGATGTTGCTGGATGGGGTGCTTTTCTAAAGAACTCGGTCAGCAAA  2280
 710  L  L  G  K  S  D  V  A  G  W  G  A  F  L  K  N  S  V  S  K    729

                          V716N

2281  AATGAATACCTTGGAGAATACACCGGTGAATTGATCTCACACCATGAGGCGGATAAGCGT  2340
 730  N  E  Y  L  G  E  Y  T  G  E  L  I  S  H  H  E  A  D  K  R    749

2341  GGGAAAATATATGACCGGGCAAATTCGTCCTTCCTCTTTGACTTGAATGATCAGTACGTC  2400
 750  G  K  I  Y  D  R  A  N  S  S  F  L  F  D  L  N  D  Q  Y  V    769

2401  CTCGATGCTCAACGCAAAGGTGACAAGCTGAAATTTGCCAATCACTCAGCTAAACCCAAT  2460
 770  L  D  A  Q  R  K  G  D  K  L  K  F  A  N  H  S  A  K  P  N    789

2461  TGCTACGCTAAGGTGATGTTTGTAGCAGGAGATCACAGGGTCGGGATTTTTGCAAACGAA  2520
 790  C  Y  A  K  V  M  F  V  A  G  D  H  R  V  G  I  F  A  N  E    809

2521  CGAATAGAAGCTAGCGAAGAGCTTTTCTATGACTATAGATATGGACCAGACCAAGCACCA  2580
 810  R  I  E  A  S  E  E  L  F  Y  D  Y  R  Y  G  P  D  Q  A  P    829

2581  GTGTGGGCTCGCAAACCTGAAGGCTCCAAGAAAGATGATTCAGCCATTACTCATCGTAGA  2640
 830  V  W  A  R  K  P  E  G  S  K  K  D  D  S  A  I  T  H  R  R    849

2641  GCCAGAAAGCACCAATCTCATTGATGATTACTGGCTAAGAGAAGTAACTTTTATAAAAAT  2700
 850  A  R  K  H  Q  S  H  *                                        857

2701  AACTTATAGAGTTGTGAGAGATGATATTTGAAGTTTGATAACTTAAGCTTGTCTTTATTA  2760

2761  ATTAATTATTATAGAGTTGAGATTTTATTTTATTTTGACATCGAGTTTGGACTTTGTATA  2820

2821  GGTGATAAAACAATTTATGAATTATTGGGGTCAATAAGTAAAAATGTATCATTTCGAAAA  2880

2881  ATAGTTTCAAATTTAT  2896
```

Figure 4.1 The cDNA and predicted amino acid sequence of *CLK1*. The *CLK1* cDNA sequence is shown (top) together with the predicted amino acid sequence (bottom). The amino acids comprising the N-terminal acidic domain are highlighted in violet while those in the cysteine-rich domain, the SANT domain, the NLS, the CXC domain and the SET domain are indicated in pink, red, green, dark blue and blue respectively. The valine residue at position 716 which was mutated to asparagine in the *in vitro* mutagenesis experiment is circled in yellow. The position of the T-DNA insertion F21 and R27 in the Wisconsin PCR screen are marked by the orange and purple upside-down triangle respectively.

the 856-amino-acid long CLK1 protein shared several conserved domains with the CLF protein. Firstly, the carboxy terminus of the CLK1 protein (residues 714-827) contained the previously mentioned SET domain. Sequence comparison showed that this 115-amino-acid region of the CLK1 protein was 84% identical to the CLF SET domain (Figure 4.2). Multiple protein alignment revealed that the SET domain is also very highly conserved in all other plant E(Z) homologues (Figure 4.3). Secondly, as in the CLF protein, two cysteine-rich domains were present in the CLK1 protein. The first of these was found near the N-terminus (residues 248-309) whereas the large CXC domain lied immediately N-terminal to the SET domain at the carboxyl end of the protein (residues 611-676). These domains were 61% and 86% identical to the corresponding regions in the CLF protein respectively (Figure 4.2). It is noteworthy that the spacing of the cysteine residues was highly conserved within this group of the SET domain protein family which includes CLF, CLK1, MEA, *Drosophila* E(Z) and the two CLF *Antirrhinum* homologues; AmCLF1 and AmCLF2 (Figure 4.3). Moreover, the arrangement of these cysteine residues in E(Z) and its five plant E(Z) homologues seemed to be unique and different from any other family of cysteine-rich zinc finger motifs. Analysis of the phylogenetic tree generated using the sequences of the CXC domains which displayed highest degree of homology among all the conserved regions clearly showed that *CLK1* and *CLF* genes are very closely related and likely to have evolved from a duplication event (Figure 4.4). The tree also indicated that the sequence of the other *Arabidopsis CLF* homologue; *MEA*, is more evolutionary diverged from both *CLF* and *CLK1*. Therefore *MEA* could have evolved in a more ancient gene duplication which also gave rise to the progenitor sequence of *CLF* and *CLK1*.

In addition to the 3 domains mentioned above, the CLK1 protein also shared the previously mentioned SANT domain with CLF and other E(Z) homologues (Figure 4.3). This domain lied N-terminal to the CXC domain (residues 493-533) and was found to be 78% identical to the CLF SANT domain (Figure 4.2). Sequence alignment also identified another relatively well conserved short stretch of amino acids between the SANT and CXC domain. This basic domain (residues 566-582) was likely a bitartite nuclear localisation signal (NLS) and was found in a similar position to the putative NLS in both the E(Z) and CLF protein (Figure 4.2). Finally an acidic domain was found near the N-terminus of the CLK1 protein (residues 121-194) (Figure 4.3). Again the degree of sequence homology was very high between this domain and the corresponding region of the CLF protein with 69% of the amino acids of the two domains being identical (Figure 4.2). Interestingly, it was observed that the distribution pattern of the charged residues seemed to be evolutionary conserved in the other E(Z)

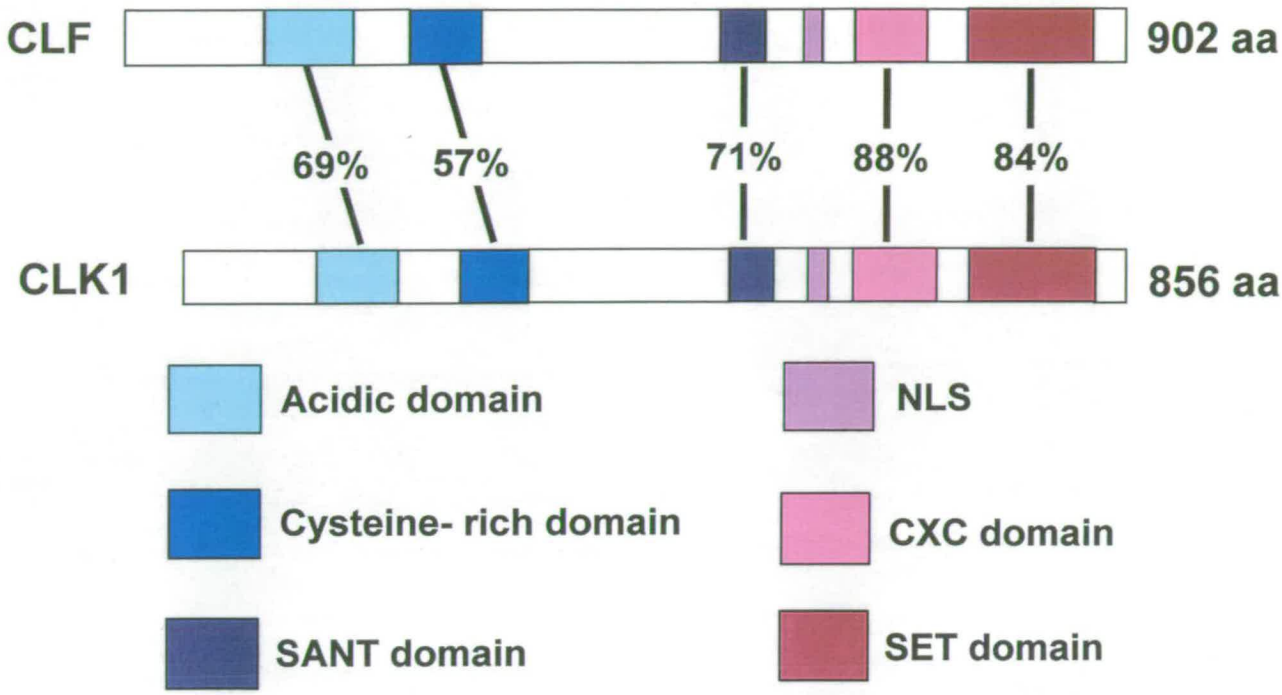
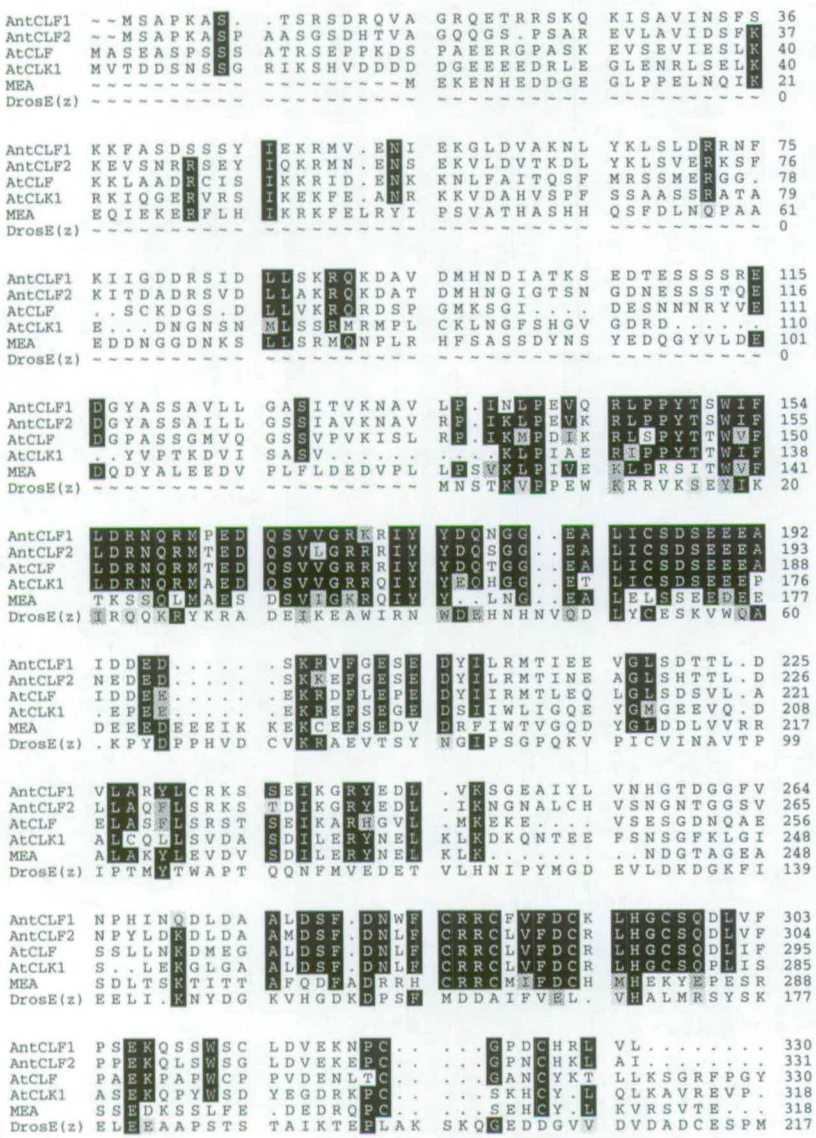
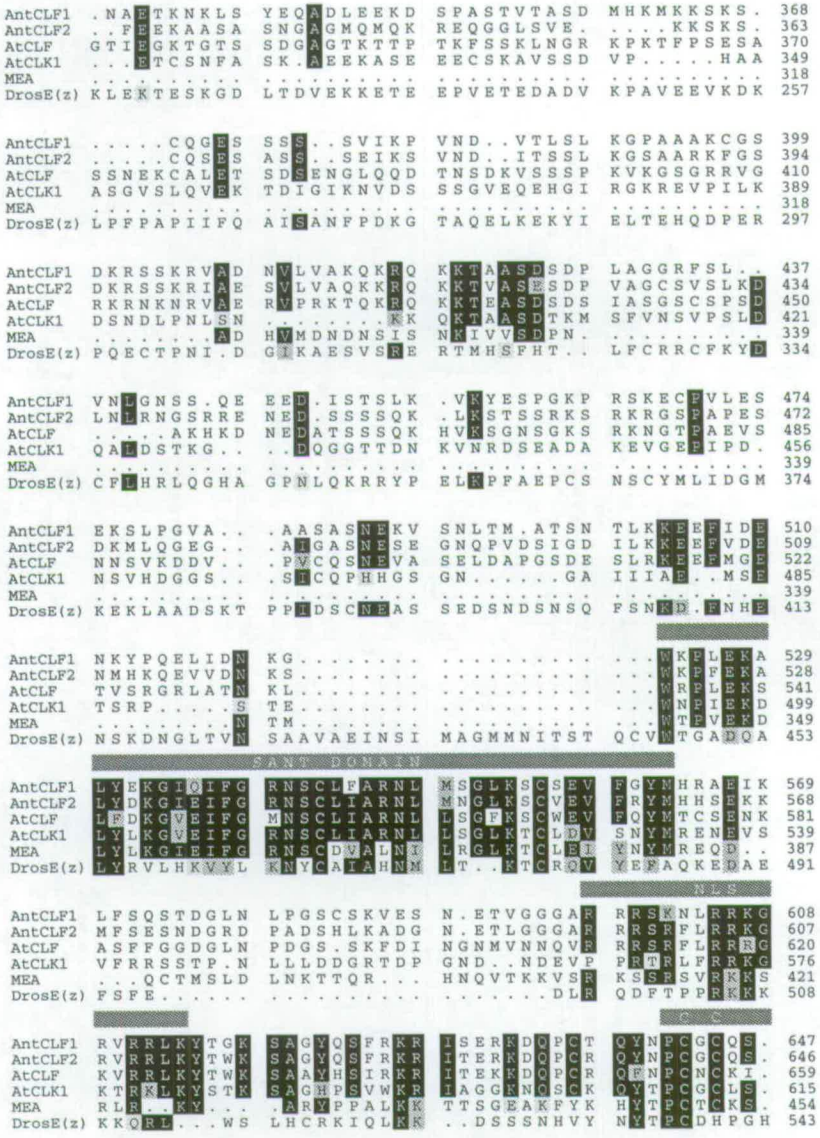


Figure 4.2 CLF and CLK1 protein domain comparison. The CLF and CLK1 protein sequences were aligned and the percentage amino acid identity within each conserved domain is indicated. The N-terminal acidic domain is boxed in light blue while the cysteine-rich domain, the SANT domain, the predicted NLS, the CXC domain and the SET domain is boxed in blue, dark blue, violet, pink and dark purple respectively.


AntCLF1 ~~MSAPKAS. .TSRSDRQVA GRQETRRSKQ KISAVINSFS 36
AntCLF2 ~~MSAPKASP AASGSDHTVA GQQGS.PSAR EVLAVIDSFK 37
AtCLF MASEASPSSS ATRSEPPKDS PAEERGPASK EVSEVIESLK 40
AtCLK1 MVTDDSNSSG RIKSHVDDDD DGEEEEDRLE GLENRLSELK 40
MEA ~~~~~~~~~~ ~~~~~~~~~M EKENHEDDGE GLPPELNQIK 21
DrosE(z) ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 0
AntCLF1 KKFASDSSSY IEKRMV.ENI EKGLDVAKNL YKLSLDRRNF 75
AntCLF2 KEVSNRRSEY IQKRMN.ENS EKVLDVTKDL YKLSVERKSF 76
AtCLF KKLAADRCIS IKKRID.ENK KNLFAITQSF MRSSMERGG. 78
AtCLK1 RKIQGERVRS IKEKFE.ANR KKVDAHVSPF SSAASSRATA 79
MEA EQIEKERFLH IKRKFELRYI PSVATHASHH QSFDLNQPAA 61
DrosE(z) ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 0
AntCLF1 KIIGDDRSID LLSKRQKDAV DMHNDIATKS EDTESSSSRE 115
AntCLF2 KITDADRSVD LLAKRQKDAT DMHNGIGTSN GDNESSSTQE 116
AtCLF ..SCKDGS.D LLVKRQRDSP GMKSGI.... GDESNNNRYVE 111
AtCLK1 E...DNGNSN MLSSRMRMPL CKLNGFSHGV GDRD..... 110
MEA EDDNGGDNKS LLSRMQNPLR HFSASSDYNS YEDQGYVLDE 101
DrosE(z) ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ 0
AntCLF1 DGYASSAVLL GASITVKNAV LP.INLPEVQ RLPPYTSWIF 154
AntCLF2 DGYASSAILL GSSIAVKNAV RP.IKLPEVK RLPPYTSWIF 155
AtCLF DGPASSGMVQ GSSVPVKISL RP.IKMPDIK RLSPYTTWVF 150
AtCLK1 ..YVPTKDVI SASV..... ...KLPIAE RIPPYTTWIF 138
MEA DQDYALEEDV PLFLDEDVPL LPSVKLPIVE KLPRSITWVF 141
DrosE(z) ~~~~~~~~~~ ~~~~~~~~~~ MNSTKVPPEW KRRVKSEYIK 20
AntCLF1 LDRNQRMPED QSVVGRKRIY YDQNGG..EA LICSDSEEEA 192
AntCLF2 LDRNQRMTED QSVLGRRRIY YDQSGG..EA LICSDSEEEA 193
AtCLF LDRNQRMTED QSVVGRRRIY YDQTGG..EA LICSDSEEEA 188
AtCLK1 LDRNQRMAED QSVVGRRQIY YEQHGG..ET LICSDSEEEP 176
MEA TKSSQLMAES DSVIGKRQIY Y..LNG..EA LELSSEEDEE 177
DrosE(z) IRQQKRYKRA DEKEAWIRN WDEHNHNVQD LYCESKVWQA 60
AntCLF1 IDDED..... .SKRVFGESE DYILRMTIEE VGLSDTTL.D 225
AntCLF2 NEDED..... .SKKEFGESE DYILRMTINE AGLSHTTL.D 226
AtCLF IDDEE..... .EKRDFLEPE DYIIRMTLEQ LGLSDSVL.A 221
AtCLK1 .EPEE..... .EKREFSEGE DSIIWLIGQE YGMGEEVQ.D 208
MEA DEEEDEEEIK KEKCEFSEDV DRFIWTVGQD YGLDDLVVRR 217
DrosE(z) .KPYBPPHVD CVKRAEVTSY NGIPSGPQKV PICVINAVTP 99
AntCLF1 VLARYLCRKS SEIKGRYEDL .VKSGEAIYL VNHGTDGGFV 264
AntCLF2 LLAQYLSRKS TDIKGRYEDL .IKNGNALCH VSNGNTGGSV 265
AtCLF ELASFLSRST SEIKARHGVL .MKEKE.... VSESGDNQAE 256
AtCLK1 ALCQLLSVDA SDILERYNEL KLKDKQNTEE FSNSGFKLGI 248
MEA ALAKYLEVDV SDILERYNEH KLK..... ..NDGTAGEA 248
DrosE(z) IPTMYTWAPT QQNFMVEDET VLHNIPYMGD EVLDKDGKFI 139
AntCLF1 NPHINQDLDA ALDSF.DNWF CRRCFVFDCK LHGCSQDLVF 303
AntCLF2 NPYLDKDLDA AMDSF.DNLF CRRCLVFDCR LHGCSQDLVF 304
AtCLF SSLLNKDMEG ALDSF.DNLF CRRCLVFDCR LHGCSQDLIF 295
AtCLK1 S..LEKGLGA ALDSF.DNLF CRRCLVFDCR LHGCSQPLIS 285
MEA SDLTSKTITT AFQDFADRRH CRRCMIFDCH MHEKYEPESR 288
DrosE(z) EELI.KNYDG KVHGDKDPSF MDDAIFVELL VHALMRSYSK 177
AntCLF1 PSEKQSSWSC LDVEKNPC.. ...GPDCHRL VL..... 330
AntCLF2 PPEKQLSWSG LDVEKEPC.. ...GPNCHKL AI..... 331
AtCLF PAEKPAPWCP PVDENLTC.. ...GANCYKT LLKSGRFPGY 330
AtCLK1 ASEKQPYWSD YEGDRKPC.. ...SKHCYIL QLKAVREVP. 318
MEA SSEDKSSLFE .DEDRQPC.. ...SEHCYIL KVRSVTE... 318
DrosE(z) ELEEAAPSTS TAIKTEPLAK SKQGEDDGVV DVDADCESPM 217


AntCLF1 .NAETKNKLS YEQADLEEKD SPASTVTASD MHKMKKSKS. 368
AntCLF2 ..FEEKAASA SNGAGMQMQK REQGGLSVE. ....KKSKS. 363
AtCLF GTIEGKTGTS SDGAGTKTTP TKFSSKLNGR KPKTFPSESA 370
AtCLK1 ...ETCSNFA SK.AEEKASE EECSKAVSSD VP.....HAA 349
MEA 318
DrosE(z) KLEKTESKGD LTDVEKKETE EPVETEDADV KPAVEEVKDK 257
AntCLF1 .....CQGES SSS..SVIKP VND..VTLSL KGPAAAKCGS 399
AntCLF2 .....CQSES ASS..SEIKS VND..ITSSL KGSAARKFGS 394
AtCLF SSNEKCALET SDSENGLQQD TNSDKVSSSP KVKGSGRRVG 410
AtCLK1 ASGVSLQVEK TDIGIKNVDS SSGVEQEHGI RGKREVPILK 389
MEA 318
DrosE(z) LPFPAPIIFQ AISANFPDKG TAQELKEKYI ELTEHQDPER 297
AntCLF1 DKRSSKRVAD NVLVAKQKRQ KKTAASDSDP LAGGRFSL.. 437
AntCLF2 DKRSSKRIAE SVLVAQKKRQ KKTVASESDP VAGCSVSLKD 434
AtCLF RKRNKNRVAE RVPRKTQKRQ KKTEASDSDS IASGSCSPSD 450
AtCLK1 DSNDLPNLSN ....KK QKTAASDTKM SFVNSVPSLD 421
MEA ....AD HVMDNDNSIS NKIVVSDPN. 339
DrosE(z) PQECTPNI.D GIKAESVSRE RTMHSFHT.. LFCRRCFKYD 334
AntCLF1 VNLGNSS.QE EED.ISTSLK .VKYESPGKP RSKECHVLES 474
AntCLF2 LNLRNGSRRE NED.SSSSQK .LKSTSSRKS RKRGSPAPES 472
AtCLF .....AKHKD NEDATSSSQK HVKSGNSGKS RKNGTPAEVS 485
AtCLK1 QALDSTKG.. ..DQGGTTDN KVNRDSEADA KEVGEHIPD. 456
MEA 339
DrosE(z) CFLHRLQGHA GPNLQKRRYP ELKPFAEPCS NSCYMLIDGM 374
AntCLF1 EKSLPGVA.. .AASASNEKV SNLTM.ATSN TLKKEEFIDE 510
AntCLF2 DKMLQGEG.. .AIGASNESE GNQPVDSIGD ILKKEEFVDE 509
AtCLF NNSVKDDV.. .PVCQSNEVA SELDAPGSDE SLRKEEFMGE 522
AtCLK1 NSVHDGGS.. .SICQPHHGS GN......GA IIIAE..MSE 485
MEA 339
DrosE(z) KEKLAADSKT PPIDSCNEAS SEDSNDSNSQ FSNKD.FNHE 413
SANT DOMAIN
AntCLF1 NKYPQELIDN KG.......... ....WKPLEKA 529
AntCLF2 NMHKQEVVDN KS.......... ....WKFFEKA 528
AtCLF TVSRGRLATN KL.......... ....WRPLEKS 541
AtCLK1 TSRP.....S TE.......... ....WNPIEKD 499
MEA ....N TM.......... ....WTPVEKD 349
DrosE(z) NSKDNGLTVN SAAVAEINSI MAGMMNITST QCVWTGADQA 453
SANT DOMAIN
AntCLF1 LYEKGIQIFG RNSCLFARNL MSGLKSCSEV FGYMHRAEIK 569
AntCLF2 LYDKGIEIFG RNSCLIARNL MNGLKSCVEV FRYMHHSEKK 568
AtCLF LFDKGVEIFG MNSCLIARNL LSGFKSCWEV FQYMTCSENK 581
AtCLK1 LYLKGVEIFG RNSCLIARNL LSGLKTCLDV SNYMRENEVS 539
MEA LYLKGIEIFG RNSCDVALNI LRGLKTCLEI YNYMREQD.. 387
DrosE(z) LYRVLHKVYL KNYCAIAHNM LT..KTCRQV YEFAQKEDAE 491
NLS
AntCLF1 LFSQSTDGLN LPGSCSKVES N.ETVGGGAR RRSKNLRRKG 608
AntCLF2 MFSESNDGRD PADSHLKADG N.ETLGGGAR RRSFFLRRKG 607
AtCLF ASFFGGDGLN PDGS.SKFDI NGNMVNNQVR RRSRFLRRRG 620
AtCLK1 VFRRSSTP.N LLLDDGRTDP GND..NDEVP PRTRLFRRKG 576
MEA ...QCTMSLD LNKTTQR... HNQVTKKVSR KSSRSVRKKS 421
DrosE(z) FSFE..... ....DLR QDFTPPRKKK 508
C C
AntCLF1 RVRRLKYTGK SAGYQSFRKR ISERKDQPCT QYNPCGCQS. 647
AntCLF2 RVRRLKYTWK SAGYQSFRKR ITERKDQPCR QYNPCGCQS. 646
AtCLF KVRRLKYTWK SAAYHSIRKR ITEKKDQPCR QFNPCNCKI. 659
AtCLK1 KTRKLKYSTK SAGHPSVWKR IAGGKNQSCK QYTPCGCLS. 615
MEA RLR..KY... .ARYPPALKK TTSGEAKFYK HYTPCTCKS. 454
DrosE(z) KKQRL...WS LHCRKIQLKK ..DSSSNHVY NYTPCDHPGH 543

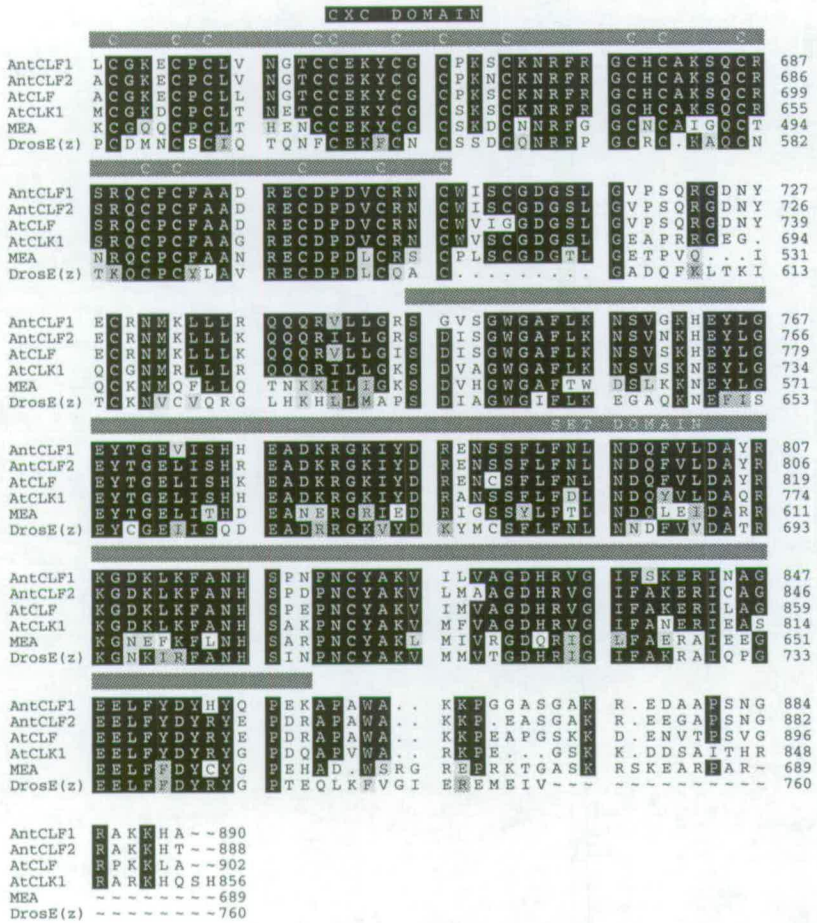


Figure 4.3 Protein sequence alignment of plant E(Z)-type SET domain proteins. CLK-1 protein sequenced is aligned with sequences of four other plant E(Z)-type SET domain proteins; the *Arabidopsis* CLF, MEA and the *Antirrhinum* CLF1 and CLF2, and also the *Drosophila* E(Z) protein. Conserved domains are indicated. Amino acid residues which are identical in the majority of these protein sequences are boxed in black and those which show conserved changes are boxed in dark grey. The conserved spacing of the cysteine residues in the CXC domain is also indicated.

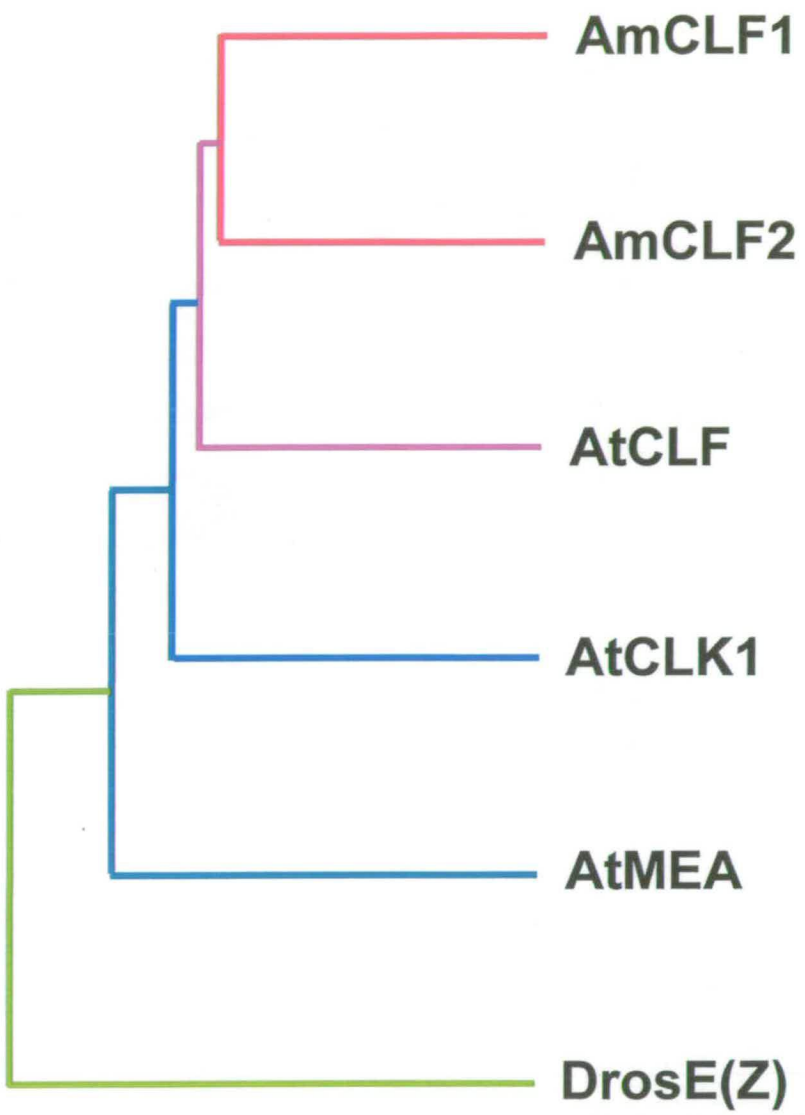


Figure 4.4 Phylogenetic tree of the E(Z)-type SET domain proteins. The tree was constructed based on the calculation of the evolutionary distance between the proteins using the sequences of the highly conserved CXC domain. The programme used to construct the tree is the PILEUP program within the GCG10 package (Devereux et al., 1984).

homologues such as MEA and animal E(Z)-like proteins even though the DNA sequence is less well conserved in those proteins. As previously mentioned, the functional significance of this acidic region and the adjacent cysteine-rich region is still not yet known.

4.2 *CLK1* expression pattern

Analysis of a gene's expression pattern often provide useful insights into its function and regulation. It has often been noted that expression patterns of genes with shared function overlap with each other. Extensive sequence homology at both the nucleotide and protein level suggested that *CLK1* might be partially or completely functionally redundant with *CLF*. In order to obtain more information on the potential genetic redundancy between the *CLF* and *CLK1* pair, characterisation of the *CLK1* mRNA expression pattern was, therefore, carried out by *in situ* hybridisation.

Due to the high sequence homology shared between the *CLF* and *CLK1* genes, a relatively poorly conserved 700 bp fragment at the N-terminus was chosen as a probe for this experiment. To obtain this fragment, full length *CLK1* cDNA was first amplified by RT-PCR using the primers YC28 and YC29 and then subcloned into the polylinker of the Invitrogen pCR®II TOPO® vector. The recombinant plasmid was digested with the restriction enzyme *Sac*I and allowed to recircularise (Figure 4.5B). The resultant plasmid pCOCK1 was digested with the enzyme *Sac*I and *Eco*RI and the 0.7 kb *CLK1* insert was then subcloned into pBluescript vector (KS$^-$) (Strategene) in the 3' to 5' orientation to give the pCOCK3 plasmid (Figure 4.5C). The pCOCK3 plasmid DNA was extracted and purified and the DNA template for *in vitro* transcription was then PCR amplified using the primer pair M13f and M13r. *In vitro* transcription was carried out using the T7 DNA polymerase and T3 DNA polymerase for the antisense and sense probe respectively (Figure 4.5C).

The *CLK1* expression pattern was investigated in both vegetative and inflorescence tissues. Hybridisation of the labelled control RNA sense probe to sections of wild-type vegetative and inflorescence meristems gave very little signal (Figure 4.6B). On the contrary, when the antisense RNA probe was used, the *CLK1* mRNA could be detected in the vegetative and inflorescence meristem. The *CLK1* mRNA was detected from very early stages of leaf development through to later developmental stages. Thus signal was detected throughout the vegetative apical meristem, leaf primordia and young leaves (Figure 4.6A). Weak signal was

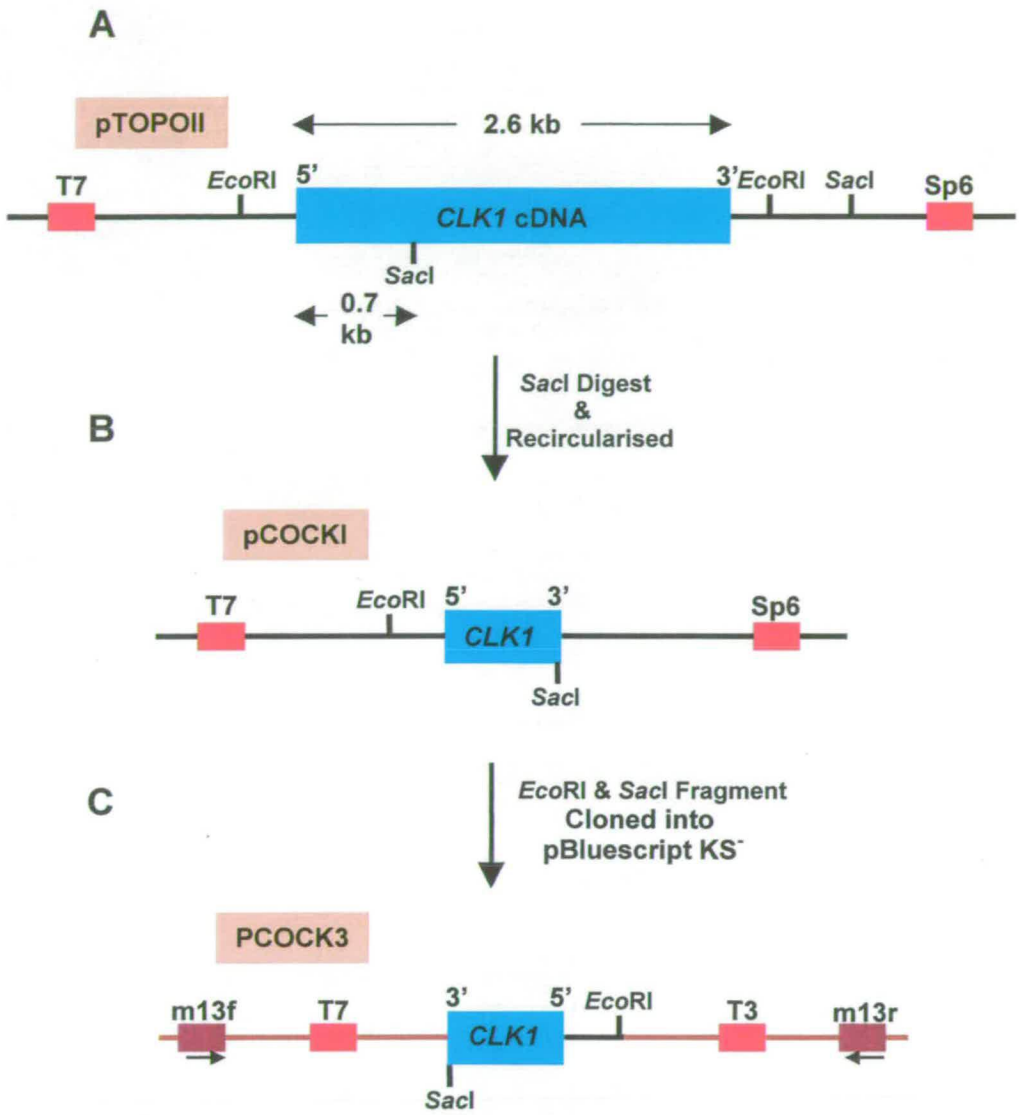


Figure 4.5 *In vitro* transcription of *CLK1* RNA. A, A 2.6 kb *CLK1* cDNA fragment amplified using the primers YC28 & YC29 was cloned into the polylinker of the pCRII TOPO vector. B, The recombinant pCRII TOPO vector containing the full length *CLK1* cDNA was double-digested with the enzyme *Sac*I and *Eco*RI and allowed to recircularise to give plasmid pCOCKI. C, The *Eco*RI-*Sac*I fragment of the insert in the pCOCKI plasmid was cloned into pBluescript KS$^-$ plasmid to give plasmid pCOCK3. The DNA template for *in vitro* transcription was amplified using the primer pair m13f and m13r. The antisense and sense probes were transcribed using the T7 and T3 DNA polymerase respectively.

also found in the vein of the hypocotyl. A similar mRNA localisation pattern in the vegetative apical meristems was obtained using the antisense *CLF* probe (Goodrich et al., 1997).

Analysis of the sections from the inflorescences showed that the *CLK1* gene is expressed throughout the inflorescence meristems. Moreover, the *CLK1* signal was also detected throughout the young floral meristems (Figure 4.6C). Subsequently, strong signal was observed in the sepal primordia during stage 3 of flower development. *CLK1* expression continued into later stages of development with signals detected in petal, stamen and carpel primordia during stages 6-9 (Figure 4.6C). At the later stages the signal, and hence the expression of *CLK1* became weaker in all four floral whorls . On the other hand, in the late stage 12, a high level of *CLK1* expression was detected in the developing ovules (Figure 4.6E). No signal was detected in the sections probed with the *CLK1* sense probe (Figure D-F).

The observed more or less ubiquitous expression pattern of the *CLK1* gene was remarkably similar to that documented for the *CLF* gene (Goodrich et al., 1997), indicating that the two genes share overlapping expression domains. It is worth mentioning that the expression level of *CLK1* seemed rather low. Thus the *CLK1* signal detected was rather weak and 3 or more days were required for the detection of the hybridised probe. The weak signal observed with the *CLK1* probe became more apparent when compared with those of the *AG* antisense probe. Thus hybridisation with the *AG* probe gave very strong signal after just 1-2 days of detection time (Figure 4.6H). Interestingly previous *in situ* hybridisation analysis also revealed a relatively low level of *CLF* expression as compared to those of the floral homeotic genes *AG* and *LFY* (Goodrich et al., 1997). These observations indicated that not only that the patterns but also the levels of expression of the *CLF* and *CLK1* genes are very much alike.

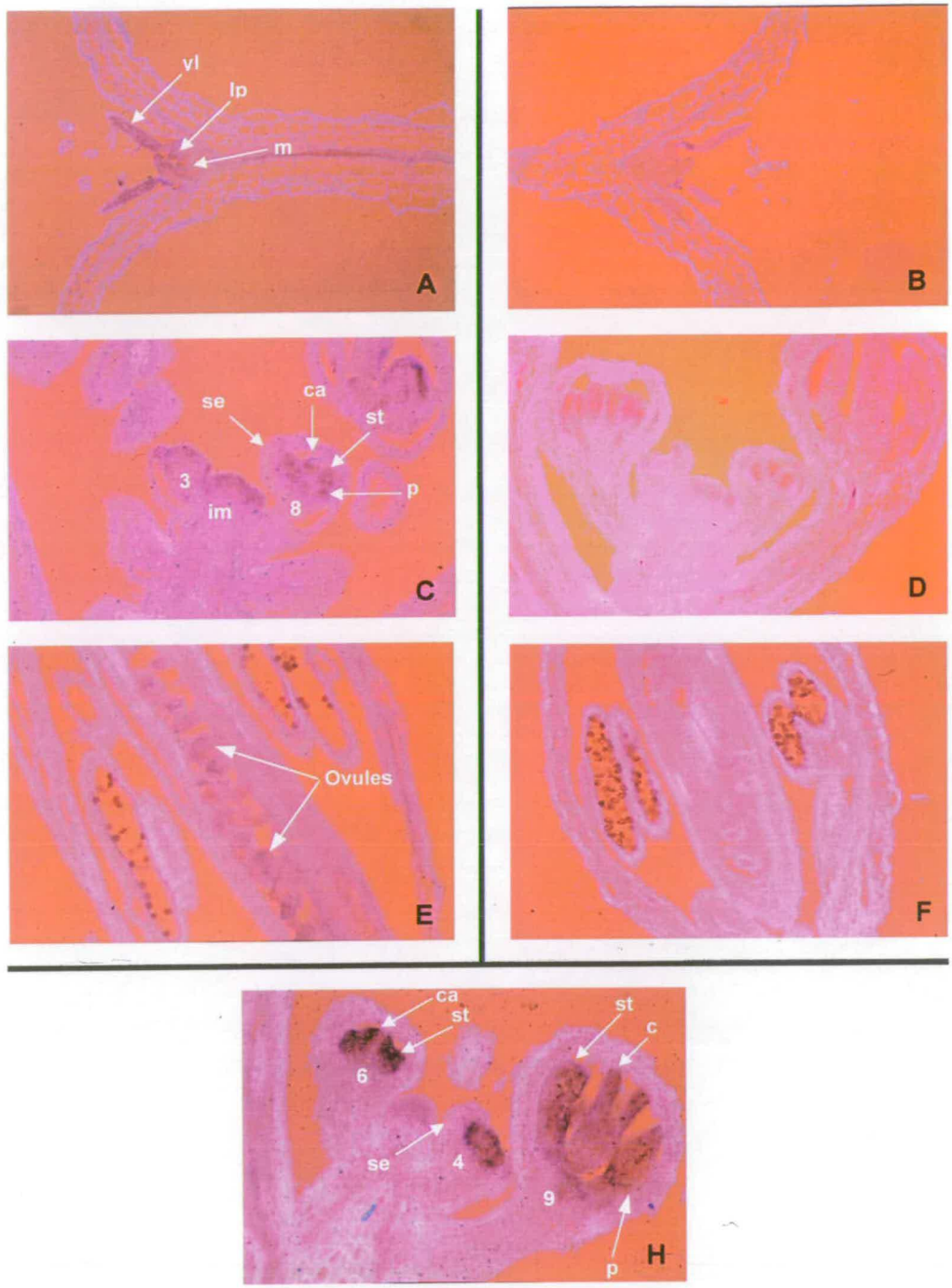


Figure 4.6 The expression of *CLK1* in the vegetative, inflorescence and floral meristems. A, A longitudinal section of 8-days old seedling passing through the vegetative shoot apical meristem hybridised with antisense *CLK1* probe. *CLK1* mRNA is detected throughout the meristem (m), leaf primordia (yp) and young leaves (yl). B, no signal is detected when similar sections were probed with the *CLK1* sense fragment. C-F, Infloresence sections probed with *CLK1* antisense and sense fragments respectively. *CLK1* is expressed throughout the inflorescence meristem (im) and the stage 3 flower. Later on in the stage 8 flower expression is detected in carpels (ca), stamens (st) and petals (p). Weak expression is also found in sepals (se), (C). In older flowers, strong *CLK1* expression is detected in the developing ovules (E). No hybridisation is detected when the sections were probed with the *CLK1* sense probe (D,F). H, *AG* expression pattern in wild-type section. In stage 4 flower *AG* transcript is found in the centre of the flower primordium and is absent from the sepals (se). In stage 6 flower, *AG* expression is detected in the stamen (st) and carpel (ca) primordia. No transcript is detected in the sepals and petals (p). Strong *AG* expression is later detected in the stamens and carpels of the stage 9 flowers of wild-type plants.

4.3 CLK1 protein interactions

As mentioned earlier, the ability of the *Drosophila* Pc-G proteins to interact with one another is well-documented. Furthermore a recent upsurge of interest in Pc-G interactions has led to the revelation that the interaction patterns have been highly conserved in evolution. In plants, the *Drosophila* ESC-E(Z) partnership is represented by the direct physical association between the *Arabidopsis* ESC homologue FIE and the E(Z) equivalent CLF and MEA (C MacDougall & C Stock, pers. comm.; Luo et al., 2000; Spillane et al., 2000; Yadegari et al., 2000). Since the *CLK1* gene encodes a Pc-G protein which showed high degree of structural conservation with the CLF protein, it was of particular interest to establish if, like CLF, the CLK1 protein is also able to interact and possibly form functional complexes with other Pc-G members particularly the ESC homologue, FIE.

In order to investigate possible physical interactions between CLK1 and other *Arabidopsis* Pc-G proteins a yeast two-hybrid analysis was carried out. In this experiment, the full length *CLK1* cDNA was amplified from total RNA from young seedlings by RT-PCR using the YC32 and YC33 primer pair which contained an in-frame *Sal*I and *Not*I site respectively. The PCR product was purified, digested with the enzymes *Sal*I and *Not*I and then subcloned in-frame into the *Sal*I/*Not*I digested yeast two hybrid vectors pBI880 and pBI881 (described in Kohalmi et al., 1998). The fusion vector pBI880 encodes a GAL4 DNA binding domain (DB) while the plasmid pBI881 encodes a GAL4 transcriptional activation (TA) domain. Competent yeast cells were co-transformed with the pBI880 *CLK1* and a pBI881 construct from another *Arabidopsis* Pc-G gene. In a complementary experiment the pBI881 *CLK1* fusion construct was co-transformed into the yeast cells with the pBI880 construct of the Pc-G gene of interest (Apart from the *CLK1* construct, all the other Pc-G pBI880 and pBI881 fusion constructs used in this study were kindly provided by C MacDougall and C Stock). If there is a physical interaction between the CLK1 fusion construct and another Pc-G construct, the GAL4 activation domain and the DNA binding domain would be brought together resulting in the transcriptional activation of the *lac*Z reporter gene. This can be scored by a histochemical reaction that gives a blue product when the X-Gal substrate is supplied (Figure 4.7A).

Result from the yeast two-hybrid analysis showed that of all the possible pairwise combinations of the GAL-DB and GAL-TA fusion proteins, only the CLK1 and FIE pair showed physical association (Figure 4.7B). No interaction was detected between CLK1 and

CLF, MEA or FIS2 (data not shown). In addition, the test involving the CLK1 GAL-TA and the CLK1 GAL-DB protein pair also yielded a negative result indicating that the CLK1 protein did not show self-association (Figure 4.7B). The interaction seen between CLK1 and FIE was likely genuine since autoactivation was not detected in the yeast cells containing either the CLK1 GAL-TA or the CLK1 GAL-DB fusion construct alone. The failure to detect any other CLK1 interacting proteins apart from FIE in this assay cannot, however, be regarded as a conclusive proof of the real lack of interaction *in vivo*. Thus some fusion proteins might have been unstable in the environment used or expression of some of the constructs might somehow have been reduced in this experiment. It is also possible that the interactions are extremely weak such that they could not be detected in the yeast two-hybrid assay and that other more sensitive assays such as co-immunoprecipitation or the GST fusion protein pull-down experiment might need carried out in order to detect the protein interactions. Another possibility is that the interaction of CLK1 with a given Pc-G member might require a pre-established complex of CLK1 and other proteins. This hypothesis could be tested by yeast three-hybrid analysis or co-immunoprecipitation experiment.

As previously mentioned, a number of studies of interaction between the SET domain E(Z) protein and the WD repeat ESC protein in *Drosophila* have implicated a 33 amino-acid region at the N-terminal region of E(Z) as essential for its interaction with ESC (Jones et al., 1998; Tie et al., 1998). This interaction domain is also conserved in the *Arabidopsis* E(Z) homologue, MEA, in which the first 110 amino acid region of the protein is both sufficient and necessary for its interaction with the ESC homologue, FIE (Luo et al., 2000; Spillane et al., 2000; Yadegari et al., 2000). A recent study has also indicated that CLF is also able to interact with FIE and has mapped the FIE-interacting domain to the N-terminal region of the CLF protein (C MacDougall, pers comm.). Since CLK1 was also able to interact with FIE and that the N-terminal region of CLK1 also shared some degree of similarity with the FIE-interacting domain of the MEA and CLF protein (37% and 22% identity respectively), it was of particular interest to find out if the N-terminal part of the CLK1 protein can also mediate the interaction with FIE on its own. In order to address this question, a truncated CLK1 construct containing just the first 120 amino acid at the N-terminus of the protein was amplified using the primers YC32 and YC40. The truncated construct was subcloned into the yeast two-hybrid vectors and the interaction analysis was then carried out. The test using the X-Gal assay showed that the truncated CLK1 protein interacted *in vitro* with the full length FIE fusion protein (Figure 4.7B). This result clearly demonstrated that the N-terminal

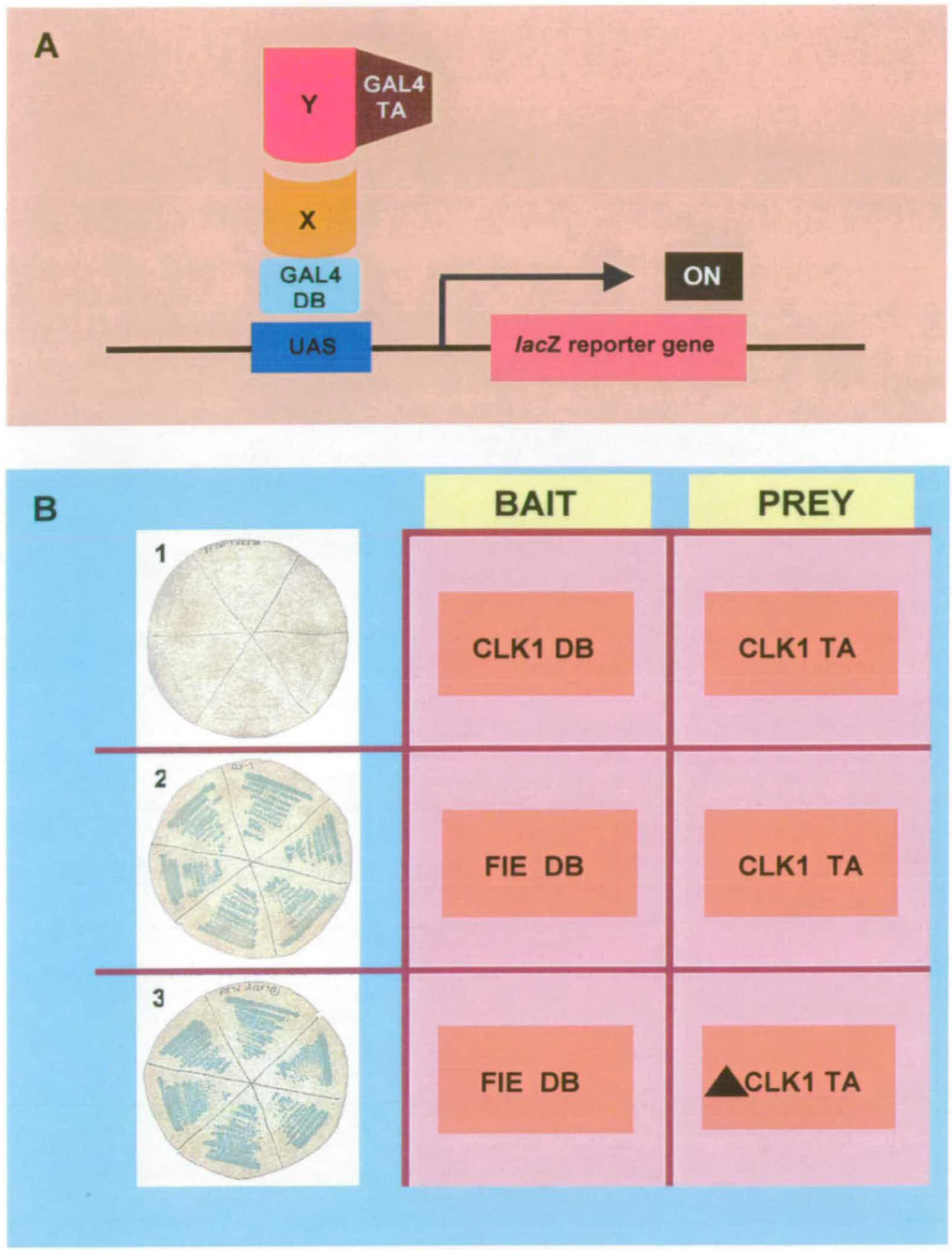


Figure 4.7 Yeast two-hybrid analysis. A, Protein X, the bait, is fused to the GAL4 DNA binding domain (DB) while the prey, protein Y, is fused to the GAL4 trans-activation domain (TA). When protein X and Y interact, the GAL4 DB and TA are brought together resulting in the activation of the downstream *lacZ* reporter gene. B, 1. No interaction is observed between the CLK1-DB and CLK1-TA fusion proteins suggesting that the CLK1 protein does not self-interact. 2. CLK1 and FIE protein interact with each other. 3. A 110 N-terminal fragment of the CLK1 protein is sufficient for its interaction with the FIE protein. Six independent yeast transformants are shown in each case.

interaction domain has also been conserved in the CLK1 protein despite the divergence in the primary sequence.

At present the significance of the interaction of MEA, CLF and CLK1 with FIE is not in all cases clear. Evidence from studies in *Drosophila* have shown that mutations that abolish ESC function in the establishment of Pc-G mediated silencing *in vivo* also disrupt its ability to interact with E(Z) suggesting that the ESC-E(Z) partnership is essential for Pc-G silencing (Jones et al., 1998; Tie et al., 1998). Recent analysis revealed that a mutation that disrupts the β-propeller structure of the WD40 repeats of the FIE protein also abolishes the ability of FIE to interact with MEA *in vitro* (Spillane et al., 2000). By analogy to the *Drosophila* paradigm one can, therefore, postulate that the MEA:FIE, CLF:FIE and CLK1:FIE association might be a pre-requisite and hence indispensable for the establishment of Pc-G mediated silencing in the *Arabidopsis* system. The inability to form the FIE-MEA partnership might lead to the failure in the establishment of the Pc-G repression network during early embryo and endosperm development. Recently, it has been demonstrated that when rescued through the embryonic stage, the *fie* mutant plants exhibit an accelerated flowering phenotype reminiscent of the phenotype of the *emf* mutants (Kinoshita et al., 2001). This suggests that, in addition to its repressor role during embryogenesis, *FIE* also functions as a floral repressor during vegetative growth. Therefore any mutations in the FIE protein leading to the abolishment of its interaction with CLF which acts as a negative regulator of floral homeotic gene could cause a major disruption in the establishment of silencing during vegetative development. Since the CLK1 protein is structurally highly similar to the CLF protein and are likely to perform redundant functions, one can speculate that the CLK1:FIE partnership could be of similar functional significance.

4.4 Isolation of *clk1* mutant alleles

The overlapping expression patterns and similar protein interactions in yeast two-hybrid assays between *CLF* and *CLK1* suggested that the two genes might perform similar roles in the Pc-G mediated silencing mechanism during *Arabidopsis* development. In order to gain more insights into the biological function of the *CLK1* gene and to further address the possibility of genetic redundancy between *CLK1* and *CLF*, reverse genetic approaches were used in an attempt to isolate mutant alleles of *CLK1*. Genetic and molecular analysis of such mutants would give invaluable information on the role of *CLK1* in development.

In recent years, mutational approaches have been widely used with great success to assign functions to a large number of newly identified genes. Several techniques have been developed to generate informative mutant alleles of a gene of interest. These include antisense and transgene co-suppression approach, site-specific mutagenesis, double-stranded RNA-mediated interference (RNAi) and insertional mutagenesis. In addition to the generation of loss-of-function alleles, gain-of-function analysis such as overexpression study can also provide vital clues about the relationship between a gene sequence and its function, particularly if the gene's function is redundant to other genes in the genome.

Two different mutagenesis techniques were employed in this study to isolate mutant alleles of the *CLK1* gene. The first was an attempt to generate a dominant negative allele of *CLK1* whilst the second approach involved screening for T-DNA insertional alleles using a large population mutagenised by random T-DNA insertion.

4.4.1 *clk1* dominant negative allele

In addition to the use of the conventional and more common gene disruption approaches, the problem of assigning a function to a gene can also be addressed by generation and characterisation of dominant negative mutations. In general, an allele is said to be dominant negative if it encodes a variant protein product which is capable of inhibiting the product of the wild-type gene such that the function of the wild-type gene becomes deficient. The dominant nature of the mutation means that its phenotype is manifested even in the presence of the wild-type gene (Herskowitz, 1987).

Several dominant negative alleles of the *Drosophila E(z)* gene have been identified in classical genetic screens (Pelegri and Lehman, 1994) and the lesions characterised (R Jones, pers. comm.) One of these alleles, designated *E(z)son3* contains an isoleucine to asparagine substitution at residue 635 in the SET domain of the E(Z) protein (Pelegri and Lehman, 1994). Analogous changes in the *CLF* gene have also been shown to result in a dominant negative phenotype in wild-type plants (J Goodrich, pers. comm.).

Since the SET domain is also highly conserved in the CLK1 protein, a putative dominant negative allele of *CLK1* was designed to mimic the dominant negative *E(z)* allele; *E(z)son3*. To achieve this, an oligonucleotide-mediated site-directed mutagenesis was carried out to direct a homologous change from valine to asparagine at residue 716 in the CLK1 protein. Thus an 8.7 kb *Eco*RI fragment containing the full length *CLK1* gene was purified from the T10M13 BAC clone DNA prepared from the agar stab stock provided by the DNA stock centre at the *Arabidopsis* Biological Resource Centre, Ohio, USA. The DNA was subcloned into pBluescript SK^- vector to give clone pYC3. Single stranded DNA was then prepared (see section 2.6.3 in Materials and Methods) and annealed with the mutagenic primer YC15 which contained the predicted dominant negative change (Figure 4.8A). After DNA synthesis using T4 DNA polymerase, the DNA was transformed into competent bacterial cells. Sequencing using primer YC18 was carried out to identify clones containing the desired change in the *CLK1* gene. Insert DNA was prepared from those clones, ligated to the *Eco*RI cut pBIN19 binary vector and transformed into *Agrobacterium* strain GV3101 mp90.

Because the function of *CLK1* was unknown and there was a possibility that dominant negative mutations might have deleterious or lethal effects, a steroid inducible from of the *clk1* dominant negative allele was also created. The use of the inducible transcription system has a number of advantages. It allows the activity of the gene of interest to commence at various time points during development and thus the developmental timing of various regulatory processes can be probed in finer detail, for instance. Conditional expression also circumvents any potential problems of embryonic lethality caused by dominant loss of *CLK1* activity. The GR inducible system was devised based on the knowledge of animal steriod receptors and has been shown to confer inducible activity on several plants and animal transcription factors (Picard 1993; Lloyd et al., 1994). Normally the GR fusion construct is sequestered in the cytoplasm by the chaperone hsp90 heatshock protein. However, when steroid is supplied externally, the GR fusion protein is released, enters the nucleus and able

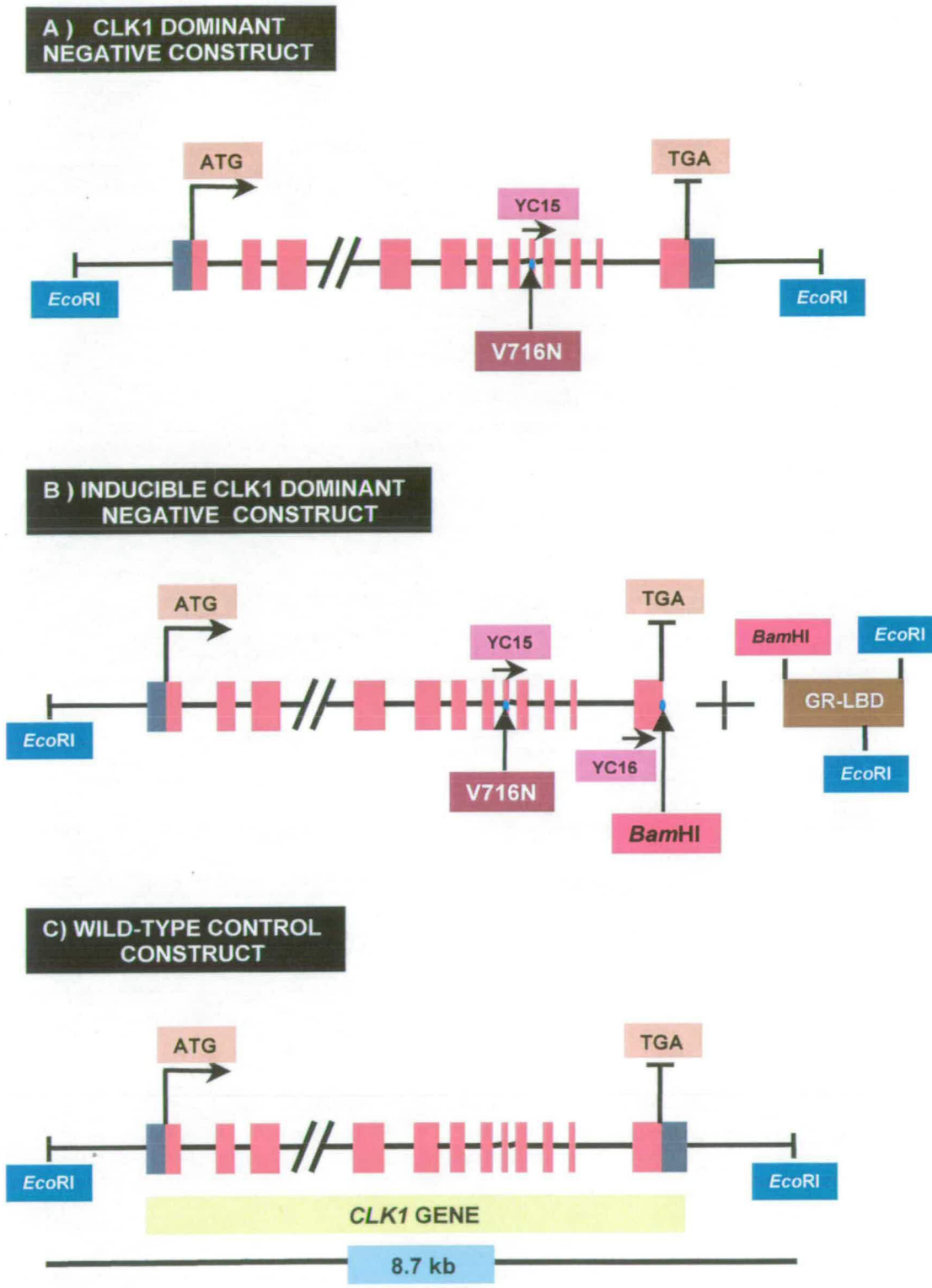


Figure 4.8 *CLK1 in vitro* mutagenesis. A, The primer YC15 was used in the mutagenesis experiment to introduce the predicted dominant negative change from valine to asparagine at residue 716 at the beginning of the SET domain of CLK1 protein. B, Creation of an inducible CLK1 dominant negative construct. A *Bam*HI site was introduced in-frame at the stop codon of the *CLK1* transcript. The *Eco*RI-*Bam*HI digested *CLK1* fragment was ligated to the 1.1 kb *Eco*RI-*Bam*HI fragment from the partial digest of p△GR containing the rat glucocorticoid ligand binding domain (LBD) (gift from Alan Lloyd). C, A wild-type control construct containing the 8.7 kb *Eco*RI fragment of the BACT10M13 containing full length *CLK1* sequence. In each case the fusion construct was cloned into *Eco*RI-cut pBIN19 vector and the recombinant plasmid introduced into plants by *Agrobacterium*-mediated transformation.

to exert its function. The synthetic glucocorticoid, dexamethasone (DEX) was used in this study as a convenient glucocorticoid analogue.

In order to generate the steroid inducible *CLK1* dominant negative construct, another mutagenic primer, YC16, which replaced the *CLK1* stop codon with *Bam*HI site that was in the same reading frame as the *Bam*HI site at the start of the ligand-binding domain (LBD) of the rat glucocorticoid receptor (GR) in the vector pΔGR (gift from Alan Lloyd) was used in the annealing step together with the dominant negative mutagenic primer YC15. Recombinant clones containing both the *CLK1* dominant negative change and the engineered *Bam*HI site were identified by DNA sequencing using the primers YC18 and YC19 respectively. Recombinant plasmid DNA was extracted and digested with *Bam*HI and *Eco*RI. A three-way ligation was then carried out with the *Eco*RI/*Bam*HI cut *CLK1* fragment, the phosphatased, *Eco*RI cut pBIN19 vector and the 1.1 kb *Bam*HI/*Eco*RI cut fragment of the pΔGR (Figure 4.8B).

As a control, a construct harbouring wild-type *CLK1* sequence was also made. Thus the 8.7 kb *Eco*RI fragment of the BAC clone T10M13 which spans the *CLK1* gene was subcloned into the pBluescript SK^- vector. The 8.7 kb insert was then cloned into the pBIN19 binary vector (Figure 4.8C).

To introduce the *CLK1* constructs into CLF^+ and clf^- backgrounds the Ws wild-type plant and also the *clf-50*/+ plants were transformed using *Agrobacterium*-mediated transformation either by vacuum infiltration or floral dipping. Because the *clf-50* homozygotes are not very fertile, transformation was performed on *clf-50*/+ heterozygotes, so that in the T1 generation many of the primary transformants were in *clf-50*/*clf-50* or *clf-50*/+ background. Seeds were collected and in the T1 generation transgenic plants were selected on GM+kanamycin plates.

In the first experiment, only 1 transformant was recovered from vacuum infiltrating *clf-50*/+ plant with the *Agrobacterium* carrying the plasmid containing the inducible dominant negative *CLK1* construct. Phenotypically this plant appeared wild-type. When selfed its progeny segregated 163 kanamycin resistant to 69 kanamycin sensitive (Figure 4.9B), consistent with it carrying a single locus T-DNA insertion. The T2 progeny segregated for wild-type, *clf* plants, and a novel enhanced phenotype like that of the *emf* and *clf moe* double mutants. The segregation pattern seen suggested that the *CLK1* transgene might have little effect in CLF^+ background but enhanced the *clf* phenotype in clf^-/clf^- background.

A

PHENOTYPE / MEDIA	Wild-type	Enhanced	*clf-50*
GM	121	31	11
DEX	31	7	-
Kan	46	18	3

B

PHENOTYPE / PLATE NO.	Wild-type	Enhanced	*clf-50*	Kanamycin sensitive
1	28	12	1	26
2	28	12	1	16
3	25	14	1	17
4	24	17	-	16
TOTAL	105	55	3	69
TOTAL KAN^R VS KAN^S	163			69

Figure 4.9 Scoring of the *CLK1* transgenic line obtained from vacuum infiltration experiment. A, Phenotypic scoring of T2 transgenic plants grown on 3 different growth media; GM, GM+DEX and GM+kanamycin. B, Phenotypic scoring of T2 transgenic plants grown on GM+DEX+ kanamycin plates.

However, the enhanced phenotype appeared regardless of whether plants were grown with or without steroid. Therefore, either this enhanced phenotype was not a result of the transgene, or alternatively the dominant negative effects were not inhibited by fusion to the steroid receptor.

Plants exhibiting the enhanced phenotype were extremely small in size with much reduced cotyledons and leaves. As the plants grew older, only a few small sessile leaves were developed and only about 10 days after germination these plants started their reproductive development producing a few very small flowers at the apex with no inflorescence structure formed (Figure 4.10A, C). Only very occasionally those plants with this so-called enhanced phenotype that were sown straight onto soil rather than the GM media and grown in mild conditions produced inflorescence structure (Figure 4.10E). The inflorescences developed were thin and small with short internode elongation. Flowers were formed at the tip of both the primary and lateral inflorescences similar to those seen in the *tfl-1* mutants. The flowers formed exhibited extreme carpelloidy with a mass of carpelloid structure developed instead of normal floral organs (Figure 4.10D).

To verify the effects of the *CLK1* transgene in a wild-type background, if any, seed were collected from several kanamycin resistant T2 plants that had a wild-type phenotype. When T3 families were sown, several were found that no longer segregated *clf* or enhanced *clf* plants, although they segregated kanamycin resistant to kanamycin sensitive seedlings in a 3:1 ratio. These T3 families were therefore the progeny of a CLF^{+}/CLF^{+} plant hemizygous for the *CLK1* transgene. Although these T3 plants appeared completely wild-type as seedlings, as plants got older, a weak phenotype was apparent. Thus several plants that were first thought to be wild-type with normal appearance produced leaves that were serrated (Figure 4.10G, H). Serration was detected in both the rosette (Figure 4.10I) and cauline (Figure 4.10J) leaves and was noticeably much more pronounced than the small degree of serration seen in the wild-type plants. Other aspects of leaf development seemed unaffected in those plants. Likewise no perturbation was seen in reproductive structures (Figure 4.10K). The observed ratio of wild-type plants to plants with serrated leaves was 31:106. Seed was collected from 15 wild-type plants and 20 plants with serrated leaves and sown on kanamycin plates. All 15 wild-type plants gave exclusively kanamycin sensitive progeny, whereas all 20 plants with serrated leaves gave either all kanamycin resistant progeny or a mixture of kanamacin resistant and sensitive progeny.

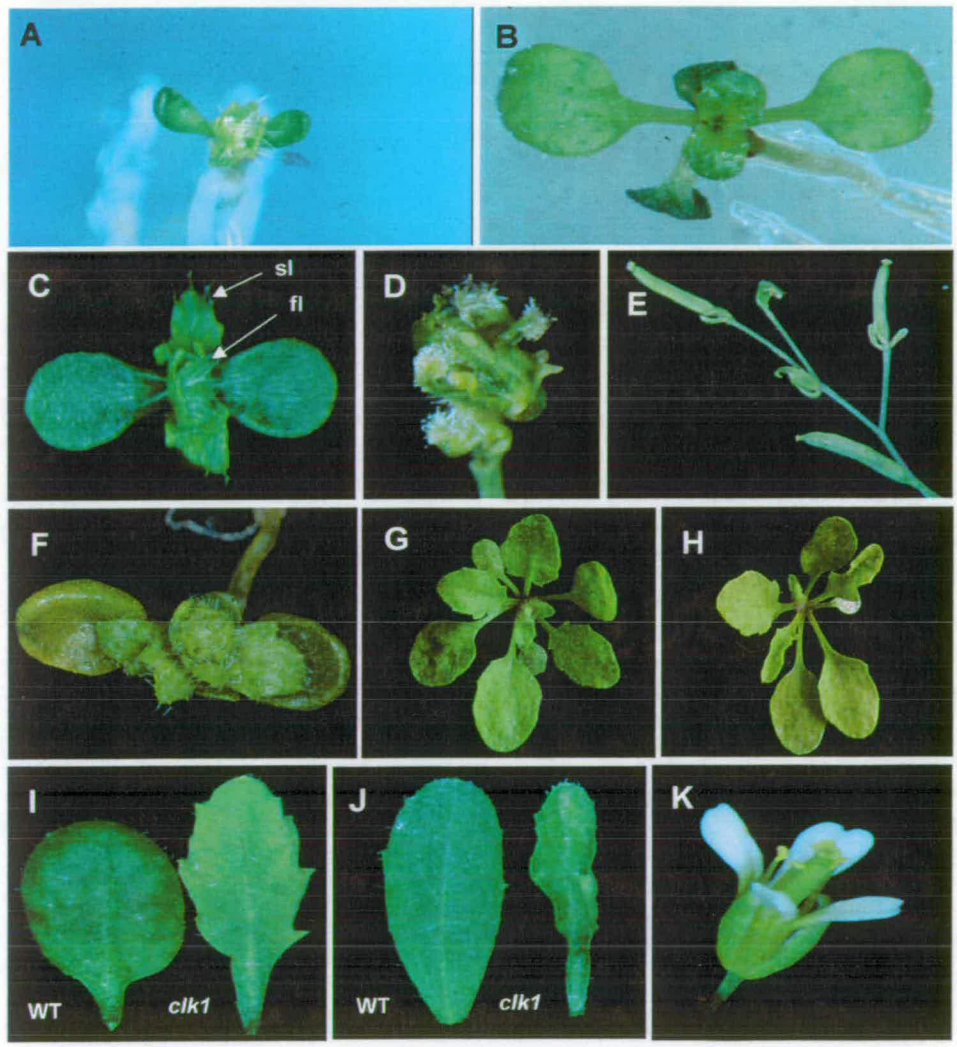


Figure 4.10 Genetic characterisation of *CLK1* transgenic lines. A-B, Homozygous *clf-50* plants carrying the *CLK1* dominant negative transgene shows an enhancement of the *clf* phenotype. Plants are very small (A) compared to wild-type (B) and only produced a number of very small sessile leaves (sl) before entering the reproductive developmental phase with floral bud (fl) formed at the apex of the plants (C). Flowers produced showed strong degree of homeotic conversions and usually a mass of carpelloid organs was formed (D). Sometimes inflorescence structures with characteristics of terminal flower mutants were seen (E). F, Similar enhanced phenotype was observed with the F2 plants from the cross of the *CLK1* transgenic line to the *clf-2* background that carried the *CLK1* transgene and were homozygous for the *clf-2* allele. G-J, Plants carrying the *CLK1* transgene in Ws background showed serrated leaf phenotype (H) compared to Ws plants (G). Both rosette (I) and cauline leaves (J) were serrated. In a number of cases the serrated leaves also showed some weak curling (J). Flowers of those plants were phenotypically normal (K).

Because all the plants with the serrated leaf phenotype were hemi- or homozygous for the transgene, these results were consistent with the hypothesis that the transgene conferred a dominant mutant phenotype by interfering with the endogenous wild-type *CLK1* gene. The phenotype was much less severe than that of null *clf* plants. However, it was difficult to know if this was because the transgene did not give a null *clk1* phenotype or because the requirement for $CLK1^+$ is less than for CLF^+ function.

Scoring of the T2 progeny on GM plates revealed a segregation ratio of 121 wild-type:11 *clf*: 31 enhanced *clf* plants (scoring was performed before the weak serrated phenotype would have been discernable) (Figure 4.9A). This was consistent with a 12:1:3 ratio as expected if *clf* plants that carried 1 or 2 copies of the *CLK1* transgene had the enhanced phenotype. If this model was correct, a prediction was that if seedlings were grown on kanamycin plates, so that only plants with 1 or more *CLK1* transgene survived, then only wild-type and enhanced *clf* phenotype should be seen (in a 9 wild-type: 3 enhanced *clf* ratio). However, this prediction was not met (Figure 4.9). On kanamycin alone, a ratio of 46 wild-type: 18 enhanced *clf*: 3 *clf* plants was observed whereas that seen on kanamycin+DEX was 105 wild-type: 55 enhanced *clf*: 3 *clf*. Thus the enhanced phenotype was linked to the transgene but did not co-segregate absolutely with it. Two possible explanations occurred. Firstly, the dominant negative effects of the *CLK1* transgene might be variable, so that *clf/clf clk1/+* hemizygotes might occasionally show normal *clf* phenotype. Alternatively, the enhanced phenotype might be caused by an independent modifier mutation linked to the T-DNA.

To distinguish these possibilities, transformation experiments were repeated by floral dipping method, and independent transformants were generated. If the enhanced phenotype was unrelated to the dominant negative *CLK1* transgene, then it was unlikely that independent transformants would also display the enhanced phenotype. The result of the scoring of the primary transformants from all 6 possible combinations of transgene constructs and backgrounds is shown in figure 4.11.

From the table it is clear that a range of phenotypes was seen in all 6 lines investigated. Some of the transgenic plants exhibited weak leaf curling phenotype while some looked like *clf-50* mutants. Also unexpectedly, when Ws wild-type plants were transformed, some T1 plants showed weak or even strong *clf* leaf curling phenotype (Figure 4.11). The fact that both the lines with the wild-type *CLK1* construct and many of the non-induced CLK::GR lines showed the leaf curling defects was rather puzzling. One possibility is that the defects

Line	Construct	Parental background	Phenotype		
			Wild-type	Weak curling	Strong curling
J207	Dom neg CLK1	*clf-50/+*	37	4	10
J208	Dom neg CLK1::GR	*clf-50/+*	16	2	5
J209	CLK1	*clf-50/+*	35	2	8
J210	Dom neg CLK1	Ws	21	1	3
J211	Dom neg CLK1::GR	Ws	23	3	2
J212	CLK1	Ws	36	3	2

Figure 4.11 Phenotypic scoring of *CLK1* transgenic lines from floral dip transformation. Three different *CLK1* transgenes; Dominant negative construct, inducible dominant negative construct and wild-type CLK1 construct, were introduced into plants in the Ws wild-type and *clf-50* background. Phenotype of the kanamycin resistant T1 plants carrying the transgene were scored according to the degree of leaf curling shown.

seen were caused by co-suppression. Co-suppression or transgene silencing phenomena is known to have widely occurred across kingdoms including plants. It is known to be a consequence of the ability of some transgenes to simultaneously silence themselves and the endogenous homologous ´loci in the genome. Co-suppression has been extensively investigated in plants in which silencing is thought to take place at both the transcriptional the post-transcriptional level. Several speculative models have been proposed to explain the mechanisms underlying the silencing process. Most of these models have indicated the aberrant RNAs as the silencing signals. Aberrant RNAs are short single stranded RNA molecules produced as a result of epigenetic changes induced by the transgene that interfere with normal transcription. It has been speculated that when exported into the cytoplasm these RNAs are recognised by the RNA-dependent RNA polymerase (RdRP) and short antisense RNAs are transcribed as a result. These antisense RNAs can then interact with the homologous mRNA molecules and in the process trigger the degradation of the single-stranded part of the RNA hybrid causing the breaking up of the complementary mRNA into small non-functional pieces and thereafter the silencing of the gene. Alternatively, large double-stranded RNA molecules might be synthesised by an RdRP using the antisense RNAs as primers. These large RNA molecules are subsequently degraded resulting in the accumulation of antisense RNAs and thereby triggering several new loops of antisense RNA annealing, RdRP-mediated double-stranded RNA synthesis and double-stranded RNA degradation. This catalytic feedback loop would eventually lead to most if not all the endogenous mRNA molecules being degraded and hence the silencing of the gene (Reviewed in Cogoni and Macino, 2000; Meins Jr., 2000). Data from several independent studies have suggested that transgene-induced gene silencing effect can be stably inherited through many generations.

In this experiment, it is possible that in some transgenic lines, the introduced *CLK1* transgene may induce silencing of itself and its endogenous homologue. The range of the phenotype observed could also be a consequence of the varying degree of co-suppression and hence the differences in the amount of the CLK1 protein synthesised in different transgenic lines. In addition to this, changes in the level of the CLK1 protein or transcript could also, to some extent, disrupt the function of the *CLF* gene due to the likely close functional relationship between the two genes. This might also result in the varying degree of leaf curling seen in the transformants. It is also possible that the *CLK1* transgene could also directly induce the silencing of the *CLF* gene since the two genes are highly similar in their sequences. Transgenic plants that were heterozygous for the *clf-50* allele might be more

sensitive to the fluctuation in the wild-type CLF protein level. This in turn means that co-suppression would have been more common in the *clf-50* background. This notion agrees with the results shown in figure 4.11.

To carry out a more detailed characterisation of the transgenic lines obtained, seeds were collected from the T1 transformants and the phenotype investigated in the T2 generation. Scoring of the T2 plants showed that several independent J208 lines segregated for the enhanced phenotype. Again as in the case of the original transformant from the vacuum infiltration experiment the occurrence of the plants with the enhanced phenotype was independent of DEX treatment. The degree of enhancement seemed varied between different plants of the same lines and also from one line to another. None of these lines exhibited a uniformity in the degree of enhancement as seen with the progeny of the original transformant obtained from vacuum infiltration. It is interesting that apart from J208, the enhanced phenotype was not detected in the progeny from the other 5 lines. Instead regardless of the type of construct present and the background that they were in, plants from the lines J207, J209, J210, J211 and J212, exhibited a range of leaf curling phenotype from weak to strong.

The specificity of the enhanced phenotype to the J208 lines and hence the CLK1::GR construct provided strong evidence against the second modifier mutation hypothesis. A more likely cause of the observed phenotypes in those lines is co-suppression. The variation in the effectiveness of co-suppression within the same plant and between different transgenic lines is well documented. In addition co-suppression could also be unstable in some cells but not others of the same tissue or organs of the same plant. The variable and unstable nature of the co-suppression phenomenon might account for the observed phenotypic variation within and between different transgenic lines.

In order to test if co-suppression was the real cause of the phenotypes seen in these transgenic lines, RT-PCR analysis was carried out to investigate if there was the expected reduction in the level of *CLK1* transcript in the lines that segregated for the enhanced phenotype. Thus total RNAs were extracted from seedlings of the T2 progeny of the original transformant from the infiltration experiment and from an independent line which also segregated for the enhanced phenotype. In addition RNA samples were also prepared from the seedlings of four other lines. One of these; the line J497-1 was hemizygous for the *CLK1::GR* transgene while the other three lines; J425-10, J497-6 and J497-7 were

homozygous for the transgene. All 4 lines were in the wild-type background and hence did not show the enhanced *clf* phenotype. In addition to the seedling tissue, RNA was also prepared from the cauline leaves of the line J425-10 and J497-6 both of which exhibited the serrated leaf phenotype.

First strand cDNA synthesis was carried out on those RNA samples and RNA from the Ws wild-type seedlings. RT-PCR was performed using cDNA specific primers for the gene *CLK1* (YC22 & YC23), *CLF* (YC24 & YC25), *AG* (YC9 & YC10) and *KNAT1* (KNAT1-F & KNAT1-R) which plays a role in the maintenance of meristem integrity. Results from the RT-PCR experiment revealed no major difference in the level of the *CLK1* and *CLF* transcripts between the 9 samples (Figure 4.12). However ectopic *AG* expression was detected in the seedlings from the lines J497-1, J497-6, J497-7, J425-10, the T2 progeny of the original transformant and also in the cauline leaf sample from the lines J497-6 and J425-10 (Figure 4.12). This suggested a link between the presence of the *CLK1* transgene and the derepression of *AG* expression and in turn implicated the role of *CLK1* as an *AG* repressor. It is therefore likely that the phenotypic defects seen in those lines are partly due to the misregulation of *AG* expression as caused by the loss of *CLK1* function. In addition to *AG*, ectopic *KNAT1* expression was also observed in the cauline leaf sample from the line J497-6 and J425-10 (Lane 2 and 3, figure 4.12). Previous studies have shown that *KNAT1* is expressed in the seedlings, flowers and inflorescence stems but not in the leaves (Lincoln et al., 1994). The low level of *KNAT1* transcript seen with the cauline leaf sample from the Ws wild-type plants could be attributed to the fact that cauline leaves can be considered as parts of the reproductive structure. Also there could be some contamination from the inflorescence tissue in the samples used. Since overexpression of *KNAT1* has been demonstrated to cause strong lobing of leaves (Lincoln et al., 1994; Chuck et al., 1996) it is possible that the pronounced serration of the rosette and cauline leaves of the plants from the line J497-6 and J425-10 which can be interpreted as mild leaf lobing could be the consequence of *KNAT1* misexpression in those organs.

The observation that the levels of *CLF* and more importantly *CLK1* transcript in the transgenic lines that exhibited the enhanced and serrated leaf phenotype were similar to those seen in wild-type was inconsistent with the co-suppression hypothesis. However, the possible effect of co-suppression cannot be ruled out since the RT-PCR analysis might not have been sensitive or quantitative enough to detect subtle differences in the expression levels between different samples.

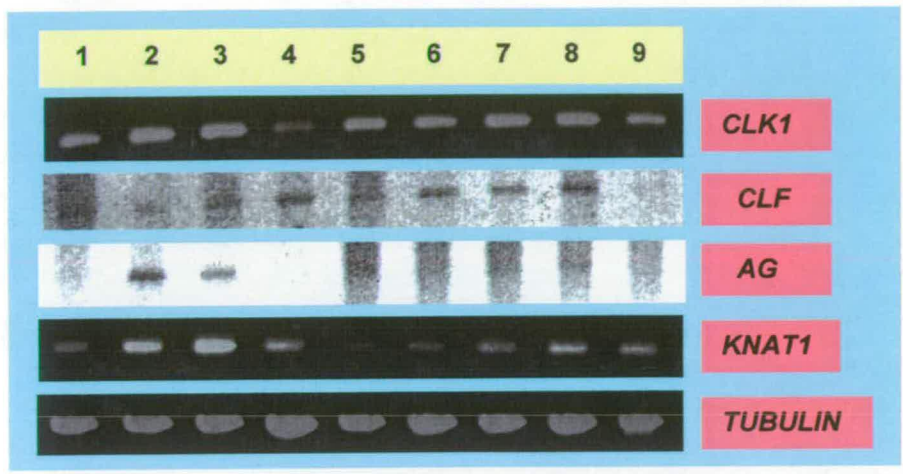


Figure 4.12 Molecular analysis of *CLK1* transgenic line. Semi-quantitative RT-PCR analysis on total RNA extracted from seedlings showed that there was no change in the level of *CLK1* or *CLF* expression in the *CLK1* transgenic lines investigated (lane 5= J425-10, lane 6= J497-1, lane 7= J497-6, lane 8= J497-7, lane 9= T2 progeny of the original *CLK1* line from the vacuum infiltration transformation) compared to wild-type (lane 4). Similarly no change was observed with the cauline leaf samples (lane 1= Ws wild-type, lane 2= J425-10, lane 3= J497-6). However, misexpression of the floral homeotic gene *AG* was detected in both the seedling (lane 5-9) and cauline leaf (lane 2-3) tissues of all the *CLK1* transgenic lines compared to the wild-type level (lane 1 for cauline leaf and lane 4 for leaf sample). In addition *KNAT1* expression was also up-regulated in the cauline leaf samples from the *CLK1* transgenic lines (lane 2 and 3 compared to lane 1).

Although large-scale investigation of several transgenic lines did not provide conclusive evidence for the molecular basis of the phenotypes seen associated with those lines, there are three lines of evidence to suggest that the phenotypes are likely caused by the *CLK1* transgene. First, the enhanced phenotype was detected in several independent *CLK1* transgenic lines making it unlikely that they were caused by a mutation in another locus elsewhere in the genome. Second, an enhanced *clf* phenotype was derived from the cross of the line J430-8 which was hemizygous for the *CLK1* transgene to another strong *clf* background, the *clf-2* mutant line (Figure 4.10F). This finding not only suggested that the *CLK1* transgene was the likely cause of the enhanced phenotype, but it also provided another line of evidence for the overlapping function of the *CLF* and *CLK1* gene. In addition to this when the *CLK1* transgene was introduced into the p*CLF*::GR inducible lines, F2 plants that were grown on non-DEX plates showed segregation for the enhanced phenotype. On the contrary, no enhanced phenotype was observed on the DEX plates in which almost all plants were wild-type in appearance. Only 1 or 2 plants showed a strong leaf curling phenotype like that seen with the *clf-2* mutant. These plants were likely *clf-2* mutants that had lost the CLF::GR construct and therefore could not respond to the DEX treatment. This result clearly indicated that in addition to the presence of the *CLK1* transgene, the loss of *CLF* function was also responsible for the enhanced phenotype seen.

In conclusion, these results suggested that introduction of a dominant negative *CLK1::GR* transgene somehow interfered with the function of the endogenous *CLK1* gene. This gave weak phenotypes in a CLF^{+} background, and dramatically enhanced the phenotype of *clf* plants, consistent with the *CLF* and *CLK1* genes having some functional overlap. However, it was difficult to draw any robust conclusions as there were many inconsistencies. Thus, the effects of the dominant negative CLK1::GR fusion were independent of steroid. Also the dominant negative construct (without the GR fusion) didn't enhance the *clf* phenotype. Furthermore the wild-type control construct occasionally gave a leaf curling phenotype. In part these may reflect the limitations and variabilities of experiments involving increased gene expression.

4.4.2 *clk1* insertion allele

So far the data accumulated from the study of *CLK1* expression and the analysis of *CLK1* transgenic lines have provided some evidence for an overlapping role of *CLK1* with its homologue, the *CLF* gene. However, the results obtained were inconclusive regarding the molecular basis of the phenotype conferred by the *CLK1* transgene. Towards the end of the *CLK1* overexpression experiment, the ABRC stock centre and the *Arabidopsis* knockout facility at The University of Wisconsin made available an excellent facility for reverse genetic screens for knockout alleles of any gene of interest. This offered an exciting opportunity to isolate recessive mutant alleles of the endogenous *CLK1* gene, and to circumvent some of the drawbacks of the approach of expressing *CLK1* transgenes. Thus in order to gain more information on the effect of *clk1* mutation on development and the genetic interaction with the *CLF* gene, attempts were therefore made to screen for *clk1* knockout alleles using a large T-DNA mutagenised population available from the University of Wisconsin, U.S.A

Transfer DNA (T-DNA) is a segment of the tumour inducing (Ti) plasmid of *Agrobacterium tumefaciens* (Azpiroz-Leehan & Feldmann, 1994). It is delimited by short imperfect-repeat border sequences. T-DNA integration into the plant genome is at random and when inserted any sequences lying between the borders would also be transferred and integrated into the genome. Several plant transformation techniques commonly employ engineered T-DNAs that have been disarmed of the tumour-promoting genes and contain an antibiotic resistance marker to allow selection of transformants. The use of T-DNA as an insertional mutagen in *Arabidopsis* offers a number of advantages over other mutagenesis systems. For example, T-DNA insertion in a gene often results in the complete disruption of the expression and hence the function of the gene leading to the creation of a knockout allele of the gene. Furthermore, mutations in essential genes which normally lead to lethality can be recovered and maintained in the heterozygous state. Another advantage of the T-DNA tagging system is that unlike post-transcriptional gene silencing approaches which may confer systemic silencing effect, single genes within a closely related gene family could be targeted and each of their function can be individually assessed. The mutations can also be combined by crossing different mutant lines together enabling the effect of multiple mutations to be investigated. Additionally, T-DNA insertion is stable and inherited in a Mendelian fashion through multiple generations. Identification of T-DNA insertion in the target gene is also relatively simple. Thus PCR-based techniques using combinations of gene-specific and T-

DNA-specific primers allows easy and rapid detection of an insertion in the target gene from a large population of plants. Subsequently DNA sequencing analysis could be carried out to pinpoint the exact insertion site. In spite of all these advantages that the T-DNA tagging system offer, there is one slight but potentially problematic drawback associated with it. Thus T-DNA often inserts itself as multiple copies. Therefore analysis of the effect of the T-DNA insertion in a gene of interest could be complicated by the insertions elsewhere in the genome and the time-consuming backcrossing to the wild-type background is required to eliminate the other insertions.

Given its many favourable properties and the vast improvement in *Agrobacterium*-mediated transformation techniques, T-DNA tagging is now the preferred choice of mutagenesis approach for reverse genetics and has been used with enormous success. Attempts have been made to saturate the *Arabidopsis* genome with T-DNA and several large collections of T-DNA transformed lines are now made available for the research community.

PCR-based screening of T-DNA transformed lines was provided by the *Arabidopsis* Knockout Facility at the University of Wisconsin, U.S.A. and was employed in this study to identify potential insertional alleles of the *CLK1* gene. The facility have a collection of 60,480 Arabidopsis lines of the Ws ecotype transformed with the T-DNA vector pD991 (Figure 4.13). In order to minimise the time and labour involved in the screening, these resources are organised into a pool architecture. Thus DNA was extracted and arranged into 270 pools of 225 lines. DNA pools were in turn organised into 30 superpools each of which can be deconvoluted into 9 subpools of 225 and then 25 pools of 9. T2 generation seeds were bulk collected from the pools of nine plants and are commercially available from the *Arabidopsis* Biological Resource Centre at the Ohio State University, U.S.A.

Three rounds of PCR-based screening are needed to identify an insertion in a gene of interest from the superpool level down to the particular pool of 225 and eventually to individual plants (Figure 4.14B). Since the probability of finding a knockout is directly proportionate to the gene size and given that the *CLK1* gene is about 5 kb, it was quite likely that one of those T-DNA insertion lines would contain a hit in the *CLK1* gene. First round screen of the 30 superpools was carried out using a T-DNA left border primer JL202 and two *CLK1* gene specific primers; one in the forward direction with a binding site at the beginning of the gene (CLK-F), and one in the reverse direction and binds to the sequence near the stop codon (CLK-R) (Figure 4.14A). Southern analysis using an 8.7 kb *Eco*RI fragment from the BAC

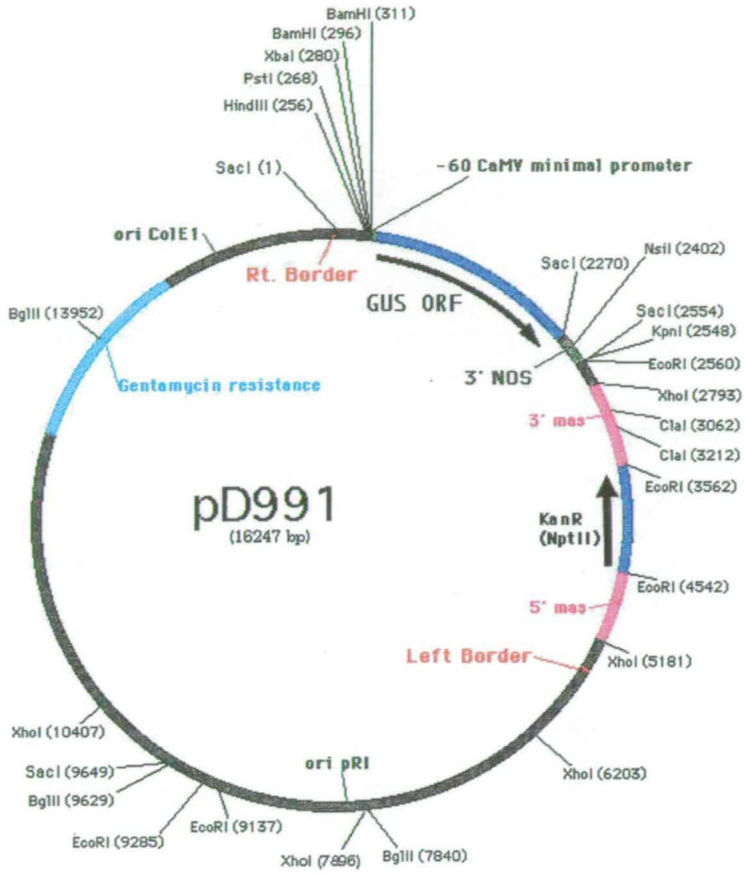


Figure 4.13 Diagrammatic representation of the T-DNA vector pD991 used to create the insertion allele of the *CLK1* gene. The integrated T-DNA fragment is delimited by the right and left border sequences. (Picture obtained from the web site http://www.dartmouth.edu/~tjack).

clone T10M13 containing full length *CLK1* sequence revealed a number of potential hits in the *CLK1* gene. These DNA bands were gel extracted, cloned into pCR®II TOPO® vector and sequenced with the Sp6 promoter specific primer which has a binding site near the multiple cloning sites in the vector. Two of these clones appeared to have a hit in the *CLK1* sequence. The first clone F21 (superpool 21, CLK-F primer) contained an insertion in exon 5 and was likely to be a knockout (Refers to Figure 4.1). An insertion in the 5' UTR region 28 bp upstream of the translation initiation codon was found in the second clone, R27 (superpool 27, CLK-R primer) (Refers to figure 4.1). This insertion was also likely to be mutagenic as it has often been the case that T-DNA insertions directly upstream of the start codon or within the 5'UTR of a gene causes a disruption of its transcription.

As the insertion site in both the F21 and R27 clones looked promising, second round screen was carried out on the corresponding subpools using the relevant primers. Again PCR products were run on an agarose gel which was then blotted and the filter hybridised with the full length *CLK1* probe. Disappointingly, the hit F21 was not recovered in this screen. Attempts to detect the desired fragment by reamplification of the original PCR products using the nested primer of JL202 and CLK-F (JL270 and CLK-Fn respectively) also failed to recover the hit. The disappearance of this insertion was not unprecedented since false positives are not uncommon in PCR-based screening. In contrast to the F21 clone, the second round screen of the R27 hit proved successful. Thus a 'hit' of a similar size of ~5 kb was recovered in one of the 9 subpools. Sequencing of the PCR band reconfirmed the existence and position of the insertion in the 5' UTR of *CLK1* (Figure 4.14B).

In order to narrow the hit down to a single pool of 9 plants, seeds from the 25 pools of 9 that comprised the subpool 237, in which the hit was found, were obtained and grown on soil. The DNA was then bulk extracted from the seedlings of each of the 25 pools and subjected to PCR screening using the T-DNA primer JL202 and a reverse *CLK1* primer YC50r. Rather encouragingly, one of the pools gave a PCR band of the expected size of about 1.7 kb (Figure 4.14B). To identify single plants that carried the insertion, seeds from this pool, 5887, were resown and DNA was extracted from individual seedlings. PCR was carried out using the primer pair JL202 and YC50r. Of the 71 T2 transgenic seedlings investigated, 6 were found to carry the T-DNA insertion in the *CLK1* gene. Three plants; 3, 27r and 37, were first identified (Figure 4.14B) and the other three; 5, 11B and 20, were later identified in a second sowing. To verify that these plants had disruptions in the *CLK1* gene, Southern analysis was carried out on *Eco*RI digested genomic DNA isolated from the leaf tissue of the

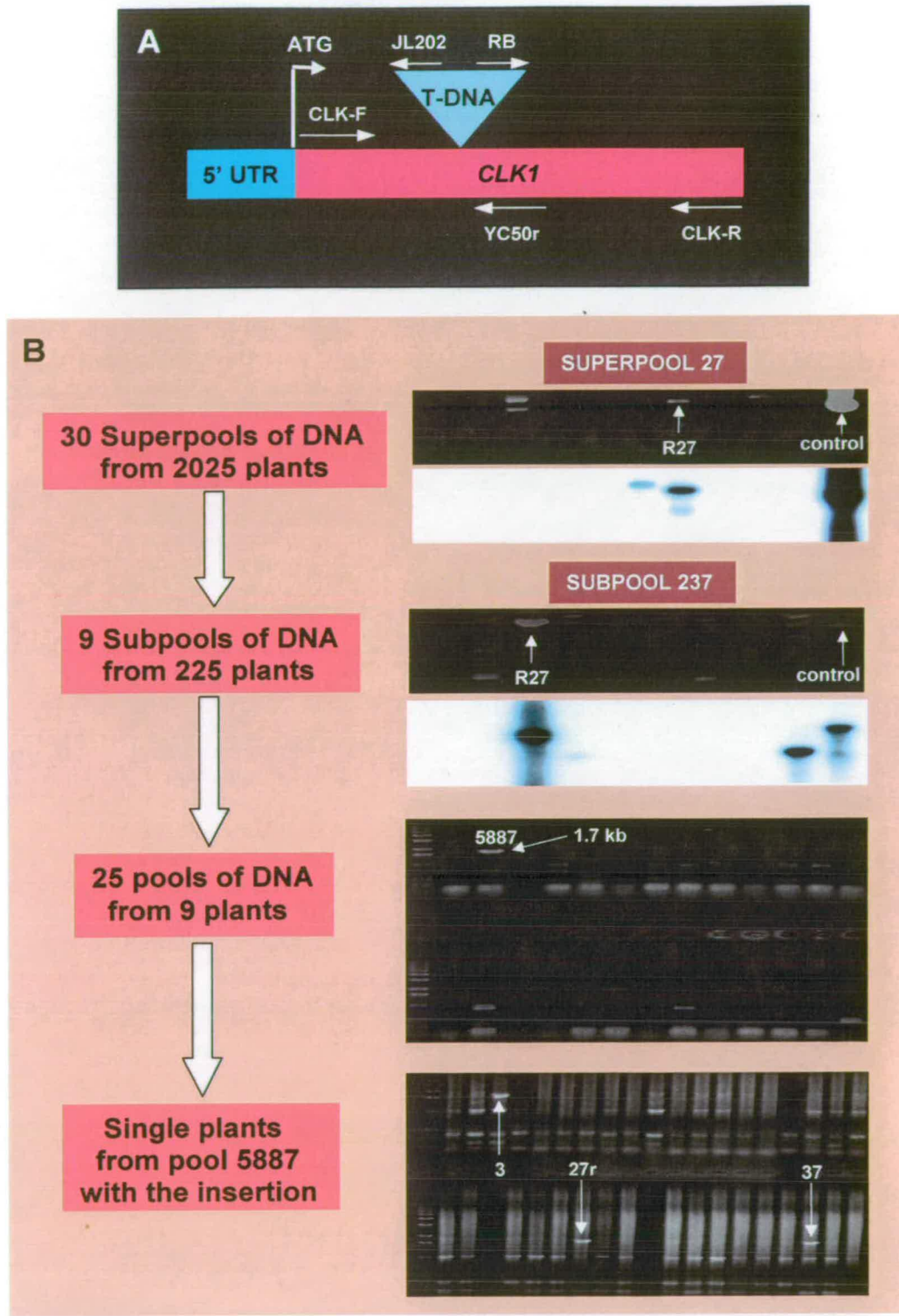


Figure 4.14 PCR-based screening of *CLK1* insertion lines. A, Screening system. Plants carrying a T-DNA insertion in the *CLK1* gene were identified by PCR screening using a T-DNA left border primer, JL202, and a forward *CLK1* specific primer, CLK-F, or a reverse *CLK1* specific primer, CLK-R. B. Screening process. DNAs from 60.480 T-DNA insertion lines were bulked together and arranged in a pool architecture. In the first round screen, a hit (R27) was identified in one of the 30 superpools each comprising of DNA from 2025 plants. The same hit also came through in the second round screen which identified the hit in one of the 9 subpools (subpool 237) each comprising of DNA from 225 plants. At the single pool level the same hit was detected from DNA extracted from plants in the pool 5887. Screening at the individual plants level identified 3 plants; 3, 27r and 37, which carried the T-DNA insertion in the *CLK1* gene.

6 insertion lines and the filters were probed with a 2 kb *CLK1* fragment PCR amplified with the primer pair CLK-U and YC50r (Figure 4.15A).

Result from the Southern analysis showed that 4 of the 6 plants were homozygous for the insertion. From figure 4.15B, the Southern analysis of plants in the T3 generation demonstrated that the T2 plant 11B and 20 were both homozygous for the T-DNA allele since their progeny only segregated the 6.7 kb T-DNA associated band whilst the progeny of plant 5 segregated for both the 8.7 endogenous *CLK1* band and the 6.7 kb band suggesting that the parental plant was heterozygous for the insertion. Plants 3, 27r and 37 were genotyped in a similar way and it appeared that plant 3 and 27r were homozygous for the T-DNA allele, now designated as *clk1-1* allele, whereas plant 37 was heterozygous for the insertion (data not shown). Detailed investigation revealed no obvious phenotypic defects exhibited by these plants apart from delayed flowering. The six *clk1-1* lines flowered after about 6-9 weeks post-sowing. The observed flowering times were significantly longer than the wild-type siblings and also the Ws wild-type line sown at the same time. This raised the possibility that the T-DNA insertion in the *CLK1* gene caused late-flowering. However, careful analysis of the progeny of plant 5 showed that this was not the case. Thus plant 5.1 which had lost the T-DNA insertion still exhibited the late-flowering phenotype suggesting that the insertion at the *CLK1* gene was unrelated to the phenotype seen.

Two other possible alternatives could account for the delayed flowering time see in the insertion lines. Thus another T-DNA copy might have been inserted into and disrupt the function of a flowering time gene or alternatively, the T0 parental plant could have carried a mutation in a flowering time locus.

In order to test the second insertion locus hypothesis the number of T-DNA insertions in the insertion lines were determined. Southern analysis was carried out on the *Eco*RI cut genomic DNA. It was decided to use the *GUS* gene sequence as a probe to detect the insertion number since the T-DNA fragment inserted contained the *GUS* ORF (Figure 4.15A). The probe used in this experiment was a 1.8 kb *Bam*HI-*Eco*RI fragment containing the *GUS* gene from the plasmid pBI121 (Clontech). The number of the bands seen on the autoradiograph would represent the T-DNA copy number since the integrated T-DNA gives a single fragment containing the *GUS* gene when cut with *Eco*RI (Figure 4.15A). The result from the Southern analysis indicated that two T-DNA copies were found in the genome of all 6 insertion lines (Figure 4.15B, data not shown for lines 3, 27r and 37). The 5.2 kb fragment represented the

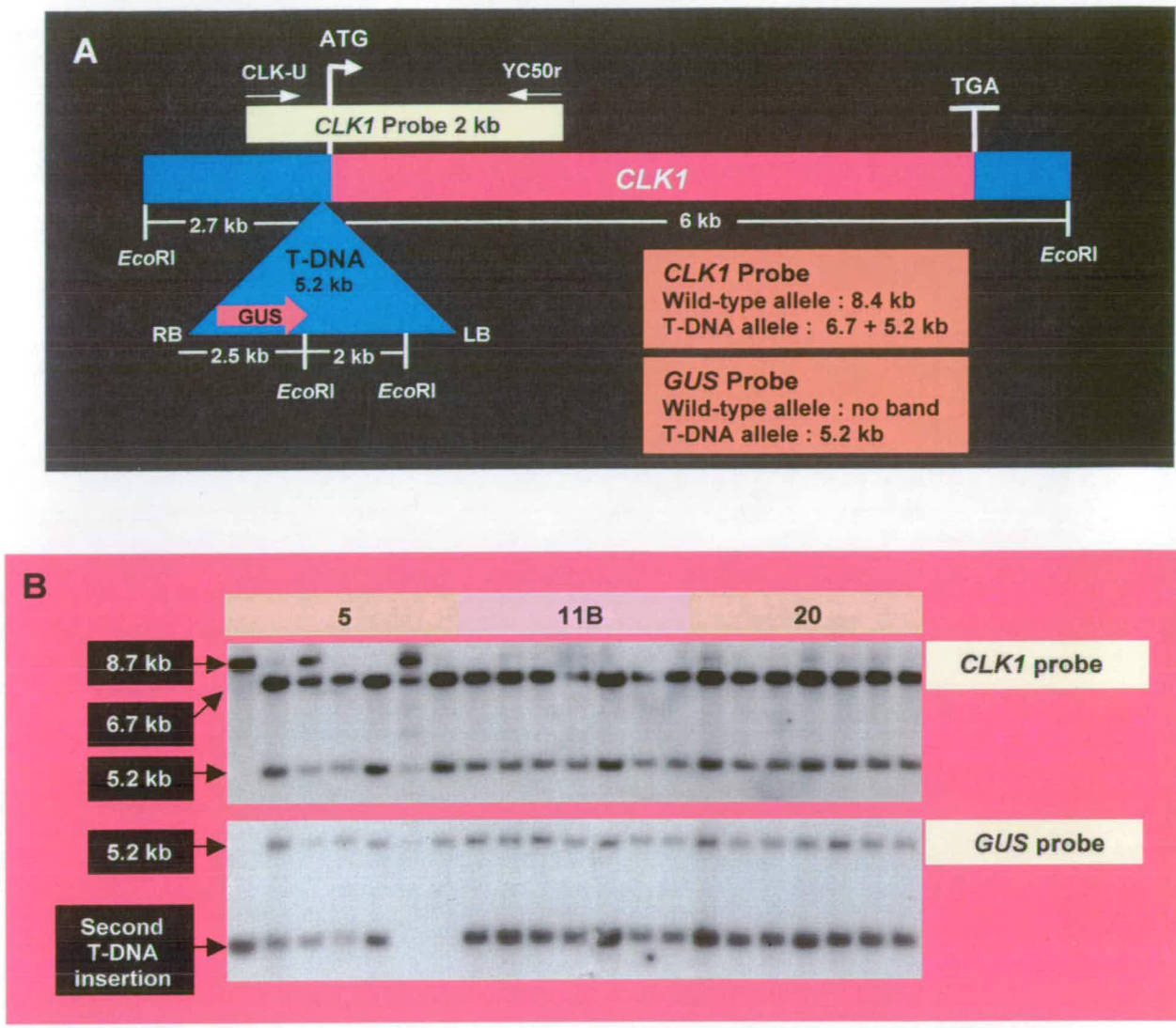


Figure 4.15 Molecular genotyping of *CLK1* T-DNA insertion lines. A, Rationale behind the experiment. The T-DNA associated *clk1-1* allele can be distinguished from wild-type allele by Southern blotting of genomic DNA using a *CLK1* probe amplified from the primer pair CLK-U and YC50r. The number of T-DNA copy in the genome can be identified by probing with *GUS* fragment. B, Southern analysis on randomly chosen 6 progeny of the T2 plants carrying the T-DNA insertion in the *CLK1* gene using *CLK1* probe showed that plant 11B and 20 were homozygous whereas plants 5 was heterozygous for the *clk1-1* allele. Southern analysis using the *GUS* probe showed that of the 3 plants, only plant 5 was heterozygous for the second T-DNA insertion and two of its progeny, plants 5.6 and 5.7 (from left), had lost the second insertion and were homozygous for the *clk1-1* allele.

insertion in the *CLK1* gene while the smaller band corresponded to an insertion elsewhere in the genome. Of the six plants, only plant 5 was heterozygous for the second T-DNA insertion since its progeny segregated for the presence of the smaller band (Figure 4.15B). Data from segregation analysis revealed that the two insertions were unlinked. Thus a near segregation ratio of 15:1 (143:11) kanamycin resistant to kanamycin sensitive seedlings was scored from the progeny of plant 5 on GM+kanamycin plates suggesting independent segregration of the two T-DNA copies. In addition data from the Southern analysis also showed that although two of the progeny of plant 5 investigated; plant 5.6 and 5.7, had lost the second site insertion (Figure 4.15B). However, the flowering time of those plants was still significantly much longer than normal with 34 and 49 leaves produced at the time of flowering respectively. This finding suggested that the second insertion did not affect the flowering time.

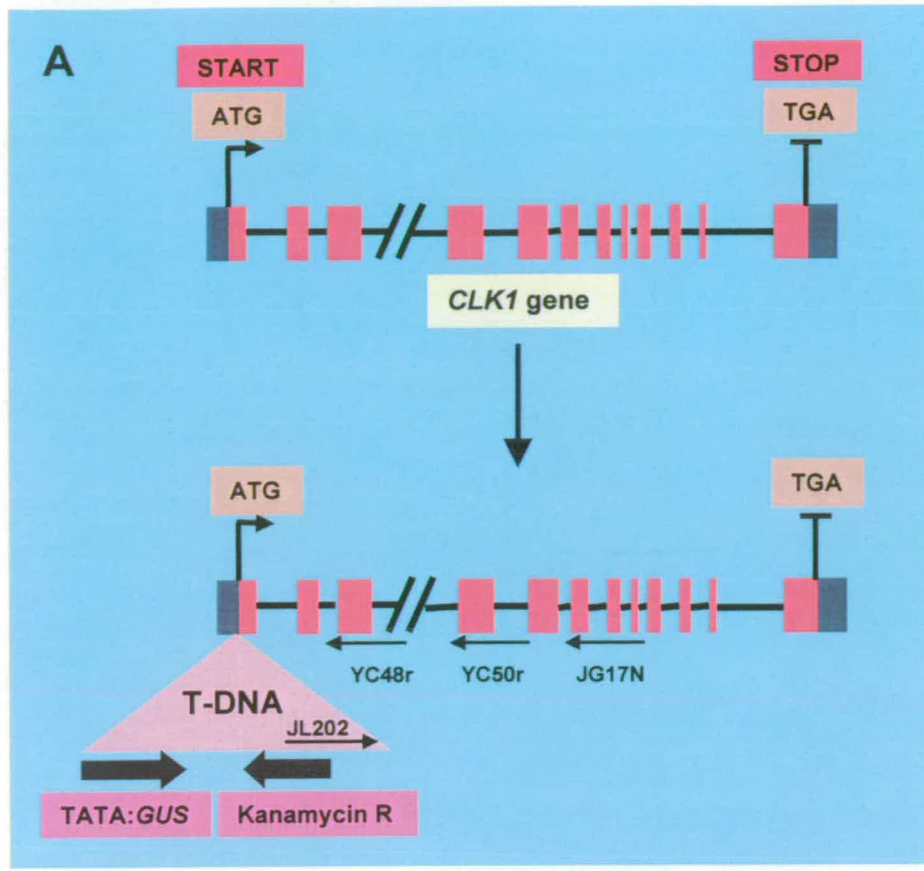
The data accumulated so far indicated that neither the T-DNA insertion at the *CLK1* gene nor the second insertion was responsible for the late flowering phenotype associated with the *clk1-1* lines. It seemed, therefore, likely that the phenotype was caused by a mutation at an unlinked locus. This is supported by the finding that when plant 5.7 which had lost the second insertion and was homozygous for the *clk1-1* allele was backcrossed to the Ws wild-type background, the late-flowering phenotype segregated as recessive and independently of the *clk1-1* allele. Previous study on the *CLF* gene indicated that the leaf curling phenotype is suppressed in some late flowering backgrounds (J Goodrich, pers. comm.). It is possible that a phenotype caused by T-DNA insertion at the *CLK1* locus might have been masked in the late flowering background. Rather disappointingly, however, the F2 plants from the backcross of plant 5.7 to Ws background which were homozygous for the T-DNA insertion and had normal flowering time did not show any defects in leaf or other aspects of development (data not shown).

This raised the question of the effect, if any, of the T-DNA insertion on the function of the *CLK1* gene. Because the insertion was not in the coding sequences, it is possible that the insertion might not have any effect on *CLK1* function or it might have only resulted in a weak mutant allele. Thus transcription might have been able to proceed to completion even in the presencé of the T-DNA and hence the gene function is not affected. Alternatively the insertion might cause only a slight perturbation of the transcriptional regulation machinery resulting in only a weak disruption of the *CLK1* function. It is also possible that transcription might be able to initiate as normal from *CLK1* promoter but terminates in the T-DNA since

the *GUS* gene contains a 3' terminator sequence (Figure 4.16A). However it is possible that the enhancer for the *NptII* gene which confers resistance to kanamycin is able to initiate transcription within the T-DNA that proceed into the *CLK1* gene resulting in the *CLK1* function being restored. To investigate the possible effect of the T-DNA insertion on CLK1 transcription, a semi-quantitative RT-PCR analysis was carried out. Thus total RNA was bulk extracted from progeny of plant 5 11B and 20. PCR was carried out using the T-DNA specific primer JL202 and a series of *CLK1* specific reverse primers. These were YC48r which binds nearest to the 5' end to YC49r, YC50r and JG17N in the middle of the gene. In all cases the cDNA bands of expected size were amplified and no significant differences in the amount of the amplified products was observed between the T-DNA heterozygous (plant 5) and homozygous (plant 11B and 20) samples and also between the different PCR reactions with different *CLK1* reverse primers (Figure 4.16B). This result indicated that a transcript was initiated in the T-DNA and extended into the *CLK1* open reading frame. However the chance that the insertion caused functional defect at the protein level cannot yet be discarded. An antibody against CLK1 protein would be needed to look at the effects on CLK1 protein size and quantity.

Given the strong likelihood of genetic redundancy between the *CLF* and *CLK1* gene as suggested by the expression data and the analysis of the *CLK1* transgene, it was of particular interest to analyse possible genetic interactions between the *clk1-1* allele and the *clf* mutation. To address this issue plant lines that carried both the *clf* mutation and the *clk1-1* allele were generated. Thus the *clk1-1*/+ heterozygote plant 5 was used as a male in the cross to *clf-2/clf-2* and *clf-50/clf-50* females. Wild-type F1 plants that carried the *clk1-1* allele were selected on kanamycin plates and allowed to self. The F2 seeds were put out on soil and their phenotype investigated.

In both cases, double mutants with striking phenotypes were recovered in the F2 generation with the expected 1/16 frequency. These plants showed strong enhancement of the *clf* phenotype indicating a synergistic interaction between the *clf* and *clk1-1* alleles. The *clf-50/clk1-1* doubles were strikingly similar to the plants with enhanced phenotype obtained with the *CLK1* dominant negative transgene in the *clf-50* background. Thus the double mutant plants were significantly smaller than the *clf-50* plants (Figure 4.17B). The plants did not produce normal rosette leaves but instead formed a few sessile leaves which characteristically resembled cauline leaves. In most cases vegetative growth was severely affected and drastically shortened. Only about 10 days after sowing these small plants started



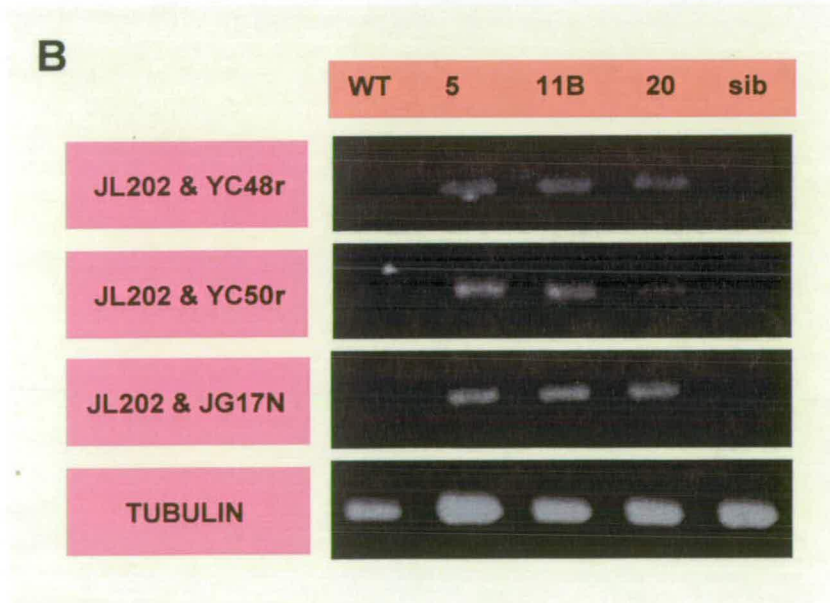


Figure 4.16 Molecular effect of T-DNA insertion on *CLK1* transcription. A, T-DNA insertion could lead to a total disruption of the *CLK1* transcription. Alternatively, it might only cause small interference with the transcription machinery and hence small effect on transcription. Transcription might be able to proceed as normal but could stop when reaching the stop codon of the *GUS* gene in the T-DNA. However, the enhancer of the kanamycin resistant gene lying nearby could initiate transcription that proceeds into the *CLK1* coding sequence resulting in the *CLK1* function being restored. Semi-quantitative RT-PCR analysis using the T-DNA specific primer, JL202 and a number of *CLK1* primers was carried out to find out if there might be a premature truncation of *CLK1* transcription on the T-DNA insertion lines. B, Result from RT-PCR analysis indicated that there was no difference in the amounts of PCR product amplified using different *CLK1* specific primers in the RNA samples from the T-DNA insertion lines suggesting that the effect of the insertion on *CLK1* transcription, if any, was minimal.

their reproductive development and produced small floral buds at the apex without bolting (Figure 4.17B). There were some exceptions, nevertheless. Thus in some cases inflorescence structure with elongated internode were observed. However, this mild phenotype could be the result of the suppression of the *clf* and possibly *clk1-1* defects by the mutation in the late flowering locus which also segregated in the F2 generation. Flowers of those plants with the mild phenotype were small but phenotypically rather normal with full sets of floral organs and no homeotic conversion was seen (Figure 4.17B). On the contrary, in most plants with the extreme phenotype only a terminal carpel structure was found and no other floral structure was seen (data not shown). Interestingly this phenotype was remarkable similar to that of the *emf2-3* plants suggesting that the *CLK1*, *CLF* and *EMF2* gene might act in a common regulatory pathway.

The phenotype of the *clf-2/clk1-1*double mutants was somewhat milder than that of the *clf-50/clk1-1* doubles (Figure 4.17A). Thus even though their development was drastically affected the organs produced such as the cotyledons and the sessile leaves were noticeably larger than those of the *clf-50/clk1-1* double mutants. Again the plants did not bolt and floral buds were often seen surrounded by the sessile leaves and the flowers produced were largely unaffected (Figure 4.17A).

In general it can be said that both classes of double mutant combination of the *clk1-1* insertion allele and the *clf* mutant allele exhibited a premature switch from the vegetative to reproductive development. Three lines of evidence suggested that the extreme enhancement of the *clf* phenotype seen was likely caused by the T-DNA insertion in the *CLK1* gene. First, molecular genotyping showed that the enhanced phenotype co-segregated with the T-DNA insertion in the progeny of the *clf/clf clk1-1/+* plants. Thus DNA was extracted from the progeny of an F2 plant which was genotypically *clk1-1/+* and *clf-50/clf-50* from the *clf-50* cross. And PCR analysis was performed using the T-DNA associated JL202 and the *CLK1* specific YC50r primer pair and the CLK-U and YC50r pair. The result showed that all the mutants with the extreme phenotype were homozygous for the *clk1-1* allele. On the other hand the *clf-50* siblings were either heterozygous for the insertion or had lost it completely (Figure 4.17C). Second, the phenotypes of the double mutants from crosses to two different *clf* alleles in different genetic backgrounds (*clf-2* is in L*er* while *clf-50* is in Ws background) were remarkably similar. It would be extremely unlikely that they were both caused by spontaneous mutations in an identical second locus. This result also reinforced the notion that *CLF* and *CLK1* are to an extent genetically redundant. Third, a similar enhanced

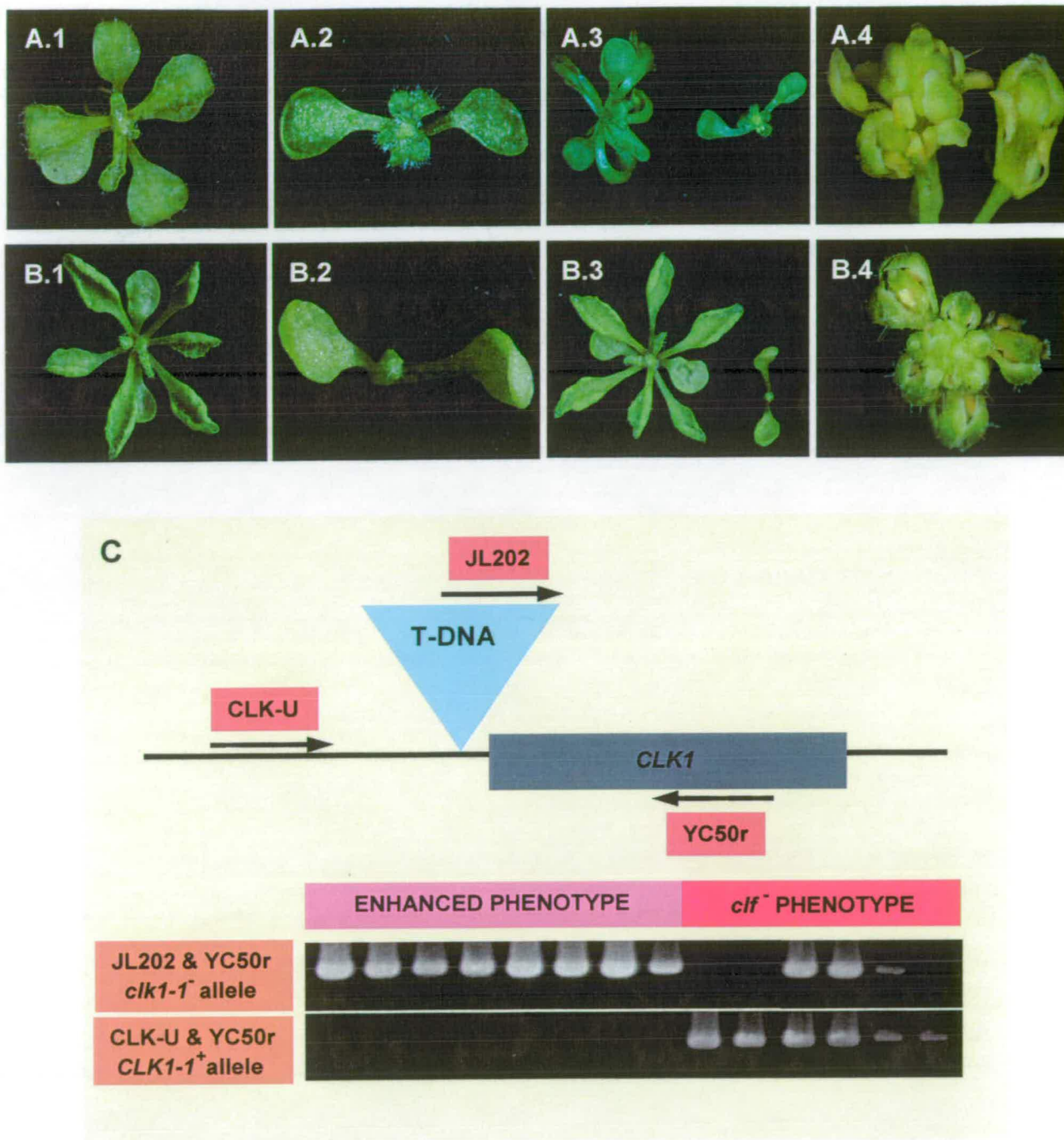


Figure 4.17 Genetic and molecular characterisation of *clf clk1* double mutants. A-B, Plants that were both homozygous for the *clf-2* or the *clf-50* allele and the *clk1-1* allele (A.2 and B.2 respectively) showed a strong enhancement of the *clf-2* and the *clf-50* phenotype (A.1 and B.1 respectively). Size comparison showed that the *clf2 clk1-1* and the *clf-50 clk1-1* double mutants were much smaller than their respective *clf* parent (A.3 and B.3). Only a few small sessile leaves were seen in the plants with the enhanced phenotype. Those plants also exhibited shortened vegetative development and flowered only about 10-14 days post-germination. Flowers produced were small but relatively phenotypically normal (A.4 and B.4). C, Progeny of a *clf-50/clf-50 clk1-1/+* plant segregated a 3:1 ratio of *clf*:enhanced phenotype. Molecular genotyping by PCR indicated that plants with the enhanced phenotype were all homozygous for the *clk1-1* allele whereas their *clf-50* siblings either lacked the T-DNA insertion or were heterozygous for the insertion but none were homozygous for the *clk1-1* allele.

phenotype was seen in *clf/clf* plants carrying the *CLK1* dominant negative construct. Together these links between the extreme mutant phenotype and the *clk1-1* insertion allele indicated that the T-DNA insertion likely had some effect on the *CLK1* function.

Chapter 5. Discussion

5.1 Genetic Redundancy between *CLF* and *CLK1*

5.2 Interaction between Pc-G Genes, Flowering Time Genes and Floral Homeotic Genes during *Arabidopsis* Development

5.3 Role of Plant Pc-G Genes in the Control of Cell Proliferation

5.4 Distinct Pc-G Complexes in Plants

5. Discussion

Forward and reverse genetic approaches were used successfully in this study to identify and characterise two *Arabidopsis* Pc-G genes. Thus characterisation of the *moe leaf* mutation indicated that it is an unusual weak allele of the recently cloned *EMF2* gene and detailed analysis of an *Arabidopsis CURLY LEAF* homologue, the *CURLY LEAF LIKE1* gene, suggested functional redundancy between the two genes. The results give novel insights into the roles of the *EMF2* and *CLK1* genes during *Arabidopsis* development, and also reveal aspects of *CLF* functions that were not previously evident from studies of null *clf* alleles. In particular, they suggest that *CLF* and *CLK1* may act in the *EMF* pathway, most likely as part of a common complex.

5.1 Genetic redundancy between *CLF* and *CLK1*

The deduced amino acid sequence of the *Arabidopsis* CLF homologue, CLK1 protein, showed very high sequence and structural homology with the CLF protein particularly in the conserved motifs such as the cysteine-rich CXC, SANT and SET domain. The high sequence conservation between the CLF and CLK1 proteins invoked the possibility that their functions are similar. In supportive of this view, *in situ* hybridisation analysis has shown that *CLK1* expression pattern is extremely similar to that of *CLF*. Thus both genes are expressed ubiquitously throughout the shoot, inflorescence and floral meristems. Furthermore protein interaction study revealed that as in the *Drosophila* paradigm both the E(Z)-like CLF and CLK1 proteins interact with the ESC homologue, FIE, suggesting that the three proteins might act very closely and impinge their regulatory effects on similar target genes.

To investigate the possible redundancy between *CLF* and *CLK1* in more detail, attempts were made to isolate and characterise various mutant alleles of *CLK1*. Although it is unclear whether either the dominant negative *CLK1* transgene or the T-DNA associated, *clk1-1*, allele are true null alleles, the synergistic interaction between both of these allele and *clf* mutation as seen in double mutant combinations strongly suggests that *CLK1* function is, to a certain extent, redundant to *CLF* activity.

The emerging theme from the genetic and molecular analyses of various plant developmental and physiological networks is the extreme complexity within these regulatory pathways. In most cases, genetic dissection of these pathways is made complicated by the fact that they are comprised of factors with highly redundant activity. In *Arabidopsis* the ancient tetraploidization and subsequent linear duplication events have resulted in a large amount of its genome being duplicated (Pickett and Meeks-Wagner, 1995). As a result many gene families exist each of which comprises of members with similar sequences. As predicted, in many cases it has been found that not only the sequences but the functions of the paralogous genes in the same gene family are also highly similar. Perhaps the most well characterised gene family in *Arabidopsis* is the MADS-box multigene family which consists of more then 40 members. According to the result from phylogenetic studies the MADS-box genes can be grouped into distinct subfamilies. Genetic and molecular analyses have shown that within a given subfamily, the gene members often perform similar and redundant roles during development. Redundancy among MADS-box genes was first demonstrated between the gene pair of *AP1* and *CAL* which act redundantly to specify floral meristem identity (Bowman et al., 1993; Kempin et al., 1995). Subsequent study has added the redundant activity of the closely related *FUL* gene to the *AP1/CAL* pathway (Ferrándiz et al., 2000). Apart from the *AP1/CAL/FU*L case, many other examples of functional redundancy among the MADS-box group have now been uncovered. For example, the homologous *SHATTERPROOF* (*SHP*) gene *1* and *2* share similar expression patterns and code for proteins that show 87% amino acid sequence identity. Genetic analysis showed that putative loss-of-function mutants of either gene has no detectable phenotypic defects. However when the alleles of the two genes are combined the resulting *shp1 shp2* double mutants bear fruit that fail to dehisce suggesting that the two *SHP* genes encode redundant proteins whose activities are required for fruit dehiscence in *Arabidopsis* (Liljegren et al., 2000). Another example of genetic redundancy among closely related MADS-box members is that illustrated by the three *SEPALLATA* genes which play a part in the specification of floral organ identities. Again insertion alleles of the *SEP1*, *SPE2* and *SEP3* display no obvious phenotypic alteration either singly or in double mutant combinations but the triple mutant of the three genes exhibit a striking phenotype with all of the floral organs in whorl 2, 3 and 4 being converted to sepals suggesting that the *SEP1*, *2* and *3* function redundantly and in combination with the B- and C-function genes to specify petal, stamen and carpel identities in whorl 2, 3 and 4 respectively (Pelaz et al., 2000).

With the exception of some of the duplicated housekeeping genes, complete functional redundancy is extremely rare. Thus redundancy between members of a gene family is usually only partial and in most cases members of a gene family with a shared set of function have also acquired new completely independent functions during evolution. The acquisition of distinct, specialised functions of a pair of duplicated genes would allow the activities of both genes being selectively maintained in evolution. Thus according to several proposed models of evolutionary pressures permitting the maintenance of genetic redundancies, as a result of positive selective pressure being impinged upon only one gene copy, deleterious mutations would accumulate in one of the duplicated genes such that redundancy between the two genes are eventually lost since one of the two genes will become pseudogenes (Tautz, 1992; Thomas, 1993; Weintraub, 1993). Pseudogenes are, to follow a common analogy, rusting wrecks of genes that have been holed below the waterline by a serious mutation and sunk. They now lie on the floor of the genomic ocean, gradually growing rustier (that is, accumulating more mutations) until they no longer even resemble the gene they once were (quoted from chapter 8, Ridley, 2000). Thus the redundant functions of two homologous genes will be maintained only when some functional independence have arose between them resulting in selectable non-redundant functions, and as a by-product the redundant activities of both genes being positively selected in evolution. Genetic and molecular studies of several of the closely related and functionally redundant MADS-box genes have provided evidence to support this view of the maintenance of redundancies by positive selection force. Thus it is emerging that in addition to the preserved shared functions with their homologues, many of these genes appear to have evolved a number of other developmental roles. For example, in addition to the floral meristem identity specification role that it shares with the highly related *AP1* and *CAL* gene, the *FUL* MADS-box factor also has other independent functions, one of which is the role in fruit valve differentiation (Gu et al., 1998). Furthermore, genetic analysis also suggests that the 3 *SEP* functions are also likely not completely redundant to one another. Thus the additional misexpression of the *SEP2* gene in the transgenic lines in which *SEP3*, *AP1*, *AP3* and *PI* are ectopically expressed result in a more complete conversion of rosette leaves to petals suggesting that *SEP2* and *SEP3* are not completely redundant i.e., *SEP2* is likely to have a function that is independent of *SEP3* (Pelaz et al., 2001). In accordance with this finding, *SEP3* appears to have a different expression pattern from that of *SEP1* and *SEP2*. Thus *SEP1* and *2* are expressed in all four floral whorls while *SEP3* transcript is only found in whorl two, three and four (Flanagan and Ma, 1994; Savidge et al., 1995; Mandel and Yanofsky, 1998). Moreover, in

the case of the *CAL* and *AP1* gene it seems, at first glance, that the role of *CAL* in the specification of floral meristem identity is fully encompassed by *AP1* since the *cal* phenotype is manifested only in an *ap1* mutant background (Bowman et al., 1993). However if *CAL* activity is totally redundant to *AP1* during the establishment of floral meristem identity and has no other function, it is likely to be strongly selected against in evolution. The fact that *CAL* activity has been maintained implies that there is likely some selective advantage associated with it. Indeed result from a population genetic study revealed that the *CAL* locus is evolving in a non-neutral fashion with considerable degree of intraspecific diversity (Purugganan and Suddith, 1998). Although the mechanisms responsible for this pattern of evolution of *CAL* and any associated selective advantages are still not known, the result obtained strongly suggests that there are positive evolutionary forces that act to maintain the diversity and hence the existence of the *CAL* locus. Another line of evidence to suggest that the redundancy between *AP1* and *CAL* is only likely partial comes from the study of the expression patterns of the two genes. Thus although their initial expression patterns are very similar with the RNA transcripts found accumulated over the floral meristem consistent with the redundant role of *CAL* and *AP1* in specifying floral meristem identity, their late expression domains become somewhat different from each other. It was found that *AP1* RNA accumulates to a high level throughout sepal and petal development reflecting its significant role in specifying sepal and petal identity in the outer two floral whorls. *CAL* RNA, on the other hand, is only detected at very low levels in these organ primordia consistent with the inability of *CAL* to substitute for *AP1* in organ identity specification. In addition, *CAL* is expressed in the vasculature of the inflorescence stem where *AP1* transcripts are not present (Kempin et al., 1995). Given the highly redundant nature of many developmental networks in *Arabidopsis*, it is rather likely that subtle functions of *CAL* might be hard to discern under normal laboratory conditions and will only become apparent in particular environments.

Many evolutionists have argued that, in addition to positive selection, genetic redundancies are also likely preserved by another type of evolutionary force, negative selection. According to this idea, negative or purifying selection acts to eliminate mutant alleles by selecting against individuals which carry the alleles. As a result, gene and hence functional divergence between duplicated gene pairs is slowed down. In reality, it is likely that the combination of both positive and negative selection is responsible for the maintenance of functional redundancies over a long period of time. The fact that many of the MADS-box genes have

evolved new, specialised function in addition to the redundant activity that they shared with their homologues could reflect the strong selective pressures on ancestral land plant lineages to acquire many novel, more complex characteristics that are essential for their survival. The presence of several genetically redundant activities in many of the plant developmental networks can also be regarded as a fail-safe mechanism that has evolved to prevent possible damage that could be introduced by mutation and also environmental vagaries. Thus it can be envisaged that if there are two redundant pathways that simultaneously operate toward the wild-type phenotype, the only scenario that would lead to major disruption of the phenotype is when both pathways are inactivated.

Although the data accumulated from this study strongly suggests that functional redundancy exists between the *CLF* and *CLK1* gene pair, one issue concerning the nature of the redundancy is still left unaccounted for. Thus it is not known as to what extent are the two functions redundant to one another. Given the likelihood of strong positive selection force that acted upon the two genes after the duplication event it is unlikely that either the *CLF* or the *CLK1* gene has been evolving neutrally. Thus both genes are likely to have acquired some functional independence to avoid being selected against. This is, by all means, true in the case of *CLF* as its single mutant exhibits defects at both phenotypic and molecular level which cannot be corrected for by other gene activities. In the case of *CLK1*, analysis of the transgenic lines carrying the *CLK1* dominant negative construct revealed a novel mild leaf serration phenotype that is not seen in *clf* mutant implying that *CLK1* has independent function of *CLF*. However study of the T-DNA insertion alleles revealed no associated marked phenotypic changes suggesting that *CLK1* activity is highly if not completely redundant with that of *CLF*. Still it is possible that these are very weak alleles of *CLK1* and only cause subtle phenotypic defect which can't be assessed under normal laboratory conditions. The lack of phenotype could also be due to the fact that *CLK1* is expressed at very low level. Thus the CLK1 protein may be less abundant than the CLF protein and when depleted would have a much lesser impact on the target genes. If the differences are indeed solely due to the expression levels, it is predicted that overexpression of *CLK1* under the control of *CLF* promoter or the strong cauliflower mosaic virus 35S promoter would complement the *clf* phenotype. Another possibility is that other aspects of *CLK1* function are mostly redundant to the activities of other factors.

In order to gain more understanding into the genetic redundancy between the *CLK1* and the *CLF* and possibly other *Arabidopsis* Pc-G genes, more alleles of the *CLK1* genes need to be isolated. In particular a definitive null *clk1* allele, such as an insertion in the coding sequence, needs identified to be certain of null phenotype. Also the generation of a good anti-CLK1 antibody would help determining if the *clk1* insertion allele causes the elimination of the CLK1 protein, and hence defines a null allele. If the null *clk1* allele has little or no phenotype, *CLK1* function is likely redundant to that of *CLF*.

5.2 Interactions between Pc-G genes, flowering time genes and floral homeotic genes during *Arabidopsis* development

Genetic analysis of the double mutant combinations of *moe* and various floral homeotic genes and the molecular study of floral homeotic gene expression in the *moe* background showed that many of the phenotypic defects associated with the *moe* allele resemble those of the *clf* mutations suggesting that they may share similar basis. As in the case of *clf* plants, the leaf curling defect in *moe* mutants requires the ectopic expression of the homeotic gene *AG* in the leaves. Therefore when the activity of *AG* was removed in the *moe ag* double mutant background the degree of leaf curling was alleviated to a large extent. The fact that there was still some leaf curling in the *moe ag* plants and that the plants still flowered early and were rather small in size compared to wild-type or *clf ag* plants suggested that other target genes were still misexpressed and their activities might contribute to the rather more severe phenotype in the *moe ag* background as compared to the *clf ag* plants. Interestingly, the *GUS* reporter gene analysis revealed that in addition to *AG*, the homeotic gene *AP3* and *PI* were also ectopically expressed, albeit to a lesser extent, in the leaves of *moe* mutants. Previous study has demonstrated that transgenic plants that constitutively express both the *AP3* and *PI* genes under the control of the cauliflower mosaic virus 35S promoter exhibit vegetative phenotypes which included leaf curling and early flowering (Krizek and Meyerowitz, 1996). In contrast ectopic expression of either *AP3* or *PI* alone does not cause those defects (Jack et al., 1994; Krizek and Meyerowitz, 1996). The finding that *AP3* but not *PI* is misexpressed in the leaves of the *clf* plants (Goodrich et al., 1997), whereas both genes are ectopically expressed in the leaves of the *moe* mutants may explain why the leaf curling is more severe in the *moe* background. If this is the case it is expected that the vegetative phenotype of the *moe* plants would be restored to nearer to wild-type in the *ag ap3 pi* triple mutant background. It is noteworthy that study of the weak *moe* allele of the *EMF2* gene has allowed the identification of the *AP3* and *PI* as its target genes which although is expected since *EMF2* has a general role as a floral repressor, had not been established in the studies of other *emf2* mutant alleles. In addition to *AP3* and *PI* other floral genes could also be misexpressed in *moe* leaves and therefore further contribute to the curling defects.

The fact that some *moe* flowers exhibited carpelloidy and staminoidy in the first and second floral whorls respectively suggested that similar to *CLF*, the *EMF2* gene also represses *AG* expression in those 2 whorls of flowers. This result agrees with previous findings that *AG* is

ectopically expressed in the outer two whorls of other *emf2* alleles (Chen et al., 1997). However, the fact that the majority of the *moe* flowers did not show homeotic conversions in whorl 1 and 2 suggest that the role of *EMF2* as an *AG* repressor in the flowers is to an extent redundant to the activities of other regulators such as *AP2*, *LUG*. It is conceivable that the role of *EMF2* as *AG* regulators during flower development was recruited at later times after their role as floral gene repressors during leaf development had been established. The presence of several *AG* repressors could mean that there is a stringent requirement for proper development of sepals and petals in the first and second floral whorls respectively. This might be of evolutionary significance particularly in the case of petals. Thus due to their importance in pollination and hence the species survival it would have been advantageous for plants to maintain several repressors of class C function in the flowers (Krizek et al., 2000). Previous analysis has shown that misexpression of *AG* is only detected in the *clf* mutant flowers from stage 9 of development onwards consistent with its role in maintaining the initial off-state of *AG* activity in the outer two floral whorls (Goodrich et al., 1997). Since *EMF2* also encodes a Pc-G protein, it would be interesting to carry out an *in situ* hybridisation experiment to investigate the pattern of *AG* expression during various stages of development in the *moe* background and compare with that of the *clf* flowers to determine if the *EMF2* might also act in a similar temporal manner as *CLF* during flower development.

As previously mentioned, the *EMF2* gene is expressed throughout the inflorescence and floral meristems (Yoshida et al., 2001). As with the case of *CLF*, the ubiquitous expression pattern of *EMF2* in all four floral whorls raises the question of how its activity as an *AG* repressor is only limited to the outer two floral whorls. One possibility is that EMF2 is recruited to the *AG* target by binding of the early repressor AP2 protein. This is supported by the finding that the weak *moe* and *ap2-1* alleles showed synergistic interaction which suggested that the two wild-type genes act in the same pathway. However, alternative explanations such as interactions of *EMF2* with whorl specific factors, post-translational modifications are not excluded.

The fact that the introduction of the weak *moe* allele of the *EMF2* gene into the weak *clf-9* background resulted in a double mutant phenotype that exhibited strong enhancement of the near wild-type *clf-9* phenotype strongly suggested that the two mutations interact synergistically. This finding is suggestive of a close interaction of the *CLF* and *EMF2* genes. This could mean that they act in a common complex. Consistent with this view, *CLF* and

EMF2 show overlapping expression patterns with transcripts of both genes found throughout leaf primordia, young leaves, inflorescence and floral meristems (Goodrich et al., 1997; Yoshida et al., 2001). Another line of evidence in favour of the common complex hypothesis comes from a recent finding that the *emf2-3* allele is epistatic to the *clf-2* allele (J Goodrich, pers. comm.). The finding that the phenotypes of the *clf clk1* double mutants obtained from this study are remarkably similar to the phenotypes of *emf2* mutants (Yang et al., 1995), and also the embryo-rescued *fie* mutant, also suggests that in wild-type the four proteins act in a common functional complex. To test this hypothesis immunoprecipitation experiments need to be carried out to determine if the CLF and EMF2 proteins co-immunoprecipitate.

An alternative explanation is that the two genes act in a common pathway with *EMF2* being upstream *CLF*. Thus *EMF2* might act as a positive regulator of *CLF* which in turn represses expression of floral homeotic targets such as *AG*. The epistatic nature of the *emf2-3* and *clf*-2 mutations and the synergistic relationship between the weak *moe* and *clf-9* alleles are both consistent with this notion. Comparison of *CLF* expression level in strong *emf2* mutant backgrounds with the wild-type level would help determine if this hypothesis is correct.

Although both of the models mentioned above are attractive and each has several lines of supportive evidence, they are inconsistent with the result from the phenotypic analysis of the double mutant of the weak *moe* and the null *clf-2* alleles. Thus unexpectedly, the *moe clf-2* double mutant phenotype, albeit slightly more severe, is remarkably similar to that of the *moe clf-9* double. The observed synergistic relationship of the *moe* and *clf-2* alleles suggests that the *EMF2* and *CLF* genes likely act in independent pathways and the corresponding proteins are unlikely to be parts of a common protein complex.

However, these apparently contradictory results that both the weak *clf-9* and null *clf-2* mutations act synergistically with *moe* can be potentially resolved by the partial redundancy between the *CLF* and *CLK1* genes. Thus the loss of or decrease in *CLF* activity can be partially compensated for by the *CLK1* function and therefore the phenotypic defects seen in the *clf* mutants are much milder compared to those in the *emf2* and embryo-rescued *fie* mutants. The redundancy between the *CLF* and *CLK1* gene pair also suggests that the null *clf-2* allele is, in fact, not a real null allele since in this background some aspects of the *CLF* function can still be supplied by the *CLK1* activity. This view is consistent with the observed synergistic interaction of the *clf-2* and *moe* alleles.

Another important feature of *clf* mutants which is often overlooked is that they flower early (Goodrich et al., 1997). The early flowering phenotype was also observed in the *moe* mutants consistent with the proposed role of the *EMF2* gene as a floral repressor (Chen et al., 1997). This early flowering phenotype associated with the *moe* and other *emf2* alleles is likely the direct consequence of ectopic expression of various floral gene targets in the *emf2* mutant background. Thus recent molecular analysis has shown that the floral meristem identity genes *LFY* and *API* are both ectopically expressed from the early stage of development (Chen et al., 1997; Chou et al., 2001), while the results from this study provides evidence for the misexpression of the *AP3* and *PI* genes from very young seedling stage. In accordance with this view, previous studies have shown that overexpression of several floral genes including *AP1*, *AG*, *AP3*, *PI*, *SEP3* and *LFY* all result in plants exhibiting early flowering defects (Mandel and Yanofsky, 1995; Weigel and Nilsson, 1995; Krizek and Meyerowitz, 1996; Mizukami and Ma, 1997; Pelaz et al., 2001). In the case of the *clf* mutant the early flowering phenotype must be partly due to *AG* misexpression since the *clf ag* double mutants flower earlier than wild-type but later than the *clf* mutants (Goodrich et al., 1997).

As previously mentioned, genetic study of the *clf* mutants revealed that the leaf curling defect is somewhat alleviated in the *co* and *fca* late-flowering backgrounds (J Goodrich, pers. comm). Similarly when the *moe* allele was introduced into the late-flowering background the leaf curling phenotype became less severe (data not shown). This invokes the possibility that the *CLF* and *EMF2* genes might interact with various flowering time genes during the onset of reproductive development. A recent genetic study has shown that, at least in the case of the *EMF2* gene, this is indeed the case. Phenotypic analysis of the double mutant combinations between *emf2* mutants and various late-flowering mutants such as *ft*, *fca*, and *fd* suggests that *EMF2* genetically interacts with and acts upstream of many of these late flowering genes (Chou et al., 1998; Haung and Yang, 2001). Thus in addition to the expected increase in flowering time, the majority of these double mutant plants are significantly larger in size and also produce larger inflorescence and more flowers compared to the *emf2* single mutants. This indicates that the late flowering mutations can, to a certain extent, suppress many of the defects associated with the *emf2* mutations. A number of genetic and molecular analyses have implicated a role of most late-flowering genes in the activation of floral meristem identity genes such as *LFY*, *AP1* and *CAL* during the inflorescence-to-flower transition (reviewed in Pidkowich et al., 1999). This raises the possible scenario that the effect of the *emf2* mutation on the expression of many floral gene targets is indirect and

rather is a consequence of the effect on the activity of many flowering time genes. In accordance with this view analysis of the expression of *AP1* and *LFY* in the double mutants of *EMF* and various late-flowering mutants using promoter assay and quantitative RT-PCR suggested that the ectopic expression of these two genes in the *emf* mutant background requires the wild-type activity of most of the late-flowering genes studied. Thus the level of both *AP1* and *LFY* expression is reduced in the *emf* late-flowering double mutants consistent with the notion that the ectopic early activity of *AP1* and *LFY* is caused by the premature release of suppression of late-flowering gene activity in the *emf* mutant background (Chou et al., 2001). The incomplete loss of *AP1* and *LFY* expression in the *emf* late-flowering double mutants implies that there are at least two existing independent pathways among these late-flowering genes that promote floral fate. It is also consistent with the possibility that *EMF* genes might also exert their inhibitory effect directly on their target genes independent of the pathways involving the late-flowering genes. As previously mentioned, the early flowering and the terminal flower phenotype are remarkably similar to the phenotype of *tfl1* mutant plants. Genetic and molecular analyses have shown that the functions of the *TFL1* gene and the floral meristem identity genes are antagonistic and that *TFL1* acts to maintain the vegetative identity in inflorescence meristems by indirectly inhibiting the activity of the meristem identity genes *LFY*, *AP1* and *CAL* (Ratcliffe et al., 1998; Ratcliffe et al., 1999). Consistent with this view, both *LFY* and *AP1* are ectopically expressed in the centre of *tfl1* mutant shoot apices (Weigel et al., 1992; Bowman et al., 1993; Gustafson-Brown et al., 1994; Bradley et al., 1997). In contrast the activity of *TFL1* is inhibited at a transcriptional level by the activity of *LFY*, *AP1* and *CAL* in the periphery meristems that have acquired competence for flowering (Ratcliffe et al., 1999). The negative interactions between the *TFL1* and floral meristem identity genes is thought to be at the centre of the mechanism underlying growth phase transition during *Arabidopsis* life cycle. The similarities in the defects at both phenotypic and molecular level between the *emf2* and *tfl1* mutants suggest that the two genes might interact with each other to maintain inflorescence growth. This is supported by the finding that *emf* mutations are epistatic to *tfl1* alleles. Thus the *EMF* genes likely act downstream of the *TFL1* gene and that *TFL1* may mediate its function during inflorescence development by positively regulating *EMF* activities (Chen et al., 1997).

Due to the many features shared by the *clf* and *moe* mutations, one can envisage that the alleviation of the leaf curling defect in *clf* mutants in late-flowering background might also share a similar molecular basis with that of the *moe* plants as described above. Thus the loss

of *clf* function is likely leading to the derepression of many late-flowering targets and thereby starting a cascade of gene action that lead to the establishment of floral fate. Thus the early ectopic expression of the late-flowering genes would trigger the expression of the meristem identity genes which in turn switches on the transcription of the floral homeotic genes such as *AP3*, *PI* and *AG*. It is expected, therefore, that mutations in the downstream late-flowering targets of *CLF* would suppress many aspects of the *clf* phenotype. Thus the flowering time and hence the vegetative growth is likely to be lengthened resulting in bigger plants with more leaves produced. In addition the loss of late flowering gene functions should also result in the delay of the onset of their floral target gene expression and therefore the severity of many of the phenotypic defects associated with the ectopic expression of these genes should be less in the *clf* late-flowering double mutant compared to the *clf* single mutant. The fact that even in the late-flowering background, albeit somewhat alleviated, the leaf curling is still noticeably apparent and that the *clf* plants still flower earlier than wild-type suggest that *CLF* likely also mediates its repressive function through pathways that are independent of the late-flowering genes, one of which could be the direct interaction with the floral gene targets.

5.3 Role of plant Pc-G genes in the control of cell proliferation

Studies of Pc-G genes in animals have suggested that in addition to their role in maintaining the repressed state of homeotic gene expression, many Pc-G members also participate in various aspects of cell proliferation control. For example in the nematode *C. elegans* the maternally derived *mes-2*, *mes3*, *mes-4* and *mes-6* Pc-G gene products are required for proper germline proliferation (Holderman et al., 1998; Korf et al., 1998). Additional evidence that links Pc-G genes with the control of cellular proliferation comes from the finding that transgenic mice that overexpress the *bmi-1* gene, a homologue of the *Drosophila psc* gene, exhibit defects in the proliferation of haematopoietic cells amongst many of their phenotypic abnormalities and that one of its targets is the *INK4a* gene, whose products act as cell-cycle inhibitors and suppressors of tumour growth (Jacobs et al., 1999).

Since Pc-G function in the control of homeotic gene expression appears to be conserved from animal to plant systems, it is of particular interest to investigate if the role of Pc-G genes in cell proliferation control has also been conserved in evolution. Indeed the data accumulated in this and many of the recent studies on plant Pc-G genes indicate that this might be the case. Thus there are now several emerging compelling pieces of evidence that link plant Pc-G genes and the regulation of cell proliferation of development.

One of the striking aspects of the phenotype of the *clf* and *moe* mutant plants is that their sizes are significantly smaller than that of the wild-type. This becomes much more pronounced when the two mutation are combined with the resulting *clf moe* double mutant being extremely small in size. A similar phenotype is also apparent in various *emf2* alleles, the *clf clk1* double mutants and the *fie* mutant plants. Thus not only is their vegetative growth phase drastically shortened but the vegetative organs of those mutants are very much reduced in size compared to wild-type, resulting in a severe impact on plant morphology. This strongly indicates that one of the most fundamental aspects of growth, that is the control of cell proliferation is drastically affected in those mutants. This in turn hints that the *Arabidopsis* Pc-G genes play an important part in such developmental control.

One of the early lines of evidence that supports the proposed role of plant Pc-G genes in cell proliferation regulation comes from the study of the *FIS* genes. As previously mentioned, phenotypic analyses of the three *FIS* genes showed that loss-of-function *fis* mutants

exhibited autonomous endosperm development without fertilisation (Ohad et al., 1996; Chaudhury et al, 1997; Kiyosue et al., 1999; Ohad et al., 1999). This suggests that the *fis* mutations activate central-cell nuclear replication and hence endosperm proliferation in the absence of fertilisation. Therefore in wild-type the *FIS* genes participate in the inhibition of central-cell nuclear replication in the female gametophyte before fertilisation perhaps by suppressing transcription of genes required for the replication of the central cell nucleus. Moreover, *FIS* activities are also required for proper cell proliferation post-fertilisation as when *fis* mutant ovules are fertilised, endosperm overproliferates relative to wild-type (Sorensen et al., 2000).

In addition to the *FIS* genes, the role in cell proliferation control has also been implied for the *CLF* gene. A recent study using an anatomical and genetic approach has indicated abnormalities in cell proliferation patterning in the *clf* mutant. Thus both the extent of cell elongation and the number of cell divisions are substantially decreased in *clf* mutant leaves suggesting that *CLF* function regulates both the cell elongation and cell division process during leaf development. The upward curling of *clf* mutant leaves could be the result of the greater reduction in cell expansion along the leaf margin compared to along the leaf midrib region. Alternatively, there could be a greater reduction in cell division in the upper leaf cell layers compared to the lower cell layers (Kim et al., 1998). Several lines of evidence from independent studies suggest that the effect of *clf* mutation on the regulation of cell expansion and cell division may be an indirect effect of the ectopic expression of the floral homeotic gene *AG*.

One of the striking features of *ag* mutant flowers is that they are indeterminate. Thus in addition to the effect on organ identity, mutations in *AG* result in the loss of determinacy of the floral meristem causing the formation of indeterminate flowers in which the fourth whorl acts as a new meristem and continues to proliferate, resulting in the replacement of carpels by interior flowers (Bowman et al., 1989; Yanofsky et al., 1990). This inability to limit the proliferation of floral stem cells in the centre of the flower indicates that in addition to its role in carpel and stamen identity specification wild-type *AG* function is also required locally to repress cell division and thereby prevent proliferation of stem cells in the centre of the flower (Meyerowitz, 1997). The role of *AG* in the termination of the floral meristem is supported by evidence from gain-of-function studies. Thus in transgenic plants in which *AG* is constitutively expressed under the control of the cauliflower mosaic virus 35S promoter

the inflorescence meristem terminated in a central flower indicating that *AG* function is sufficient to convert indeterminate to determinate meristems by preventing the accumulation of a pool of undifferentiated stem cells in the centre of the flowers (Mizukami and Ma, 1997). Study of the partial loss-of-function *ag-4* allele suggests that different *AG* functions in the flowers require varying amount of the gene product and that floral determinacy might requires a high level of *AG*. Alternatively different *AG* functions could be mediated by its interaction with region-specific factors (Sieburth et al., 1995).

The apparent role of *AG* in the regulation of cell proliferation during *Arabidopsis* development is further highlighted by the finding that *AG* is required to regulate cell division patterns by controlling both the rate and the orientation of cell divisions during floral development. Thus it has been demonstrated that the transformed third whorl and the organs of the internal flowers of *ag-1* mutant exhibited a contribution of the derivatives of the meristematic layers corrresponding to the new transformed identity and not to their position in the flower. In addition, clonal analysis of the severe *ag-1* mutant indicates inappropriate orientations of cell divisions in the L2 layer which contributes to the mesophyll of the organs. Furthermore, the *ag-1* mutant flowers also exhibit excessive contribution of the L3 layer ,which contribute to the receptacle and some of the mesophyll in the basal part of some floral organs, to several organs including those in the second floral whorls where *AG* is not normally expressed. These findings suggest a non-autonomous role for *AG* in regulating genes involving in the specification of cell division patterns in different cell layers within the floral meristem (Jenik and Irish, 2000).

Interestingly the indeterminate nature of *ag* flowers is somewhat opposite to the premature termination of floral meristems caused by loss-of-function of the *WUSCHEL* (*WUS*) homeobox gene. Genetic and molecular characterisation revealed that *WUS* activity is required for the maintenance of central stem cell population in both shoots and floral meristem, so that in *wus* mutants the main shoot terminates after producing only a few leaves and the rare flowers that develop do not have a full complement of organs and lack carpels and most stamens (Laux et al., 1996; Meyer et al., 1998). Genetic analysis has shown that the control of the size of the stem cell population in the shoot apical and floral meristems is achieved by the negative feedback loop formed between *WUS* and the previously mentioned *CLAVATA* (*CLV*) genes. Thus *WUS* activity induces the expression of the *CLV3* gene which in turn limits the *WUS* expression domain via the *CLV* complex dependent signalling

pathway (Brand et al., 2000; Schoof et al., 2000). The fact that *WUS* plays a key role in stem cell maintenance together with the opposite phenotype of *wus* and *ag* flowers and the overlapping expression of *WUS* and *AG* in the centre of the flowers suggest that the two gene might regulate each other and that their antagonistic relationship plays a central role in the termination of the floral meristem in the centre of wild-type flowers. Indeed data from recent studies by two research groups supports this. Thus it has been shown that *WUS* expression persists in *ag* mutant background and that overexpression of *WUS* in an *ag* background results in the failure of tissues in the inner two floral whorls to differentiate. Conversely a decrease in *WUS* activity results in a smaller *AG* expression domain and when *WUS* is expressed under the control of the *LFY*, *AP3* or *AG* promoter, supernumerary stamens and carpels are formed. Together these findings indicate that *WUS* is a direct activator of *AG* expression and after *AG* expression is established in the flower, *AG* in turn represses *WUS* activity, thereby generating a negative feedback loop required for the termination of the floral meristems (Lenhard et al., 2001; Lohmann et al., 2001). It has also been demonstrated that the role of *WUS* as a regulator of *AG* activity in the flower requires the activity of another *AG* activator, the *LFY* gene (Lenhard et al., 2001; Lohmann et al., 2001). This finding explains why *WUS* activity is not sufficient to activate *AG* expression in the vegetative meristems where *LFY* is not expressed and terminate their activity. Alternatively the *WUS* activity might be repressed by another gene function in the leaves. A candidate for such a repressor function is the *CLF* gene. It is possible that the *CLF* activity represses the *AG* function indirectly by inhibiting the function of its activator *WUS*. The *WUS*-repressing role of *CLF* might require other co-factors which are leaf-specific such that *CLF* is only able to repress *WUS* and hence *AG* in the leaves but not the flowers. However, an *in situ* analysis showed that *WUS* expression pattern did not expand in *clf* mutant seedlings suggesting that *WUS* is not a target of *CLF* (J Goodrich, pers. comm.). However this might be an overstatement since it is also possible that apart from *CLF* there are other genes that also act in parallel to *CLF* to repress *WUS* function in the leaves. One such factor might be the *EMF2* gene.

Results from an early study of the *emf2* mutant provides a strong piece of evidence for the role of *EMF2* in cell proliferation. Thus a mutation, designated *cyr1* was isolated in a screen for mutants which are resistant to elevated levels of cytokinins and subsequently found to be allelic to *emf2* mutations (Deikman and Ulrich, 1995; Z.R. Sung, pers. comm.). It is well documented that cytokinins are required in many aspects of plant growth and development

including cell division and organogenesis. The apparent defect in cytokinin response exhibited by *emf2* mutants would inevitably lead to detrimental consequences on many developmental processes including the control of cell division and hence the phenotype of the *emf2* plants. A further link of *EMF2* to cell proliferation control comes from the studies of two of its targets, the *AP3* and *PI* genes. Thus clonal analysis in *ap3-3* mutant plants indicates that, as in the case of *AG*, *AP3* activity is required in the regulation of cell proliferation patterns during floral development (Jenik and Irish, 2000). In addition genetic and molecular analyses of a recently identified target of *AP3* and *PI*, the *NAP* (*NO APICAL MERISTEM-LIKE, ACTIVATED BY AP3/PI*) gene, revealed that its activity is required in the transition between growth by cell division and cell expansion in stamens and carpels (Sablowski and Meyerowitz, 1998).

The results from this study and a number of recent studies on plant Pc-G group genes and their targets have provided several strong lines of evidence for the role of Pc-G genes in the regulation of cell proliferation in developing plants. It is noteworthy that the *CLF* and *EMF2* genes seem to have opposite effects on cell proliferation during vegetative development to that of the *FIS* genes in the endosperm. Thus the phenotypes of the *clf* and in particular the *emf2* mutants suggest that the wild-type genes function to promote cell proliferation and hence growth while the overproliferation defects seen in *fis* mutants suggests that the *FIS* genes play a role in repressing cell proliferation. Proper regulation of Pc-G target genes is required to ensure that many different aspects of growth including cell division and cell expansion are correctly regulated and co-ordinated. Pc-G mediated control, is thus one of the regulatory mechanisms, amongst others, that play a significant role in determining the overall morphology of the plants.

5.4 Distinct Pc-G complexes in plants

As previously mentioned, *Drosophila* Pc-G proteins act in large multimeric protein complexes. Two biochemically distinct *Drosophila* Pc-G complexes have recently been purified. Genetic and molecular studies suggest the ESC-E(Z) complex acts early to establish silencing while the PRC1 complex containing several Pc-G proteins including PC, PSC, PH and SCM is recruited later and functions to maintain silencing throughout development (Shao et al., 1999; Ng et al., 2000 ; Tie et al., 2001). In contrast to the *Drosophila* system, no plant Pc-G complexes have yet been purified. However, there are a number of strong lines of evidence to suggest that plant Pc-G proteins also form and function as parts of protein complexes. Firstly, although no homologues of Pc-G proteins in the PRC1 complex have been documented, the homologues of both the E(Z) and ESC proteins have been identified and their interaction appear to be conserved. Thus recent studies have shown that the *Arabidopsis* E(Z) homologues; CLF and MEA, are able to interact with the ESC homologue, the FIE protein (Luo et al., 2000; Spillane et al., 2000; Yadegari et al., 2000; J Goodrich, pers. comm.). In addition, characterisation of the third *Arabidopsis* E(Z) homologue, the CLK1 protein, carried out in this study showed that the E(Z)-ESC interaction is also conserved between the CLK1 and FIE proteins. More strikingly, in each of the E(Z) homologues, the FIE-interacting domain has also been mapped to the N-terminal part of the protein similar to that found in the E(Z) protein indicating that the molecular partnership between the WD repeat and SET domain proteins has been highly conserved in evolution. By analogy to the *Drosophila* paradigm it is quite likely therefore that the direct physical associations seen between the *Arabidopsis* WD repeat and SET domain proteins reflect the fact that they act in complexes and different complexes may have different functions.

Another finding that is consistent with plant Pc-G proteins form functional protein complexes comes from the studies of the *Arabidopsis* Pc-G gene expression patterns. Thus the *CLF*, *CLK1* and *EMF2* genes share overlapping expression domains with ubiquitous expression throughout the vegetative, inflorescence and floral meristems (Goodrich et al., 1997; Yoshida et al., 2001). In addition, the expression patterns of the three *FIS* genes also coincide with one another in the female gametophyte prior to fertilisation and later in the embryo proper and specific regions of the endosperm (Kinoshita et al., 1999; Vielle-Calzada et al., 1999; Luo et al., 2000; Spillane et al., 2000).

More support for the for the existence of plant Pc-G protein complexes is provided by the results of the phenotypic analysis of various *Arabidopsis* Pc-G mutants. Thus the *emf2* (Yang et al., 1995), embryonically rescued *fie* (Kinoshita et al., 2001), *clf emf2* and *clf clk1-1* plants all exhibit similar mutant phenotype. In each case the vegetative growth is drastically shortened and the plants are extremely small in size and only produce a few sessile leaves before entering the reproductive development. The shared common phenotype seen in these Pc-G mutants suggests that they are likely to have similar molecular basis. One possibility is that the wild-type genes function very closely together and perhaps form functional protein complexes that regulate their common target genes.

In *Drosophila*, the stable maintenance of the silent state homeotic gene expression by Pc-G genes involves a series of distinct temporal recruitment events. Thus the ESC protein is thought to be recruited first by sequence-specific DNA binding factors such as the HUNCHBACK (HB) zinc finger transcription factor bound to the PRE elements (Kehle et al., 1998; Zhang et al., 1998). The lack of physical interaction between ESC and HB suggests recruitment of ESC by HB is indirect and is mediated by the activity of other factors, one of which is the HB-interacting protein dMi2 (Kehle et al., 1998). Once recruited to its target sites, ESC forms a protein complex with its partner E(Z) and other proteins including the histone deacetylase RPD3 and the histone binding protein p55 (Tie et al., 2001). The intrinsic histone deacetylase activity of the ESC complex and that associated with the dMi-2 protein is thought to influence the local chromatin structure required for the establishment of Pc-G silencing mechanism. The change in local chromatin conformation is the prerequisite for the binding of a second Pc-G complex, the PRC1 complex whose activity is required for the stable maintenance of long-term Pc-G silencing (Poux et al., 2001; Tie et al., 2001). By analogy, it is possible that a similar process might take place in plants. Thus the binding of the zinc finger proteins, EMF2 and FIS2, might initiate the recruitment of the silencing initiation complex containing the ESC and E(Z) homologues; FIE, CLF and CLK1 which acts to modify local chromatin conformation. The fact that no homologues of the Pc-G components in the PRC1 complex has been found in plants suggests that different mechanisms involving a different set of proteins might be employed in plants to maintain the repressive chromatin state.

The finding that *Arabidopsis* Pc-G genes play important roles in diverse developmental processes suggests that several Pc-G complexes might be operating in plants. In fact the data

accumulated so far indicates that there are at least three distinct Pc-G complexes that function during distinct stages of *Arabidopsis* development. Thus during embryogenesis it is likely that the three *FIS* genes form a functional complex that represses expression of the target genes. The putative DNA binding FIS2 protein might bind the DNA first and then indirectly recruit the MEA-FIE protein complex via the interaction with another protein. A candidate for such mediator of recruitment of the MEA-FIE complex by FIS2 is the GYMNOS (GYM) protein. Cloning of *GYM* whose activity is required to repress embryonic identity and hence the transition from embryonic to postembryonic development revealed that it encodes an *Arabidopsis* homologue of the dMi-2 protein (Eshed et al., 1999; Ogas et al., 1999). In addition to its role in embryo development *FIE* is also required later on during the vegetative growth phase. FIE might in this case be recruited by the zinc finger EMF*2* protein to the homeotic targets such as *AG* and *AP1*. The FIE complex containing the SET domain CLF and CLK1 protein then acts to maintain the stable state of repression through later stages of development, perhaps by modifying local chromatin organisation using the activity of putative associated histone deacetylase proteins or by methylating the target genes using the potential methyltransferase activity of the CLF and CLK1 SET domains. The suggested early and crucial role of both the EMF2 and FIE protein in the initial phase of Pc-G mediated silencing mechanism is in accordance with the severe phenotypic defects exhibited by the *emf2* mutant and the embryo-rescued *fie* mutant.

Although there are many lines of evidence to support the hypothesis that plant Pc-G proteins form functional complexes, at present other possibilities still cannot be excluded. Thus as previously mentioned it is possible that *EMF2* and *CLF* act in a common pathway with the *EMF2* gene being a positive regulator of *CLF* function. Likewise, in the absence of the evidence that FIS2 physically interacts with either MEA or FIE, one can speculate that *FIS2* might exert its repression effect by regulating *MEA* and *FIE* expression.

In addition to the requirement of Pc-G function to repress cell proliferation during embryogenesis and homeotic gene expression during vegetative and reproductive growth, the revelation that the *VERNALIZATION 2* gene encodes a zinc finger Pc-G protein suggests that another distinct Pc-G complex exists in plants and plays an important role in the epigenetic regulation of the expression of the gene *FLC* during vernalization (Gendall et al., 2001).

Plants and animals are thought to have acquired their multicellularity independently of each other during evolution. The adoption of Pc-G-mediated repression mechanism to counteract the common problem of the maintenance of the established expression patterns of their targets through cell divisions during later development was, therefore, likely to have taken place independently between the two kingdoms. The ancestor sequence of the *ESC* and *E(Z)* genes must have existed prior to the divergence of plant and animal kingdom. Later on the polyploidisation events in plants resulted in most of the plant genes being found in multi-copies. Due to the selective pressure caused by the demand for more elaborate regulatory mechanisms to cope with the increase in the complexity of the organisms, many of the duplicated genes such as the three *Arabidopsis E(z)* homologues; *CLF*, *CLK1* and *MEA*, have acquired novel functional specialisation and are required in diverse regulatory networks. The potential of these E(Z) homologues to form distinct protein complexes with the ESC homologue, FIE, to generate multiplicity of plant Pc-G complexes also provide opportunities for the evolution of specialised roles in regulating gene expression. In the *Drosophila* and mammalian systems, the lack of totipotency and the potential lethality caused by slight perturbation to the cellular patterning process implies a more rigid control of chromosomal state which requires the evolution of more complex regulatory machanisms. This would explain why additional Pc-G functions have been acquired in the *Drosophila* and mammalian systems.

References

Aasland, R., Stewart, A.F., and Gibson, T. (1996). The SANT domain: a putative DNA-binding domain in the SWI-SNF and, ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. *Trends Biochem. Sci.* **21**, 87-88.

Adler, P.N, Martin, E.C., Charlton, J., and Jones, K., (1991). Phenotypic consequences and genetic interaction of a null mutation in the *Drosophila Posterior sex combs* gene. *Dev. Genet.* **12**, 349-361.

Akasaka, T., Kanno, M., Balling, R., Miezu, A., Taniguchi, M., and Koseki, H. (1996). A role for *mel-18*, a Polycomb-group related vertebrate gene, during the anteroposterior specification of the axial skeleton. *Development* **122**, 1513-1522.

Alvarez, J., Guli, C.L., Yu, X.-H., and Smyth, D.R. (1992). *TERMINAL FLOWER*: a gene affecting inflorescence development in *Arabidopsis thaliana*. *Plant J.* **2**, 103-116.

Aubert, D., Chen, L., Moon, Y.-H., Martin, D., Castle, L.A., Yang, C.-H., and Sung, Z.R. (2001). *EMF1*, A novel protein involved in the control of shoot architecture and flowering in *Arabidopsis*. *Plant Cell* **13**, 1865-1875.

Azpiroz-Leehan, R., and Feldmann, K.A. (1997). T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth. *Trends Genet.* **13**, 152-156.

Bai, S., and Sung, Z.R. (1995). The role of *EMF1* in regulating the vegetative and reproductive transition in *Arabidopsis thaliana* (*Brassicaceae*). *Am. J. Bot.* **82**, 1095-1103.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective regulation of methylated lysine 9 on histone H3 by the HP1 chromodomain. *Nature* **410**, 120-124.

Beuchle, D., Struhl, G., and Muller, J. (2001). Polycomb group proteins and heritable silencing of *Drosophila* Hox genes. *Development* **128**, 993-1004.

Biggins, M.D., and Tjian, R. (1988). Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts. *Cell* **53**, 699-711.

Birve, A., Sengupta, A.K., Beuchle, D., Larsson, J., Kennison, J.A., Rasmuson-Lestander, A., and Muller, J. (2001). *Su(z)12*, a novel *Drosophila* Polycomb group gene that is conserved in vertebrates and plants. *Development* **128**, 3371-3379.

Blanc, G., Barakat, A., Guyot, R., Cooke, R., and Delseny, I. (2000). Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* **12**, 1093-1101.

Bowman, J.L., Alvarez, J., Weigel., D., Meyerowitz, E.M., and Smyth, D.R. (1993). Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development* **119**, 721-743.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1989). Genes directing flower development in *Arabidopsis*. *Plant Cell* **1**, 37-52.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1991). Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112**, 1-20.

Bradley, D., Radcliffe, O., Vincent, C., Carpenter, R., and Coen, E.S. (1997). Inflorescence commitment and architecture in *Arabidopsis*. *Science* **275**, 80-83.

Brand, U., Fletcher, J., Hobe, M., Meyerowitz, E.M., and Simon, R. (2000). Dependence of stem cell fate in *Arabidopsis* on a regulatory loop regulates by *CLV3* activity. *Science* **289**, 617-619.

Breene, T.R., and Harte, P.J. (1993). *trithorax* regulates multiple homeotic genes in the bithorax and Antennapedia complexes and exerts different tissue-specific, parasegment-specific and promoter-specific effects on each. *Development* **117**, 119-134.

Breiling, A., Bonte, E., Ferrari, S., Beeker, P.B., and Paro, R. (1999). The *Drosophila* Polycomb protein interacts with nucleosomal core particles *in vitro* via its repression domain. *Mol. Cell. Biol.* **19**, 8451-8460.

Brock, H.W., and van Lohuizen, M. (2001). The Polycomb group- no longer an exclusive club? *Curr. Opin. Genet. Dev.* **11**, 175-181.

Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.-L., and Kassis, J.A. (1998). The *Drosophila* Polycomb Group gene *pleiohomeotic* encodes a DNA binding protein with homology to the transcription factor YY1. *Mol. Cell* **1**, 1057-1064.

Brown, R.C., Lemmon, B.E., Nguyen, H., and Olsen, O.-A. (1999). Development of endosperm in *Arabidopsis thaliana*. *Sex Plant Reprod.* **12**, 32-42.

Buchenau, P., Hodgson, J., Strutt, H., and Arndt-Jovin, D.J. (1998). The distribution of Polycomb-group proteins during cell division and development in *Drosophila* embryo: impact on models for silencing. *J. Cell Biol.* **141**, 469-481.

Busch, M.A., Bomblies, K., and Weigel, D. (1999). Activation of a floral homeotic gene in *Arabidopsis*. *Science* **285**, 585-587.

Busturia, A., and Bienz, M. (1993). Silencers in *abdomimal-B*, a homeotic *Drosophila* gene. *EMBO J.* **12**, 1415-1425.

Busturia, A., Wightman, C.D., and Sakonju, S. (1997). A silencer is required for maintenance of transcriptional repression throughout *Drosophila* development. *Development* **124**, 4343-4350.

Byzova, M.V., Franken, J., Aarts, M.G.M., de Almeida-Engler, J., Engler, G., Mariani, C., van Lookeren Campagne, M.M., and Angeneut, G.C. (1999). *Arabidopsis STERILE APETALA*, a multifunctional gene regulating inflorescence, flower and ovule development. *Genes Dev.* **13**, 1002-1014.

Campbell, R.B., Sinclair, D.A.R., Couling, M., and Brock, H.W. (1995). Genetic interactions and dosage effects of *Polycomb* group genes of *Drosophila*. *Mol. Gen. Genet.* **246**, 291-300.

Carrington, E.A., and Jones, R.S. (1996). The *Drosophila Enhancer of zeste* gene encodes a chromosomal protein: examination of wild-type and mutant protein distribution. *Development* **122**, 4073-4083.

Cavalli, G., and Paro, R. (1998). The *Drosophila Fab-7* chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* **93**, 505-518.

Chan, C.S., Rastelli, L., and Pirrotta, V. (1994). A *Polycomb* response element in the *Ubx* gene that determines an epigenetically inherited state of repression. *EMBO J.* **13**, 2553-2564.

Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S., and Peacock, W.J. (1997). Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4223-4228.

Chen, X., and Meyerowitz, E.M. (1999). *HUA1* and *HUA2* are two members of the floral homeotic *AGAMOUS* pathway. *Mol. Cell* **3**, 349-360.

Chen, L., Cheng, J.-C., Castle, L., and Sung, Z.R. (1997). *EMF* genes regulate *Arabidopsis* inflorescence development. *Plant Cell* **9**, 2011-2024.

Chiang, A., O'Connor, M.B., Paro, R., Simon, J., and Bender, W. (1995). Discrete Polycomb-binding sites in each parasegmental domain of the bithorax complex. *Development* **121**, 1681-1689.

Chou, M.-L., Haung, M.-D., and Yang, C.-H. (2001). *EMF* genes interact with late-flowering genes in regulating floral initiation genes during shoot development in *Arabidopsis thaliana*. *Plant Cell Physiol.* **42**, 499-507.

Chuck, G., Lincoln, C., and Hake, S. (1996). *KNAT1* induces lobed leaves with ectopic meristems when overexpressed in *Arabidopsis*. *Plant Cell* **8**, 1277-1289.

Clark, S.E., Williams, R.W., and Meyerowitz, E.M. (1997). The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* **89**, 575-585.

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

Cogoni, C., and Macino, G. (2000). Post-transcriptional gene silencing across kingdoms. *Curr. Opin. Genet. Dev.* **10**, 638-643.

Connor, J., and Liu, Z. (2000). *LEUNIG*, a putative transcription corepressor that regulates *AGAMOUS* repression during flower development. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12902-12907.

Core, N., Bel, S., Gaunt, S.J., Aurrand-Lions, M., Pearce, J., Fisher, A., and Djabali, M. (1997). Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. *Development* **124**, 721-729.

Deikman, J., and Ulrich, M. (1995). A novel cytokinin-resistent mutant of *Arabidopsis* with abbreviated shoot development. *Planta* **195**, 440-449.

Devereaux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387-395.

Deyholos, M.K., and Sieburth, L.E. (2000). Separable whorl-specific expression and negative regulation by enhancer elements within the *AGAMOUS* second intron. *Plant Cell* **12**, 1799-1810.

Dietzel, S., Niemann, H., Bruckner, B., Maurange, C., and Paro, R. (1999). The nuclear distribution of Polycomb during *Drosophila melanogaster* development shown with a GFP fusion protein. *Chromosoma*, **108**, 83-94.

Drews, G.N., Bowman, J.L., and Meyerowitz, E.M. (1991). Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product. *Cell* **65**, 991-1002.

Eshed, Y., Baum, S.F.,and Bowman, J.L. (1999). Distinct mechanisms promote polarity establishment in carpels of *Arabidopsis*. *Cell* **99**, 199-209.

Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H., and Karch, F. (1994). The *trithorax*-like gene encodes the *Drosophila* GAGA factor. *Nature* **371**, 806-808.

Ferrándiz, C., Gu, Q., Martienssen, R., and Yanofsky, M.F. (2000). Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. *Development* **127**, 725-734.

Ferris, S.D., and Whitt, G.S. (1979). Evolution of the differential regulation of duplicated genes after polyploidization. *J. Mol. Evol.* **12**, 267-317.

Finnegan, E.J., Peacock, W.J., and Dennis, E.S. (1996). Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8449-8454.

Firestein, R., Chi, X., Huie, P., and Cleary, M.L. (2000). Set domain-dependent regulation of transcriptional silencing and growth control by SUV39H1, a mammalian ortholog of *Drosophila* Su(var)3-9. *Mol. Cell. Biol.* **20**, 4900-4909.

Flanagan, C.A., and Ma, H. (1994). Spatially and temporally regulated expression of the MADS-box gene *AGL2* in wild-type and mutant *Arabidopsis* flowers. *Plant. Mol. Biol.* **26**, 581-595.

Fletcher, J.C., Brand, U., Running, M.P., Simon, R., and Meyerowitz, E.M. (1999). Signaling of cell fate decisions by *CLAVATA3* in *Arabidopsis* shoot meristems. *Science* **283**, 1911-1914.

Franke, A., DeCamillis, M., Zink, D., Cheng, N., Brock, H.W., and Paro, R. (1992). *Polycomb* and *polyhomeotic* are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*. *EMBO J.* **11**, 2941-2950.

Frei, D., Bopp, D., Burri, M., Baumgartner, S., Edstrom, J.E., and Noll., M. (1985). Isolation and structural analysis of the *extra sex combs* gene of *Drosophila*. *Cold Spring Harbor Symp. Quant. Biol.* **50**, 127-134.

Fritsch, C., Brown, J.C., Kassis, J.A., and Muller, J. (1999). The DNA-binding Polycomb group protein Pleiohomeotic mediates silencing of a *Drosophila* homeotic gene. *Development* **126**, 3905-3913.

Garvin, C., Holderman, R., and Strome, S. (1998). The phenotype of *mes-2*, *mes-3*, *mes-4*, and *mes-6*, maternal-effect genes required for survival of the germline in *Caenorhabditis elegans*, is sensitive to chromosome dosage. *Genetics* **148**, 167-185.

Gendall, A,R., Levy, Y.Y., Wilson, A., and Dean, C. (2001). The *VERNALIZATION 2* gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. *Cell* **107**, 525-535.

Gindhart, J.G., Jr., and Kaufman, T.C. (1995). Identification of Polycomb and trithorax group responsive elements in the regulatory region of the *Drosophila* homeotic gene *Sex combs reduced*. *Genetics* **139**, 797-814.

Goldberg, R.B., de Paiva, G., and Yadegari, R. (1994). Plant embryogenesis: Zygote to seed. *Science* **266**, 605-614.

Goodner, B., and Quatrano, R.S. (1993). *Fucus* embryogenesis: A model to study the establishment of polarity. *Plant Cell* **5**, 1471-1481.

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

Goto, K., and Meyerowitz, E.M (1994), Function and regulation of the *Arabidopsis* floral homoetic gene *PISTILLATA*. *Genes Dev* **8**, 1548-1560.

Goto, K., Kyozuka, J., and Bowman, J.L. (2001). Turning floral organs into leaves, leaves into floral organs. *Curr. Opin. Genet. Dev.* **11**, 449-456.

Gratti, M., and Baker, B.S. (1989). Genes controlling essential cell cycle functions in *Drosophila melanogaster*. *Genes Dev.* **3**, 438-453.

Grossniklaus, U., Vielle-Calzada, J.-P., Hoeppner, M.A., and Gagliano, W.B. (1998). Maternal control of embryogenesis by *MEDEA*, a polycomb-group gene in *Arabidopsis*. *Science* **280**, 446-450.

Grossniklaus, U., Spillane, C., Page, D.R., and Kohler, C. (2001). Genomic imprinting and seed development: endosperm formation with and without sex. *Curr. Opin. Genet. Dev.* **4**, 21-27.

Gu, Q., Ferrandiz, C., Yanofsky, M.F., and Martienssen, R. (1998). The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* **125**, 1509-1517.

Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1994). Regulation of the floral homeotic gene *APETALA1*. *Cell* **76**, 131-143.

Hagstrom, K., Muller, M., and Schedl, P. (1997). A *Polycomb* and GAGA dependent silencer adjoin the *Fab-7* boundary in the *Drosophila* bithorax complex. *Genetics* **146**, 1365-1380.

Hagstrom, K., and Schedl, P. (1997). Remembrance of things past: maintaining gene expression patterns with altered chromatin. *Curr. Opin. Genet. Dev.* **7**, 814-821.

Haig, D., and Westoby, M. (1989). Parent-specific gene expression and the triploid endosperm. *Am. Nat.* **134**, 147-155.

Haig, D., and Westoby, M. (1991). Genomic imprinting in endosperm: Its effect on seed development in crosses between species, and between different ploidies of the same species and its implications in the evolution of apomixis. *Philos. Trans. R. Soc. Lond.* **333**, 1-14.

Hamilton, A.J., and Baulcombe, D.C. (1999). A novel species of small antisense RNA in post-transcriptional gene silencing. *Science* **286**, 950-952.

Herskowitz, I. (1987). Functional inactivation of genes by dominant negative mutations. *Nature* **329**, 219-222.

Holdeman, K., Nehrt, S., and Strome, S. (1998). MES-2, a maternal protein essential for viability of the germline in *Caenorhabditis elegans*, is homologous to a *Drosophila* Polycomb group protein. *Development* **125**, 2457-2467.

Honma, T., and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409**, 525-529.

Horard, B., Tatout, C., Poux, S., and Pirrotta, V. (2000). Structure of a Polycomb response element and *in vitro* binding of Polycomb group complexes containing GAGA factor. *Mol. Cell. Biol.* **20**, 3187-3197.

Huala, E., and Sussex, I.M. (1992). *LEAFY* interacts with floral homeotic genes to regulate *Arabidopsis* floral development. *Plant Cell* **4**, 901-913.

Huang, M.-D., and Yang, C.-H. (1998). *EMF* genes interact with late-flowering genes to regulate *Arabidopsis* shoot development. *Plant Cell Physiol.* **39**, 382-393.

Hughes, M.K., and Hughes, A.C. (1993). Evolution of duplicated genes in a tetraploid animal *Xenopus laevis*. *Mol. Biol. Evol.* **10**, 1360-1369.

Ingham, P.W. (1988). The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature* **335**, 25-33.

Irish, V.F., and Sussex, I.M. (1990). Function of the *APETALA1* gene during *Arabidopsis* floral development. *Plant Cell* **2**, 741-753.

Ivanova, A.V., Bonaduce, M.J., Ivanov, S.V., and Klar, A.J.S. (1998). The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast. *Nat. Gen.* **19**, 192-195.

Jack, T. (2001). Relearning our ABCs: new twists on an old model. *Trends Plant Science* **6**, 310-316.

Jack, T., Brockman, L.L., and Meyerowitz, E.M. (1992). The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* **68**, 683-688.

Jack, T., Fox, G.L., and Meyerowitz, E.M. (1994). *Arabidopsis* homeotic gene *APETALA3* ectopic expression: Transcriptional and Posttranscriptional regulation determine floral organ identity. *Cell* **76**, 703-716.

Jacobs, J.J., Kieboom, K., Marino, S., DePinho, R.A., and van Lohuizen, M. (1999). The oncogene and Polycomb-group gene *bmi-1* regulates cell proliferation and senescence through the *ink4a* locus. *Nature* **397**, 164-168.

Jacobsen, S.E., and Meyerowitz, E.M. (1997). Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. *Science* **277**, 1100-1103.

Jenik, P.D., and Irish, V.F. (2000). Regulation of cell proliferation patterns by homeotic genes during *Arabidopsis* floral development. *Development* **127**, 1267-1276.

Jeong, S., Trotochaud, A.E. and Clark, S.E. (1999). The *Arabidopsis CLAVATA2* gene encodes a receptor-like protein required for the stability of the *CLAVATA1* receptor-like kinase. *Plant Cell* **11**, 1925-1934.

Jimenez, G., Paroush, Z., and Ish-Horowicz, D. (1997). Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. *Genes Dev.* **11**, 3072-3117.

Jofuku, K.D., den Boer, B.G.W., van Montagu, M., and Okamuro, J.K. (1994). Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. *Plant Cell* **6**, 1211-1225.

Jones, R.S., and Gelbart, W.M. (1990). Genetic analysis of the *Enhancer of zeste* locus and its role in gene regulation in *Drosophila melanogaster*. *Genetics* **126**, 185-199.

Jones, R.S., and Gelbart, W.M. (1993). The *Drosophila* Polycomb-Group gene *Enhancer of zeste* shares a domain of sequence similarity with *trithorax*. *Mol. Cell. Biol.* **13**, 6357-6366.

Jones, C.A., Ng, J., Peterson, A.J., Morgan, K., Simon, J., and Jones, R.S. (1998). The Drosophila esc and E(z) proteins are direct partners in Polycomb-Group-mediated repression. *Mol. Cell. Biol* **13**, 2825-2834.

Jurgen, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. *Nature* **316**, 153-155.

Karch, F., Galloni, M., Sipos, L., Gansz, J., Gyurkovics, H., and Schedl, P. (1994). *Mcp* and *Fab-7*: molecular analysis of putative boundaries of *cis*-regulatory domains in the bithorax complex of *Drosophila melanogaster*. *Nucleic Acids Res.* **22**, 3138-3146.

Kassis, J.A. (1994). Unusual properties of regulatory DNA from the *Drossophila engrailed* gene: three "pairing-sensitive" sites within a 1.6 kb region. *Genetics* **136**, 1025-1038.

Kassis, J.A., Desplan, C., Wright, D.K., and O'Farrell, P.H. (1989). Evolutionary conservation of homeodomain binding sites and other sequences upstream and within the major transcription unit of the *Drosophila* segmentation gene *engrailed*. *Mol. Cell. Biol.* **9**, 4304-4311.

Kaufman, T.C., Seeger, M.A., and Olsen, G. (1990). Molecular and genetic organization of the Antennapedia gene complex of *Drosophila melanogaster*. *Adv. Genet.* **27**, 309-362.

Kehle, J., Beuchle, D., Trenheit, S., Christen, B., Kennison, J.A., Bienz, M., and Muller, J. (1998). dMi-2, a Hunchback-interacting protein that function in Polycomb repression. *Science* **282**, 1897-1900.

Kelly, W.G., and Fire, A. (1998). Chromatin silencing and the maintenance of a functional germline. *Development* **125**, 2451-2456.

Kempin, S.A, Savidge, B., and Yanofsky, M.F. (1995). Molecular basis of the cauliflower phenotype in *Arabidopsis*. *Science* **267**, 522-525.

Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of *Polycomb* and *Antennapedia* mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8136-8140.

Kidner, C., Sundaresan, V., Hendriks, G., Weisbeek, P., and Scheres, B. (2000). Clonal analysis of the *Arabidopsis* root confirms that position, not lineage, determines cell fate. *Planta* **211**, 191-199.

Kim, G.-T., Tsukaya, H., and Uchimiya, H. (1998). The *CURLY LEAF* gene controls both division and elongation of cells during the expansion of the leaf blade in *Arabidopsis thaliana*. *Planta* **206**, 175-183.

Kinoshita, T., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2001). Polycomb repression of flowering during early plant development. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14156-14161.

Kinoshita, T., Yadegari, R., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Imprinting of the *MEDEA* Polycomb gene in the *Arabidopsis* endosperm. *Plant Cell* **11**, 1945-1952.

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J., Goldberg, R.B., and Fischer, R.L. (1999). Control of fertilization-independent endosperm development by the *MEDEA* polycomb-gene in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4186-4191.

Kohalmi, S.E., Leader, L.J.V., Samach, A., Nowak, J., Haughn, G.W., and Crosby, W.L. (1998). Identification and characterisation of protein interactions using the yeast 2-hybrid system. *Plant Mol. Biol. Man.* **M1**, 1-30.

Komachi, K., Redd, M.J., and Johnson, A.D. (1994). The WD repeats of tup1 interact with the homeodomain protein α2. *Genes Dev.* **8**, 2857-2867.

Koornneef, M., de Bruine, J.H., and Goettsch, P. (1980). A provisional map of chromosome 4 of *Arabidopsis*. *Arabidopsis Inf. Serv.* **17**, 11-18.

Koornneef, M., Dellaert, L.M.M., and van der Veen, J.H. (1982). EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana (L.) Heynh. Mut. Res.* **93**, 109-123.

Korf, I., Fan, Y., and Strome, S. (1998). The Polycomb group in *Caenorhabditis elegans* and maternal control of germline development. *Development* **125**, 2469-2478.

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

Krizek, B.A., Prost, V., and Macias, A. (2000). *AINTEGUMENTA* promotes petal identity and acts as a negative regulator of *AGAMOUS*. *Plant Cell* **12**, 1357-1366.

Krysan, P.J., Young, J.C., and Sussman, M.R. (1999). T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* **11**, 2283-2290.

Kunkel,, T.A., Roberts, J.D., and Zakour, R.A. (1987). Rapid and specific site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**, 367-379.

Kyba, M., and Brock, H.W. (1998). The *Drosophila* Polycomb group protein Psc contacts ph and Pc through specific conserved domains. *Mol. Cell. Biol.* **18**, 2712-2720.

Lachner, M., O'Carroll, D.O., Rea, S., Mecutler, K., and Jenuwein, T. (2001). Methylation of histine H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116-120.

Laible, G., Wolf, A., Dorn, R., Reuter, G., Nislow, C., Lebersorger, A., Popkin, D., Pillus, L., and Jenuwein, T. (1997). Mammalian homologues of the Polycomb-group gene *Enhancer of zeste* mediate gene silencing in *Drosophila* heterochromatin and at *S.cerevisiae* telomeres. *EMBO J.* **16**, 3219-3232.

LaJeunesse, D., and Shearn, A. (1996). *E(z)*: a polycomb group gene or a trithorax group gene? *Development* **122**, 2189-2197.

Landecker, H.L., Sinclair, D.A.R., and Brock, H.W. (1994). Screen for enhancers of *Polycomb* and *Polycomb-like* in *Drosophila melanogaster*. *Dev. Genet.* **15**, 425-434.

Lang, A. (1965). The physiology of flower initiation. In *Handbuch d. Pflanzenphysiologie (Encyclopedia of Plant Physiology)* (Vol. XV/1) (Ruhland, W., ed.), pp. 1379-1536, Springer-Verlag.

Lapowski, E.E., Martin, P., Garvin, C., and Strome, S. (1991). Identification of grandchildless loci whose products are required for normal germ-line development in the nematode *Caenorhabditis elegans*. *Genetics* **129**, 1061-1072.

Laux, T., Mayer, K.F.X., Berger, J., and Jurgens, G. (1996). The *WUSCHEL* gene is required for shoot and floral meristem integrity in *Arabidopsis*. *Development* **122**, 87-96.

Lawrence, P.A. (1992). *The Making of a fly*. Oxford: Blackwell Scientific Publications.

Lenhard, M., Bohnert, A., Jurgens, G., and Laux, T. (2001). Termination of stem cell maintenance in *Arabidopsis* floral meristem by interactions between *WUSCHEL* and *AGAMOUS*. *Cell* **105**, 805-814.

Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565-570.

Li, J., Jia, D. and Chen, X. (2001). *HUA1*, a regulator of stamen and carpel identities in *Arabidopsis*, codes for a nuclear RNA binding protein. *Plant Cell* **13**, 2269-2281.

Liljegren, S., Ditta, .S., Eshed, Y., Savidge, B., Bowman, J.L, and Yanofsky, M.F. (2000). *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* **404**, 766-770.

Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K., and Hake, S. (1994). A *knotted1-like* homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *Plant Cell* **6**, 1859-1876.

Liu, Z., and Meyerowitz, E.M. (1995). *LEUNIG* regulates *AGAMOUS* expression in *Arabidopsis* flowers. *Development* **121**, 975-991.

Lloyd, A.M., Schena, M., Walbot, V., and Davis, R.W. (1994). Epidermal cell fate determination in *Arabidopsis*: patterns defined by a steroid-inducible regulator. *Science* **266**, 436-439.

Lohmann, J., Hong, R., Hobe, M., Busch, M., Parcy, F., Simon, R., and Weigel, D. (2001). A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. *Cell* **105**, 793-803.

Long, D., Martin, M., Sundberg, E., Swinburne, J., Puangsomlee, P., and Coupland, G. (1993). The maize transposable element system *Ac/Ds* as a mutagen in *Arabidopsis*: Identification of an *albino* mutation induced by *Ds* insertion. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10370-10374.

Lonie, A., D'Andrea, R., Paro, R., and Saint, R. (1994). Molecular characterization of the *Polycomblike* gene of *Drosophila melanogaster*, a *trans*-acting negative regulator of homeotic gene expression. *Development* **120**, 2629-2636.

Lopes, M.A., and Larkins, B.A. (1993). Endosperm origin, development, and function. *Plant Cell* **5**, 1383-1399.

Luo, M., Bilodeau, P., Dennis, E.S., Peacock, W.J., and Chaudhury, A. (2000). Expression and parent-of-origin effects for *FIS2*, *MEA*, and *FIE* in the endosperm and embryo of developing *Arabidopsis* seeds. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10637-10642.

Luo, M., Bilodeau, P., Koltunow, A., Dennis, E.S., and Peacock, W.J. (1999). Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 296-301.

Lucas, W.S., Bouche-Pillon, S., Jackson, D.P., Nguyen, L., Baker, L., Ding, B., and Hake, S. (1995). Selective trafficking of KNOTTED 1 homeodomain protein and its mRNA through plasmodesmata. *Science* **270**, 1980-1983.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.G. (1997). Crystal structure of the nucleosome core particle at 2.8 A° resolution. *Nature* **389**, 251-260.

Mandel, M.A., and Yanofsky, M.F. (1995). A gene triggering flower formation in *Arabidopsis*. *Nature* **377**, 522-524.

Mandel, M.A., and Yanofsky, M.F. (1998). The *Arabidopsis AGL9* MADS-box gene is expressed in young flower primordia. *Sex Plant Reprod.* **11**, 22-28.

Mandel, M.A., Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* **360**, 273-277.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Martin, E.C., and Adler, P.N. (1993). The *Polycomb* group gene *Posterior sex combs* encodes a chromosomal protein. *Development* **117**, 641-655.

Martinez-Balbas, M.A., Tsukiyama, T., Gdula, D., and Wu, C. (1998). Drosophila NURF-55, a WD repeat protein involved in histone metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 132-137.

McKeon, J. and Brock, H.W. (1991). Interactions of the *Polycomb* group of genes with homeotic loci of *Drosophila*. *Roux Arch. Dev. Biol.* **199**, 387-396.

McKeon, J., Slade, E., Sinclair, D.A.R., Cheng, N., Couling, M., and Brock, H.W. (1994). Mutations in some *Polycomb* group genes of *Drosophila* interfere with regulation of segmentation genes. *Mol. Gen. Genet.* **244**, 474-483.

Meins, Jr., F. (2000). RNA degradation and models for post-transcriptional gene silencing. *Plant Mol. Biol.* **43**, 261-273.

Meyer, K.F.X., Shoof, H., Haecker, A., Lenhard, M., Jurgen, G., and Laux, T. (1998). Role of *WUSCHEL* in regulating stem cell fate in *Arabidopsis* shoot meristem. *Cell* **95**, 805-815.

Meyerowitz, E.M. (1997). Genetic control of cell division patterns in developing plants. *Cell* **88**, 299-308.

Mihaly, J., Mishra, R.K., and Karch, F. (1998). A conserved sequence motif in Polycomb-response elements. *Mol. Cell.* **1**, 1065-1066.

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

Mizukami, Y., and Ma, H. (1997). Determination of *Arabidopsis* floral meristem identity by *AGAMOUS*. *Plant Cell* **9**, 393-408.

Moazed, D., and O'Farrell, P.H. (1992). Maintenance of the *engrailed* expression pattern by Polycomb group genes in *Drosophila*. *Development* **116**, 805-810.

Muller, J. (1995). Transcriptional silencing by the Polycomb protein in *Drosophila* embryos. *EMBO J.* **14**, 1209-1220.

Muller, J., and Bienz, M. (1991). Long range repression conferring boundaries of *Ultrabithorax* expression in the *Drosophila* embryo. *EMBO J.* **19**, 3147-3155.

Murzina, N., Verreault, A., Laue, E., and Stillman, B. (1999). Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. *Mol. Cell* **4**, 529-540.

Neer, E.J., Schmidt, C.J., Nambudripad, R., and Smith, T.F. (1994). The ancient regulatory-protein family of WD repeat proteins. *Nature* **371**, 297-300.

Ng, J., Hart, C.N., Morgan, K., and Simon, J.A. (2000). A *Drosophila* ESC-E(Z) protein complex is distinct from other polycomb group complexes and contains covalently modified ESC. *Mol. Cell. Biol.* **20**, 3069-3078.

Ng, J., Li, R., Morgan, K., and Simon, J. (1997). Evolutionary conservation and predicted structure of the *Drosophila* extra sex combs repression protein. *Mol. Cell. Biol.* **17**, 6663-6672.

Nielsen, A.L., Omad-Abdelghani, M., Ortiz, J.A., Rembentsika, E., Chambon, P., and Losson, R. (2001). Heterochromatin formation in mammamlian cells: Interaction between histones and HP1 proteins. *Mol. Cell* **7**, 729-739.

Nislow, C., Ray, E., and Pillus, L. (1997). SET1, a yeast member of the Trithorax family, functions in transcriptional silencing and diverse cellular processes. *Mol. Biol. Cell.* **8**, 2421-2431.

Nusse, R., and Varmus, H.E. (1992). *Wnt* genes. *Cell* **69**, 1073-1087.

Ogas, J., Kaufmann, S., Henderson, J., and Somerville, C. (1999). PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13839-13844.

Ohad, N., Margossian, L., Hsu, Y.-C., Willaims,C., Repetti, P., and Fischer, R.L. (1996). A mutation that allows endosperm development without fertilization. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5319-5324.

Ohad, N., Yadegari, R., Margossian,L., Hannon, M., Michaeli, D., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Mutations in *FIE*, a WD polycomb group gene, allow endosperm development without fertilization. *Plant Cell* **11**, 407-415.

Ohno, S. (1970). Evolution by gene duplication. (New York: Springer-Verlag).

Ohno, S., Wolf, U., and Atkin, N.B. (1968). Evolution from fish to mammals by gene duplication. *Hereditas* **59**, 169-187.

Parcy, F., Nilsson, D., Busch, M.A., Lee, I., and Weigel, D. (1998). A genetic framework for floral patterning. *Nature* **395**, 561-566.

Paro, R. (1990). Imprinting a determined state into the chromatin of *Drosophila. Trends Genet.* **6**, 416-421.

Paro, R., and Hogness, D.S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila. Proc. Natl. Acad. Sci. U.S.A.* **88**, 263-267.

Paro, R., and Zink, B. (1993). The *Polycomb* gene is differentially regulates during oogenesis and embryogenesis of *Drosophila melanogaster. Mech. Dev.* **40**, 37-46.

Pattatucci, A.M., and Kaugman, T.C. (1991). The homeotic gene *Sex combs reduced* of *Drosophila melanogaster* is differentially regulated in the embryonic and imaginal stages of development. *Genetics* **129**, 443-461.

Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E., and Yanofsky, M.F. (2000). B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* **405**, 200-203.

Pelaz, S., Gustafson-Brown, C., Kohalmi, S.E., Crosby, W.C., and Yanofsky, M.F. (2001). *APETALA1* and *SEPALLATA3* interact to promote flower development. *Plant J.* **26**, 385-394.

Pelaz, S., Tapia-Lopez, R., Alvarez-Buylla, E.R., and Yanofsky, M.F. (2001). Conversion of leaves into petals in *Arabidopsis. Curr. Biol.* **11**, 182-184.

Pelegri, F., and Lehmann, R. (1994). A role of *Polycomb* group genes in the regulation of gap gene expression in *Drosophila. Genetics* **136**, 1341-1353.

Peters, A.H.F.M., O'Carroll, D., Scherthan, H., Mechter, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., Opravil, S., Doyle, M., Sibilia, M., and Jenuwein, T. (2001). Loss of the *Suv39h* histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323-337.

Peterson, A.J., Kyba, M., Boinemann, D., Morgan, K., Brock, H.W., and Simon, J. (1997). A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. *Mol. Cell. Biol.* **17**, 6683-6692.

Picard, D. (1993). Steroid-binding domains for regulating the functions of heterologous proteins in *cis*. *Trends Cell Biol.* **3**, 278-280.

Pickett, F.B., and Meeks-Wagner, D.R. (1995). Seeing double: Appreciating genetic redundancy. *Plant Cell* **7**, 1347-1356.

Pidkowich, M.S., Klenz, J.E., and Haughn, G.W. (1999). The making of a flower: Control of floral meristem identity in *Arabidopsis*. *Trends Plant Sci.* **4**, 64-70.

Pirrotta, V. (1997a). PcG complexes and chromatin silencing. *Curr. Opin. Genet. Dev.* **7**, 249-258.

Pirrotta, V. (1997b). Chromatin-silencing mechanisms in *Drosophila* maintain patterns of gene expression. *Trends Genet.* **13**, 314-318.

Pirrotta, V. (1998). Polycombing the genome: PcG, trxG and chromatin silencing. *Cell* **93**, 333-336.

Pirrotta, V., and Rastelli, L. (1994). *white* gene expression, repressive chromatin domains and homeotic gene regulation. *Bioessays* **16**, 549-556.

Phillips, M.D. and Shearn, A. (1990). Mutations in *polycombeotic*, a *Drosophila* polycomb-group gene, cause a wide range of maternal and zygotic phenotypes. *Genetics* **125**, 91-101.

Platero, J.S., Hartnett, T., and Eissenberg, J.C. (1995). Functional analysis of the chromodomain of HP1. *EMBO J.* **14**, 3977-3986.

Platero, J.S., Sharp, E.J., Adler, P.N., and Eissenberg, J.C. (1996). *In vivo* assay for protein-protein interactions using *Drosophila* chromosomes. *Chromosoma* **104**, 393-404.

Poux, S., McCabe, D., and Pirrotta, V. (2001). Recruitment of components of Polycomb group chromatin complexes in *Drosophila*. *Development* **128**, 75-85.

Poux, S., Melfi, R., and Pirrotta, V. (2001). Establishment of Polycomb silencing requires a transient interaction between PC and ESC. *Genes Dev.* **15**, 2509-2514.

Purugganan, M.D., and Suddith, J.I. (1998). Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene: Non-neutral evolution and naturally occurring variation in floral homeotic function. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8130-8134.

Radcliffe, O.J., Amaya, I., Vincent, C.A., Rothstein, S., Carpenter, R., Coen, E.S., and Bradley, D.J. (1998). A common mechanism controls the life cycle and architecture of plants. *Development* **125**, 1609-1615.

Radcliffe, O.J., Bradley, D.J., and Coen, E.S. (1999). Separation of shoot and floral identity in *Arabidopsis*. *Development* **126**, 1109-1120.

Rastelli, L., Chan, C.S., and Pirrotta, V. (1993). Related chromosome binding sites for *zeste, suppressors of zeste* and Polycomb group proteins in *Drosophila* and their dependence on *Enhancer of zeste* function. *EMBO J.* **12**, 1513-1522.

Rea, S., Elsenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-599.

Ridley, M. (2000). *Genome: The autobiography of a species in 23 chapters*. Fourth Estate Publications.

Riechmann, J.L., and Meyerowitz, E.M. (1997). MADS domain proteins in plant development. *Biol. Chem.* **378**, 1079-1101.

Ruht, M.-L., Braun, A., Papoulas, O., Tamkin, J.W., Randsholt, N., and Meister, M. (2001). The *domino* gene of *Drosophila* encodes novel members of the SWI2/SNF2 family of DNA-dependent ATPases, which contribute to the silencing of homeotic genes. *Development* **128**, 1429-1441.

Sablowski, R.W.M., and Meyerowitz, E.M. (1998). A homolog of *NO APICAL MERISTEMS* is an immediate target of the floral homeotic genes *APETALA3/PISTILLATA*. *Cell* **92**, 93-103.

Sathe, S.S., and Harte, P.J. (1995). The *Drosophila* Extra sex combs protein contains WD repeats essential for its function as a repressor of homeotic genes. *Mech. Dev.* **53**, 77-87.

Saurin, A.J., Shao, Z., Erdjument-Bromage, H., Tempst, P., and Kingston, R.E. (2001). A *Drosophila* Polycomb group complex includes zeste and dTAFII proteins. *Nature* **412**, 655-660.

Savidge, B., Roundsley, S.D., and Yanofsky, M.F. (1995). Temporal relationship between the transcription of two *Arabidopsis* MADS-box genes and the floral organ identity genes. *Plant Cell* **7**, 721-733.

Schoof, H., Lenhard, M., Haecker, A., Meyer, K., Jurgen, G., and Laux, T. (2000). The stem cell population of *Arabidopsis* shoot meristems is maintained by a regulatory loop between the *CLAVATA* and *WUSCHEL* genes. *Cell* **100**, 635-644.

Schultz, E.A., and Haughn, G.W. (1991). *LEAFY*, a homeotic gene that regulates inflorescence development in *Arabidopsis*. *Plant Cell* **3**, 771-781.

Schultz, E.A., and Haughn, G.W. (1993). Genetic analysis of the floral initiation (FLIP) in *Arabidopsis*. *Development* **110**, 745-765.

Scmiarti, E., Lleno, Y., Tsukaya, H., Iwakawa, H., Machid, C., and Machida, Y. (2001). The *ASYMMETRIC LEAVES 2* gene of *Arabidopsis thaliana* regulates formation of a symmetric lamina, establishment of venation and repression of meristem-related homeobox genes in leaves. *Development* **128**, 1771-1783.

Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G. (1998). Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* **125**, 3329-3341.

Seeber, J.S., Marchio, A., Sitterin, D., Transy, C., and Dejean, A. (1998). Interaction of SP100 within HP1 proteins: a link between the promyelocytic leukemia-associated nuclear bodies and the chromatin compartment. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7316-7321.

Sewalt, R.G., van der Vlag, J., Gunster, M.J., Hamer, K.M., der Blaauwen, J.L., Satijn, D.P.E., Hendrix, T., van Driel, R., and Otte, A.P. (1998). Characterization of interactions between the mammalian Polycomb-Group proteins Enx1/EZH2 and EED suggests the existence of different mammalian Polycomb-Group protein complexes. *Mol. Cell. Biol.* **18**, 3586-3595.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.-T., Bender, W., and Kingston, R.E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**, 37-46.

Shearn, A., Hersperger, G., and Hersperger, E. (1978). Genetic analysis of two allelic temperature-sensitive mutants of *Drosophila melanogaster* both of which are zygotic and maternal-effect lethals. *Genetics* **89**, 341-353.

Shimell, M.J., Simon, J., Bender, W., and O'Connor, M.B. (1994). Enhancer point mutation results in a homeotic transformation in *Drosophila*. *Science* **264**, 968-971.

Sieburth, L.E., and Meyerowitz, E.M. (1997). Molecular dissection of the *AGAMOUS* control region shows that *cis* elements for spatial regulation are located intragenically. *Plant Cell* **9**, 355-365.

Sieburth, L.E., Running, M.P., and Meyerowitz, E.M. (1995). Genetic separation of third and fourth whorl functions of *AGAMOUS*. *Plant Cell* **7**, 1249-1258.

Sigrist, C.J.A., and Pirrotta, V. (1997). Chromatin insulator elements block the silencing of a target gene by the *Drosophila* Polycomb response element (PRE) but allows trans interactions between PREs on different chromosomes. *Genetics* **147**, 209-221.

Simon, J. (1995). Locking in stable states of gene expression: transcriptional control during *Drosophila* development. *Curr. Opin. Cell. Biol.* **7**, 376-385.

Simon, J., Bornemann, D., Lunde, K., and Schwartz, C. (1995). The *extra sex combs* product contains WD40 repeats and its time of action implies a role distinct from other Polycomb group gene products. *Mech. Dev.* **53**, 197-208.

Simon, J., Chiang, A., and Bender, W. (1992). Ten different *Polycomb* group genes are required for spatial control of the *abdA* and *AbdB* homeotic products. *Development* **114**, 493-505.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the *Drosophila* bithorax complex that mediate repression by Polycomb group products. *Dev. Biol.* **158**, 131-144.

Sinclair, D.A.R., Clegg, N.J., Antonchuk, J., Milne, T.A., Stankunas, K., Ruse, C., Grigliatti, T.A., Kassis, J.A., and Brock, H.W. (1997). *Enhancer of Polycomb* is a suppressor of position-effect variegation in *Drosophila melanogaster*. *Genetics* **148**, 211-220.

Soppe, W.J.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J.M. (2000). The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6**, 791-802.

Sorensen, M.B., Chaudhury, A.M., Robert, H., Bancharel, E., and Berger, F. (2000). Polycomb group genes control pattern formation in plant seed. *Curr. Biol.* **11**, 277-281.

Soto, M.C., Chou, T.B., and Bender, W. (1995). Comparison of germline mosaics of genes in the Polycomb group of *Drosophila melanogaster*. *Genetics* **140**, 231-243.

Spillane, C., MacDougall, C., Stock, C., Kohler, C., Vielle-Calzada, J.-P., Nunes, S.M., Grossniklaus, U., and Goodrich, J., (2000). Interaction of the *Arabidopsis* Polycomb group proteins FIE and MEA mediates their common phenotypes. *Curr. Biol.* **10**, 1535-1538.

Stewart, F.C. (1958). Growth and development of cultivated cells III. Interpretations of the growth from free cell to carrot plant. *Am. J. Bot.* **45**, 709-713.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* **403**, 41-45.

Struhl, G., and Akam, M.E. (1985). Altered distributions of *Ultrabithorax* transcripts in *extra sex combs* mutant embryos in *Drosophila*. *EMBO J.* **4**, 3259-3264.

Struhl, G., and Brower, D. (1982). Early role of the esc^{+} gene product in the determination of segments in *Drosophila*. *Cell* **31**, 285-292.

Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* **12**, 599-606.

Strutt, H., Cavalli, G., and Paro, R. (1997). Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.* **16**, 3621-3633.

Strutt, H., and Paro, R. (1997). The Polycomb group protein complex of *Drosophila melanogaster* has different compositions at different target genes. *Mol. Cell. Biol.* **12**, 6777-6783.

Sung, Z.R., Belachew, A., Shunong, B., and Bertrand-Garcia, R. (1992). *EMF*, an *Arabidopsis* gene required for vegetative shoot development. *Science* **258**, 1645-1647.

Tautz, D. (1992). Redundancies, development and the flow of information. *Bioessays* **14**, 263-266.

Tie, F., Furuyama, T., and Harte, P.J. (1998). The *Drosophila* Polycomb Group proteins ESC and E(Z) bind directly to each other and co-localize at multiple chromosomal sites. *Development* **125**, 33483-3496.

Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E., and Harte, P.J. (2001). The *Drosophila* Polycomb group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development* **128**, 275-286.

Theissen, G., Becker, A., Di Rosa, A., Kano, A., Kim, J.T., Munster, T., Winter, K.-U., and Saedler, H. (2000). A short history of MADS-box genes in plants. *Plant Mol. Biol.* **42**, 115-149.

Thomas, J.H. (1993). Thinking about genetic redundancy. *Trends Genet.* **9**, 395-399.

Trotochaud, A.E., Hao, T., Wu, G., Yang, Z., and Clark, S.E. (1999). The *CLAVATA1* receptor-like kinase requires *CLAVATA3* for its assembly into a signaling complex that includes KAPP and a rho-related protein. *Plant Cell* **11**, 393-405.

Trotochaud, A.E., Jeong, S., and Clark, S.E. (2000). *CLAVATA3*, a multimeric ligand for the *CLAVATA1* receptor-kinase. *Science* **289**, 613-617.

Tschiersch, B., Hotmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the *Drosophila* position-effect variegation suppressor gene *Su(var)3-9* combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* **13**, 3822-3831.

Tsukiyama, T., Becker, P.B., and Wu, C. (1994). ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367**, 525-532.

Tyler, J.K., Bulger, M., Kanakaka, R.T., Kobayashi, R., and Kadonaga, J.T. (1996). The p55 subunit of *Drosophila* chromatin assembly factor 1 is homologous to a histone deacetylase-associated protein. *Mol. Cell. Biol.* **16**, 6149-6159.

Tyler, J.K., and Kadonaga, J.T. (1999). The 'dark side' of chromatin remodeling: repressive effects on transcription. *Cell* **99**, 443-446.

van den Berg, L., Willemsen, V., Hage, W., Weisbeek, P., and Scheres, B. (1995). Determination of cell fate in the *Arabidopsis* root meristem by directional signalling. *Nature* **378**, 62-65.

van der Lugt, N.M., Domen, J., Linders, K., van Roon, H., Robanus-Maandag, E., Riele, H., Valk, M., Deschamps, J., Sofroniew, M., and van Lohuizen, M. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogenes. *Genes Dev* **8**, 757-769.

van der Vlag, J., and Otte, A.P. (1999). Transcriptional repression mediated by the human Polycomb-Group protein EED involves histone deacetylation. *Nat. Genet.* **23**, 474-478.

van Lohuizen, M. (1999). The trithorax-group and Polycomb-group chromatin modifiers: implications for disease. *Curr. Opin. Genet. Dev.* **9**, 355-361.

Vielle-Calzada, J.-P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M.A., and Grossniklaus, U. (1999). Maintenance of genomic imprinting at the *Arabidopsis medea* locus requires zygotic DDM1 activity. *Genes Dev.* **13**, 2971-2982.

Vincent, C.A., Carpenter, R., and Coen, E.S. (1995). Cell lineage patterns and homeotic gene activity during flower development of *Antirrhinum*. *Curr. Biol.* **5**, 1449-1458.

Vincent, J.-P., and Lawrence, P.A. (1994). *Drosophila wingless* sustains *engrailed* expression only in adjoining cells: evidence from mosaic embryos. *Cell* **77**, 909-915.

Voncken, J.-W., Schweizer, D., Aagard, L., Sattler, L., Jantsch, M.F., and von Lohuizen, M. (1999). Chromatin-association of the Polycomb group protein BMI1 is cell cycle-regulated and correlates with its phosphorylation status. *J. Cell Sci.* **12**, 4627-4639.

Wade, P.A., Jones, P.L., Vermaak, D., and Wolffe, A.P. (1998). A multiple subunit Mi-2 histone deacetylase from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPases. *Curr. Biol.* **8**, 843-846.

Walbot, V. (2000). A green chapter in the book of life. *Nature* **408**, 794-795.

Wall, M.A., Coleman, D.E., Lee, E., Iniguez-Lluni, J.A., Posner, B.A., Gulman, A.G., and Sprang, S.K. (1995). The structure of the G protein heterodimer $G_{i\alpha1}\beta_{1\gamma2}$. *Cell* **83**, 1047-1058.

Weigel, D., and Meyerowitz, E.M. (1993). Activation of floral homeotic genes in *Arabidopsis*. *Science* **261**, 1723-1726.

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

Weigel, D., and Nilsson, O. (1995). A developmental switch sufficient for flower initiation in diverse plants. *Nature* **377**, 495-500.

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F., and Meyerowitz, E.M. (1992). *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* **69**, 843-859.

Weintraub, H. (1993). The MyoD family and myogenesis- redundancy, networks and thresholds. *Cell* **75**, 1241-1244.

Williams, R.W., Wilson, J.M., and Meyerowitz, E.M. (1997). A possible role for kinase-associated protein phosphatase in the *Arabidopsis CLAVATA1* signaling pathway. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10467-10472.

Wolpert, L., Beddington, R., Brockes, J., Jessell, T., Lawrence, P., and Meyerowitz, E. (1998). *Principles of development.* (Wolpert, L. ed.) Oxford: Oxford University Press.

Xu, L., Fong, Y., and Strome, S. (2001). The *Caenorhabditis elegans* maternal-effect sterile proteins, MES-2, MES-3, and MES-6 are associated in a complex in embryos. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5061-5066.

Yadegari, R., Kinoshita, T., Lotan, O., Lohen, G., Katz, A., Choi, Y., Katz, A., Nakashima, K., Harada, J.J., Goldberg, R.B., Fischer, R.L., and Ohad, N. (2000). Mutations in the *FIE* and *MEA* genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. *Plant Cell* **12**, 2367-2381.

Yang, C.-H., Chen, L.-J., and Sung, Z.R. (1995). Genetic regulation of shoot development in *Arabidopsis*: Role of the *EMF* genes. *Dev. Biol.* **169**, 421-435.

Yang, W.-M., Inouye, C., Zeng, Y., Bearss, D., and Seto, E. (1996). Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulators RPD3. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12845-12850.

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., and Meyerowitz, E.M. (1990). The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* **346**, 35-39.

Yoshida, N., Yanai, Y., Chen, L., Kato, Y., Hiratsuka, J., Miwa, T., Sung, Z.R., and Takahashi, S. (2001). *EMBRYONIC FLOWER2*, a Novel Polycomb group protein homolog, mediates shoot development and flowering in *Arabidopsis*. *Plant cell* **13**, 2471-2481.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucloetide intervals. *Cell* **101**, 25-33.

Zhang, C.-C., and Bienz, M. (1992). Segmental determination in *Drosophila* conferred by *hunchback*, a direct repressor of the homeotic gene *Ultrabithorax*. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7511-7515.

Zhang, Y., LeRoy, G., Seelig, H.P., Lane, W.S., and Reinberg, D. (1998). The dermatomyositis-specific auto antigen Mi-2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95**, 279-289.

Zinc, B., and Paro, R. (1989). *In vivo* binding pattern of a trans-regulator of homeotic genes in *Drosophila melanogaster*. *Nature* **337**, 468-473.

Zink, D., and Paro, R. (1995). *Drosophila* Polycomb-group regulated chromatin inhibits the accessibility of a *trans*-activator to its target DNA. *EMBO J.* **14**, 5660-5671.