Specificities of Polycomb-Group proteins controlling flowering in *Arabidopsis*

Anthony Bishopp

Supervisor: Dr. Justin Goodrich

Thesis submitted for the degree of Doctor of Philosophy

University of Edinburgh

May 2004

Declaration

This is my own work and any contributions by others are clearly cited.

Abstract

In both plants and animals Polycomb group (Pc-G) proteins act to maintain silencing of key developmental genes. A number of Pc-G genes have been identified in *Arabidopsis* through independent genetic screens and can be placed in three groups based upon homology with the *Drosophila* proteins, E(Z), ESC and SU(Z)12. In *Arabidopsis*, there are multiple *E(Z)* homologues and these act in different pathways (*CURLY LEAF* represses flowering in the immature plant; *MEDEA* controls certain aspects of seed development and *SWINGER* has recently been shown to be partially redundant to *CLF*). Despite the high level of conservation between CLF, SWN and MEA proteins, I have shown that they are not equivalent. Both MEA and SWN are unable to rescue a *clf* mutant when expressed under either the *CLF* endogenous promoter or the constitutative 35S promoter.

In *Drosophila* it is known that E(Z) acts in a multimeric protein complex, which includes the Pc-G proteins ESC and SU(Z)12. Through yeast-two hybrid assays, I have shown that CLF physically interacts with the SU(Z)12 homologue EMF2 and the ESC homologue FIE, providing evidence of a similar complex acting to repress flowering in the immature plant. I have tested the specificity of this interaction and shown that CLF and SWN are able to interact with all three *Arabidopsis* SU(Z)12 homologues including VERNALISATION2, suggesting that these genes may function in other pathways. I have not been able to observe loss of vernalisation response equivalent to that of *vrn2* in either *clf* or *swn* single mutants, and conclude that the two genes are likely to act redundantly to control the vernalisation response.

Embryos carrying a maternal inherited *fie* allele abort during seed development, making it impossible to observe the role of *FIE* post seed devlopment. I have created a transgenic steroid-inducible *FIE* line, which I have used to rescue *fie*⁻ embryos. The resulting plants show severe developmental abnormalities, with highly disorganised organ growth.

Acknowledgments

I would firstly like to thank my supervisor, Dr. Justin Goodrich, whose support and encouragement throughout my PhD have been invaluable and whose good humour has made the PhD experience less stressful.

I am also deeply indebted to my parents who have supported me both emotionally and financially for 27 years. Although you will probably never read this thesis, I hope that you will enjoy the pictures and find it a worthy conclusion to years of being a student.

I would also like to thank many colleagues who have offered help throughout the course of my studies. Christine Stock taught me many of the techniques which have been used routinely through this study. I had numerous fascinating conversations with Daniel Schubert (many of which were work related), and he was a great source of technical knowledge. Amelia Green helped with some of the yeast two-hybrid analysis and genotyped some of the *fie* tilling lines. A minha namorada, Cecília, ouviu-me quando eu não tive paciência nenhuma com a Universidade, cozinhou-me comida boa e fez muitas outras coisas mais.

Thank you all.

List of abbreviations

a, A	adenine
ATP	adenosine triphosphate
ATP	*Arabidopsis* TILLING project
bp	base pair(s)
BSA	bovine serum albumin
c	cysteine
C	cysteine
CAP	cleaved amplified polymorphism
cDNA	complementary DNA
CIAP	calf intestinal alkaline phosphotase
CSPD	Disodium 3-(4-methoxyspiro{1,2-dioxetane-3,2'-(5-chloro)tricycle [3.3.1.1] decan}-4-yl)phenyl phosphate
dCAP	derived cleaved amplified polymorphism
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphate
dsDNA	double-stranded DNA
DTT	dithiothreiotol
EDTA	ethylenediaminetetraacteic acid (disodium salt)
EGTA	ethylene glycol-bis(β-aminoethyl ether)
EMS	ethylmethanesulfonate
EST	expressed sequence tags
g	gravity
g	guanine
H	histidine
IPTG	β-D-isopropyl-thiogalactopyranoside
kb	kilobase(s)
L	leucine
M	molar
mM	millimolar
nM	nanomolar
mRNA	messenger RNA
OE-PCR	overlap extension PCR
OD	optical density
ONGP	o-nitrophenyl β-$_D$-galactopyranoside
PCR	polymerase chain reaction
PEG	polyethylene glycol
pM	picomolar
RNA	ribonucleic acid
RNAase	ribonuclease

rpm	revolutions per minute
S-D IVM	site-directed *in-vitro* mutagenesis
SDS	sodium dodecyl sulfate
t	thymine
TE	Tris-EDTA buffer
TILLING	Targeting Induced Local Lessions In Genomes
T_m	melting temperature
tween	polyethylenesorbitan monolaurate
u	unit
µM	micromolar
UTR	untranslated region
UV	ultraviolet
W	tryptophan
X-Gal	5-bromo-4-chloro-3-indolyl-β-galactopyranoside

1 Introduction

1.1 General overview of the role of the Polycomb group

The Polycomb group (Pc-G) genes were first discovered in *Drosophila* as long-term regulators of homeotic gene repression and therefore segment identity, although their targets are now known to include many other genes. A common factor is that in Pc-G mutants, the expression patterns of the target genes are initiated normally, but later in development break down leading to expression beyond their usual domains. Therefore, the Pc-G genes are not required in setting up the initial expression patterns, but rather to maintain transcriptional repression in cells where the target was originally repressed (Summarised in Goodrich and Tweedie, 2002).

Since their discovery in animals, Pc-G genes have been identified in *Arabidopsis* through genetic screens. These have demonstrated roles in diverse pathways and shown, that like in animals, Pc-G genes play important roles in the regulation of various developmental pathways. These include control of flowering time, seed development and maintaining an epigenetic memory of a substantial period of cold treatment (vernalization). The processes by which these proteins operate are not identical in plants and animals, however, many aspects of Pc-G function are conserved between the two kingdoms.

In plants cell fate is relatively labile (compared with that in animals) and can be re-specified following changes in environment. It is likely that cell fate in plants is mitotically heritable as a result of chromatin remodelling genes maintaining expression patterns of key genes, but unlike in animals it is relatively easy to reset, for example during tissue culture.

In the following chapters, the direct targets of the *Drosophila* Pc-G and the roles that these proteins have on development are discussed first. This in turn leads to a discussion

of how the Pc-G genes were isolated as regulators of these genes, how the Pc-G function as part of multimeric complexes and finally how the Pc-G proteins act on chromatin catalytically. The Pc-G genes are then introduced in *Arabidopsis* and the cloning, complexes and biochemical activity are discussed.

1.2 Developmental processes in animals

In animals, development largely occurs during embryogenesis during which stage all major organs are formed. Although the processes of development continue throughout the life of the animal, these consist mainly of growth and renewal of tissue. This can be seen in mammals where a newly born human, largely resembles an adult. In *Drosophila* this is less apparent because their lifecycle involves a stage of metamorphosis. However, in the larvae although structures such as the wing have not formed, they are present in the form of imaginal discs. These imaginal discs are small sheets of epithelial cells which give rise to the adult structures, wings, legs, halteres, genitalia, eyes, mouthparts and antennas during metamorphosis (Lawrence, 1992).

1.3 Control of segment identity by homeotic genes in *Drosophila*

The *Drosophila* embryo is divided into 14 parasegments plus the head and tail regions. Each of these parasegments has a different fate; forming three offset head segments, three offset thoracic segments and eight offset abdominal segments in the adult. Each segment produces organs associated with its position in the fly, for example cells from the second thoracic segment form the muscles set associated with the wing disc and the second leg (figure 1.1) (Lawrence, 1992).

Homeotic genes of the Antennapedia and bithorax complexes encode transcription factors that control segment identity, or cell fate, along the anterior-posterior body axis in the developing fly (Lewis, 1978; Akam, 1989). These homeotic genes contain a highly conserved DNA-binding homeodomain (Thali et al, 1988). They were discovered based on mutant phenotypes where segments, or groups of segments, are transformed to show likeness of another. For example, mutations in the

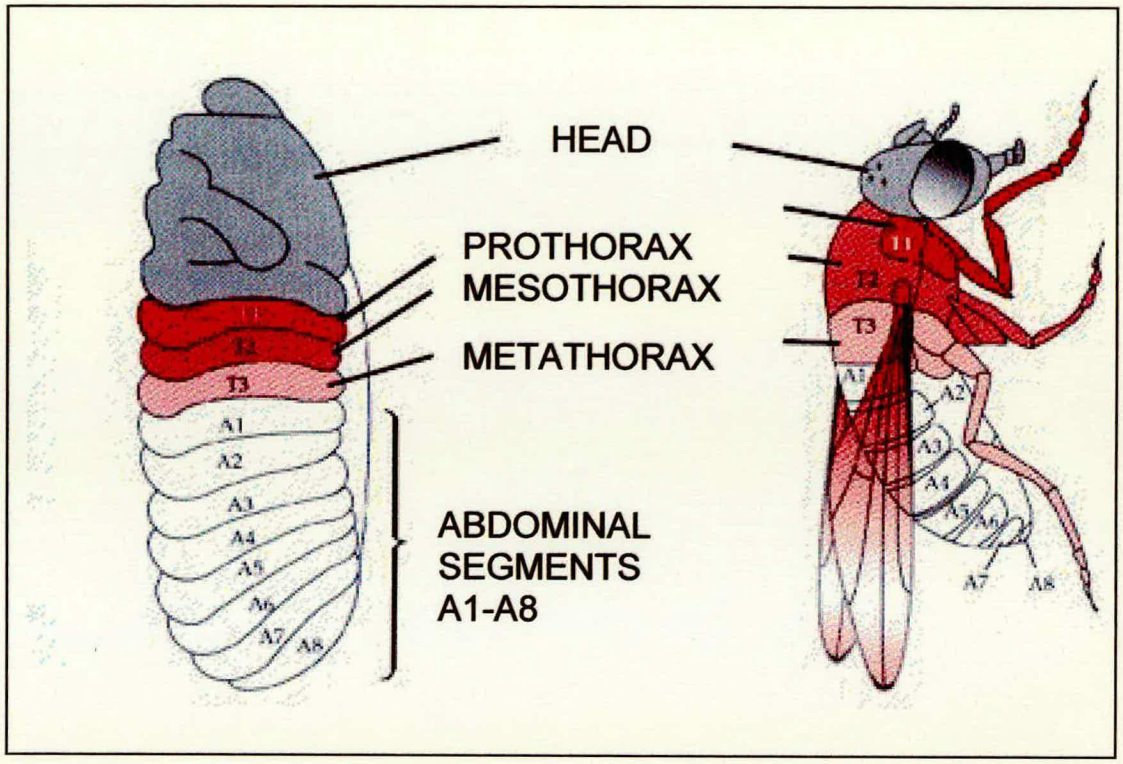


Figure 1.1 Segment identity in the adult *Drosophila* fly
On the left is a *Drosophila* larva with all abdominal and thoracic segments labelled (T1-3 and A1-8). In the adult fly the structures arising from each segment are labelled or colour coded. For example, the second thoracic segment gives rise to the wings and the second pair of legs. The head segments are not labelled but give rise to the eyes, mouthparts and antenna. Diagram reproduced from http://www.zum.de/Faecher/Materialien/hupfeld/index.htm?/Faecher/Materialien/hupfeld/Entwicklung/Drosophila.

ULTRABITHORAX (*UBX*) gene result in transformations of the appendages of the third thoracic segment (usually the haltere and third leg) into appendages associated with the second thoracic segment (wing and second leg) (Struhl, 1982).

Each homeotic gene is transcribed at specific locations within the anterior-posterior body axis and the particular combinations of homeotic proteins in each segment determine its identity. *UBX* expression patterns have been visualised by *in-situ* hybridisation and by using antibodies to localise protein products, and are distributed in parasegments 5-13 towards the end of embryogenesis, with the strongest expression in the sixth parasegment (Akam and Martinez-Arias 1985; White and Wilcox, 1985). These expression patterns are set up early in development and maintained during the embryonic and larval life in specific locations. For example, *UBX* is expressed in the imaginal discs of the third thoracic segment (haltere and third leg) of the adult fly (Akam, 1983; White and Wilcox, 1984).

Changes in the expression patterns of these genes can have profound effects on morphology, such as the alteration of body plans. Normally *Drosophila* have just one pair of wings, however many other insects such as butterflies have a second pair as well as larval abdominal limbs. Butterflies and *Drosophila* have diverged in their abdominal expression of *UBX*, in butterflies early *UBX* expression is absent in the abdomen allowing the derepression of *DISTAL-LESS* and the specification of abdominal limbs (Warren et al., 1994). Other factors are likely to be involved in the production of a second pair of wings in butterflies because there are no significant differences in *UBX* expression in the third thoracic imaginal discs which form halteres in *Drosophila* and hindwings in butterflies (Warren et al., 1994; White and Wilcox, 1984).

Clearly the expression patterns of homeotic genes need to be tightly controlled throughout development. In *Drosophila* this is achieved through the gap and pair-rule proteins. Most gap and pair-rule proteins are DNA binding regulators with spatially restricted distributions; they bind directly to their target genes and regulate transcription.

In the case of *UBX*, the gap gene *HUNCHBACK* (*HB*) plays an important role in specifying the boundaries of *UBX* expression. In *hb* mutants, *UBX* is ectopically expressed in regions both anterior and posterior to the boundaries observed in wild-type embryos (White and Lehmann, 1986). However, *HB* is only transiently expressed, and expression disappears after gastrulation, whilst homeotic gene expression is maintained throughout the adult phase (Lawrence and Morata 1994; Tautz, 1998).

1.4 Identification of the Pc-G as regulators of homeotic gene expression

The Pc-G genes were originally identified in *Drosophila* through their role in maintaining the expression patterns of homeotic genes, and since have been characterised in diverse organisms including mammals and plants. The transcription factors which initially set up patterns of homeotic gene expression in *Drosophila* are only transiently expressed. However, cells require a mechanism of permanently recording these homeotic expression patterns so cell identity can be maintained and passed on to daughter cells following mitotic division, thereby creating an epigenetic memory of cell identity. This is achieved through the Pc-G and Trithorax group (Trx-G) complexes. Together they maintain the repressed state of homeotic gene expression in cells where homeotic genes were previously inactive, and maintain the active state in cells where homeotic genes were expressed.

Twelve Pc-G genes have been characterised genetically in *Drosophila* (although there are potentially many more) and loss-of-function mutations shown to cause mis-expression of homeotic genes in embryos and larvae. Most of these *Drosophila* Polycomb proteins are conserved in both structure and function in mammals. However only EXTRA SEX COMBS (ESC), ENHANCER OF ZESTE (E(Z)), SUPPRESSOR OF ZESTE (SU(Z)12) and p55 are more widely conserved. In *C. elegans* homologues of E(Z) and ESC function in germline development, but do not have any role in the regulation of homeotic gene expression (Holdeman et al., 1998; Korf et al., 1998). It has recently been discovered that in *C. elegans* homeotic gene repression is carried out in

part by SOP2, a protein related to, but not orthologous with any Pc-G members (Zhang et al., 2003). *sop2* mutants show widespread ectopic expression of homeotic genes as well as homeotic transformations. This suggests that in nematode worms, global repression of homeotic genes has been taken over by a different family closely related to the Pc-G. In plants, homologues of ESC, E(Z) and SU(Z)12 exist and these are discussed in detail in the next sections. For this reason, discussions about the *Drosophila* Pc-G genes will focus on these three.

The *Drosophila EXTRA SEX COMBS* (*ESC*) gene is required early during embryogenesis to ensure the correct spatial expression of genes belonging to the Bithorax and Antennapedia complexes during subsequent development. Mutant *esc* embryos (lacking both maternal and paternal ESC^+ activity) show severe homeotic transformations where all abdominal and thoracic segments and some head segments develop into eighth abdominal segments by the first instar larvae stage, not surprisingly these embryos fail to hatch and die (Struhl, 1981). Homozygous *esc* mutant embryos can be rescued fairly completely by *ESC* product provided maternally during oogenesis (i.e. esc^-/esc^- progeny of ESC^+/esc^- mother), progeny are viable and show only defects in the positioning of sex combs. This suggests that the gene product is persistent long enough after fertilisation to allow the correct formation of almost all segments, or that the effects of ESC^+ are persistent, i.e. acts early to initiate a change but is not needed for its maintenance (Struhl, 1981). However, zygotically produced E(Z) is required for larval development and continuously throughout development (Jones and Gelbart, 1990).

The expression of the homeotic gene *UBX* has been described in both wild-type and *esc* mutant embryos (Struhl and Akam, 1985). Initially the pattern of *UBX* expression is indistinguishable between wild-type and *esc* embryos, but shortly thereafter *UBX* transcripts began to accumulate indiscriminately in all segment primordia. This suggests the role of *ESC* is not to specify where homeotic genes are to be expressed, but rather to

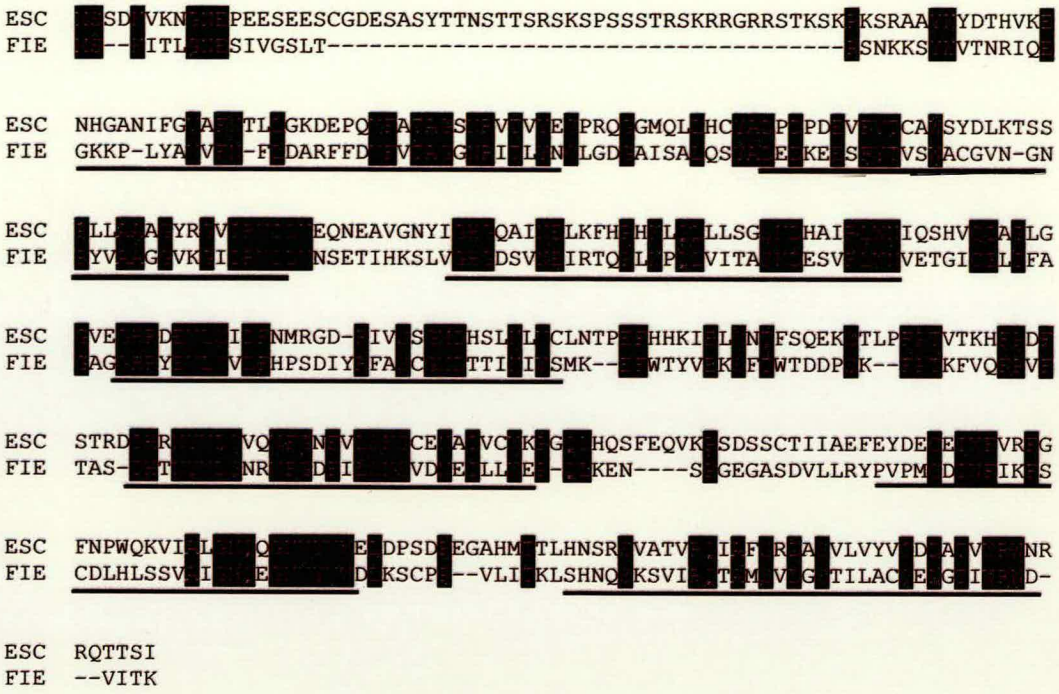
act subsequently to ensure homeotic genes remain off in primordia where they are inactive.

A complete lack of *ESC*, *E(Z)* or *SU(Z)12* activity results in similar phenotypes (Struhl, 1981; Jones and Gelbart, 1990; Birve et al., 2001); embryos show severe homeotic phenotypes with all abdominal, thoracic and some head segments converted to eighth abdominal segments. A combination of genetic techniques have been used to produce flies showing only weak homeotic changes for all three genes; mutants showed similar phenotypes with the second and third legs resembling the first leg (ie. with additional sex combs) (Slifer, 1942; Jones and Gelbart, 1990; Birve et al., 2001). Furthermore, *UBX* is ectopically expressed in *esc*, *e(z)* and *su(z)12* mutant embryos demonstrating that these phenotypes are associated with the deregulation of homeotic gene expression (Struhl and Akam, 1985; Jones and Gelbart, 1990; Birve et al., 2001). An additional feature of SU(Z)12 which distinguishes it from other Pc-G genes is it suppresses position-effect variegation (PEV) (Birve et al., 2001). PEV is observed in chromosomal rearrangements in which a euchromatic gene is placed near heterochromatin. The translocated gene may then become inactivated in a fraction of cells, presumably because transcription of the gene is silenced by spreading of heterochromatin.

Cloning of these three genes has revealed that their protein products all have distinct structures, and it has subsequently been shown that these proteins are well conserved in diverse species including plants.

ESC encodes a nuclear localised protein with seven conserved WD-40 repeats (Gutjahr et al., 1995). Based upon homology with another WD-40 protein, it is predicted that these seven WD-40 repeats fold into a β-propellor structure (Figure 1.2) (Ng et al., 1997). Similar proteins are known to work as components of multimeric protein complexes, where they act as scaffolds upon which other components are assembled (Neer et al., 1994). ESC shows extensive homology with the EED in mammals, MES6

A



B

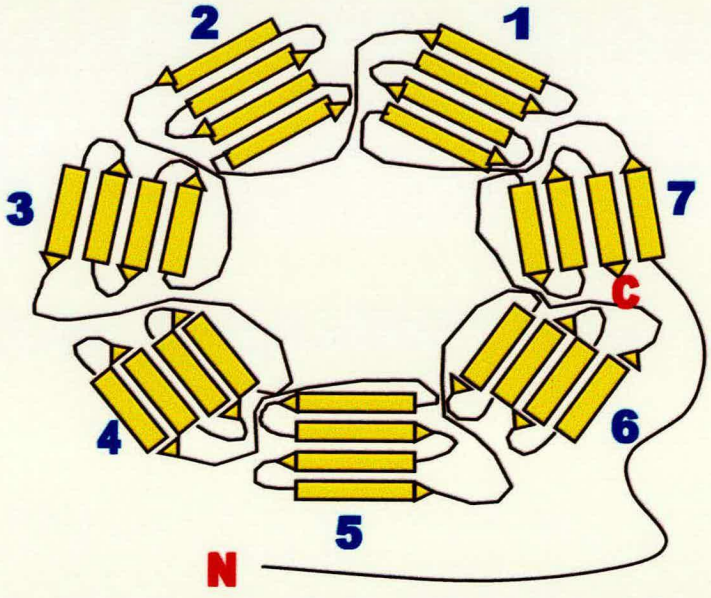


Figure 1.2 Alignment of ESC with the *Arabidopsis* homologue, FIE
A. Alignment of *Drosophila* ESC with *Arabidopsis* FIE. Black boxes illustrate residues which are identical in both sequences. Black lines show the 7 WD40 repeats. B. Schematic diagram showing predicted structure for ESC (adapted from Ng et al., 1997). The 7 WD-40 repeats are organised into a β-propellor structure.

in *C. elegans* and FIE in plants (Schumacher et al., 1996; Korf et al., 1998; Ohad et al., 1999)

E(Z) encodes a protein with a conserved SET domain and adjacent cysteine rich regions (figure 1.3) (Jones and Gelbart, 1993). The SET domain was defined as a region of conservation between the founding members of the family (*SUPPRESSOR OF VARIEGATION* 3-9, *E(Z)* and *TRITHORAX*) (Tschiersch et al., 1994). Subsequently, the SET domain was shown to have a biochemical function in showing a H3-specific methyltransferase activity (Rea et al., 2000; Nakayama et al., 2001; Cao et al., 2002). E(Z) is homologous to MES2 in *C. elegans*, EZH1 and EZH2 in mammals and CLF, SWN and MEA in *Arabidopsis* (Holdeman et al., 1998; Laible et al., 1997; Goodrich et al., 1997; Grossniklaus et al., 1998).

SU(Z)12 encoded a novel type of protein with homologues in *Arabidopsis* and mammals, but not in *C. elegans* (figure 1.4) (Birve et al., 2001). The characteristic feature of these proteins is a zinc finger motif similar to those found in DNA binding proteins and a VEFS box. The first gene of this family to be cloned was the *Arabidopsis* Pc-G gene *FERTILISATION-INDEPENDENT SEED 2* (*FIS2*). The family was named after the *Arabidopsis* genes *VERNALIZATION 2* (*VRN2*), *EMBRYONIC FLOWER 2* (*EMF2*), *FIS2*, and the *Drosophila* gene *SU(Z)12*. These genes are discussed later.

1.5 Pc-G proteins act in multimeric complexes

Based on similar mutant phenotypes it was proposed that the structurally disparate Pc-G proteins genes act together in a complex to repress gene expression. This was supported by cytological evidence through observations that all Pc-G proteins analysed often co-localise on polytene chromosomes (Carrington and Jones, 1996). However, a closer relationship between *ESC* and *E(Z)* was demonstrated, based on the observation that the

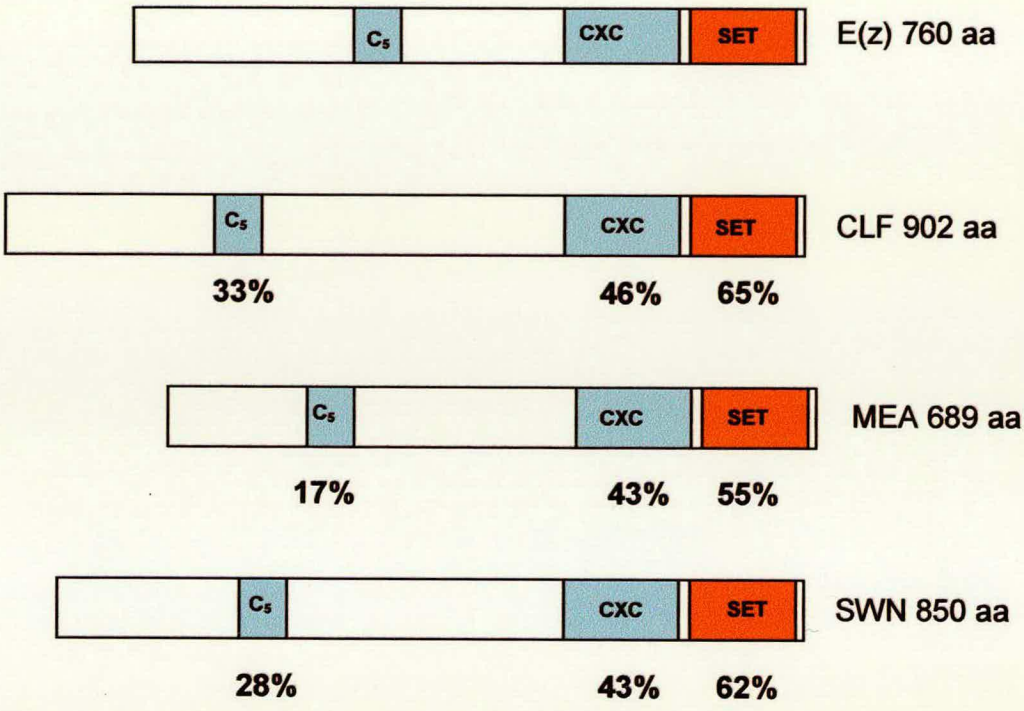


Figure 1.3 Schematic alignment of E(Z) with *Arabidopsis* homologues
There are three main conserved regions: the C-terminal SET domain (orange), and two additional cysteine rich regions (blue) consisting of the CXC domain and five additional highly conserved cysteine residues (C_5). Numbers show percentage of amino-acid identity with *Drosophila* E(Z). Protein lengths are shown on the right.

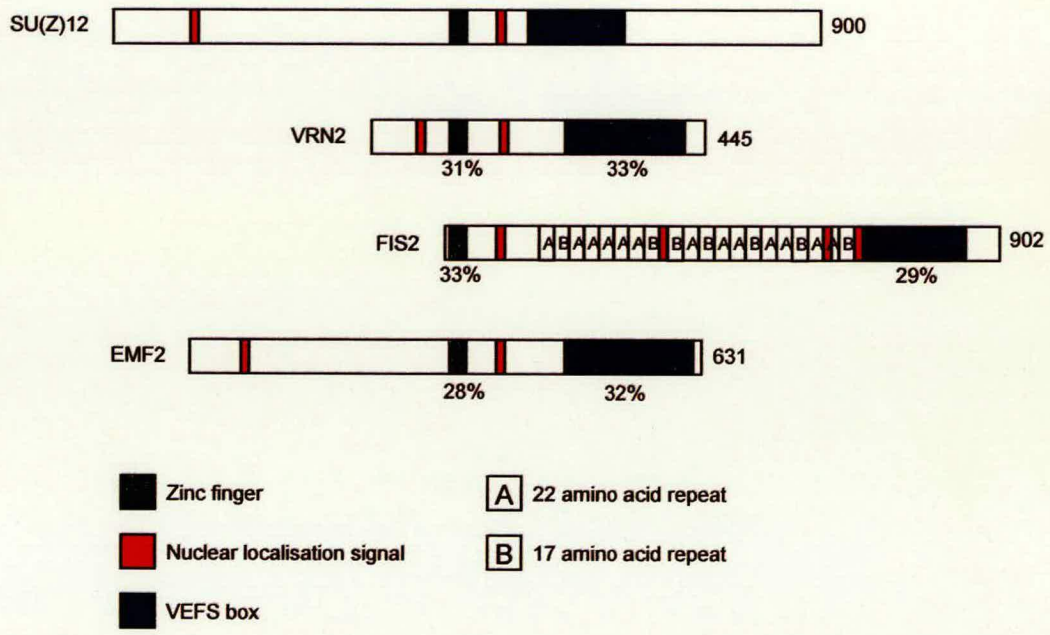


Figure 1.4 Schematic alignment of SU(Z)12 with *Arabidopsis* homologues
There are 2 main areas of conservation between these proteins, the VEFS domain (dark blue) and the zinc finger (black), although these proteins also some similarity in the N-terminal region. The percentage values shown below these domains represent the amino acid identity with *Drosophila* SU(Z)12. Figure adapted from Gendall et al (2001).

maternal effect of *esc* is exacerbated by decreasing or increasing the zygotic dosage of *E(Z)* (Cambell et al., 1995).

Consistent with these lines of evidence a direct physical interaction was shown between ESC and E(Z) through yeast two-hybrid and *in-vitro* binding assays (Jones et al., 1998). An *in-vitro* interaction was also shown between EED and EZH1 (the human homologues) showing that the ESC-E(Z) partnership is evolutionarily conserved. In addition, co-immunoprecipitation from embryo extracts demonstrated an association of ESC and E(Z) *in-vivo*. This partnership is also conserved between the *Arabidopsis* homologues (see later sections).

More recent and compelling evidence has been provided by the biochemical purification of two complexes containing *Drosophila* Pc-G proteins. The first complex, Polycomb Repressive Complex 1 (PRC1) contains POLYCOMB (PC), POSTERIOR SEX COMBS (PSC), POLYHOMEOTIC (PH), dRING1 and various other polypeptides (Saurin et al., 2001; Shao et al., 1999). A second complex, PRC2, contains ESC, E(Z), SU(Z)12 and p55 (Müller et al., 2002). Components of the two complexes can be seen in table 1.1.

1.6 Mechanism of Pc-G action in animals

These two complexes have discrete biological roles; the PRC2 complex is involved in marking histones to maintain silencing at certain loci, in turn this mark is believed to be interpreted and more stably silenced by the PRC1 complex. In *Drosophila*, homeotic gene expression is initially set up by transiently expressed transcription factors which determine the status of the homeotic gene promoters. Pc-G proteins are then recruited to their targets by binding to specialised switchable modular *cis*-acting DNA elements, known as Polycomb Response Elements (PREs) (Lyko and Paro, 1999). Because most Pc-G proteins lack DNA binding specificity they are probably recruited by partners such as *PHO*, *ZESTE* and *GAGA* to bind with PREs, PREs contain binding sites for these factors and deletion of these sites can affect their function (Hur et al., 2002). The ability

Table 1.1 ***Drosophila*** **Pc-G genes form two separate complexes**

Polycomb Repressive Complex 1			
Pc-G gene	Human Homologues	*Arabidopsis* Homologues	Motifs
POLYCOMB PC	*HPC1, HPC2, HPC3*	-	Chromodomain
POLYHOMEOTIC PH	*HPH1, HPH2*	-	Zinc finger SPM domain
SEX COMBS ON MIDLEG SCM	*SCML1, SCML2, SCMH1*	-	Zinc finger SPM domain
POSTERIOR SEX COMBS PSC	*BMI1, MEL18*	-	RING finger
dRING	-	-	RING finger

Polycomb Repressive Complex 2			
Pc-G gene	Human Homologues	*Arabidopsis* Homologues	Motifs
EXTRA SEX COMBS ESC	*EED*	*FIE*	WD-40 repeats
ENHANCER OF ZESTE E(Z)	*EZH1, EZH2*	*CLF, MEA, SWN*	SET domain
SUPPRESSOR OF ZESTE 12 SU(Z)12	*HsSU(Z)12*	*FIS2, VRN2, EMF2*	VEFS box
p55	RbAp48	*MSI1-5*	WD-40 repeats

Drosophila Pc-G proteins have been biochemically purified and shown to be present in two complexes (see text). The data is taken from Flybase www.flybase.bio.indiana.edu, with the exception of the *Arabidopsis* homologues which are from Goodrich and Tweedie (2002).

of Pc-G complexes to bind with PREs is also determined by the activity of target gene, which has already been determined by the original transcription factors (Poux et al., 2001). The PRC2 complex is then able to exert a methyl transferase activity on specific lysine residues on the N-terminal tails of histone H3 at these loci. This is achieved through the SET domain of E(Z) leaving tri-methylation marks at K9 and a di- and tri methylation marks at K27 (Cao et al., 2002; Czermin et al., 2002; Müller et al., 2002; Kuzmichev et al., 2002). After the *Drosophila* embryo has reached the blastoderm stage the PRC2 complex disassociates from its role in maintaining long-term silencing of homeotic genes, although some components such as E(Z) are persistently required. This is most likely due to additional requirements of E(Z) in maintaining higher order chromatin structure (Orlando, 2003).

Methylation of the K9 residue on histone H3 has been shown to create a binding site for the chromo domain of HETEROCHROMATIN PROTEIN 1 (HP1) (Lachner et al., 2001; Bannister et al., 2001). HP1 is able to coat the chromatin to form heterochromatically silenced regions. Recently, it has been shown that the chromo domain of PC shows specific binding to methylated K27 residues (Fischle et al., 2003). This may provide a mechanism through which the PRC1 complex is able to interpret the marks left by E(Z) and prevent transcription of repressed homeotic genes.

Although Polycomb-group mediated repression of homeotic genes begins early in embryogenesis and continues through embryonic, larval and pupal development, ESC differs from other Pc-G proteins because it is only required during early embryogenesis. Analysis of a temperature sensitive *esc* allele reveals that *ESC* activity is only required during a discrete period of 3-4 hours starting at the end of the blastoderm stage and ending at the germ-band extension stage (Struhl and Brower, 1982). This has been confirmed through the use of a transgenic heat-inducible line (Simon et al., 1996a). In accordance with this, *ESC* mRNA is expressed at the greatest levels early in development but expression disappears after germband extension (Gutjahr et al., 1995; Sathe and Harte, 1995).

Contrastingly, analysis of temperature *e(z)* sensitive alleles reveal that although it is also required at this early stage, it is also continuously required through the pupal stage to maintain repression of homeotic genes (Jones and Gelbart, 1990).

1.7 Growth and development in flowering plants

In flowering plants most developmental growth occurs post-embryonically from localised meristems through the continuous production of undifferentiated stem cells at the root and shoot. Adult structures such as shoots, leaves, roots and flowers arise from the flanks of these meristems. During embryogenesis two meristems are established, one at the root and the other at the tip of the shoot, all the adult meristems arise from these. This mode of growth means that developmental patterning not only occurs within the embryo but also in the meristems throughout the life of the plant.

After a period of vegetative growth, plants are capable of producing flowers. The transition from vegetative growth to reproductive growth requires the reorganisation of the vegetative meristem into an inflorescence meristem. In *Arabidopsis* this results in the production a primary inflorescence onto which axillary inflorescences and flowers form. Flower primordia or floral meristems form in a spiral pattern on the flanks of the inflorescence meristem. Floral organ primordia form within this floral meristem in four distinct whorls where they differentiate into the appropriate organs producing a flower.

1.8 Homeotic genes control organ identity in plants

As in *Drosophila*, homeotic genes are among the targets of the *Arabidopsis* Pc-G proteins. In plants these homeotic genes play a key role in specifying organ identity, e.g. within the flower. Flowers usually comprise four concentrically arranged whorls of organs with sepals outermost, followed by petals, stamens and carpels forming the innermost whorl. *Arabidopsis thaliana* has four sepals, four petals arranged in the

typical cruciform phyllotaxy, six stamens and an ovary of two united carpels (figure 1.5). Homeotic mutations have been discovered in *Arabidopsis thaliana* and *Antirrhinum majus* which change the identity of the floral organs. The corresponding wild-type genes have been shown to be required for directly controlling organ identity in a manner equivalent to *Drosophila* homeotic genes (organ identity genes), regulating the spatial expression pattern of organ identity genes (cadastral genes) or for the initiation of organ identity genes (meristem identity genes).

Organ identity genes can further be classified into three classes, each gene altering organ identity in two adjacent whorls. Complete loss of the A class genes, *APETALA 1* and *2* (*AP1* and *AP2*) causes transformations of the first whorl sepals into carpels and the second whorl of petals into stamens. This is somewhat complicated as severe A class mutants lack organs in the second whorl. Loss of B class gene activity, *APETALA 3* (*AP3*) and *PISTILLATA* (*PI*), causes sepals to replace petals in the second whorl and carpels to replace stamens in the third whorl. Finally, loss-of-C-function, *agamous* (*ag*) mutants show transformation of third whorl stamens into petals and fourth whorl carpels into sepals (reviewed in Weigel and Meyerowitz, 1994). Figure 1.6 shows an *ag* (C class) mutant with homeotic transformations in the third and fourth whorls. Analysis of double mutants revealed that class B activity is independent of classes A and C, however class A activity is ectopically present in class C mutants and *vice versa* (Bowman et al., 1991a). A common model for floral development was provided by the ABC model for homeotic gene activity (figure 1.7).

Molecular cloning has revealed that with the exception of *AP2* all the floral organ identity genes belong to the MADS group of transcription factors. These proteins are characterised by the presence of a MADS box (named after conservation in the founding members *MCM1* from yeast, *AG* from *Arabidopsis*, *DEFICIENS* from *Antirrhinum* and *SRF* from mammals), and a second domain showing similarities to the coiled coil domains of keratins. These genes are all expressed in regions coinciding with their domain of function, i.e. those exhibiting homeotic transformations in the loss-of-

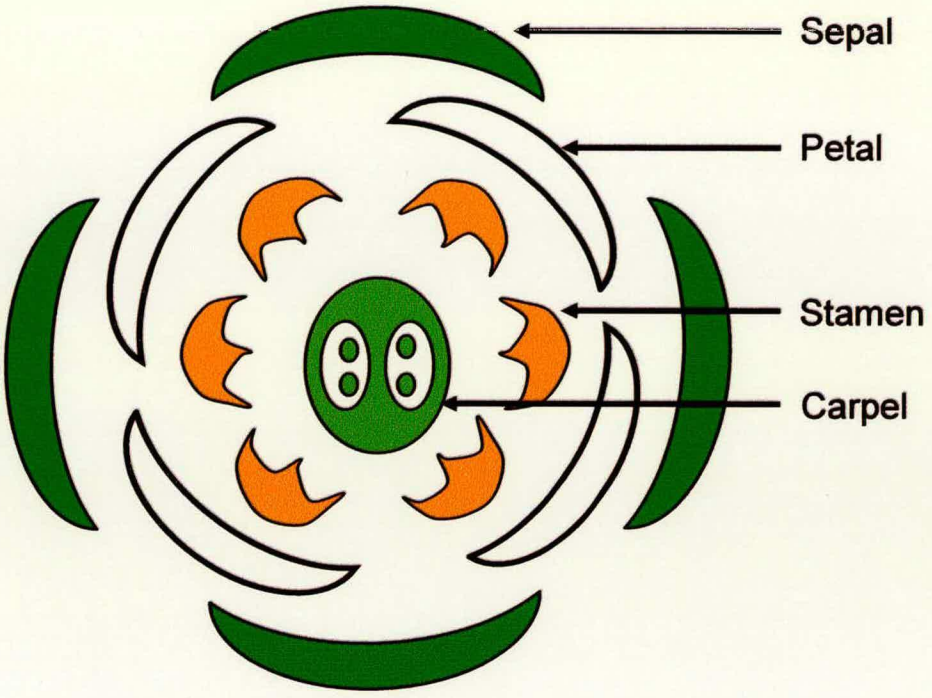


Figure 1.5 Schematic diagram of an *Arabidopsis* flower
The flower comprises of concentric whorls of floral organs as labelled. *Arabidopsis* has two fused carpels, six stamens, four petals and four sepals.

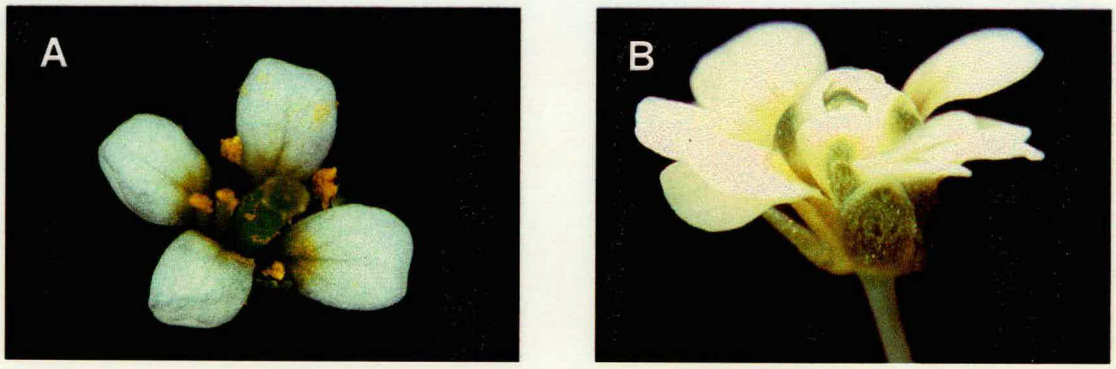


Figure 1.6 Homeotic changes in floral organ identity in an *agamous* mutant
A. A wild-type flower is shown (note concentric whorls of sepals, petals, stamens and two fused carpels). B. The flower of an *agamous* mutant is shown; the inner whorls of stamens and carpels have been replaced with additional whorls of sepals and petals.

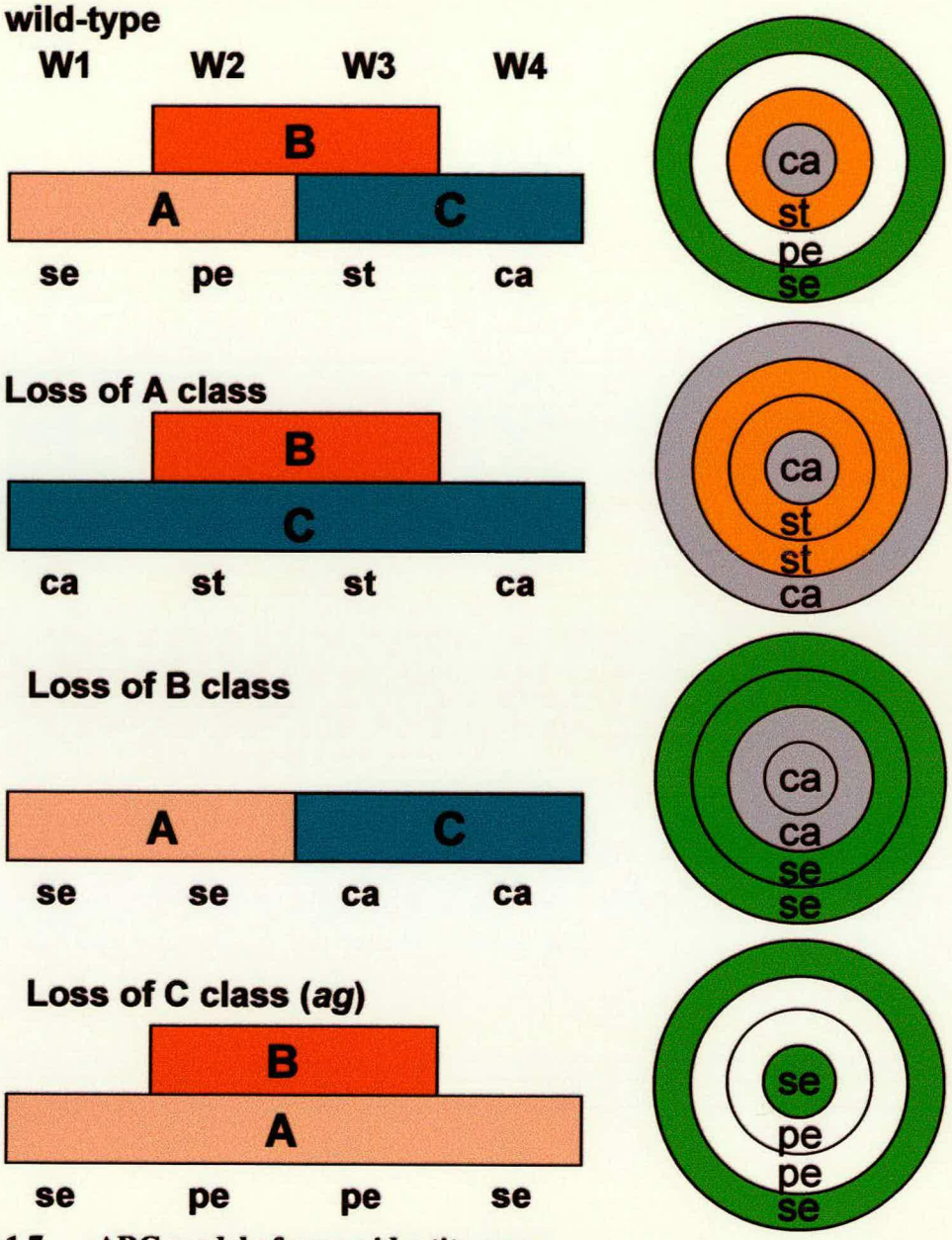


Figure 1.7 ABC model of organ identity genes

The block diagrams represent cross section through a half flower with the 4 whorls marked W1-4. In wild-type these comprise of sepals (se), petals (pe), stamens (st) and carpels (ca). A schematic aerial view of the flower is shown right with sepals shown in green, petals in white, stamens in yellow and carpels in grey. The presence of class A genes alone determines the development of sepals, classes A and B that of petals, classes B and C that of stamens and class C alone that of carpels. The presence of class A genes repress class C genes in whorls 1 and 2, and class C genes repress class A genes in whorls 3 and 4. The organ specification in mutants lacking activity of A, B and C class genes are shown. For more details see Coen and Meyerowitz (1991).

function mutants. As with homeotic genes in *Drosophila*, activation of floral identity genes precedes organ specification, and the expression of these genes persist until organogenesis is largely complete (Weigel and Meyerowitz, 1994).

The requirement to tightly control the expression patterns of these floral identity genes is paramount to correct development. Mis-expression of the floral homeotic genes in the flower often results in mutants becoming sterile. For example, early in flower development *AG* is expressed evenly within the third and fourth whorl organ primordia, but is not present in the organ primordia of the first and second whorls. Later in development *AG* is restricted to particular cell types within the stamens and carpels, such as stigmatic papillae and ovules (Bowman et al., 1991b). Studies with the temperature sensitive *def-101* allele (the *Antirrhinum AG* homologue) show that homeotic genes are needed throughout flower development for normal organ differentiation (Zachgo et al., 1995). Transgenic lines expressing *AG* ectopically produce flowers resembling those of *ap2* mutants, with sepals and petals showing homeotic transformations into carpels and stamens. Furthermore, leaf morphology is altered in these lines with leaves being smaller, curled and cauline leaves occasionally containing stigmatic papillae (Mizukami and Ma, 1992).

1.9 Genes controlling flowering time are targets of the Pc-G

In *Arabidopsis*, some Pc-G target genes are involved in the transition from vegetative to floral development. The timing of this transition is a key stage in the life cycle of a plant. It is regulated by a complex genetic network which monitors endogenous developmental inputs as well as external environmental cues, such as day length and temperature (Simpson and Dean, 2002). These multiple inputs converge to regulate the switch between vegetative growth and reproductive development (figure 1.8).

In *Arabidopsis* one of these pathways, the requirement for vernalization can be mapped with dominant alleles of the coiled-coiled protein FRIGIDA (FRI). FRI functions to

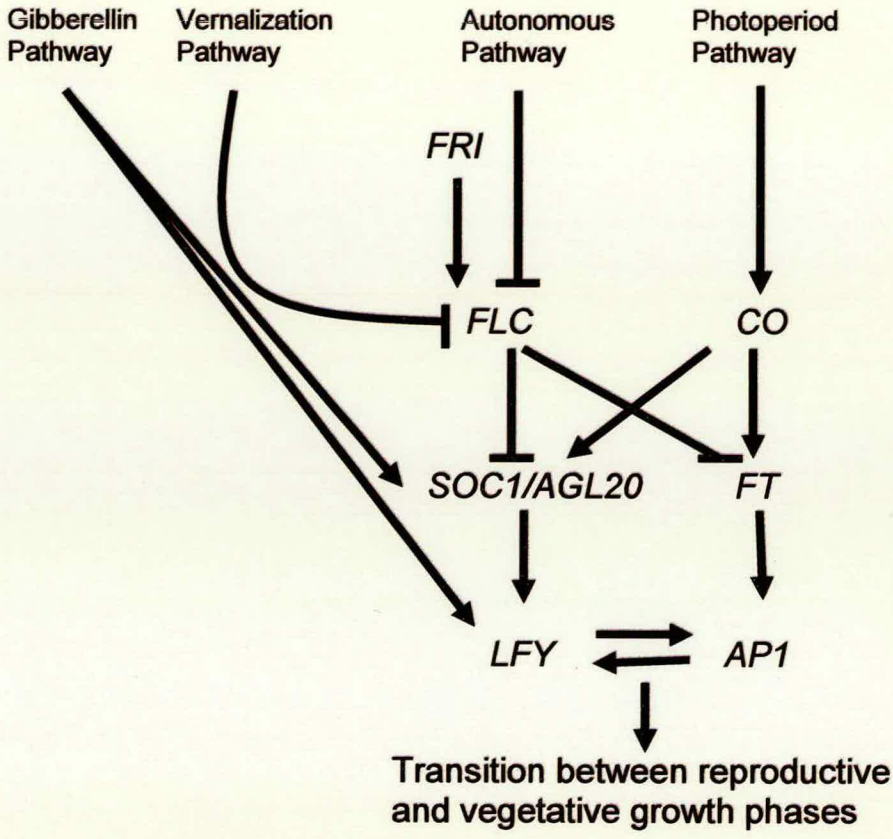


Figure 1.8 Integration of several pathways that promote early flowering
This figure is modified from Putterill et al. (2004). The inputs promote or repress the expression of *SUPPRESSOR OF CONSTANS 1* (*SOC1*) and *FLOWERING LOCUS T* (*FT*). These genes act to integrate the different inputs and together up-regulate the expression of the floral meristem identity genes *AP1* and *LEAFY* (*LFY*), as well as other genes including *CAULIFLOWER* (*CAL*), and *FRUITFUL* (*FUL*). The vernalization pathway and the autonomous pathway promote the transition to flowering by reducing the levels of the floral repressor *FLOWERING LOCUS C* (*FLC*) (Gendall et al., 2001; Michaels and Amasino 2001). The photoperiod pathway recognises long-day photoperiods and activates the transcription factor *CONSTANS*, *CO* (Samach and Coupland, 2000). The gibberellin pathway also affects the transition to flowering under short-day conditions (Wilson et al., 1992).

promote the accumulation of FLOWERING LOCUS C (FLC) (Johanson et al., 2000; Michaels and Amasino, 2001). FLC is a MADS box transcription factor that represses the transition from vegetative to reproductive development (Michaels and Amasino, 1999). The net result is that an increased level of FRI represses the transition to flowering despite environmental conditions being favourable (Johanson et al., 2000). These elevated levels of FLC can be reduced by a vernalization treatment (Sheldon et al., 2000; Gendall et al., 2001) (see section 1.14).

The autonomous pathway genes *FCA*, *FY*, *FPA*, *LUMINIDEPENDENS* (*LD*), *FLOWERING LOCUS D*, *FVE* and *FLK* normally function to repress the accumulation of *FLC* (Michaels and Amasino 2001; Simpson and Dean, 2002; Lim et al., 2004). The photoperiod pathway requires the combination of two components, detection of light and the measurement of day length. These components converge on the transcription factor *CONSTANS* (*CO*) which up-regulates the expression of *FT* and the MADS box gene *SOC1* (Samach et al., 2000). Application of gibberellins accelerates flowering time particularly under short day conditions by activation of the *LFY* promoter (Blázquez and Weigel, 2000; Wilson et al., 1992). These pathways converge and together up-regulate the expression of the floral meristem identity genes which initiate the transition from vegetative to reproductive development (Simpson and Dean, 2002).

1.10 *Arabidopsis* Polycomb group genes control the transition to flowering

In *Arabidopsis*, Pc-G genes were identified independently through genetic screens. These revealed three broad biological roles. The first role was in controlling the transition to flowering and the repression of floral homeotic genes in embryos and seedlings. The second role was in the repression of certain aspects of seed development in the absence of fertilisation. The final role was in creating an epigenetic memory of vernalization.

The first Polycomb group gene to be identified in plants was *CURLY LEAF* (*CLF*) (Goodrich et al., 1997). Mutations in *CLF* cause pleiotropic effects on flowering time

and leaf and flower morphology but do not affect the development of roots, hypocotyls or cotyledons. Mutants flower slightly earlier than wild-type under long day conditions and about three weeks earlier than wild-type under short day conditions. The stem internodes which elongate after flowering are shorter in *clf* mutants, reducing the height of the inflorescence. However, the most severely affected organs are the leaves. These are narrow and curled upwards at the margins; the extent of curling increases from the base of the plant to the apex, so that cauline leaves are the most severely affected. The *clf* mutation also has weak effects on floral morphology including partial homeotic transformations of sepals and petals towards carpels and stamens (Figure 1.9).

The *clf* phenotype had striking similarity to transgenic lines where the *AG* gene was ectopically expressed (Goodrich et al., 1997; Mizukami and Ma, 1992). Northern blot analysis confirmed that *AG* was expressed strongly in the leaves of *clf* mutants but not in wild-type leaves (Goodrich et al., 1997). Similarly weak expression of *AP3* was detected in *clf* leaves, however this unlikely to be responsible for the *clf* phenotype because constitutive expression of *AP3* does not give a vegetative phenotype and the floral phenotype is different from those observed in *clf* (Goodrich et al., 1997; Jack et al., 1994). The pattern of *AG* expression was localised by in-situ hybridisation in wild-type and *clf* plants. A weak *AG* signal was detected in young leaf primordia of *clf* mutants and this became more intense in older leaf primordia and in young leaves; no signal above background was observed in similar wild-type sections. *AG* expression in *clf* and wild-type plants was similar during the early stages of floral development, however late during petal development ectopic *AG* expression could be observed in the petals. *AG* could also be detected in the inflorescence stem and pedicels of *clf* mutants. These results demonstrate a role for the *CLF* gene in repressing *AG* expression in hypocotyls, cotyledons, leaves, inflorescence stems and petals. However, *CLF* only acts late in the flowers as demonstrated by the weak floral phenotype and late mis-expression of *AG* when compared with *ap2* mutants, this may suggest a role in maintenance rather than the initial setup of floral organ identity genes (Goodrich et al., 1997; Drews et al., 1991). Alternatively it could suggest that CLF acts redundantly in flower development.

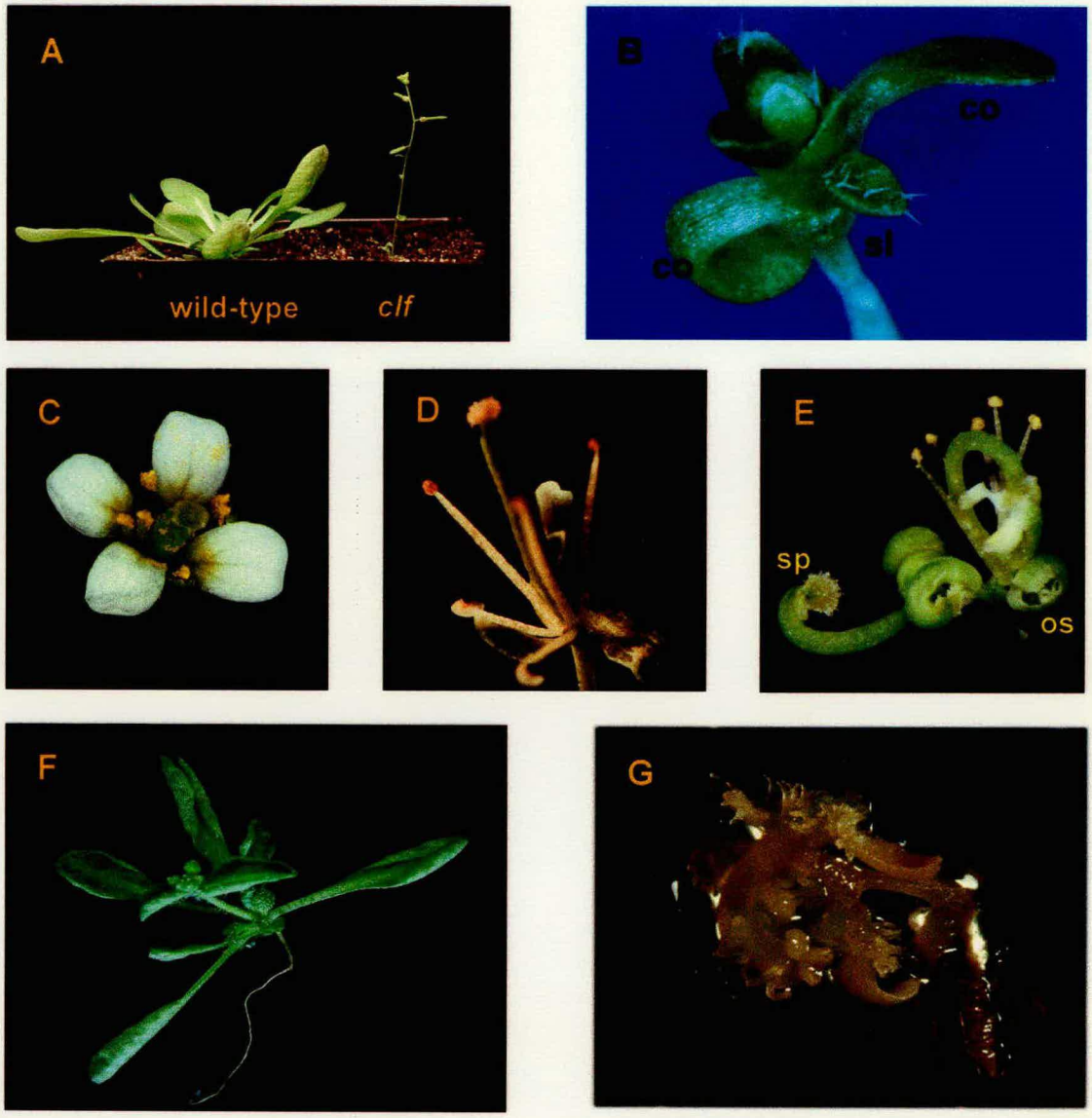


Figure 1.9 Phenotypic defects in Pc-G genes affecting the transition to flowering A. Early flowering of *clf-2* under short day conditions. B. Null *emf2-3* mutants show an extreme early flowering phenotype; a short inflorescence shoot develops prematurely after the production of the 2 cotyledons (co) and a few sessile leaves (sl). C. The wild-type flower comprises of four concentric whorls of sepals, petals, stamens and carpels. D. Occasionally late in development *clf-2* flowers show abnormal development, and sepals and petals take on carpoloid and stamonoid appearances. E. This is more pronounced in *emf2* mutants. The sepals here have clear regions with stigmatic papillae (sp) and ovary-like structures (os). F. In *clf* mutants, leaves show a characteristic strong upward curling phenotype. G. Although *swn* mutants are aphenotypic, the *clf swn* doubles show highly disorganised growth. The seedling shown here is 3 weeks old. Figures B, C, D, E, G are courtesy of Yindee Chanvivattana, Justin Goodrich and Daniel Schubert.

ag mutants were shown to be epistatic to *clf* for leaf and floral morphology, however plants were still earlier flowering under short day conditions than wild-type. This suggested that *AG* is the key target of *CLF* with respect to changes in leaf and floral morphology. However, CLF is likely to repress other genes affecting flowering time.

Cloning of *CLF* revealed a protein with extensive homology to the *Drosophila* Polycomb protein E(Z). The greatest homology was in the SET domain (65 % amino acid identity). The CLF protein also contained a CXC domain (a cysteine rich region located next to the SET domain) which shared 46 % amino acid identity with E(Z), and towards the N-terminus five additional highly conserved cysteine residues which are unique to E(Z) and its homologues (figure 1.3) (Goodrich et al., 1997).

CLF is expressed ubiquitously throughout vegetative and floral organs (Goodrich et al., 1997). These results demonstrate that CLF is insufficient to repress *AG* in the flowers as expression overlaps in the third and fourth whorls or may require other factors. There must be other mechanisms controlling *AG* expression within the flower, likely candidates include the *AP2* and *LUG* genes. Double *clf-2* weak *ap2-1* mutants had flowers with a more severe phenotype than either parent, but weaker than those of strong *ap2* (Goodrich et al., 1997). Furthermore, the level of AP2 is poised at a critical level in the *clf* background as recessive *ap2* mutations become semi-dominant. *clf-2 lug-8* mutants also show an enhancement of both floral phenotypes (Goodrich et al., 1997; Lui and Meyerowitz, 1995). Taken together these results suggest that *CLF*, *AP2* and *LUG* all act in the same pathway. More recently it has been shown that *seuss* mutants show ectopic expression of *AG* mRNA leading to partial homeotic transformations of floral organs in the outer two whorls, and more complete transformations are observed in *lug seuss* double mutants (Franks et al., 2002).

The shoot system of higher plants passes through several distinct phases during postembryonic development (Reviewed in Poethig, 1990). Each of these is characterised by a unique set of morphological features which are produced by the shoot

apical meristem (SAM). The major phase change undertaken by the SAM involves the transition from a vegetative phase to a reproductive phase. In *Arabidopsis*, this occurs when the vegetative SAM converts to the inflorescence SAM after producing several rosette leaves.

The *EMBRYONIC FLOWER* (*EMF*) genes were identified through mutagenic screens based on extremely early flowering phenotypes (Sung et al., 1992). *emf* mutants do not show rosette development and produce only small inflorescences whose lateral buds produce only flowers and no additional inflorescences (Sung et al., 1992). In addition to the early flowering phenotype, *emf1* and *emf2* mutants also show homeotic transformations of floral organs (Yang et al., 1995). In both mutants sepals and petals are either absent or take a strongly carpelloid or stamonoid appearance (figure 1.9). The *emf* mutants were shown to express *AP1* precociously and *AG* ectopically in young seedlings suggesting a role in suppressing these genes in wild-type seedlings (Chen et al., 1997). Genetic analysis using double mutants with genes regulating the switch to flowering revealed that *emf* mutants were epistatic with respect to the transition from vegetative growth to inflorescence development (Chen et al., 1997; Haung and Yang 1998). These data show that neither *co*, *lfy* nor *ft* could effectively rescue rosette development in *emf* mutants, suggesting that either the mechanism of *EMF* mediated repression of floral organ identity genes is independent of *CO*, *LFY* and *FT* or that ectopic expression of additional targets contributes to the *emf* phenotype.

The targets of the EMF proteins have been examined using microarray analysis (Moon et al., 2003). Due to the embryonic nature of the *emf* mutations it is likely that many of the genes which are mis-regulated in the microaaray are not direct targets of the EMF proteins. It was discovered that a great number of genes were deregulated in the germinating *emf* mutants, in particular floral organ identity genes. This suggests that vegetative development in wild-type plants results from *EMF* repression of the floral programme. Seed maturation genes were also ectopically expressed. However, the floral promoting genes *CO*, *FT*, *LFY* and *SOC1* were not expressed at significantly

elevated levels; this is consistent with previous experiments showing that *emf* genes are epistatic to these flowering mutants, and suggests that the *EMF* genes target the floral organ identity genes independently of the floral promoting genes. A model has been proposed whereby the *EMF* genes and floral promoting genes converge on the same targets, the floral organ identity genes. It is likely that the role of the *EMF* genes is to maintain a Polycomb-mediated silenced state of these shared targets.

Subsequent cloning of *EMF2* revealed a protein showing homology to the SU(Z)12 group of Polycomb proteins (Yoshida et al., 2001). Like the other members of this group EMF2 contains a zinc finger motif and a VEFS box. Transgenic lines constitutively expressing antisense *EMF2* produced weak mutants with a phenotype similar to *clf*, suggesting that the two genes may act in the same pathway (Yoshida et al., 2001). *EMF2* expression was detected throughout the life-cycle of the plant, but was especially abundant in actively proliferating tissues (Yoshida et al, 2001). *EMF1* was also cloned and revealed a protein showing no homology with any protein of known function except a putative protein in the rice genome (Aubert et al., 2001).

1.11 An *Arabidopsis E(Z)* homologue shows partial redundancy with *CLF*

Completion of the *Arabidopsis* genome sequence revealed the presence of a third, previously unidentified *E(Z)* homologue (Riechmann et al., 2000). This has been named SWINGER (SWN) based on its ability to share partners with CLF; previously it has been referred to as both *EZA1* and *CLK* (Luo et al., 2000; Chanvivattana, 2002). The SWN protein shares extensive homology with E(Z) in the SET and CXC domains and also contains a region near the N-terminus containing five conserved cysteine residues. *SWN* expression can be detected from early stages in leaf development and persists through to later stages, in the vegetative meristem, throughout the inflorescence meristem, in the young floral meristems and weakly in all four floral whorls (Chanvivattana, 2002). This expression pattern overlaps considerably with that described for *CLF* (Chanvivattana, 2002; Goodrich et al., 1997). Mutants have been

identified where a T-DNA insert has disrupted the *SWN* locus (Chanvivattana 2002; Justin Goodrich pers. comm.). In all cases the *swn* mutants have no discernable phenotype.

Analysis of *swn-1 clf* double mutants showed a strong enhancement of the *clf* phenotype (Chanvivattana et al., 2002). Double mutants were significantly smaller and did not produce rosette leaves, but instead formed a few sessile leaves which characteristically resembled cauline leaves. Vegetative growth was severely shortened; after 10 days the plants flowered and produced small buds at the apex but did not bolt. The phenotype strongly resembles that of the *emf* mutants, suggesting the genes could act in a common pathway.

The T-DNA insertion in the *swn-1* line lies in the 5'-UTR just upstream of the transcriptional start (Chanvivattana et al., 2002). RT-PCR analysis of *swn-1* revealed the presence of some *SWN* transcripts and so this does not represent a null allele (Chanvivattana, 2002). More recently identified *swn* mutants with T-DNA insertions located in exons, and therefore likely to be null, are still aphenotypic. However additional roles of *SWN* can be seen in the *clf swn* double mutants (Daniel Schubert pers. comm.). These mutants show severe developmental abnormalities; growth is slow and results in the production of undifferentiated organs. All organs are affected including the roots (figure 1.9).

These results demonstrate additional roles for both CLF and SWN in diverse aspects of development. Furthermore the two proteins show partial redundancy with respect to many of their targets. The molecular basis of the *clf swn* phenotype has yet to be investigated, based on the severity of the phenotype it is likely that many genes are deregulated in these mutants.

1.12 *Arabidopsis* Polycomb group genes function in reproductive development

Pollination in *Arabidopsis*, like other angiosperms, results in a double fertilisation event. The ovule contains the female gametophyte, which comprises the egg, central, synergid and antipodal cells (Reiser and Fischer 1993). All these cells are haploid with the exception of the central cell which contains two polar nuclei. One sperm nucleus fuses with the egg nucleus and the resulting diploid zygote gives rise to the diploid embryo, and a second sperm nucleus fuses with the homodiploid central cell nucleus to form the triploid endosperm nucleus. Both fertilisation products pass through distinct developmental patterns. In *Arabidopsis* the embryo passes through a series of morphologically defined stages, which have been termed the preglobular, globular, heart, cotyledon and mature stages. Endosperm development is initiated by a series of rapid and synchronous nuclear divisions that are not followed by cell divisions, to form a multinucleated syncytium. This syncytial endosperm then undergoes cellularisation into individual cells to produce a multicellular tissue which eventually degenerates to provide nourishment for the developing embryo (Sørenson et al., 2000). Fertilisation also initiates the development of the maternal tissues of the inner and outer integuments surrounding the embryo sac to become the seed coat, and the ovary to grow and form a fruit.

To understand the dependence of seed development on fertilisation and more specifically identify components of apomictic seed development, two groups performed genetic screens to identify mutants which are capable of initiating some aspects of seed development in the absence of fertilisation (Ohad et al., 1996; Chaudhury et al., 1997). Both screens were performed in a male-sterile mutant background and individuals were identified which showed silique development. These screens yielded several mutants defining three genes *FERTILISATION INDEPENDENT SEED 1* (*FIS1*), *FIS2* and *FIS3/FIE* (Ohad et al., 1996; Chaudhury et al., 1997).

An additional screen to identify mutants causing parent-of-origin effects on seed development identified the Polycomb group gene *MEA* (Grossniklaus et al., 1998) which

was later shown to be allelic to *FIS1* (Luo et al., 1999). A fourth *FIS* gene *MSI1* was identified based on homology with *Drosophila* p55 and the mutant phenotype subsequently characterised (Köhler et al., 2003a)

Mutations in any of these *FIS* genes have similar effects on autonomous endosperm development (Ohad et al., 1996; Chaudhury et al., 1997; Grossniklaus et al., 1998; Ohad et al., 1999; Köhler et al., 2003a). When fertilisation is prevented in plants heterozygous for one of these mutations, 50 % of ovules show no further development of embryo, endosperm or any other tissue associated with seed development post-fertilisation. However, up to 50 % of the remaining ovules show fertilisation-independent growth and develop into seed-like structures causing the silique to elongate. Closer examination of these structures reveals that the seed-like structures contain an endosperm but no embryo. In *fie* mutant ovules the endosperm develops up to the syncytium stage (Ohad et al., 1996; Ohad et al., 1999) and in *fis2* or *mea* mutant ovules the endosperm develops further and undergoes cellularisation (Chaudhury et al., 1997; Kiyosue et al., 1999). In addition to endosperm development in the absence of fertilisation, in all *fis* mutants the diploid integuments develop into a seed coat and the carpels elongate to form siliques.

In addition to their effects on autonomous endosperm development, the *fis* mutants also share another common phenotype, namely that they confer maternal effects on seed development. When fertilised, ovules inheriting a mutant *fis* allele abort at either the torpedo or heart stage of embryo development (Ohad et al., 1996; Chaudhury et al., 1997; Grossniklaus et al., 1998). The main effect is endosperm over-proliferation; the embryo arrests with fewer cells than the mature wild-type embryo (Sørenson et al., 2000). In all cases this abnormal development is unable to be corrected by the presence of a wild-type paternal allele.

These phenotypes reveal that the *FIS* genes play an important role by controlling central cell proliferation and endosperm development. Furthermore, they are required either

directly or indirectly in the female gametophyte to attain the normal levels of cell proliferation during embryogenesis.

Molecular cloning revealed that the three genes share extensive homology with the PRC2 group of Polycomb proteins. The *FIE* gene encodes a protein with about 40 % amino acid sequence identity to ESC from *Drosophila* and EED from mouse and humans (Ohad et al., 1999; Gutjahr et al., 1995; Sathe and Harte, 1995; Schumacher et al., 1996; Sewalt et al., 1998). Like ESC and EED, FIE contains seven predicted repeats of the WD motif (Ohad et al., 1999; Ng et al., 1997). These WD repeats are believed to form a β-propellor structure that functions in promoting protein-protein interactions (Ng et al., 1997).

The *MEA* gene encodes a protein similar to the *Drosophila* E(Z) protein and the *Arabidopsis* CLF and SWN proteins (Grossniklaus et al., 1998; Kiyosue et al., 1999). The highest degree of similarity (55 % amino acid identity) is found in the SET domain. MEA shows 43 % amino acid identity with CLF in the CXC domain and also contains the five additional highly conserved cysteine residues unique to E(Z) and its vertebrate and plant homologues (Grossniklaus et al., 1998; Goodrich et al., 1997).

The *FIS2* gene encodes a novel protein containing a C_2H_2 zinc finger signature (Luo et al., 1999). It was later found that the *Drosophila* SU(Z)12 protein and the *Arabidopsis* EMF2 and VRN2 proteins encode homologues of FIS2 (Birve et al., 2001). A conserved region, the VEFS domain, was defined as being unique to this group.

Studies of expression patterns confirmed that all *FIS* genes have overlapping expression patterns. *FIE* and *MEA* are expressed in the embryo sac, in each of the polar nuclei and in the central cell nucleus prior to fertilisation, and after fertilisation in both fertilisation products (Vielle Calzada et al., 1999; Spillane et al., 2000; Luo et al., 2000; Yadegeri et al., 2000). However, in addition, *FIE* is expressed persistently in additional sporophytic and gametophytic tissues including anthers and vegetative tissues (Ohad et al., 1999;

Spillane et al., 2000). The fact that they are expressed primarily in the endosperm and central cell is consistent with their mutant phenotypes. The expression pattern of *FIS2* is slightly different, as reporter gene analysis has shown it to be expressed transiently and specifically to the central cell and endosperm (Luo et al., 2000).

Based upon the overlap in expression patterns in the endosperm, the molecular identities and similar mutant phenotypes it was proposed that these genes act in a complex in the central cell to prevent transcription of target genes necessary to initiate the development of the endosperm. Yeast two-hybrid and *in-vitro* pull-down assays confirmed the interaction between FIE and MEA but were unable to detect an interaction between either FIE or MEA with FIS2 (Spillane et al., 2000; Yadegari et al., 2000; Luo et al., 2000). More recently FIE, MEA and MSI1 have been associated together *in-vivo* by co-immunoprecipitation in protein extracts from flowers and young siliques (Köhler et al., 2003a). This provided support for the presence of a Pc-G complex controlling seed development in *Arabidopsis*, although the exact role of *FIS2* is unknown.

As the FIS genes encode regulators, not cell cycle factors, they probably act on endosperm by regulating other genes. One direct target of the MEA-FIE complex has been identified by RNA profiling using high-density oligonucleotide microarrays as *PHERES1* (*PHE1*) (Köhler et al., 2003b). *PHE1* encodes a MADS-box protein which is expressed transiently in wild-type endosperm after fertilisation. However, *PHE1* is commonly deregulated in *mea*, *fie* or *fis2* mutants and is expressed at levels exceeding those in wild-type and more persistently until mutant embryos abort. In wild-type plants *PHE1* expression is never observed before fertilisation, however in *fie* mutants, *PHE1* expression can be detected in the absence of fertilisation shortly before the autonomous formation of endosperm nuclei. Reduced expression levels of *PHE1* in *mea* mutant seed can suppress *mea* seed abortion, indicating a key role for *PHE1* in *mea* seed abortion. The role for *PHE1* in wild-type seed is less clear, as there is no mutant phenotype associated with *phe1*, its function must be implied from expression patterns and the effects of mis-expression in *mea* seed.

The maternal effects of the *mea* mutation are at least partly due to *MEA* being imprinted in the endosperm (Vielle-Calzada et al., 1999; Kinoshita et al., 1999; Luo et al., 2000). Only maternal *MEA* expression can be detected in the endosperm of developing seed at the torpedo stage and later, whereas both parental *MEA* alleles are expressed in embryos at this stage (Kinoshita et al., 1999). Both maternal and paternal *FIE* can be detected in embryo and endosperm at the torpedo stage (Yadegari et al., 2000). This suggests that *MEA* is imprinted for longer than *FIE*. However, if there is an early requirement for FIE and MEA in the embryo, maternal effect lethality would be observed in both cases. It has been reported that paternal alleles of *FIS2* are silenced in the endosperm through the analysis of transgenic lines incorporating a GUS reporter construct (Luo et al., 2000). However, the degree to which the endogenous *FIS2* gene is imprinted has not been investigated

1.13 FIE also controls the transition from vegetative to floral development

Due to the maternal effect lethality of *fie* it had not been possible to investigate the role of *FIE* beyond seed development by analysing a homozygous mutant plant. However, based on its molecular identity and its status a single copy gene it was always considered a potential partner for *CLF*. Unlike other *FIS* genes *FIE* is expressed persistently in the whole plant (Ohad et al., 1999). It is known in animals that E(Z) only confers its biological activity towards histones as part of a larger complex containing ESC. Based on this evidence, a role alongside *CLF* and *EMF2* in the floral repression pathway seemed likely.

A role for FIE in vegetative development was demonstrated by the use of a transgenic line providing active *FIE* function in the endosperm to rescue the lethality of homozygous *fie* mutants (Kinoshita et al., 2001). The transgene lacked later expression, so revealing FIE^{+} function in later stages of development. In these *fie* mutants, the seedling shoot bypasses vegetative development and produces flower-like structures and organs. Flower-like structures were also observed on the hypocotyls and roots, organs

not normally associated with reproductive development. The phenotype appeared similar to those observed in strong *clf swn* doubles. The floral homeotic genes *AP1*, *LFY* and *AG* were deregulated in *fie* mutants; however like *emf* mutants the floral promoting genes *CO* and *FT* were unaffected (Kinoshita et al., 2001). A transgenic line where *FIE* expression was partially eliminated (by co-suppression) gave a weaker phenotype showing deregulation of *AG* and producing plants with morphological changes similar to *clf* mutants (Katz et al., 2004). Together these results suggest a role for *FIE* outside of seed development in repressing flowering and floral organ identity genes in the seedlings.

1.14 *Arabidopsis* Polycomb genes control the vernalization response

In many plant species including some *Arabidopsis* accessions the transition from vegetative growth to reproductive development is accelerated after a sustained period of cold temperature, known as vernalization. In the natural habitat this response ensures that plants flower only after a period of winter, therefore synchronising flowering between individuals and protecting delicate floral buds from forming before winter is over.

The molecular basis of this response to cold treatment was investigated through a genetic screen in *Arabidopsis* designed to identify mutants showing a reduced vernalization response in a genetic background (*fca*$^{-}$) showing a strong vernalization requirement (Chandler et al., 1996). Mutations were identified in three genes conferring altered vernalization responses, these were designated *vrn1-3*. Molecular cloning of the *VRN2* gene revealed that it encoded a protein showing homology with the SU(Z)12 family of Polycomb group proteins (Gendall et al., 2001). The structures of VRN2, EMF2, FIS2 and SU(Z)12 are similar in terms of conservation of a zinc finger and VEFS box. The molecular identity of VRN2 suggested that the response to vernalization may be regulated in a PRC2-like manner.

It has been shown that vernalization response is quantitative so that increased periods of cold cause progressively earlier flowering. Plants which are transferred from cold conditions to ambient temperatures do not immediately flower, demonstrating that an epigenetic memory of the cold treatment is created. Vernalization promotes flowering by decreasing levels of *FLC* transcript in *Arabidopsis*, (Sheldon et al., 1999; Gendall et al., 2001), this process acts antagonistically to FRI which increases the levels of *FLC* mRNA thereby delaying flowering (Michaels and Amasino, 1999).

By examining a profile of *FLC* mRNA expression in *fca-1* and *fca-1 vrn2-1* double mutants it was observed following a period of vernalization *FLC* levels were reduced in both *fca* single and *vrn2 fca* double mutants (Gendall et al., 2001). In *fca* mutants these low levels of *FLC* mRNA were maintained upon transfer to ambient conditions. In the *fca vrn2* double mutants on return to ambient conditions *FLC* mRNA levels began to rise and by 10-20 days were equivalent to those observed in non-vernalized individuals. These results indicated that VRN2 is not required for the regulation of *FLC* that occurs during a cold period, but rather for the stable repression of *FLC*, a trait commonly associated with the Pc-G genes (Gendall et al., 2001).

1.15 Modification of chromatin by Pc-G proteins in *Arabidopsis*

Although in animals, a biological function of the PRC2 complex has been demonstrated towards histones, there is little evidence of the effect that this has on the accessibility of target genes. This was demonstrated in plants by showing altered sensitivity to DNase I of the *FLC* locus in *fca* and *vrn2 fca* seedlings subjected to a four week period of vernalization (Gendall et al., 2001). Sensitivity of loci to DNase I is believed to be correlated with the chromatin organisation, so that loci which have been modified in such a way that they are inaccessible to transcription factors show reduced accessibility to DNase I. *FLC* sensitivity was enhanced in *vrn2* mutants, suggesting that VRN2 mediates changes in chromatin organisation of *FLC* following vernalization.

Although members of the PRC2 complex are conserved between plants and animals other components necessary for Polycomb-mediated silencing in *Drosophila* are absent from *Arabidopsis*. There are no PLEIOHOMEOTIC, ZESTE and GAGA homologues in *Arabidopsis* suggesting that new DNA-binding proteins and *cis*-acting sequence motifs have been recruited (Goodrich and Tweedie, 2002). It is also likely that plants have inherited different factors to give POLYCOMB-like function as there is no homologue in *Arabidopsis*. One aspect of Polycomb group function which is conserved between *Drosophila* and *Arabidopsis* is dependence of Pc-G complexes on methyltransferase activity (Bastow et al., 2004; Sung and Amasino, 2004). However, it would appear that plants have evolved new methods by which to interpret these methylation marks or that they may use different proteins.

Analysis of the histone methylation status of *FLC* in vernalized and non-vernalized plants using chromatin immunoprecipitation (ChIP) analysis has demonstrated a link between vernalization and chromatin modification. ChIP involves the isolation of chromatin from cells, followed by shearing of DNA to a constant size range and immunoprecipitation with specific antibodies, in this case against different histone modifications. Immunoprecipitates are examined by PCR for enrichment of specific regulatory regions specific to the gene of interest. Coupled with the identification and characterisation of new mutants showing reduced vernalization, this has greatly increased the understanding of the molecular mechanisms regulating the vernalization response.

VRN1 encodes a non-specific DNA binding protein with two B3 domains which like VRN2 is not responsible for the initial vernalization-mediated repression of *FLC* but rather acts to maintain *FLC* repression when plants are transferred to an ambient growth environment (Levy et al., 2002). Over-expression of the *VRN1* gene (under the control of the constitutive 35S promoter) results in plants which flower early without exposure to vernalization. This early flowering phenotype can be accelerated further by exposure to a period of vernalization; however, flowering is delayed under short-day conditions

indicating that the acceleration of flowering is independent of the photoperiod pathway. Additionally, 35S::*VRN1* plants have alterations in flower and silique morphology, suggesting that VRN1 also regulates processes outside vernalization. Expression of the floral integrator genes *SOC1* and *FT* is elevated in 35S::*VRN1* without vernalization. *VRN1* can therefore also act on a flowering time pathway not affecting *FLC* at least when mis-expressed.

The *VERNALIZATION INSENSITIVE 3* (*VIN3*) gene was recently identified through mutagenesis and subsequent phenotypic screen for mutants which were insensitive to vernalization (Sung and Amasino, 2004). VIN3 functions in the establishment of *FLC* repression during vernalization, unlike VRN1 and VRN2 which are involved in the maintenance of *FLC* repression (Sung and Amasino, 2004; Levy et al., 2002; Gendall et al., 2001). Expression studies show that *FLC* repression mediated by VIN3 acts upstream of VRN1 and VRN2 and acts as a marker for cold exposures of a period which are effective for vernalization (Sung and Amasino, 2004). *VIN3* expression is induced by cold exposure in *vrn1* and *vrn2* mutants and therefore its expression is not dependent on these genes (Sung and Amasino, 2004). This suggests that VIN3 is unlike VRN1 and VRN2 which are required for to maintain an epigenetic memory of vernalization, and instead VIN3 functions in the early establishment of *FLC* repression.

ChIP analysis revealed that the first intron and a region upstream of the transcriptional start of *FLC* showed decreased acetylation during vernalization, and that this decrease was maintained on transfer to ambient growth conditions (Sung and Amasino, 2004). In *vin3* mutants vernalization-mediated changes in acetylation of the chromatin environment surrounding *FLC* locus do not occur (Sung and Amasino, 2004). However, in *vrn1* and *vrn2* mutants, acetylation surrounding the *FLC* locus is decreased during vernalization, but this decrease is not maintained following transfer to warm growth conditions. These data indicate a role of VIN3 in establishing transient repression of *FLC* involving histone deacetylation.

Like homeotic gene repression in animals, the longer-term maintenance of *FLC* repression is associated with di-methylation of K9 and K27 residues on histone H3 (Bastow et al., 2004; Sung and Amasino, 2004). In wild-type, di-methylation the K9 and K27 residues of histone H3 at the *FLC* locus is gained during vernalization, and these marks are maintained after the subsequent transfer to ambient growth conditions (Bastow et al., 2004; Sung and Amasino, 2004). In the *vrn1*, *vrn2* and *vin3* the vernalization-mediated increase in K9 methylation does not occur. However, the vernalization-mediated increase in K27 methylation occurs in *vrn1* mutants but not *vrn2* or *vin3* mutants. This suggests that VRN1 may represent a novel specific plant Pc-G member, which is perhaps analogous to the animal PRC1, as it is not required for the K27 mark and may be required in the interpretation of these modifications.

Together these data suggest that VIN3 acts upstream of VRN1 and VRN2 to signal that an adequate exposure to cold has been provided. VIN3 is necessary to make a transient modification to histones via deacetylation (Sung and Amasino, 2004). In turn this leads to a more permanent modification involving a VRN2-containing complex marking histones with a di-methylation mark at K27 on H3 (Bastow et al., 2004; Sung and Amasino, 2004). This mark in turn could provide the vernalization specific factor necessary to recruit VRN1 on *FLC*. VRN1 may act as part of a complex mediating the long-term expression of *FLC* involving additional methylation of the K9 residue (Bastow et al., 2004).

1.16 Questions addressed in this thesis

At the start of my studies the tools existed to investigate the specificity of the Pc-G proteins in *Arabidopsis*. Completion of the *Arabidopsis* genome sequence had revealed that E(Z) was represented by a small gene family in *Arabidopsis*. Characterisation of the *mea* and *clf* phenotypes had shown that in *Arabidopsis* that the *E(Z)* homologues had discrete functions. During the course of my studies, two important advances were made. Firstly, a new family of *Arabidopsis* Pc-G genes (containing *EMF2* and *VRN2*) were

discovered. Secondly, the PRC2 complex was purified from *Drosophila* embryos and found to contain homologues of these new proteins. It had already been demonstrated that the ESC-E(Z) partnership was conserved between plants and animals. What remained uncertain was whether other partnerships in the PRC2 complex were conserved and what factors were responsible for the diversification of these *Arabidopsis* Pc-G genes into small families. I therefore set out to answer the following questions.

1 How have the different *Arabidopsis E(Z)* homologues acquired discrete functions? Is it due to changes in expression patterns alone, or have the proteins acquired discrete functions?

2 Are the VEFS box proteins part of the plant Pc-G complexes? With which Pc-G members do they interact? What is the significance of conserved regions of E(Z) and the VEFS proteins?

3 Do other Pc-G proteins act with VRN2 to mediate the vernalization response?

4 Can we dissect the function of FIE to reveal new roles post seed development, for example, in the repression of flowering or in the vernalization response?

2 Materials and Methods

2.1 Plant materials

clf-2

The *clf-2* mutant was derived from a transposon mutagenesis experiment in the Landsberg *erecta* background (Long et al., 1993). The *Ds* element is inserted at the junction between the first intron and second exon. Consequently a cryptic splice site is used and the mRNA carries a five nucleotide insertion.

clf-50

This allele was identified in a T-DNA mutagenesis screen in the Wassilewskija (Ws) background. Molecular evidence has shown that *CLF* and adjacent loci have been deleted in this line; therefore this represents a null allele. The *clf-50* line was provided by Dr. E. Huala.

swn-3, *swn-5* and *swn-7*

These mutations were all obtained from the Salk Institute Genome Analysis Laboratory (SIGnAL) collection and contain T-DNA inserts. All three are in the Columbia background and have inserts within exons. It is likely, although not proven, that these represent null alleles.

fis3-1

The *fis3-1* mutant was discovered in an EMS mutagenesis screen for mutants showing aspects of fertilisation-independent seed development in a male sterile background (Chaudhury et al., 1997). The mutation lies at the 3'-border of the third intron where a G is substituted by an A (Ohad et al., 1999). The mutation is presumably null as the phenotype is as severe as other mutations, including those encoding premature stop codons.

mea-2

The *mea-2* mutation was generated by a *Ds* transposable insertion into the CXC domain (Grossniklaus et al., 1998). This has left a seven bp footprint which introduces two stop codons upstream of the SET domain.

fca-1

The *fca-1* mutant was isolated from an EMS mutagenesis screen in the landsberg *erecta* background based on a late flowering phenotype (Korneef et al., 1991). Cloning of the *FCA* gene revealed that the *fca-1* mutant contained a polymorphism converting a glutamine codon to a stop codon in the thirteenth exon (Macknight et al., 1997).

fca-9

The *fca-9* mutant is in the Columbia ecotype and creates an alternative splicing product. The seed was a generous donation from Prof. Caroline Dean (JIC, Norwich).

Columbia *FRI*$^{+}$

A Columbia line carrying a copy of the dominant *FRI*$^{+}$ allele from the H51 ecotype from Stockholm was generously donated by Prof. Caroline Dean (JIC, Norwich). This line is late flowering without vernalization.

2.2 Growth and genetic analysis of *Arabidopsis*

2.2.1 Growing *Arabidopsis* in sterile culture and on soil

For growth in sterile tissue culture (used for selecting transgenic plants that carry antibiotic resistance genes), seed were first sterilised by exposure to chlorine vapour for 16 h. This was performed by mixing 100 ml sodium hypochlorite solution with 3 ml concentrated HCl in an air-tight bell jar. The sterile seed were sown on plates containing GM media and the appropriate antibiotics (kanamycin was used at 50 μg/ml and hygromycin B at 40μg/ml). The seed were stratified by placing at 4 °C for two days

to synchronise germination and then moved to growth chambers where they were grown at 23 °C under either short (12 h) or long day (16 h) conditions. To grow plants on soil, seed was simply sprinkled onto a 2:1:1 mixture of Levington F2 compost, sand and fine grit. These seed were also stratified as above, and were covered for seven days subsequently whilst the seed germinated.

2.2.2 Genetic crosses

The female partner was emasculated by the removal of sepals, petals and anthers from flowers just before the anthers reached maturity. Mature stamens from the paternal plant were brushed against the exposed carpel so that the papillae were yellow with pollen. This pollination was repeated the following day.

2.3 DNA analysis

2.5.2 Small scale plant DNA extraction (for PCR analysis)

100 mg of plant tissue was placed in 96 well collection microtubes with a 3 mm tungsten carbide bead in each. The tubes were frozen in liquid nitrogen and macerated twice using a Mixer Mill MM300 (Qiagen) for 45 secs at 30 Hz. 200 µl of Extraction Buffer (100 mM Tris.HCl pH 8, 1.4 M NaCl, 20 mM EDTA, 2 % CTAB, 0.2 % β-mercaptothanol) was added and the samples incubated at 65°C for 10 mins. The samples were then extracted once with chloroform, and the DNA precipitated with ethanol (see 2.4.6 DNA purification). Pellets were washed in 70 % ethanol and dissolved in 50 µl TE (10 M Tris.HCl pH 7.4, 10 mM EDTA). The Tris-hydrochloride buffers the solution and the EDTA chelates any Mg^{2+} ions in the solution, protecting the DNA against degradation by nucleases, most of which require magnesium.

2.5.3 Larger scale plant DNA extraction (for Southern analysis)

500 mg of young leaf tissue was collected into a 1.5 ml microcentrifuge tube. The tissue was then frozen in liquid nitrogen and ground manually with a small pestle, before

adding 500 µl chilled extraction buffer (50 mM EDTA, 100 mM NaCl, 100 mM Tris.HCl pH 8, 1 % SDS). The tissue was then further macerated using an Ultra Turrax-T8 homogeniser. A phenol/chloroform extraction was then performed twice (see 2.4.6 DNA purification). The DNA was precipitated with 0.1 volume 3 M NaOAc (pH 5.2) and 0.7 volume isopropanol, washed in 70 % ethanol, air dried and dissolved in 50 µl TE.

2.5.4 PCR analysis to determine plant genotype

The gene of interest was amplified using PCR, and the product was either sequenced or cut with a restriction endonuclease and analysed by gel electrophoresis to determine genotype. The following components were mixed on ice in a thin walled PCR tube and transferred to a MJ Systems DNA engine thermocycler preheated to 94 °C.

2.0 µl 10 x Magnesium-free PCR buffer (Promega)
2.0 µl $MgCl_2$ (2 mM)
0.4 µl dNTPs (10 mM)
0.2 µl Taq polymerase (5 u/µl Promega)
0.8 µl primer 1 (10 µM)
0.8 µl primer 2 (10 µM)
13.6 µl sterile water
1.0 µl DNA (approximately 50 ng of genomic template)

A typical cycle used was -

Step 1	94 °C	for 2 mins	
Step 2	94 °C	for 30 secs	Repeat 29 times
	55 °C	for 30 secs	
	72 °C	for 1 min	
Step 3	72 °C	for 5 mins	
Step 4	4 °C	for ever	

After PCR cycling reactions were run on a 1 % agarose gel.

2.3.4 DNA sequence analysis

PCR products were purified prior to sequencing using a QIAquick PCR Purification Kit (Qiagen). 10 ng of PCR product or 150 ng of plasmid template were mixed in a PCR tube with 4µl of Big Dye v3.1 (Applied Biosystems), 1 µl of primer (1.6 µM) and made

up to 10µl with sterile water. These reactions were then cycled using the following parameters.

Step 1	96 °C	1 min	
Step 2	96 °C	10 secs	⎫
	50 °C	5 secs	⎬ Repeat 24 times
	60 °C	4 mins	⎭
Step 3	4 °C	for ever	

The sequencing products were purified, precipitated and analysed on an ABI377 sequencer by Jill Lovell at the ICAPB sequencing service, Edinburgh University.

2.3.5 Digestion of DNA with restriction endonucleases

Single Digest

Enzymatic digestions were typically performed in 10µl reactions. The following ingredients were added to a microcentrifuge tube in the order stated.

Autoclaved distilled water	5.5 µl
Restriction enzyme 10 x buffer	1 µl
BSA, acetylated (1 mg/ml)	1 µl
DNA sample in TE	2 µl (containing 0.1-1 µg DNA)
Restriction enzyme 2-10 u	0.5 µl

The reaction was mixed by gentle pipetting and incubated at the recommended temperature for 1-3 h. Variations in the quantity of DNA added were compensated by varying the quantity of water. When large quantities of cut DNA were required, for example when purifying a particular fragment by gel electrophoresis, the reaction volume was increased to 50 µl.

Double Digests

The restriction enzyme buffer compatibility charts supplied by the manufacturer were consulted to compare the activities of the two enzymes in each buffer. When there was a suitable buffer that both enzymes had 100 % activity in, the reaction was set up exactly as above using 0.5 µl of both enzymes. When there was no suitable buffer sequential digests were performed. After digestion with the first enzyme, the cut DNA was purified by phenol/chloroform extraction and ethanol precipitation to completely remove

any enzyme. The pellet was dissolved in a buffer appropriate to the second digest and the new enzyme added.

2.3.6 Agarose gel electrophoresis

The required agarose was weighed (typically 0.7-2 % of final gel volume) and added to 0.5 x TBE (4 mM Tris.borate, 1 mM EDTA) buffer in a conical flask. The mixture was heated in a microwave oven for the minimum time for the agarose to completely dissolve, with frequent swirling. The mixture was cooled to approximately 60 °C and poured in a casting tray. The gel was allowed to set for 30 mins, then placed in a gel tank and covered with 0.5 x TBE buffer. The samples were mixed with 10 x loading buffer (0.25 % bromophenol blue, 0.25 % xylene cyanol FF, 15 % ficoll (type 400)) to a final concentration of 1x, loaded in the wells and the appropriate voltage applied (typically 80 v for 1 h). After electrophoresis was complete the gel was stained in 0.5 µg/ml ethidium bromide solution; the ethidium bromide intercalates between DNA base-pairs and fluoresces when activated with ultraviolet radiation allowing visualisation of DNA bands. The gel was placed on a UV light source and photographed.

2.3.7 Analysis of genomic DNA by Southern blotting

Southern blotting

Genomic DNA was first digested with restriction endonucleases. 15.5 µl of DNA (about 2 µg), 2.5 µl 1 M spermidine, 2.5 µl of buffer, 2.5 µl RNAase (200 µg/ml) and 2 µl (10-20 u) enzyme were mixed in a microcentrifuge tube and incubated at 37 °C for 3 h. When double digests were required, only 14.5 µl of DNA and 1.5 µl of each enzyme were used. The samples were then loaded on to a 0.7 % agarose gel and run at 35 v for 16 h to separate the fragments according to size. The gel was stained in ethidium bromide to visualise DNA and confirm digestion, and photographed. After trimming to a convenient size the gel was immersed twice in 0.25 M HCl for 15 mins, to partially depurinate the DNA which aids in the subsequent transfer of large DNA molecules. The gel was then immersed twice in denaturation solution (1.5 M NaCl, 0.5 M NaOH) for 15

mins to denature DNA prior to hybridisation, and twice in neutralisation solution (1.5 M NaCl, 0.5 M Tris.HCl pH 7.2, 1 mM EDTA) for 15 mins. DNA was transferred to nylon membrane (Hybond-N) by southern blotting (as shown in figure 2.1) and left overnight to allow capillary transfer. The following day the membrane was washed for 10 mins in 2 x SSC and DNA was crossed-linked to the membrane using a UV transilluminator at 40 $mJcm^{-2}$.

DNA labelling with Digoxigenin

Molecular probes were made by labelling DNA with digoxigenin in PCR reactions using dNTPs incorporating Digoxigenin (Dig). 50 µl reactions were prepared in the same way as PCR reactions using 5 µl of Dig labelling mix. Incorporation can be seen by slower migration of a small aliquot of Dig labelled product next to a standard reaction.

Hybridisation of DNA probes to Southern blots

To limit unspecific hybridisation of DNA probe to the membrane, it was first blocked by pre-hybridisation at 65 °C in 70 ml hybe buffer (5 x SSC, 0.1 % N-lauryl sarcosine, 0.02 % SDS and 1 % w/v milk powder) for 2 h with shaking. The probe was then denatured by boiling for 3 mins and hybridised to the membrane by incubation in 50 ml of hybe buffer at 65°C overnight. The next day the membrane was washed twice for 5 mins at room temperature in usual wash (2 x SSC, 0.1 % SDS), and twice in stringent wash (0.5 x SSC, 0.1 % SDS) at 68 °C.

Detection of Dig-labelled DNA

The membrane was then blocked in 50 ml maleic acid blocker (10 % w/v blocking reagent (Roche), 100 mM maleic acid, 150 mM, NaCl pH 7) for 1 h at room temperature to prevent unspecific binding of antibodies to the membrane, and incubated for 30 mins in 20 ml maleic acid blocker at room temperature with 2 µl anti-Dig-AP antibody. The membrane was washed twice for 15 mins at room temperature in 100 ml 1 x maleic acid buffer (100 mM maleic acid pH 7.5, 150 mM NaCl, 0.3 % tween,) and equilibrated in detection buffer for 3 mins at room temperature, wetted with the chemiluminescent

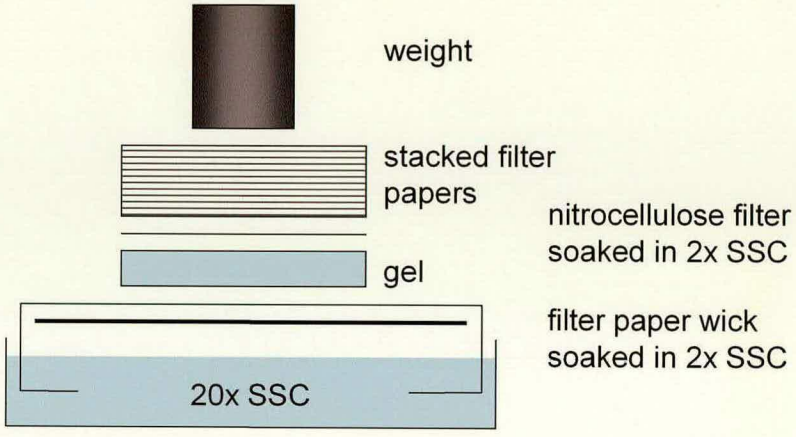


Figure 2.1 Southern blot apparatus

detection substrate, CSPD (250 μM in detection buffer), sealed in a bag and exposed for 2 h.

2.4 Cloning into plasmid vectors

2.4.1 Transformation of *E. coli* with plasmid DNA

DH5α competent cells were gently thawed on ice. 1-50 ng of plasmid DNA was placed in a 1.5 ml microcentrifuge tube, the quantity was increased for ligation and mutagenesis reactions. 100 μl of competent cells were gently mixed with the DNA and incubated on ice for 30 mins. The cells were then subjected to a heat-shock (45 secs in a 42 °C water bath) then incubation for 2 mins in an ice water bath. 0.5 ml of SOC medium was added and the cells incubated at 37 °C forl h with shaking at 200 rpm to allow for expression of antibiotic resistance genes. The cells were then pelleted by centrifugation at 10 000 x *g* for 10 secs, the supernatant removed and the cells resuspended in 50 μl SOC. The entire volume was then plated on LB agar plates with the appropriate antibiotic.

2.4.2 *E. coli* culture on plates

LB agar was autoclaved, cooled to approximately 55 °C, the appropriate antibiotics were added and the agar poured into petri dishes. When the agar had set *E. coli* cells were spread over the surface. The *E. coli* was then incubated at 37 °C for 16 h. Kanamycin was used at a final concentration of 50 μg/ml, and ampicillin and spectinomycin at 100 μg/ml.

On occasions where the screening of recombinant plasmids was required and the insertion of a fragment disrupted the *LacZ* gene, blue-white screening was used. The *LacZ* gene encodes a subunit of the enzyme β-galactosidase and is under the control of the *Lac* promoter. If the *E. coli* colony is expressing the *Lac* repressor then expression of the vector may be induced using IPTG and X-gal to yield a blue product. Insertional inactivation of *LacZ* in the production of a recombinant plasmid would prevent the development of a blue colour. In these cases the white colonies can be selected, which

have no β-galactosidase and are hence likely to contain the inserted target fragment. In these cases X-gal (40 µg/ml) and IPTG (0.1 mM) were included in the LB-agar plates.

2.4.3 *E. coli* culture in liquid medium

Antibiotics were used at the same concentrations as used for agar plates. Liquid cultures were incubated at 37 °C with shaking at 200 rpm; 5 ml-100 ml cultures were inoculated with a single colony and grown for 16 h.

2.4.4 *E. coli* plasmid preparation (small scale)

1.5 ml of bacterial culture was poured into a microcentrifuge tube and centrifuged at 12 000 x *g* for 1 min. The supernatant was removed and the bacterial pellet resuspended by vigorous vortexing in 150 µl resuspension buffer (5 M Tris.HCl pH 8.0, 10 mM EDTA pH 8.0, 100 µg/ml RNAase). 150 µl of lysis solution (0.2 N NaOH, 1 % SDS) was added and mixed by gentle inversion, followed by 200 µl of ice cold neutralisation buffer (3 M potassium acetate, 11.5 % glacial acetic acid). The tube was gently inverted to mix contents, centrifuged at 12 000 x *g* for 10 mins and the supernatant removed to a fresh tube.

The SDS disrupts the cell membrane and denatures the proteins; the alkaline conditions denature the DNA and begin hydrolysis of RNA. The neutralisation buffer then precipitates the denatured proteins, along with the chromosomal DNA and most of the detergent. The supernatant contains plasmid DNA, which, being small and closed-circular, is easily renatured after the alkaline treatment.

Crude DNA preparations were extracted using phenol/chloroform to remove contaminating proteins, and the DNA precipitated by addition of 2.5 volumes of ethanol and 0.3 volumes of 3 M sodium acetate. The precipitated DNA was then pelleted by

centrifugation (18 000 x *g* for 15 mins), the ethanol removed and the pellet air dried and dissolved in 50 µl TE.

2.4.5 *E. coli* plasmid purification (large scale)

Large scale preparations (50 ml bacterial cultures) were made using the Qiagen midi plasmid preparation kit following the manufacturer's instructions.

2.4.6 DNA purification (phenol/chloroform extraction)

DNA solutions were made up to a more manageable 0.5 ml with distilled water in a microcentrifuge tube and an equal volume of phenol/chloroform/isoamyl alcohol (25/24/1) added. The mixture was vortexed for 1 min, allowed to stand for 5 mins, vortexed for an additional 1 min and centrifuged at 12 000 x *g* for 5 mins. The aqueous phase was removed to a fresh tube and 1 ml of 100 % ethanol was added. The sample was placed on ice for 20 mins, and then centrifuged at 14 000 x *g* for 20 mins. The supernatant was then removed, 200µl of 70 % ethanol added, this was centrifuged at 12 000 x *g* for 10 mins and the 70 % ethanol removed. The DNA pellet was dried in a vacuum desiccator, dissolved in TE and stored at -20°C.

2.4.7 Partial digestion of DNA

On occasions it was necessary to excise a section of DNA from a plasmid using a restriction endonuclease which could cut at multiple sites within the fragment. In these cases it was desirable to limit the effectiveness with which the enzyme cut, so as to achieve only one cleavage per molecule. In these cases ethidium bromide was added to the digest at final concentrations of 0.02, 0.01, 0.005 and 0.002 mg/ml. After the plasmid was cleaved and hence linearised the ethidium bromide was able to intercalate with the DNA. This has the effect of slightly unwinding the helix, increasing the distance between adjacent base-pairs and therefore preventing recognition by restriction endonucleases. Incubation times were decreased to 20 mins. When digestion with a second enzyme was required the DNA was purified and re-cut.

2.4.8 Site-directed *in-vitro* mutagenesis to introduce novel restriction sites

Sometimes it was necessary to insert a novel restriction site within DNA when no suitable site existed, or when a fragment of DNA had to be excised in an exact place or in a particular reading frame. In these cases site-directed *in-vitro* mutagenesis (S-D IVM) was used. This method has the advantage that PCR-induced mutations are unlikely as all molecules amplified are copied from the original template.

Two complimentary mutagenic oligonucleotide primers were designed containing the desired mutation flanked by unmodified nucleotide sequence. Primers were between 25 and 40 bp in length and had a melting temperature (T_m) above 78 °C. The equation used to calculate T_m was:

$T_m = 81.5 + 0.41$ (%GC) - 675/N - % mismatch

N is the primer length in bases. Values for % GC and % mismatch are whole numbers.

The reaction was set up in a thin walled PCR tube as follows:

5 µl 10 x *Pfu Turbo* buffer (Stratagene)
10 ng dsDNA template
1.25 µl oligonucleotide mutagenic primer #1 (100 ng/µl)
1.25 µl oligonucleotide mutagenic primer #2 (100 ng/µl)
1 µl dNTP mix (10 mM for each NTP)
dH_2O to 49 µl
and then 1 µl of *Pfu Turbo* polymerase was added (5 u/µl)

The reaction was then cycled in a MJ systems DNA engine using the following programme:

Step 1	95 °C	for 1 min	
Step 2	95 °C	for 50 secs	Repeat 17 times
	60 °C	for 50 secs	
	68 °C	for 12 mins*	
Step 3	68 °C	for 7min	
Step 4	4 °C	for ever	
		*2 mins per kb of plasmid	

Following temperature cycling 1 µl of *Dpn* I (NEB) restriction enzyme was added directly to the reaction and incubated at 37 °C for 3 h. *Dpn* I is a unique restriction endonuclease which cuts only when the recognition sequence is methylated; therefore the bacterially derived template is digested, leaving only the non-methylated amplified product. The reaction was then transformed into competent *E. coli* cells.

2.4.9 Converting a 5'-overhang to a blunt ended terminus

Where possible when two DNA molecules were cut and then joined together, restriction endonucleases which produced cohesive ends were selected. These can bind to any other end with the same overhanging sequence by annealing of the single-stranded tails, and may subsequently be covalently joined by ligation (see following sections). However this was not always possible, on these occasions 5'-overhangs could be converted to blunt ended termini by filling in the overhangs with dNTPs using the Klenow enzyme (DNA polymerase I large-fragment). These blunt ended termini may then be covalently joined with other blunt ended termini, for example those produced by restriction enzymes such as *Sma* I or by amplification by certain proof reading polymerases, such as *Tgo* polymerase.

The DNA was initially digested with the relevant restriction endonuclease, purified by phenol/chloroform extraction, ethanol precipitation and dissolved in 20 µl 1 x Klenow buffer (50 mM Tris-HCl pH 7.2, 10 mM $MgSO_4$, 100 µM DTT, 40 µM of each dNTP and 0.1 mg/ml of acetylated BSA). 1 u of Klenow (DNA polymerase I large-fragment) was added per µg of DNA and the reaction incubated for 10 mins at room temperature.

The reaction was stopped by incubation at 75 °C for 10 mins, to denature the Klenow enzyme.

2.4.10 Dephosphorylation of 5'-ends

To prevent ligation of vector sequences without the addition of recombinant DNA, the 5'-ends of the vector molecules can be modified. Alkaline phosphotase removes the phosphate group from these ends, meaning that the linear vector will be unable to self-ligate, since no phosphates are available for the ligation. A ligation with a target DNA insert can still proceed, since one phosphate is present to ligate one strand at each cut site

Following digestion with restriction endonuclese, the DNA was purified, and dissolved in 50 µl of 1 x calf intestinal alkaline phosphotase (CIAP) buffer (50 mM Tris-HCl pH 9.3, 1 mM $MgCl_2$, 100 µM $ZnCl_2$ and 1 mM spermadine). 0.01 u of CIAP enzyme was added for each pmol of ends to be dephosphorylated and the reaction incubated at 37 °C for 30 mins. Another 0.01 u of CIAP/pmol ends was added and incubated for a further 30 mins. To stop the reaction 5 µl 0.5 M EGTA was added and the reaction heated at 65 °C for 20 mins. The product was then purified with phenol/chloroform and ethanol precipitated.

A general formula for calculating pmol of ends of linear double-stranded DNA is:

(µg DNA/ length of DNA (kb)) x 3.04 = pmol ends.

2.4.11 Gel purification of DNA fragments

When a particular fragment of DNA was required, it was separated on TAE (4 mM Tris.borate, 1 mM EDTA) gels and stained with ethidium bromide. The band of interest was excised with a clean scalpel, and DNA was extracted using a Qiagen qiaex II kit following the manufacturer's instructions.

2.1.12 Ligation of plasmid vector and insert DNA

The concentration of digested vector and insert DNA was estimated by agarose gel electrophoresis with molecular weight standards. A 10 µl reaction was set up containing a roughly 1:3 molar ratio of vector:insert mixed in 1 x ligation buffer (30 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 10 mM ATP) with 1 u of T4 DNA ligase. This was incubated at 16 °C overnight and introduced into *E. coli* DH5α by transformation.

2.4.13 Assembly of promoter swap constructs

Constructs were built using the pART7 and pART27 vector systems (Gleave, 1992). The expression cartridge of the primary cloning vector, pART7, comprises the 35S promoter of the cauliflower mosaic virus and the octopine synthase gene (*ocs*) 3'-UTR, between which lies a multiple cloning site (MCS) containing eight unique restriction sites (*Xho* I, *Eco*R I, *Asp* 718, *Sma* I, *Hind* III, *Cla* I, *Bam*H I and *Xba* I) for cloning the gene of interest downstream of the 35S promoter. This design allows the initiation of transcription at the promoter's natural mRNA start, and the absence of any ATG translational initiation signals in the MCS ensures that translation will begin at the first ATG of the cloned DNA. The ocs 3'-UTR then acts to terminate transcription. The entire expression cartridge (35S-mcs-*ocs* 3') is flanked by *Not* I sites to shuttle into the binary vector pART27. In addition pART7 also contains the f1 origin of replication and Amp^R gene to allow replication and selection by ampicillin in *E. coli*.

The binary vector pART27 contains the T-DNA comprising right border sequences, followed by the *LacZ* gene (with *Not* I site into which the expression cartridge is cloned), a chimaeric plant kanamycin selectable marker and finally the left border. The insertion of the expression cartridge directly into the *LacZ* gene allows blue/white screening of pART27 recombinants. The vector also contains origins of replication for

both *E. coli* and *Agrobacteria* as well as the Tn7 spectinomycin/streptomycin resistance gene.

35S::*MEA*

Full length *MEA* cDNA was excised from the plasmid pKS::*MEA* (kind gift of Prof. U. Grossniklaus), by digestion with *Eco*R I and *Asp* 718. This was cloned into pART7 using the same restriction sites to form pART7::*MEA*. The 35S promoter-*MEA* cDNA-*ocs* cartridge was cut out using *Not* I and cloned into pART27 which had previously been cut with *Not* I and had the 5'-ends de-phosphorylated (figure 2.2). The resulting binary construct, pART27::*MEA*, was then introduced into the *Agrobacterium tumefaciens* strain GV3101.

35S::*CLF*

Full length *CLF* cDNA was cut from the plasmid p*CLF*mutA4 by digestion with *Asp* 718 and *Bam*H I. This was fragment was cloned into pART7 which had previously been cut using the same restriction enzymes to give pART7::*CLF*. The 35S promoter-*CLF* cDNA-*ocs* cartridge was subcloned into pART27 after digestion with *Not* I.

35S::*SWN*

A full length *SWN* cDNA clone (pda05864) was obtained from Riken Science Centre (Seki et al., 2002). An *Eco*R I site was introduced upstream of the ATG start codon by site-directed *in-vitro* mutagenesis using the primers AB-EcoRIswn5f and AB-EcoRIswn5r to make p*SWN*mut1. The *SWN* cDNA was then excised using a partial digest with *Bam*H I followed by complete digestion with *Eco*R I. The plasmid p*SWN*mut1 contained two *BamH* I sites, one within the cDNA and the other in the vector cloning site. The 2.7 kb fragment corresponding to full length *SWN* was separated by gel electrophoresis, purified and then cloned into *Eco*R I and *Bam*H I cut pART7 to form pART7::*SWN*. The 35S-*SWN-ocs* cartridge was finally subcloned as a *Not* I fragment into pART27 to form pART27::*SWN*.

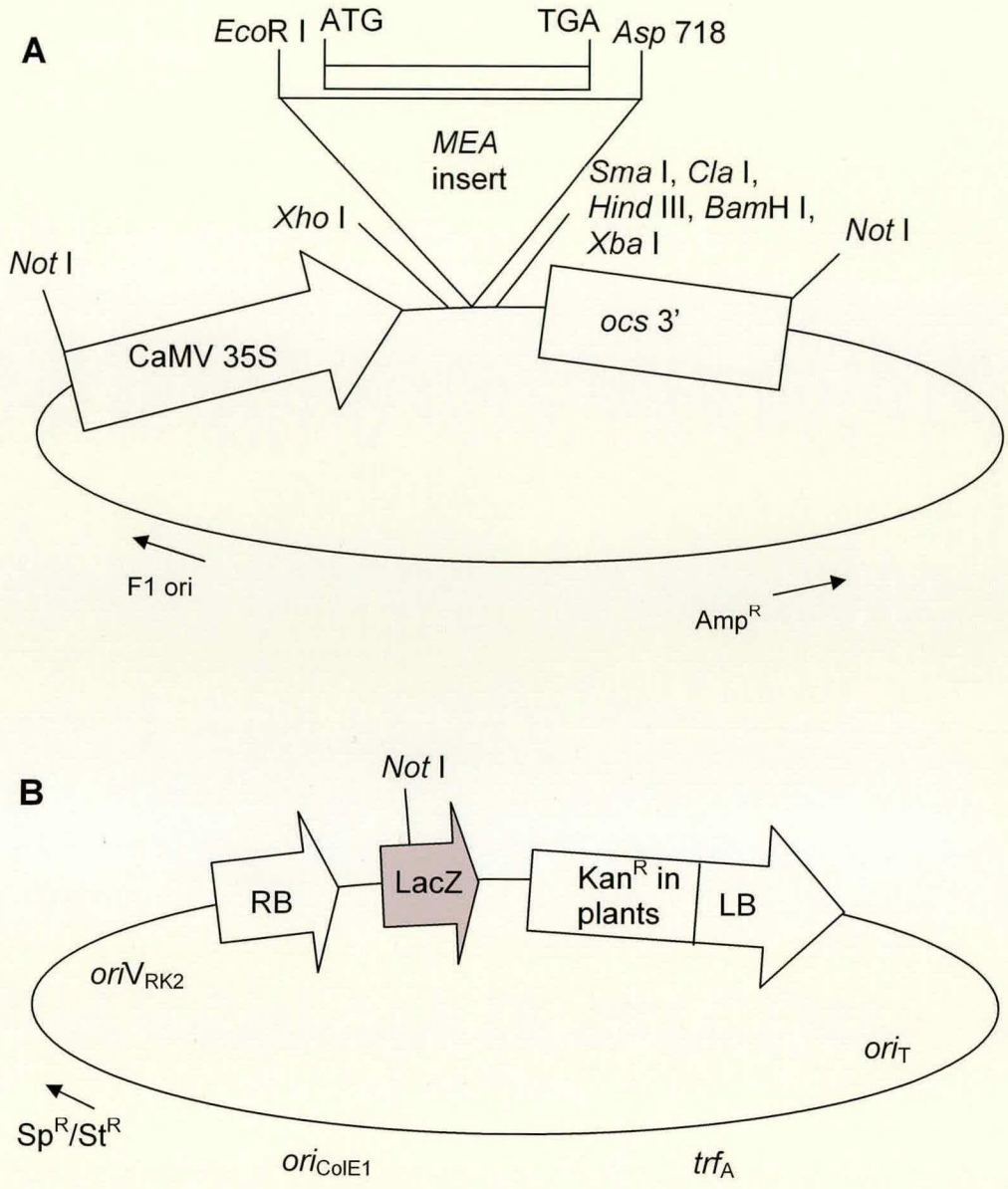


Figure 2.2 35S::*MEA* constructs

A. Full length *MEA* cDNA has been cloned into the MCS of pART7. The entire 35S cartridge is removed by digestion with *Not* I and inserted into the *LacZ* gene of pART27 (shown in B). The T-DNA inserted into plants is that from the right boarder (RB) to the left boarder (LB). Origins of replication and antibiotic resistance genes for both *E. coli* and *Agrobacteria* have been included in both diagrams.

p(*CLF*)::*CLF*, p(*CLF*)::*MEA* and p(*CLF*)::*SWN*

These constructs were made identically to those with the 35S promoter but using the vector pART7AB instead of pART7. pART7AB was a derivative of pART7 in which the 35S promoter had been replaced with a 1.4 kb section of genomic DNA upstream from the *CLF* transcriptional start and was made by Lucia Primavesi.

2.4.13 Assembly of a steroid-inducible *FIE* construct

The plasmid pBSgen*FIE* contained an 11 kb *Pst* I fragment of genomic DNA spanning the *FIE* locus, inserted into the *Pst* I site of pBluescript KS (Stratagene). In addition to containing the *FIE* coding region including introns, it also contained 4 kb of DNA upstream of the *FIE* transcriptional start and 3 kb of DNA 3'-of the stop codon. This was believed to be sufficient to initiate *FIE* expression *in-vivo*.

In order to render *FIE* activity under the control of the mammalian steroid dexamethasone (DEX), a translational fusion protein was created with the rat glucocorticoid receptor (GR). The GR-*FIE* complex exists in the cytoplasm with the heatshock protein Hsp90 and is unable to enter the nucleus. When DEX traverses the plasma membrane by diffusion and binds to the GR, this ligand binding leads to the disassociation of Hsp90 and allows the rapid localisation of the liganded form of GR-*FIE* into the nucleus (Yamamoto, 1985).

The GR insert was kindly provided by Dr. Robert Sablowski (pRS202). It already had a *Bam*H I site at the 5'-terminus; however it was further modified by the addition of a second *Bam*H I site at the 3'-terminus by S-D IVM using the primers AB-GRBam1f and AB-GRBam1r. This could then be cloned into pBSgen*FIE* as a *Bam*H I fragment. However modification of the pBSgen*FIE* plasmid was also required. This firstly involved removing an existing *Bam*H I site in the pBS polylinker; this was achieved by IVM using the primers AB-KOBam1 and AB-KOBam2 (figure 2.3). S-D IVM was also used to insert a new in *Bam*H I site (using the primers AB-FIEBam1F and 1R) just

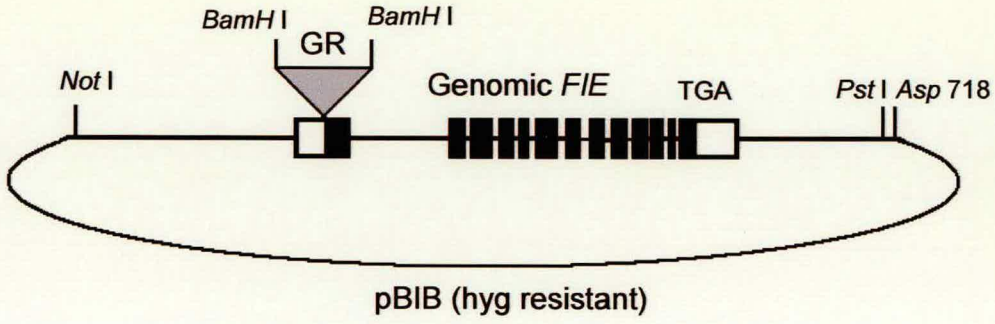


Figure 2.3 Steroid Inducible *FIE* construct
Genomic *FIE* locus shown by black horizontal line
Black boxes represent exons
White boxes represent untranslated region

downstream of the *FIE* transcriptional start in the modified pBSgen*FIE* into which the GR fragment could be cloned to form pBSgen*FIE*mut2. The locations of these sites were carefully designed so the GR-*FIE* fusion would be in the same reading frame relative to the *FIE* DNA and provide a continuous fusion protein. The modified pBSgen*FIE*mut2 vector was then linearised by digestion with *Bam*H I and the 5'-ends modified by removal of phosphate groups to prevent self ligation. The GR fragment was then cloned in as a *Bam*H I fragment and the recombinant colonies sequenced to ensure the correct orientation of the GR insert.

The GR-*FIE* genomic fragment was then excised from pBS by digestion with *Not* I and *Asp* 718 and cloned into the binary vector pBIB hyg (Becker, 1990). This allowed introduction into *Agrobacteria* and finally plants, using hygromycin as a selectable in plants.

2.5 Construction of chimaeric fusion genes by Overlap Extension PCR

2.5.1 Principle of OE-PCR technique

This technique was used to make chimaeric fusion proteins and is summarised in figure 2.4. Two preliminary PCR reactions were performed to amplify two non-contiguous regions of DNA one region from *CLF* the other from *MEA*. The example shown in the figure amplifies one fragment corresponding to the 5'-half of *MEA*, and the other to the 3'-half of *CLF*. The primers AB7 and AB8 were specifically designed to contain 5'-extensions which are homologous to a portion of the other target gene. PCR was performed using *Tgo* (Roche) proof reading polymerase. The manufacture's guidelines were modified to reduce the chance of PCR-induced mutations. This was achieved by increasing the reaction volume to 100 µl and the quantity of template used by 10-20 fold. The number of cycles was reduced to 12. This had the effect of increasing the proportion of amplified fragments that were copied from the original template.

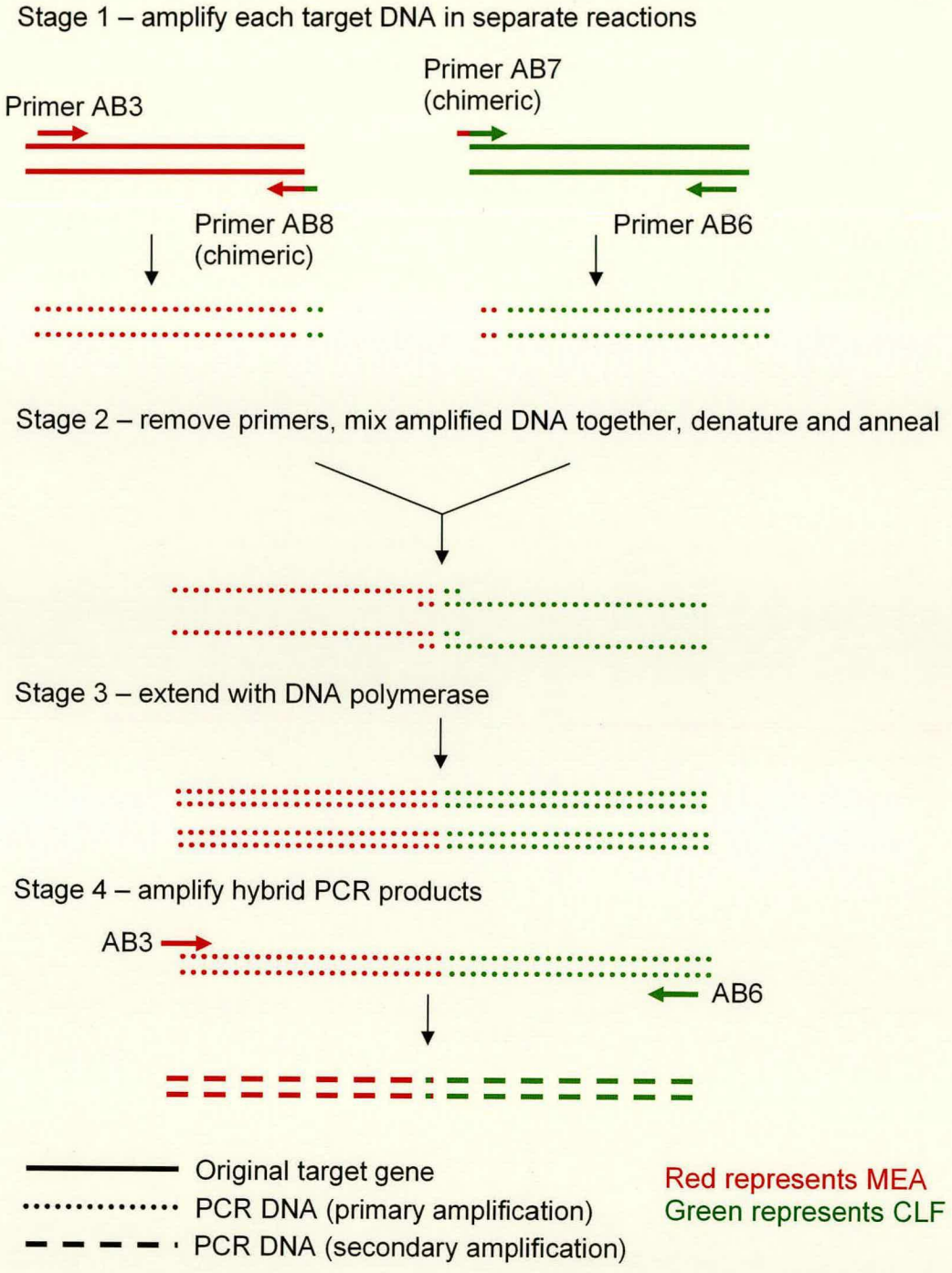


Figure 2.4 Construction of a recombinant DNA molecule by OE-PCR
The creation of a fusion chimaera between *CLF* and *MEA* cDNA was a 4 stage process. For a complete description see text or Ausubel et al. (1996).

The two PCR products were purified from template and primer by gel electrophoresis separation, the bands excised and the DNA purified. Because the primers AB7 and AB8 contain complementary 5'-extensions, the two PCR products contain a region of overlapping homology and can be annealed by denaturation and renaturation. Three different DNA species are generated in this reaction: a heteroduplex associating at the region of overlapping homology between the two parental sequences and two parental homoduplexes. The recessed ends of the heteroduplexes were extended with DNA polymerase to produce a single fragment that is equal in length to the sum of the two overlapping fragments. In a second PCR reaction, the combined heteroduplex was amplified using the outside primers AB3 and AB6. These primers will have complete homology with the fusion heteroduplex, but will fail to amplify either of the parental homoduplexes because only one of outside primer will anneal to each parental homoduplex.

In practice it was found that the extension reaction with DNA polymerase and the second amplification could be merged into one reaction. This was performed by setting up the PCR reaction without primers. This reaction was cycled five times, the outside primers added and the reaction cycled a further 14 times. The recombinant fusion product could then be cut with restriction enzymes and cloned into plasmid vectors. After cloning, the chimaeric constructs were sequenced to ensure there were no PCR-induced mutations.

2.5.2 Construction of MEDEA-CURLY LEAF protein chimaeras

Four constructs were made encoding chimaeric fusion proteins by the OE-PCR method. These were cloned into pART27 for transformation into *Arabidopsis*. Details of protein fusions can be seen in figure 2.5.

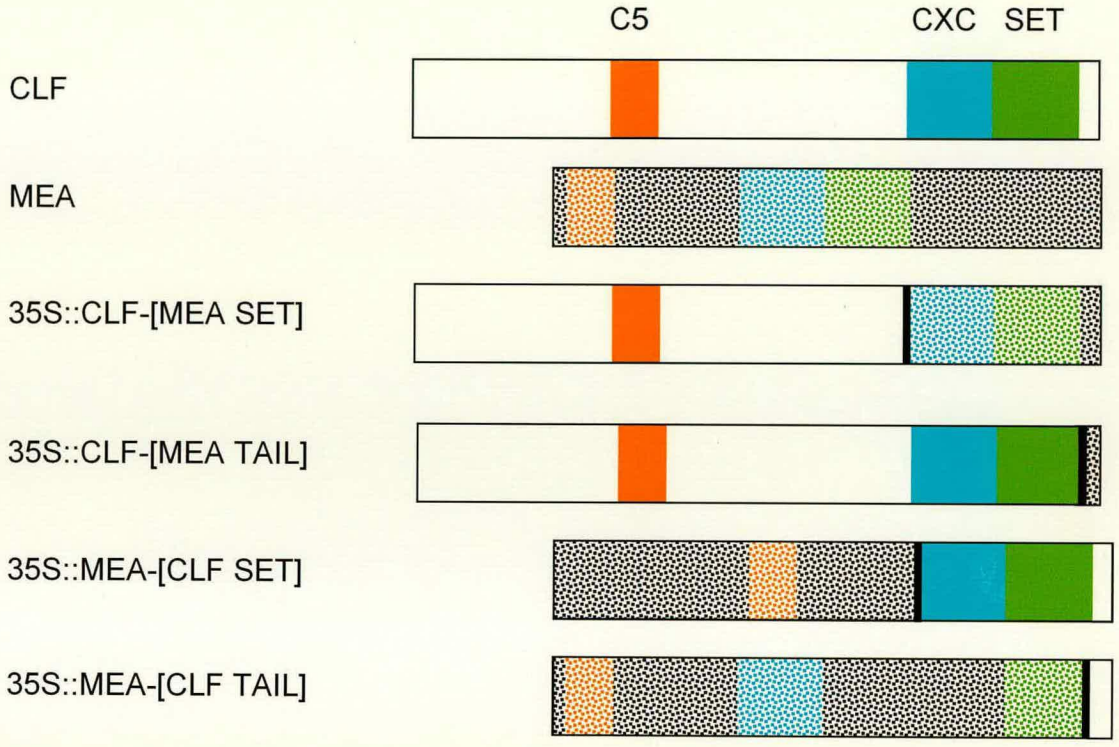


Figure 2.5 Chimaeric protein fusions between MEA and CLF
MEA is shown by patterned boxes and CLF by plain boxes. The heavy black line denotes junctions between CLF and MEA.

35S::*MEA*-[*CLF* SET]

A 5'-fragment of *MEA* was amplified from pKS::*MEA* using the primers AB3 and AB8. A fragment corresponding to the 3'-half of *CLF* was amplified from the cDNA clone using the primers AB7 and AB8. These purified fragments were then used as the template in an OE-PCR reaction with the primers AB3 and AB6. The chimaeric PCR product was then cut with *Bgl* I, gel purified and cloned into pART7::*MEA* cut with *Bgl* I and *Sma* I to form pART7::MCS. Because the *Tgo* DNA polymerase leaves a blunt end it was possible to ligate this directly to *Sma* I-cut vector. The 35S-chimaera-terminator fragment was then subcloned into pART27 to form pART27::*MEA*-[*CLF* SET].

35S::*MEA*-[*CLF* TAIL]

A 5'-*MEA* fragment was amplified using AB3 and AB12 and a 3'-*CLF* fragment with the primers AB11 and AB6. These were fused in an OE-PCR reaction with the primers AB3 and AB12. The fusion protein was cut with *Cla* I and cloned into pART7::*MEA* that had been cut with *Cla* I and *Sma* I to form pART7::*MEA*-[*CLF* TAIL]. This construct was sequenced and subcloned into pART27 to form pART27::*MEA*-[*CLF* TAIL]. .

35S::*CLF*-[*MEA* SET]

Initial PCR reactions amplified a 5'-region of *CLF* (AB5 and AB10) and a 3'-fragment of *MEA* (AB9 and AB4). These were fused and amplified with AB5 and AB4. The product cut with *Xho* I and *Bam*H I and cloned into pART7::*CLF*, followed by subcloning to pART27.

35S::*CLF*-[*MEA* TAIL]

AB5 and AB14 were used to amplify 5'-*CLF* and AB13 and AB4 for 3'-*MEA*. Fusion was followed by amplification with AB5 and AB4. The fusion protein was cut with *Xho* I and *Bam*H I, cloned into pART7::*CLF* and subcloned into pART27.

2.6 Gene transfer into *Arabidopsis*

2.6.1 Culture and use of *Agrobacteria*

DNA was transferred to *Arabidopsis* via the bacterium *Agrobacterium tumefaciens*. This is a soil dwelling organism which transforms normal plant cells into tumour-forming cells through the insertion of a piece of bacterial DNA (the transfer or T-DNA) into the plant genome. Normally *Agrobacteria* carry a tumour inducing plasmid (Ti) which contains the T-DNA sequence flanked by repetitive left and right border sequences. In addition to this the Ti also carries many transfer functions necessary for mobilising the T-DNA. *Agrobacterium* strain GV3101 was used in this study with binary vectors pART27 and pBIB. This strain carries the pMP90 plasmid, a disarmed Ti which does not harbour the T-DNA sequence (Koncs and Schell, 1986). Instead the T-DNA resides on a separate plasmid, the binary vector (either pART27 or pBIB-HYG). The advantage of using this system is the ease in which plasmids of relatively modest size (ca. 10 kb) can be manipulated in *E. coli* prior to transfer to *Agrobacteria*.

Agrobacteria were grown in a similar way to *E. coli*, however, YEP medium was used instead of LB, and cultures were incubated at 28 °C. Kanamycin (selects for pBIB plasmid), gentamycin (selects for pMP90 plasmid) and spectinomycin (selects for pART27 plasmid) were all used at 50 μg/ml and rifampicin at 100 μg/ml (prevents the growth of *E. coli*). Growth was slower than *E. coli* with cells taking 2-3 days to form visible colonies.

2.6.2 Transformation of *Agrobacteria* with plasmid DNA

Preparation of competent cells

A single colony of *Agrobacteria* was grown in 2 ml of YEP at 28 °C for 16 h. This was used to inoculate a 50 ml culture which was grown to an OD_{600} of about 0.5. The cells were centrifuged for 5 mins at 3 500 x *g*, washed in 10 ml 0.15 M NaCl and resuspended in a final volume of 1 ml of ice-cold 20 mM $CaCl_2$. Cells were then dispensed into

200µl aliquots and those that were not required were frozen in liquid nitrogen and stored at -80 °C.

Transformation with plasmid DNA

1µg of DNA was added to 200µl of competent cells and the mixture incubated on ice for 30 mins. The cells were frozen for 1 min in liquid nitrogen and thawed in a 37 °C water bath. 1 ml of YEP liquid medium was added to cells and they were incubated at 28 °C for 2-4 h with shaking. The cells were then plated on YEP agar plates containing the appropriate antibiotics.

2.6.3 Plasmid preparation from *Agrobacteria*

A 5 ml culture was grown overnight, and cells were harvested from a 1.5 ml aliquot by centrifugation at 12 000 x *g* for 3 mins in a microcentrifuge tube. The bacterial pellet was resuspended in 100 µl of resuspension buffer (5 M Tris.HCl pH 8.0, 10 mM EDTA pH 8.0, 100 µg/ml RNAase, 10 % lysozyme) and incubated at room temperature for 30 mins. 200 µl of lysis buffer (0.2 N NaOH, 1 % SDS) was added; the solution was mixed gently and incubated on ice for 5 mins. 150 µl of ice-cold neutralisation buffer added (3 M potassium acetate, 11.5 % glacial acetic acid), the solution was vortexed upside down for 10 secs and incubated on ice for a further 5 mins. Denatured proteins and chromosomal DNA were precipitated by alkaline lysis and pelleted by centrifugation at 12 000 x *g* for 5 mins. The supernatant was transferred to a fresh tube, DNA precipitated with 350 µl of isopropanol, washed in 70 % ethanol and dissolved in 10 µl TE. A 3 µl aliquot was cut with diagnostic restriction endonucleases and analysed by agarose gel electrophoresis to confirm the *Agrobacteria* contained the correct plasmid.

2.6.4 Floral dip transformation of *Arabidopsis*

Transgenes were introduced into *Arabidopsis* using the floral dip method. 1 L cultures of recombinant *Agrobacteria* were grown to stationary phase. These were spun down at 3 500 x *g* for 20 mins and the pellet resuspended in 500 ml 0.5 x MS salts (Sigma) with 0.02 % silwet and 5 % sucrose. Inflorescences were dipped into this bacteria suspension for 30 secs. This dipping was repeated at a one week interval.

2.7 RNA analysis

2.7.1 Plant RNA extraction

Plant tissue was harvested and immediately frozen in liquid nitrogen. 50 mg of seedlings were placed in a microcentrifuge tube with 1 ml of Trizol reagent (Gibco BRL). The tissue was macerated using a homogeniser. The samples were centrifuged for 2 mins at 18 000 x *g* and the supernatant removed to a fresh tube. 200 µl of chloroform was added, the samples vortexed for 15 secs and then spun at 18 000 x *g* for 15 mins at 4 °C. The aqueous phase was removed to a fresh tube, mixed with 500 µl of isopropanol and finally RNA precipitated by centrifugation for 10 mins at 18 000 x *g*. The supernatant was removed and the pellet washed with 70 % ethanol. The RNA pellet was dried and dissolved in water that had been treated with DEPC to inactivate RNAase.

2.7.2 Agarose gel electrophoresis of RNA

Agarose gels containing formaldehyde were prepared in the fume hood. 1 g of agarose was mixed with 10 ml 10 x MOPS buffer (200 mM MOPS, 50 mM NaOAc, 10 mM EDTA, pH 7) and 72 ml water. This was heated in a microwave oven until the agarose had dissolved. The molten solution and a small bottle of 37 % formaldehyde were placed in an oven at 60 °C to equilibrate. When both liquids were at 60 °C, 18 ml of hot formaldehyde was added to the agarose solution, mixed and the gel cast in the fume hood for 1 h. The RNA samples were denatured by mixing 7.5µl of sample RNA (5-15

µg) with 8.25 µl deionised formamide, 2.5 µl 10 x MOPS buffer, and 1 µl 37 % formaldehyde. These were heated to 80 °C for 10 mins and then transferred to ice. 0.75 µl of ethidium bromide (10 mg/ml) and 5 µl 5 x loading dye (0.25 % bromophenol blue, 0.25 % xylene cyanol FF, 50 % glycerol, 1 mM EDTA, pH 8.0) were added. The samples were loaded and the gel run in 1 x MOPS buffer initially at 100 v for 5 mins then at 20 v for 16 hours.

2.7.3 RNA transfer by Northern blotting

RNA was transferred directly to Hybond-N nylon membrane by capillary action as performed for Southern blots. After blotting the RNA was cross-linked to the membrane as for Southern blots.

Radiolabelling of DNA probes with ^{32}P dCTP

A cDNA fragment of the gene of interest was excised from a plasmid vector, and purified by gel electrophoresis. Approximately 10-50 ng of DNA was diluted in water to a final volume of 20 µl, boiled for 5 mins to denature the DNA, then cooled on ice for 10 mins and used as template for probe synthesis. A 50 µl reaction was assembled comprising 10 µl 5 x labelling buffer (250 mM Tris.HCl pH 6.9, 50 mM $MgSO_4$, 500 µM DTT, 300 µM dATP, dTTP and dGTP), 3 µl ^{32}P dCTP (approximately 30 µCi) to allow synthesis of radiolabelled DNA, 1 µl of random hexamer primers (3 µg/µl) and 10 u Klenow polymerase. This was incubated at 37°C for 1 h. To separate unincorporated dNTPs from the probe, the labelling reaction was made up to 100 µl with TE, added to 15 µl of nucleotide separation dye (1 % dextran blue, 0.1 % orange G in TE) and passed over a column containing sephadex G50 beads soaked in TE. Radiolabelled probe was collected as the fraction migrating with the blue dye and unincorporated nucleotides as the fraction with the orange dye. The probe was then boiled for 10 mins and used as a probe.

Hybridisation of probe to membrane

The RNA filter was first prehybridised in 20 ml Church Buffer (10 g/L BSA, 1 mM EDTA, 0.5 M $NaPO_4$ pH 7.2, 7 % SDS) for 2 h at 65 °C in a hybridisation tube with rotation. The volume of Church buffer was reduced to 10 ml and the denatured probe added directly to the hybridisation tube. This was incubated at 65°C overnight. The filter was washed twice in 2 x SSC 1 % SDS for 15 mins at 65°C and twice in 0.2 x SSC 1 % SDS for 15 mins at 65°C. The filter was then wrapped in clingfilm and autoradiography was performed by exposing membrane to an x-ray film at -70 °C.

2.7.4 RT-PCR analysis of gene expression

500 ng of RNA was reverse transcribed using the ImProm-II Reverse Transcription kit (Promega) following the manufacturer's instructions. Oligo $(dT)_{15}$ was used to prime the reaction and $MgCl_2$ concentration optimised at 6 mM. 1 µl of the cDNA product from the reverse transcription reaction was used in a 20 µl PCR reaction with 16 cycles. 3 µl of this reaction was run on an agarose gel and blotted on to nylon membrane (as previously described). This was then hybridised with a Dig-labelled molecular probe for the gene of interest.

2.8 Yeast two-hybrid assay for protein-protein interactions

2.8.1 General Principles

The two-hybrid system is a yeast-based genetic assay to detect protein-protein interactions *in-vivo* and has been described in detail (Fields and Song, 1989). The yeast two-hybrid method has advantages over other techniques for studying protein-protein interactions. It can detect weak and transient interactions, which may be important with the Polycomb-group proteins as they are expressed at low levels and sometimes only at specific times. The assay is performed *in-vivo*, therefore, the proteins tested are likely to be in their native conformations.

The yeast two-hybrid assay exploits the modular nature of eukaryotic transcription factors to indicate the interaction of two proteins. This technique is based on the ability to separate the DNA-binding domain from the transcriptional activation domain of many transcriptional activators, such as yeast GAL4 (Ma and Ptashne, 1987). The DNA binding domain (BD) is fused to one of the proteins of interest and the activation domain (AD) to the other. These are co-transformed into yeast cells carrying reporter genes that contain upstream GAL4 binding sites and are transcriptionally dependent upon reconstituted GAL4 activity. When these two proteins interact, the two GAL4 functional domains are tethered immediately upstream of reporter gene promoters, resulting in transcriptional activation and expression of reporter genes.

In these analyses the Matchmaker two-hybrid system (Clontech) was used. The DNA-BD cloning vector pGBT9 was used to generate fusions of the target protein with the GAL4 binding domain, termed bait, and the DNA-AD cloning vector pGAD424 for GAL4 activation domain fusions, termed prey (see figure 2.6). These were introduced into either *Saccharomyces cerevisiae* HF7c or AH109 cells. The HF7c strain carried *LacZ* and *HIS3* reporter genes (Bartel et al., 1993) and AH109 *ADE2*, *LacZ*, *HIS3* and *MEL1* reporter genes (Aho et al., 1997).

2.8.2 Preparation of constructs for two-hybrid analysis

A number of two-hybrid clones were generated in either pGBT9 or pGAD424.

CLF constructs

A series of 11 constructs were made in pGBT9 encoding truncations of the CLF protein (see figure 2.7 for schematic diagram showing peptides). pCMGB1-5 were supplied by Colin McDougall.

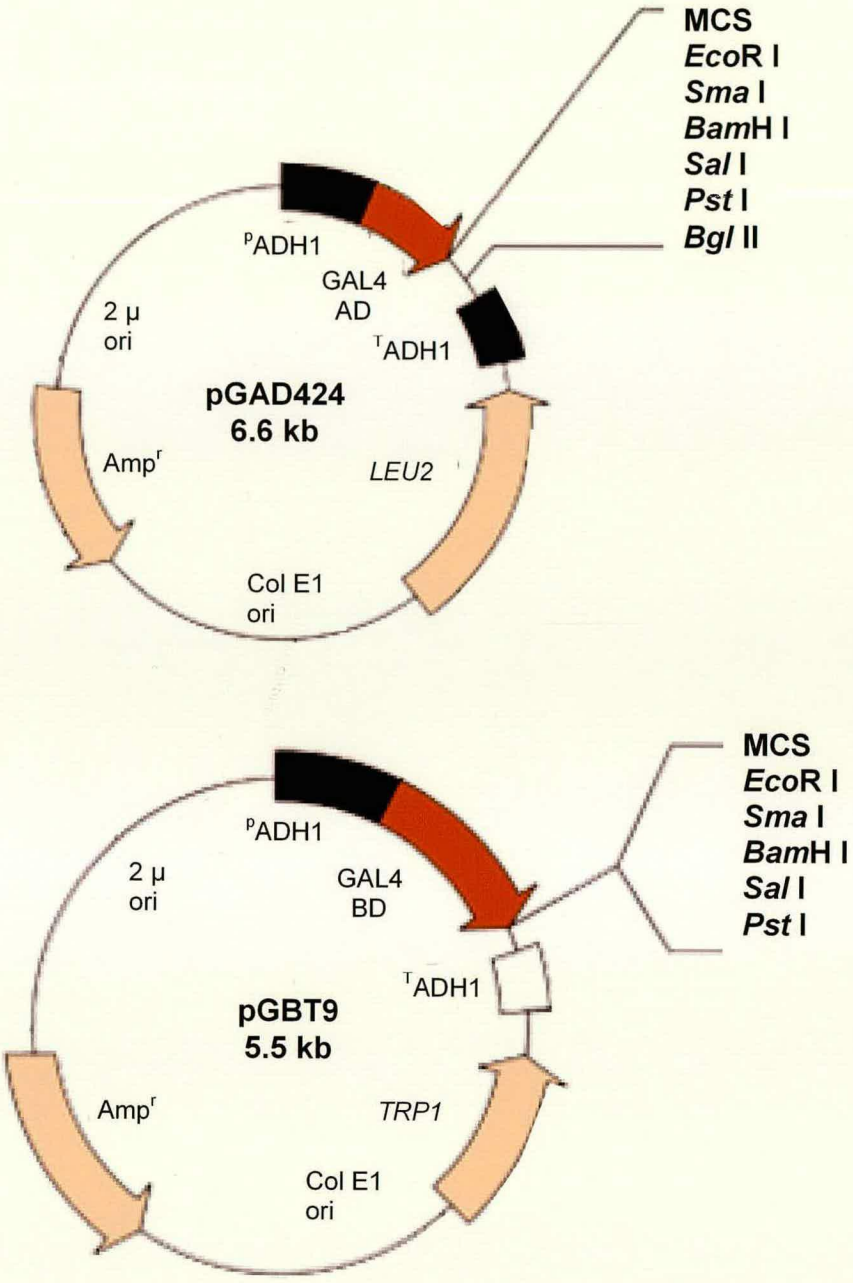


Figure 2.6 Clontech Matchmaker vectors used for yeast-two hybrid analysis
Both vectors carry Amp^r for selection in *E. coli*, the bait carries *TRP1* for selection in yeast and the prey *LEU2*. The GAL4 DNA transcription factor is split between the two plasmids the pGBT9 contains the binding domain (BD) and pGAD424 carries the activation domain (AD). Restriction sites within the multiple cloning site (MCS) are also shown.

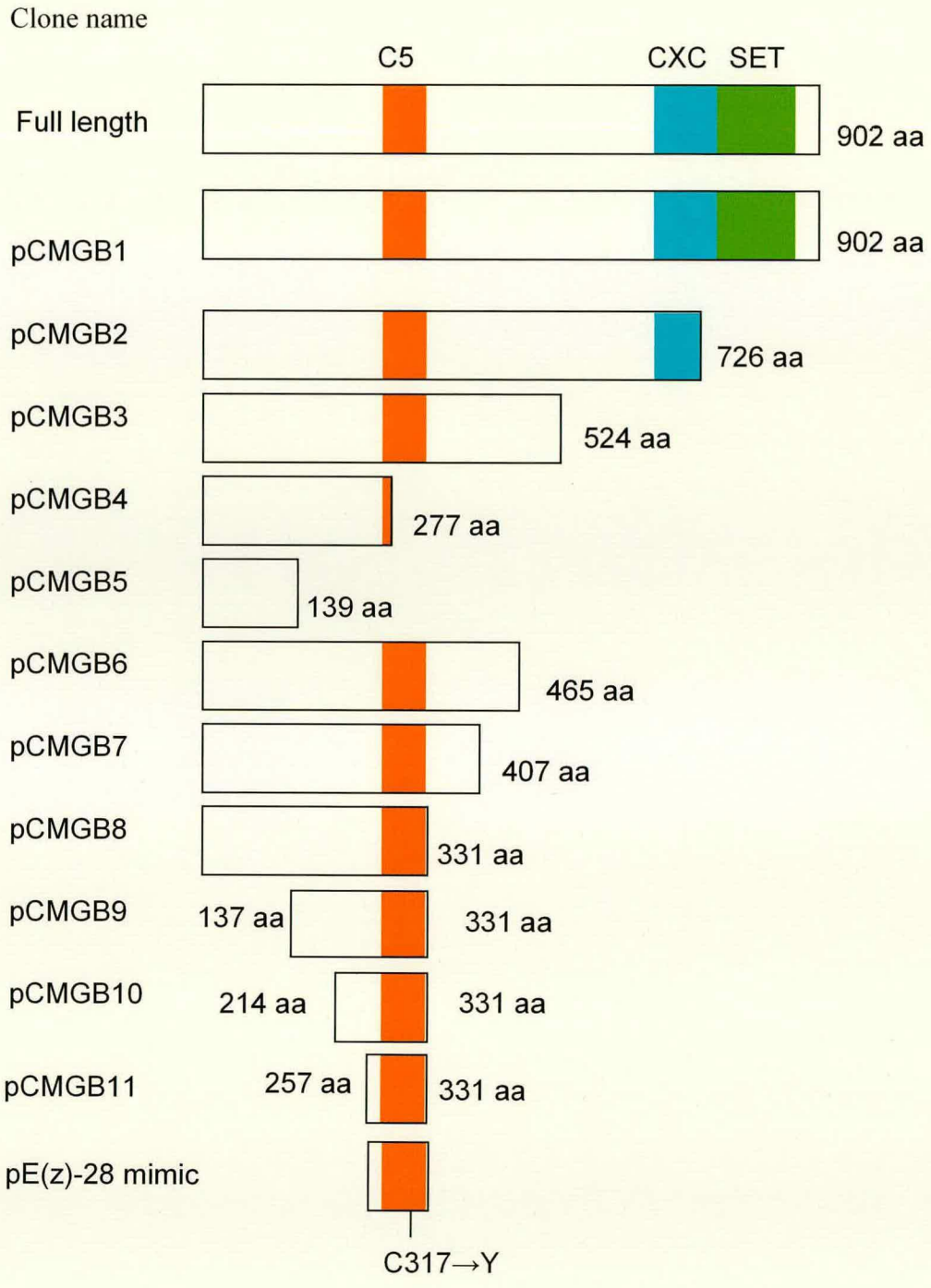


Figure 2.7 Diagram showing region of CLF protein expressed in pGBT9

pCMGB6 , pCMGB7 and pCMGB8

pCMGB3 was mutated by S-D IVM with primers JG43 and JG44, JG45 and JG46 and JG 47 and JG48 to introduce *Sal* I sites at points corresponding to amino acids 465, 407 and 331 respectively. The new plasmids were then cut with *Sal* I to lose a small 3'-fragment, the large fragments were gel purified and ligated at low concentrations to favour re-circularisation.

pCMGB9, pCMGB10 and pCMGB11

pCMGB8 was mutated by S-D IVM with primers clftrunc1 and clftrunc1r, clftrunc2 and clftrunc2r and clftrunc3 and clftrunc3r to introduce in-frame *Eco*R I sites at amino acids 137, 214 and 257 respectively. The plasmids bearing these new mutations were cut with *Eco*R I, and circularised as before, loosing a small 5'-fragment each time.

pCLF-C317Y

This construct was designed to mimic the *Drosophila e(z)-28* temperature sensitive allele described by Carrington and Jones (1996), which carries a mutation converting the fifth conserved cysteine residue in the C5 region to a tyrosine. The corresponding mutation in *Arabidopsis* would be C317 to Y. The pCMGB11 DNA sequence was mutated using S-D IVM with the primers AB-CLF-E(z)mimicf and AB-CLF-E(z)mimicr to produce this change.

pCLF11AD

The *CLF* fragment was cut out of pCMGB11 using *Eco*R I and *Sal* I and cloned into pGAD424 using the same restriction sites.

pCLF108

This construct encoded a 0-108 amino acid truncation of the CLF protein in vector pBI-88I. The starting plasmid was pBI-88I CLF 0-330; this contained a larger CLF fragment. S-D IVM was used to insert a *Sac* I restriction site at the position corresponding to amino acid 108 in the *CLF* cDNA with AB-CLF108f and AB-

CLF108r. The product was then cut with *Sac* I, the large fragment gel purified and ligated to re-circularise without the 109-330 fragment.

SWN construct

The construct pSWNc5 contained the C5 region of *SWN* cloned into pGBT9 for use as a prey vector. The C5 region of *SWN* was PCR amplified from cDNA using the primers AB-SWNC5f and AB-SWNC5r. These introduced *Eco*R I and *Sal* I sites at positions corresponding to amino acids 252 and 331 respectively. The PCR fragment was cleaved with *Eco*R I and *Sal* I, gel purified and cloned into pGBT9. The insert was sequenced prior to transformation into yeast (see figure 2.8).

MEA constructs

5 clones were made in pGBT9 for use as prey vectors. pMEAC5-1, pMEAC5-2 and pMEAC5-3 all consisted of the C5 region with pMEAC5-2 and pMEAC5-3 being slightly larger. pMEA-4 encoded a truncation missing the CXC and SET domains approximately equal to pCMGB6 and pMEA-full consisted of full length *MEA* cDNA (see figure 2.9)

pMEAC5-1, pMEAC5-2 and pMEAC5-3

These were produced in the same way as pSWNC5 using *MEA* cDNA, using the primers AB-MEAC5f and AB-MEAC5r for pMEAC5-1, MEAC52f and MEAC52r for pMEAC5-2, and MEAC53f and MEAC53r for MEAC5-3. These also introduced *Eco*R I sites at positions corresponding to amino acids 250, 237 and 225 and *Sal* I sites at positions corresponding to amino acids 323, 336 and 344. The *Eco*R I - *Sal* I fragment was cloned into pGBT9.

pMEA-4 and pMEA-full

pMEA-4 and pMEA-full were generated by S-D IVM of pKS::*MEA* with MEA-Eco-5'f and MEA-Eco-5'r to introduce an in-frame *Eco*R I site at the 5'-end of the cDNA. This was then cut with *Spe* I for pMEA-4 and *Asp* 718 for pMEA-full and blunt ended with

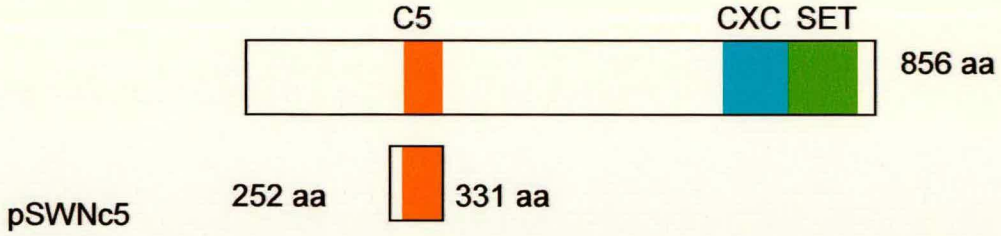


Figure 2.8 **Region of SWN expressed in pGBT9**

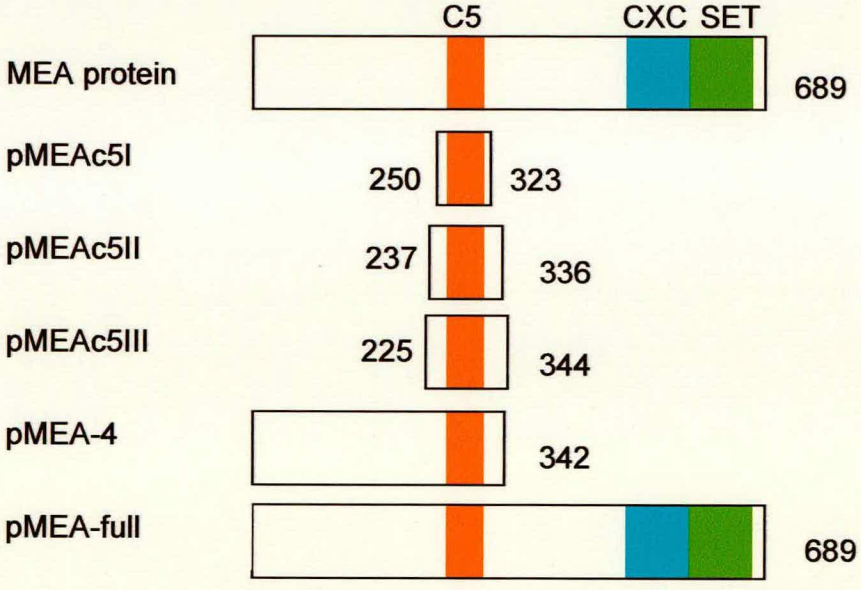


Figure 2.9 **region of MEA expressed in pGBT9**

Klenow polymerase. These products were then digested with *Eco*R I and gel purified. The vector was prepared by cutting pGBT9 with *Sal* I, filling-in with Klenow, digesting with *Eco*R I and de-phosphorylating the 5'-ends, before the fragments could finally be ligated.

EMF2 constructs

Five constructs were made in total: pEMF2-AD 1-4 by Yong Hwan Moon and pEMF2-5AD was isolated through a yeast two-hybrid screen by Daniel Schubert. Figure 2.10 shows the relative sizes of the EMF2 inserts.

pEMF2-4BD

The EMF2-4 insert was cut out of EMF2-4AD using the restriction sites *Eco*R I and *Sal* I and cloned into pGBT9 using the same sites.

pFIS2-VEFS

A *FIS2* fragment was amplified from a cDNA library produced from siliques by Christine Stock using the primers FIS2VEFS and FIS2VEFSr. FIS2VEFS introduced an *Eco*R I site at a point corresponding to amino acid 394 and FIS2VEFSr was sited just after an existing *Xho* I site in the cDNA. The PCR product was subcloned into pGEM-T (Promega). The FIS2 fragment was liberated by partial digestion with *Eco*R I and full digestion with *Xho* I. This was then cloned into *Eco*R I and *Sal* I sites of pGAD424. Before transformation into yeast the insert was sequenced and found to have a mutation causing an amino acid substitution in the cDNA, this was corrected using S-D IVM (figure 2.11).

pVRN2-VEFS

This was isolated in a yeast two hybrid screen by Daniel Schubert. The plasmid was extracted from yeast cells and used in these analyses. Figure 2.12 shows the size of the VRN2 insert.

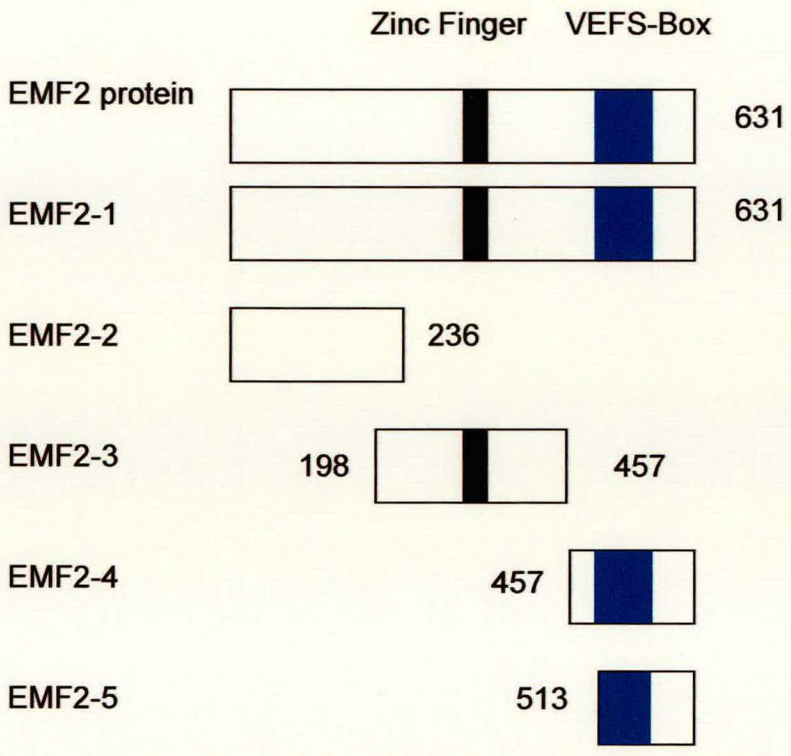


Figure 2.10 Regions of EMF2 expressed in pGAD424 clones

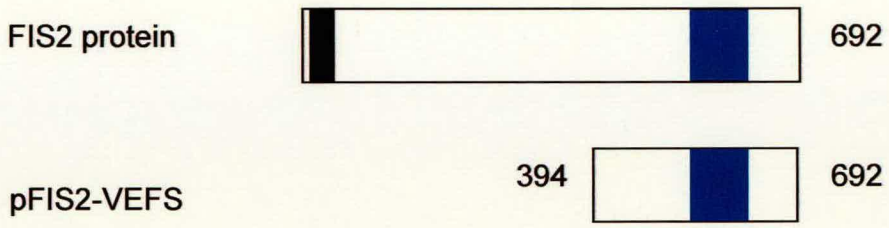


Figure 2.11 region of FIS2 protein expressed in pGAD424

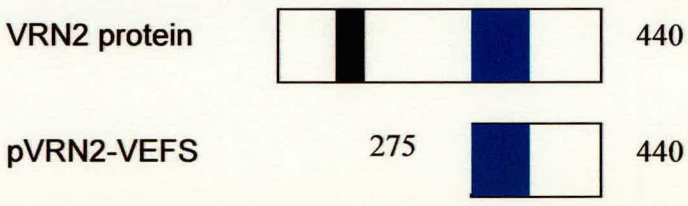


Figure 2.12 region of VRN2 protein expressed in pGAD424

2.8.3 Yeast Culture

Yeast without plasmid inserts were grown in YPDA media at 30 °C with shaking at 200 rpm or on YPDA plates at 30 °C. Selective media was used for yeast cells containing plasmids; this was YMM media with all essential amino acids added except for those used for selection. This was done conveniently with drop out powder which contained all essential amino acids except leucine, tryptophan, histidine and adenine. For example, yeast cells with pGBT9 and pGAD424 plasmids were grown on media lacking both leucine and tryptophan (-L,-W), and this was prepared by adding drop out powder supplemented with histidine and adenine.

2.8.4 Transformation of yeast with plasmid DNA

3 ml YPDA was inoculated with several yeast colonies (2-3 mm in diameter) of strains HF7c or AH109. This was vortexed vigorously to disperse any clumps and incubated at 30°C overnight with shaking at 250 rpm until the cells had reached the stationary phase (OD_{600}>1.5). 1 ml of overnight culture was transferred to a flask containing 50 ml of YPDA and incubated at 30 °C for approximately 3 h with shaking (250 rpm) until the OD_{600} was 0.4-0.6. Cells were placed in 50 ml tubes and centrifuged at 1 000 x *g* for 5 mins at room temperature. The supernatants were discarded and the cell pellets resuspended in 1m l sterile 1 x LiOAc/TE (10 mM Tris.HCl pH 7.4, 1 mM EDTA, 10 mM LiOAc). The cells were washed once more in 1 x LiOAc/TE and the cell pellet finally resuspended in 1 ml of 1 x LiOAc/TE. 1 µg each plasmid and 20 µl of salmon sperm carrier DNA (2 mg/ml) were added to a fresh 1.5 ml tube and mixed. 100 µl of cells and 0.6 ml of sterile PEG/LiOAc (40 % PEG, 10 mM Tris.HCl pH 7.4, 1 mM EDTA, 100 mM LiOAc) were added to each tube and mixed by pipetting. The cells were incubated at 30 °C for 30 mins followed by a 42 °C heat shock for 15 mins. The cells were centrifuged for 5 secs at 14 000 x *g* at room temperature, resuspended in 100 µl of sterile TE and plated on selective plates.

2.8.5 Nutritional assays

HF7c cells carry the *his3-1* mutation, conferring histidine auxotrophy, and AH109 cells also contain the *his3-1* and *ade2* mutations conferring both histidine and adenine auxotrophy. *HIS3* and *ADE2* function is provided if the reporter genes are activated by a physical interaction between the two protein fusions being tested. Four independent transformants were transferred to -L,-W selective plates and to -L,-W,-H or -L,-W,-A selective plates. Interactions were seen by growth on selective plates lacking either adenine or histidine.

2.8.6 *LacZ* reporter gene assay

2 ml overnight cultures were grown in -L,-W media to select for bait and prey plasmids. These were vortexed for 30 secs to disperse cell clumps. 1 ml of the overnight culture was transferred to 4 ml of YPDA. The fresh culture was incubated at 30 °C for 3-5 h with shaking (250 rpm) until the cells were in the mid-log phase (OD_{600} between 0.5-0.8). The exact OD_{600} was recorded when the cells were harvested. 0.5 ml of culture was placed into a 1.5 ml microcentrifuge tube, and centrifuged at 18 000 x *g* for 30 secs. Cells were washed by resuspension in 500 µl of Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$), harvested, and finally resuspended in 100 µl of Z buffer. Tubes were frozen in liquid nitrogen and thawed in a 37 °C waterbath. This freeze-thaw cycle was repeated twice more. A blank tube was set up containing only 100 µl of Z buffer, and then 0.7 ml of Z buffer with 0.27 % β-mercaptoethanol and 160 µl of o-nitrophenyl β-D-galactopyranoside (ONGP) (4 mg/ml in Z buffer) was added to all tubes. The tubes were incubated at 30 °C for 5-16 h, and then 0.4 ml of 1 M Na_2CO_3 was added to stop the reaction. The tubes were centrifuged for 10 mins at 18 000 x *g* to pellet cell debris. The spectrophotometer was calibrated against the blank and the OD_{420} of the samples measured. The β-galactosidase units could then be measured using the following equation. One unit of β-galactosidase is defined as the amount which hydrolyzes 1 pmol of ONGP to o-nitrophenol and D-galactose per min per cell (Miller, 1972).

β-galactosidase units = 1 000 x OD_{420} /(t x V x OD_{600})

where: t = elapsed time (in min) of incubation

V = concentration factor (in this case 1)

2.8.8 Yeast plasmid preparation

5 ml yeast cultures were grown overnight in YPDA. Cells were harvested from 3 ml of culture by centrifugation in a microcentrifuge tube, the supernatant removed and the pellet resuspended in the residual liquid. 200µl of yeast lysis solution (0.2 % Triton X-100, 1 % SDS, 100m M NaCl, 100 mM Tris.HCl pH 8.0, 10 mM EDTA), 200 µl of phenol/chloroform and 300 mg of acid washed beads were added and the tube vortexed violently. The two phases were then separated by centrifugation and the aqueous phase precipitated with 20 µl 3 M NaOAc and 500 µl ethanol. The pellet was washed in 70 % ethanol, dissolved in 10 µl TE and the plasmid transferred into *E. coli* DH5α cells.

3 The Specificity of *Arabidopsis E(Z)* Homologues

3.1 Introduction

This chapter investigates the relationships between and specificity of the three *Arabidopsis E(Z)* homologues. Previous studies have shown significant conservation between plant and animal *E(Z)* homologues, particularly in the SET domain.

In animals, the SET domain of E(Z) has been show to provide histone lysine methyltransferase activity to the PRC2 complex and this is central to its role in the epigenetic control of gene expression (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). The first aim, therefore, is to ascertain whether the structural conservation between *E(Z)* and the *Arabidopsis* homologues is indicative of conserved biochemical function. To do this, novel mutations which affect conserved residues in the CLF SET domain (and may affect the ability of this domain to function) were characterised.

The second aim of this chapter is to examine the evolutionary relationships between CLF, SWN and MEA. This is achieved using a phylogenetic approach to model the evolutionary history between these three proteins, and other E(Z) homologues identified in diverse plant species. The phylogeny could be informative as to function, i.e. the most closely related proteins may be more likely to have similar functions.

Despite their similar protein structures, CLF and MEA act in different pathways to regulate diverse aspects of plant development. The final aim of this chapter is to identify which changes confer this specialisation, in particular testing whether the involvement of these two proteins in separate pathways is simply due to alterations in their expression patterns or if the proteins have acquired distinct functions, for example targeting or interacting with different proteins.

3.2 Is biochemical function conserved between E(Z) and CLF?

E(Z), like its *Arabidopsis* homologues is essential for the mitotically stable repression of homeotic and other genes. It has recently been shown that PRC2 complex (containing ESC and E(Z)) has a histone methyltransferase activity towards lysine 9 and lysine 27 on the N-terminal tail of histone H3 (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). All enzymes identified to date which methylate specific lysine residues on histone tails belong to the SET domain family (Jenuwein, 2001) with the exception of the DOT1p protein from *Schizosaccharomyces pombe* which methylates K79 of H3 at the top and bottom of the nucleosome core (Feng et al., 2002; van Leeuwen et al., 2002). Consistent with this, it has been shown that the *in-vitro* methyltransferase activity of the PRC2 complex is lost when the SET domain of *E(Z)* is mutated (Czermin et al., 2002).

The recent surge in interest in epigenetics has meant that many SET domain proteins have been characterised biochemically and their structures studied by x-ray crystallography. Structures have been resolved for the human protein SET7/9 (Jacobs et al., 2002; Wilson et al., 2002), the DIM5 protein from *N. crassa* (Zhang et al., 2002), Rubisco large subunit methyltransferase from pea (Trievel et al., 2002) and CLR4 from *S. pombe* (Min et al., 2002). This has allowed valuable insights into the biochemical mechanisms by which these proteins operate.

To test, firstly, whether the SET domain of CLF was functionally important and, secondly, if it confers histone methyltransferase activity, a screen was performed to identify novel mutations affecting the CLF SET domain. The effect of these mutations on the ability of the SET domain to confer histone methyltransferase activity could be predicted based upon the crystal structures of other SET domains. This could be compared with the effects that they confer *in-vivo* to test the importance of the SET domain.

3.3 The use of TILLING to identify novel mutations in the SET domain of CLF

Targerting Induced Local Lessions In Genomes (TILLING) is a reverse-genetic strategy which provides an allelic series of point mutations in a gene of interest (Till et al., 2003a). High-throughput TILLING is a service offered by the *Arabidopsis* TILLING Project (ATP) combining effective generation of point mutations with PCR-based screening, allowing rapid and low-cost discovery of induced point mutations in a gene of interest (Till et al., 2003b).

Seed are initially mutagenised with ethylmethanesulfonate (EMS), this adds alkyl groups to nucleotides causing them to mis-pair with non-complementary bases and introduces point mutations after replication. To avoid ambiguities caused by chimaerism of mutant plants in the M1 generation, these are self-fertilised and the M2 generation is screened for mutations. The ATP screens DNA from 3072 M2 plants in pools of eight plants. These are then amplified with primers supplied by the user which amplify a fragment of about 1 kb in the gene of interest. After denaturing and re-annealing of this PCR product, heteroduplexes form between wild-type fragments and fragments bearing point mutations. The celery mismatch cleavage endonuclease CEL 1 specifically cleaves these heteroduplexes. Cleaved products are then resolved using denaturing polyacrylamide gel electrophoresis. The size of the products indicates the position of the mutation; mutations can then be characterised by sequence analysis.

Primers were designed to amplify a 1 kb region of *CLF* that encodes a region spanning the SET domain for use by the ATP to screen for mutations. The programme CODDLe (Codons Optimised to Discover Deleterious Lesions - http://www.proweb.org/coddle/) was used to identify regions where anticipated point mutations are most likely to cause deleterious effects on gene function. This separately predicts changes which will truncate the protein or destabilise the RNA (nonsense changes and splice junction changes), silent changes which do not affect the gene product and missense changes which alter the sequence of the protein product and may affect function. Primers were

chosen to optimise the probability of identifying missense mutations within the SET domain.

The ATP TILLING screen identified seven lesions in the *CLF* locus (table 3.1). Of these, one was within an intron and another was predicted to be a "silent" change, so that the protein product is unaltered. Of the remaining five, one lesion affected an intron/exon junction, and is likely to affect splicing leading to changes in the protein product. Four lesions produced missense changes (three of which were in the SET domain and the last in the C-terminal tail). Figure 3.1 shows the location of these missense mutations. Seed stocks from these five lines were ordered from the *Arabidopsis* Biological Resource Centre (ABRC) and the novel alleles were designated *clf54-58* (table 3.1). In each of the five lines, 14 individuals were genotyped using a combination of CAP markers, dCAP markers and sequencing (table 3.2).

Homozygous mutants were identified in all five lines (shown in figure 3.2). Plants homozygous for the *clf-55* allele, which changed a splice site at the start penultimate exon, showed a null phenotype. Null phenotypes were defined as those equal to those of the *clf-50* allele, which has a deletion of the entire *CLF* locus. Plants homozygous for either the *clf-56* or *clf-57* alleles also had null phenotypes (figure 3.2). The third allele affecting the SET domain, *clf-54,* showed weaker effects, with plants showing greater size, more lateral branches and cauline leaves and reduced leaf curling when compared with the null alleles. Plants homozygous for *clf-58*, a lesion in the post-SET domain show normal development and were indistinguishable from wild-type.

In order to predict the effects that these mutations would have on the ability of the CLF SET domain to confer methyltransferase activity, CLF was compared with other SET domain proteins which have had their structures resolved. CLF was aligned with CLR4 from *S. pombe*, human SUV39h1, human G9a and DIM5 from *N. crassa* (figure 3.3). Of the three mutations in the SET domain only *clf-57* (R794H) affects a well conserved residue; this arginine is conserved in all the SET domain proteins aligned and is clearly

Table 3.1 Mutations identified in the *CLF* locus through TILLING

#	Nucleotide Change [a]	Effect [b]	Genotype of parent	Allele designation
1	C132→ T	Intron (silent)	Hetero	
2	G374→ A	G779→ E	Hetero	*clf 56*
3	A406→ T	R794→ H	Hetero	*clf 57*
4	G422→ A	G795→ E	Hetero	*clf 54*
5	C438→ T	(silent)	Homo	
6	G568→ A	Splice Junction	Hetero	*clf 55*
7	C899→ T	P893→ L	Hetero	*clf 58*

[a] Nucleotide number relates to position within area TILLED (1-903 bp).

[b] Changes within coding region are shown by their effect on the protein. Residues are numbered according to position in protein.

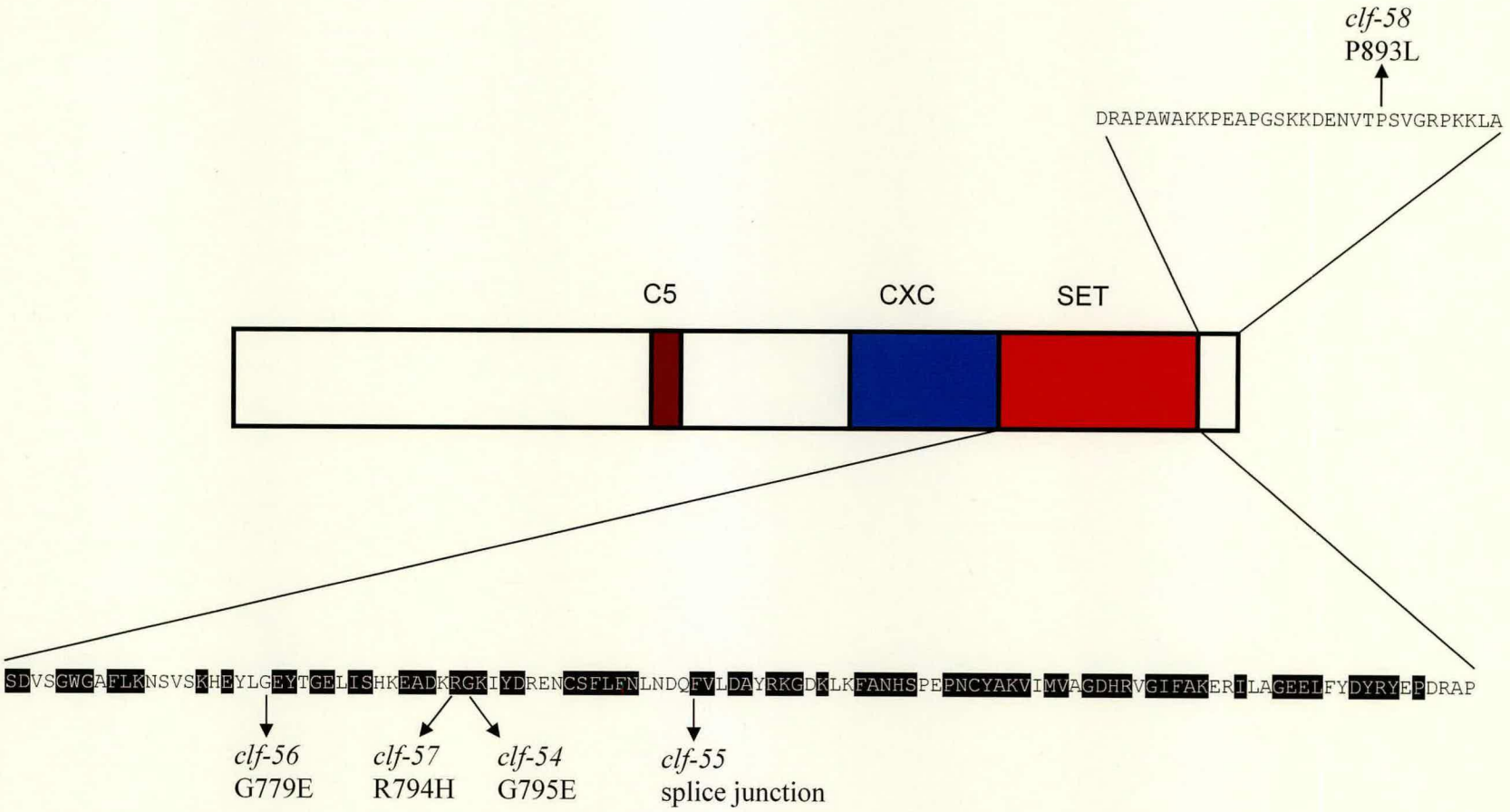


Figure 3.1 Schematic diagram of CLF showing the locations of the induced mutations
The sequences of the SET and post-SET domains are shown. Black boxes indicate residues which are identical with E(Z). Arrows indicate the altered residues and show new identity.

Table 3.2 Methods employed to genotype *clf* alleles

clf allele	Method of analysis	Amplification with primers	PCR product cleaved by	Cleaved band corresponds to
clf-54	dCAP	AB34 and AB35	*Sac* II	wild-type
clf-55	CAP	AB6 and H1	*Pst* I	wild-type
clf-56	dCAP	ABG779E and AB35	*Nco* I	wild-type
clf-57	dCAP	AB32 and AB33	*Hha* I	wild-type
clf-58	Sequence	AB32 and AB6		

CAP markers were used when the polymorphic lesion either creates or removes a recognition site for a restriction endonuclease. A fragment surrounding the polymorphism was amplified by PCR, and the product cleaved with the appropriate enzyme. When no suitable restriction site was created or lost in the mutant, or when the relevant site was only cleaved by an expensive or unreliable endonuclease, derived CAP markers were used. Here a mutagenic primer was used, which when combined with the sequence of genomic DNA amplified, either creates or removes a restriction site in either the wild-type or mutant band. Generally, the enzyme chosen had a recognition site consisting of 6 bp. Where possible the sixth bp corresponded to the polymorphism and the mutation within the primer was as far away as possible from this position (i.e. the first or second bp within the recognition site). Occasionally it was not possible to design effective CAP or dCAP markers, and in these instances DNA sequencing was used to genotype plants.

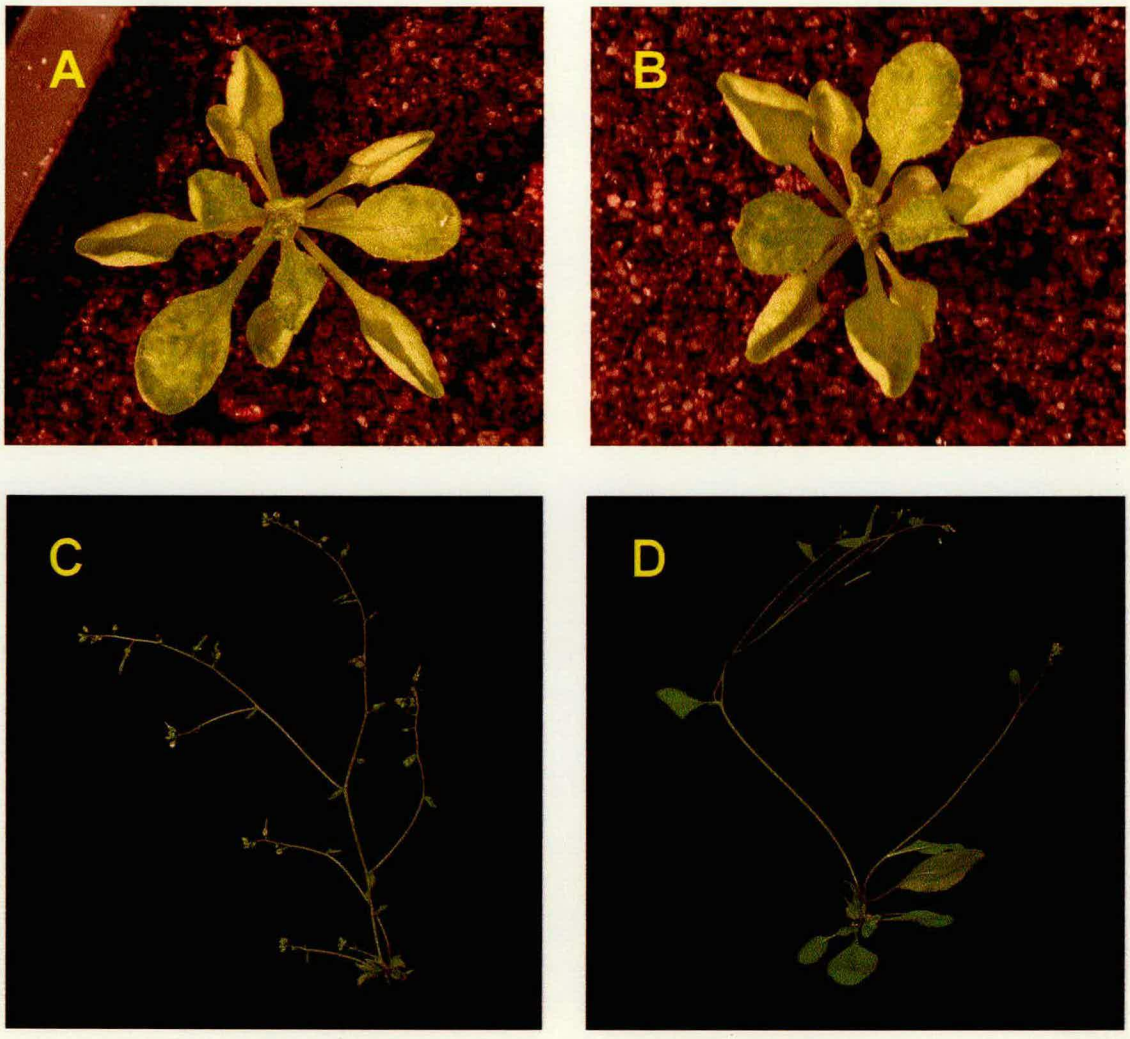


Figure 3.2 Phenotypes associated with *clf* mutants
Both *clf-56* (A) and *clf-57* (B) mutations lie in the SET domain. Plants homozygous for either of these mutations have a severe phenotype equivalent to that of the null *clf-50* allele. *clf-54* (C) also lies within the SET domain, but mutant plants have a significantly weaker phenotype. Leaves are still curled but the plant is a comparable size to that of wild-type. In *clf-58* (D) the lesion lies in the post-SET domain. Plants homozygous for this mutation show no phenotype and look like wild-type.

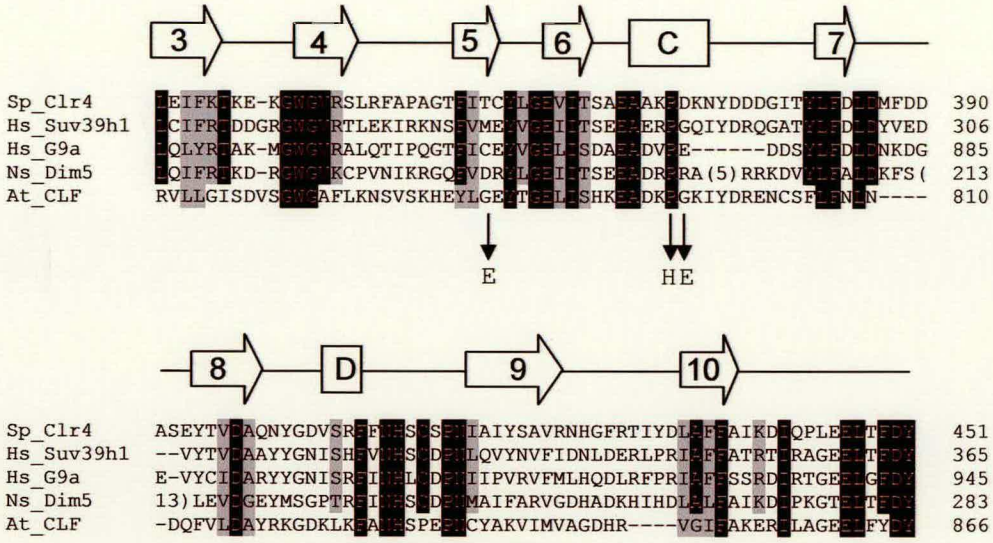


Figure 3.3 Alignment showing CLF with other methyltransferases
Alignment of SET domain of Clr4 from *Schizosaccharomyces pombe* with human Suv39h1, human G9a and Dim-5 from *Neurospora crassa* (adapted from Min et al., 2002). Identical residues are shown in white text over a black box. Similar residues are shown with grey box. Secondary protein structure elements for Clr4 are shown above (from Min et al., 2002). Strands are shown as arrows and numbered sequentially; helices are shown as boxes and labelled alphabetically.

Arabidopsis CLF has been added to the alignment. Black arrows show polymorphisms in TILLING mutants.

an important residue. It affects a residue corresponding to R369 of CLR4 which is located within one of two helices termed C in Min et al (2002). This helix is exposed and contains mostly charged and polar residues. It has been recently shown that this region is part of a groove which accommodates histone tails and is near the binding pocket for, S-adenosyl methionine (the necessary cofactor which donates the methyl group) (Xiao et al., 2003). In the human SET7/9 protein there is likely to be a direct role for the carbonyl group backbone of R258 (equivalent to CLF R794) in interacting with the side chain of R2 on the tail of histone H3 (S. Khorasanizadeh pers. comm.). Mutating this residue would, likely, abolish the ability of the SET domain to bind to its histone substrate resulting in the null phenotype observed in *clf-57* mutants.

The serious effects caused by missense mutations within the SET domain show, firstly, that it is essential for *CLF* activity *in-vivo*. In addition, the effect caused by a relatively conservative change of an arginine to a histidine in producing a null phenotype, illustrates the great importance of that particular residue. The fact that this residue has been shown in another SET domain protein to be necessary for a key aspect of histone modification, provides some evidence that CLF too acts as a histone methyltransferase.

It is difficult to evaluate whether the C-terminal tail is important to CLF function *in-vivo* as only one mutation was obtained. In addition, the *clf-58* lesion affects a non conserved residue.

3.2 Evolutionary relationships between plant *E(Z)* homologues

In *Arabidopsis* there are 30 putative SET domain genes and seven with partially truncated SET domains; of these at least 29 are active in diverse tissues (Baumbusch et al., 2001). A previous phylogenetic study has separated these SET domain proteins into four classes (Baumbusch et al., 2001). All four classes show extensive homology with proteins from the animal kingdom, not only in the SET domain but also in other conserved regions.

The class I SET domain proteins includes E(Z). In *Arabidopsis* this group is represented by three proteins compared with only one in *Drosophila* and *C. elegans* (Korf et al., 1998) and two in mouse and humans (Laible et al., 1997). This group has been defined by Baumbusch et al. (2001) as having 16-18 cysteine residues spaced in a given pattern in front of the C-terminal SET domain and also a conserved cysteine-rich stretch (known as the C5 or EZII domain), although these observations had previously been made (Goodrich et al., 1997; Grossniklaus et al., 1998).

This section aims to more closely investigate the relationships within this group by comparing the *Arabidopsis* class I SET domain proteins with those that have been isolated from other plants. The intention is to further characterise differences between *CLF*, *SWN* and *MEA*, with the ultimate aim of understanding what factors have recruited these genes into different pathways. To do this it is first necessary to examine homologues which have been identified in diverse plant species.

Three *Zea mays* genes (*MEZ1-3*) have been identified as *E(Z)* homologues from EST databases (Springer et al., 2001). *MEZ1* shows the closest homology to *CLF*, and *MEZ2* and *MEZ3* show close homology with *SWN* (Springer et al., 2002). These observations were based on sequence similarities within the SET domain as well as overall protein organisation. All three proteins have been shown to be expressed in diverse tissues, including roots, leaves and kernels. The *MEZ2* transcript is alternatively spliced in a tissue-dependent manner suggesting that different functions could be achieved by the same gene in different tissues. Interestingly, no *MEA* homologue has been identified in maize. This could be because MEA function can be achieved through one of the other MEZ proteins, perhaps using alternative splice sites, or that *MEA* is simply not represented by an EST. In *Arabidopsis*, *MEA* is only expressed at a low level in a small number of cells at distinct stages of gametophytic and seed development (Vielle-Calzada et al., 1999; Luo et al., 2000). This restricted expression pattern is consistent with the possibility that there is a maize *MEA* orthologue but it is not represented in the EST databases.

The rice *E(Z)* homologue *OsiEZ1* was identified in a cDNA library made from post-fertilised flowers of *Oryza sativa* ssp. *Indica*. The library was screened at low stringency using radiolabelled *Arabidopsis MEA* as a probe (Thakur et al., 2003). On sequence analysis only one clone showed significant similarity with the class I SET domain proteins, and this was designated *OsiEZ1*. Expression studies revealed that the *OsiEZ1* transcript level was highest in rice flowers, almost undetectable in developing seed 1-2 days post-fertilisation but increased slightly in young seed 3-5 days post-fertilisation (Thakur et al., 2003), an expression pattern not unlike that of *MEA* (Luo et al., 2000; Vielle-Calzada et al., 1999). *OsiEZ1* expression was also detected in vegetative tissues and roots.

Three *CLF* and *SWN* homologues were identified in *Petunia hybrida* by RT-PCR amplification of cDNA from inflorescence tissue using degenerate primers based upon CLF, SWN, MEA, E(Z), EZH1 and EZH2 proteins (Mayama et al., 2003). This study failed to identify any *MEA* homologues. This may have been because the degenerate primers showed less homology to *Arabidopsis MEA* than to *Arabidopsis CLF* and *SWN* cDNAs (six out of seven residues compared with seven out of seven), or because the tissue was harvested at such a stage when little or no expression of the *MEA* orthologue would be predicted (i.e. before mature development of the female gametophyte).

Two *Antirrhinum CLF* homologues were identified by Carol Wilson (J. Goodrich lab) by screening *Antirrhinum* cDNA and genomic DNA libraries at low stringency using the SET domain of *Arabidopsis CLF* as a probe (Wilson, 2001). At low stringency, one may expect to isolate not only the immediate homologues of *CLF*, but also *SWN* and possibly *MEA* homologues. However, it is not possible to predict whether the failure to identify these proteins was due to the method used for screening or, less likely, their absence in the *Antirrhinum* genome.

A search of the public Genbank database provided protein sequences for all these accessions except for those from *Antirrhinum* which were supplied by Carol Wilson. A Blast search of the Genbank database revealed another rice *E(Z)* homologue which has been included in the data as *OsiEZ2*. It is important to note that of the five species which are being examined, we can only be certain that all *E(Z)* homologues have been identified in *Arabidopsis thaliana*. Data from maize and rice suggest that there may be no direct *MEA* homologue in monocotyledonous plants. However, failure to identify *MEA* homologues in *Antirrhinum majus* and *Petunia hybrida* is more likely to be due to the methods by which the sequences were discovered, rather than the absence of these genes.

Sequences were aligned using the MegAlign programme; it was apparent that differences in protein length would complicate alignment of these proteins and could artificially group larger proteins together due to long branch attraction. Therefore, it was decided that only the conserved SET domains would be analysed (figure 3.4). These data were entered in a PAUP 3.1 nexus file and the programme executed.

A heuristic search was then performed using simple Tree Bisection-Reconnection (TBR) swapping and three most parsimonious trees were produced. In order to ensure there were no "islands" of trees of greater or equal parsimony, the search was repeated using a random addition sequence (1000 replicates). Again, this search yielded the same three trees. Statistical analysis comprised of 50 % majority-rule bootstrap values, which were calculated by a full heuristic bootstrap search based on 1000 replicates with simple addition and TBR swapping on an unrooted tree.

Drosophila E(Z) was selected as an outgroup based upon its vast evolutionary distance from the plant homologues. A strict-consensus of the three most parsimonious trees is shown in figure 3.5.

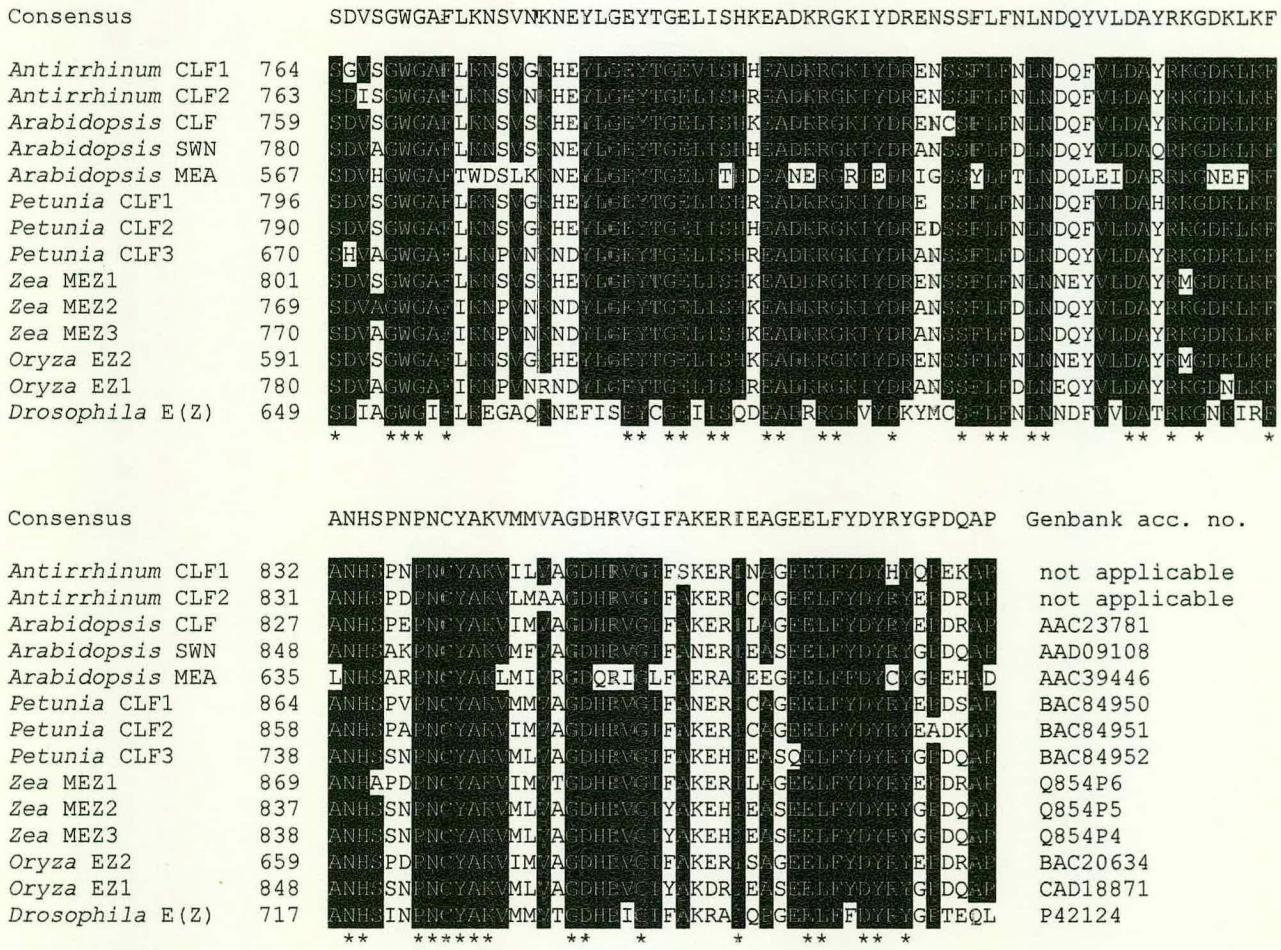


Figure 3.4 Alignment of SET domain of plant E(Z) homologues

Residues conserved in the majority of sequences (at least 12 out of 14) are shown by black boxes, and those conserved in all sequences are indicated with an asterisk. Genbank accession numbers are provided after sequences. A consensus sequence is shown above.

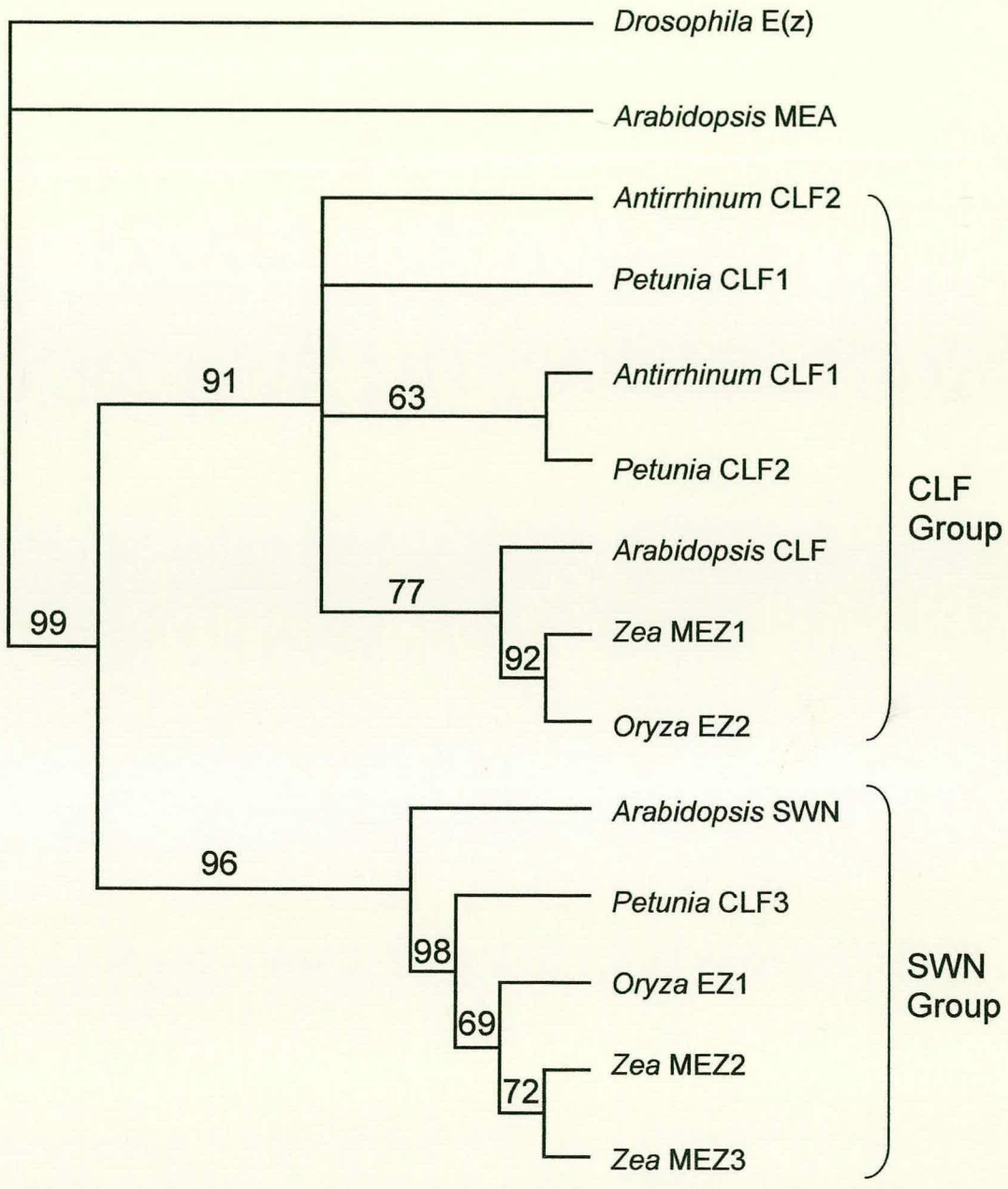


Figure 3.5 Phylogeny of plant E(Z) homologues
A strict consensus based upon 3 most parsimonious trees is shown. 50% majority-rule bootstrap values are shown above nodes (based on 1000 replicates). Bootstrapping provides a statistical estimate of the reliability of groupings. This involves taking a subsample of data (amino acids from alignment) and creates trees based on those subsamples. Values shown represent the percentage of bootstrapped trees in which the group is preserved. Generally 90% bootstrap values inspire high levels of confidence in the clade, whereas levels below 50% are considered poorly supported.

Three clear observations can be made from the data. Firstly, the plant E(Z) homologues form strongly supported groups incorporating proteins from diverse plant taxa, suggesting some degree of specialisation. Secondly, MEA is clearly different from CLF and SWN. Strong support is provided (bootstrap value of 99 %) for a clade excluding MEA but containing the *Arabidopsis* proteins CLF and SWN, along with numerous homologues from *Antirrhinum*, *Petunia*, rice and maize. Thirdly, the *Arabidopsis* proteins CLF and SWN form two distinct orthology groups; a bootstrap value of 91 % supports the presence of a CLF group and a value of 96 % for a SWN group. The inclusion of diverse taxa into these clades suggests that the divergence of CLF and SWN must have been at the least fairly basal during the evolution of the angiosperms. The data also imply that MEA is likely to be the most different in function, but that CLF and SWN may also have discrete functions.

In *Arabidopsis*, *CLF* and *SWN* act partially redundantly to control various aspects of plant development (Chanvivattana, 2002). It is difficult to imagine why changes in two redundant genes would be selected against so strongly. It is likely then that in *Arabidopsis SWN* has discrete functions that have yet to be identified. These might not be readily apparent unless screened in particular ways. For example, autonomous endosperm development of the *fis* mutants and the impaired vernalization response of *vrn2* mutants are phenotypes only observed under certain physiological conditions. Little is known about the role which Polycomb-group genes play in other plants. If the *CLF* and *SWN* homologues have distinct roles in other species as well, then selection may act strongly to maintain the two orthology groups. This hypothesis is especially attractive in monocotyledonous plants, where it looks likely that there are no *MEA* homologues, and genes showing homology to *CLF* and *SWN* are expressed in endosperm and related tissues.

Questions concerning the role of Polycomb group genes in other plant species can only be addressed by the cloning of all *E(Z)* homologues in a number of species (the data from *Petunia* and *Antirrhinum* has not done this), and by characterising the role of these

genes by knock-out mutagenesis. Doing this will address important issues such as, whether CLF and SWN-like proteins supply MEA function in monocots, and what the importance of the separate CLF and SWN orthology groups are.

3.5 The specificity of MEA, CLF and SWN *in-vivo*

The main aim of this chapter is to test whether the three *Arabidopsis* E(Z) proteins have discrete functions *in-vivo*. The phylogenetic analysis suggested that whilst the three proteins are homologous, CLF and SWN are more similar to each one another than either is to MEA. This is also supported by the analysis of *clf swn* double mutants, which suggests partially overlapping functions (Chanvivattana, 2002; D. Schubert, pers. comm.). In *Arabidopsis*, *MEA* has evolved a separate role in seed development. For monocots, some evidence exists suggesting that there are no *MEA* homologues. Presumably *MEA*-like function is provided by proteins showing homology with either CLF or SWN; and consistent with this, they are expressed within the female gametophyte pre-fertilisation and the developing seed post-fertilisation. Therefore, in monocots it is likely that there is considerable overlap between the roles of the *E(Z)* homologues. This would mean that in monocots at least, the CLF or SWN orthologues would have similar function to MEA.

In *Arabidopsis*, however, there is some evidence to suggest that the three proteins have at least partly discrete functions. Firstly, *SWN* can not complement a *clf* mutant with respect to regulation of *AG*, as *CLF* and *SWN* have similar expression patterns yet *AG* is de-repressed in *clf* SWN^+ plants. Secondly, it was recently shown that *MEA* is strongly expressed in the leaves of *clf* mutant plants (Katz et al., 2004). This suggests that *MEA* also is unable to complement *clf* mutants with respect to *AG* silencing.

To further resolve the extent to which differences between the activities of these three genes reflect the differences in the level or pattern of expression, rather than a change in protein function, a series of promoter swap experiments was initiated. Two classes of

constructs were designed. The first group contained ones in which *CLF*, *SWN* or *MEA* expression was driven by the constitutive 35S promoter. These were used to test whether merely expressing high-levels of RNA throughout the plant could provide *CLF*-like function. The 35S viral promoter has been used to drive gene expression in a wide range of plant species and reliably provides constitutive expression at high-levels. The use of the 35S promoter might be particularly informative for *SWN*, where we know that the expression pattern is very similar to *CLF*, but we do not know if there are subtle differences in expression levels. The second group contains *CLF*, *SWN* and *MEA* fused to the endogenous *CLF* promoter. The endogenous *CLF* promoter used in these experiments was a 1.4 kb section of DNA upstream from the translational start; it was previously found that a fragment of genomic DNA spanning the *CLF* locus and containing this 1.4 kb promoter region fully complemented *clf-50* mutants. The use of the endogenous *CLF* promoter will recreate the expression pattern of *CLF* and therefore test whether changes in expression patterns alone enable the complementation of *clf* mutants by *SWN* and *MEA*.

These constructs were introduced into *Agrobacterium tumefaciens* by plasmid transformation. 1 L cultures of these *Agrobacteria* were grown and used to transform the progeny of an *Arabidopsis clf-50* heterozygous plant by floral dip infiltration. Individuals homozygous for the *clf-50* allele were considered too small and too infertile for effective transfer of the transgenes. The primary transformants were grown in sterile culture on plates containing 50μg/ml kanamycin to select for those individuals carrying the transgene. The kanamycin resistant primary transformants were screened for individuals where the *clf* mutant phenotype was rescued by the presence of the transgene. Based on Mendelian inheritance half the transformants would be expected to be wild-type, one third heterozygous and one sixth homozygous mutants with respect to the *clf-50* mutation. If there was no complementation of *clf-50* by the transgene then one sixth of primary transformants would show the *clf* phenotype. If, however, the transgene was able to rescue *clf-50* then all, or virtually all, primary transformants would appear wild-type.

3.6 Ectopic expression of MEA and SWN is unable to complement *clf* mutants

The primary transformants obtained using the 35S::*CLF*, 35S::*SWN* and 35S::*MEA* constructs were analysed to look for complementation of the *clf* phenotype (table 3.3). Not one of the primary transformants in experiments using the 35S::*CLF* or 35S::*SWN* transgenes had a phenotype resembling the *clf* mutant (table 3.3). Whereas in the 35S::*MEA* line, approximately 1 in 6 plants looked phenotypically like homozygous *clf* mutants. Initially, this suggested that 35S::*CLF* and 35S::*SWN* were able to complement *clf* mutants, whereas *MEA* was not. To confirm this, 14 plants in each line were genotyped by Southern blot hybridisation to identify individuals homozygous for the *clf* mutation and to confirm presence of transgenes.

In the 35S::*CLF* primary transformants, two of the 14 plants genotyped by Southern blot hybridisation carried the transgene but were homozygous for the *clf-50* deletion (figure 3.6) and these individuals had a wild-type phenotype (figure 3.7 C). This confirmed that 35S::*CLF* complements *clf*. Also, the fact that no *clf* mutants were found in 34 primary transformants indicates that co-suppression is not an issue. In the 35S::*MEA* line, 4 of 23 primary transformants had a *clf* phenotype. Three plants with a *clf* phenotype and 11 with a wild-type phenotype were genotyped for both the genomic *CLF* locus and the *MEA* transgene by Southern blot hybridisation (figure 3.8). All 14 transformants contained the *MEA* transgene; only *clf* homozygotes had a *clf* mutant phenotype, indicating that the 35S::*MEA* transgene did not give rise to the *clf* phenotype (i.e. co-suppression). *MEA* was therefore unable to complement *clf* mutants when expressed under the control of the 35S promoter. A plant carrying the 35S::*MEA* transgene and homozygous for the *clf* mutation can be seen in figure 3.7 D. In the 35S::*SWN* line all 37 individuals had a CLF^{+} phenotype, suggesting that like 35S::*CLF*, 35S::*SWN* could complement *clf* mutants. Fourteen were genotyped by Southern blot hybridisation, however, no *clf-50* homozygous plants were found; all plants were *CLF+/CLF+* or *CLF+/clf*-50 in genotype (data not shown).

Table 3.3 Phenotype numbers of primary transformants

Construct	Number of primary transformants	Number with *clf* phenotype	Approximate ratio
35S::*CLF*	34	0	-
35S::*MEA*	23	4	1 in 6
35S::*SWN*	37	0	-
p(*CLF*)::*CLF*	51	5*	1 in 10
p(*CLF*)::*MEA*	54	8	1 in 7
p(*CLF*)::*SWN*	123	12	1 in 10

The p(*CLF*)::*CLF* line has a weak phenotypes and this is indicated with an asterisk. Although the ratios of mutants appear to be lower than expected in the p(*CLF*)::*CLF* and p(*CLF*)::*SWN* lines, statistical analysis through a chi squared test confirms that this deviation from the expected value is not significant. In both cases p=>0.05.

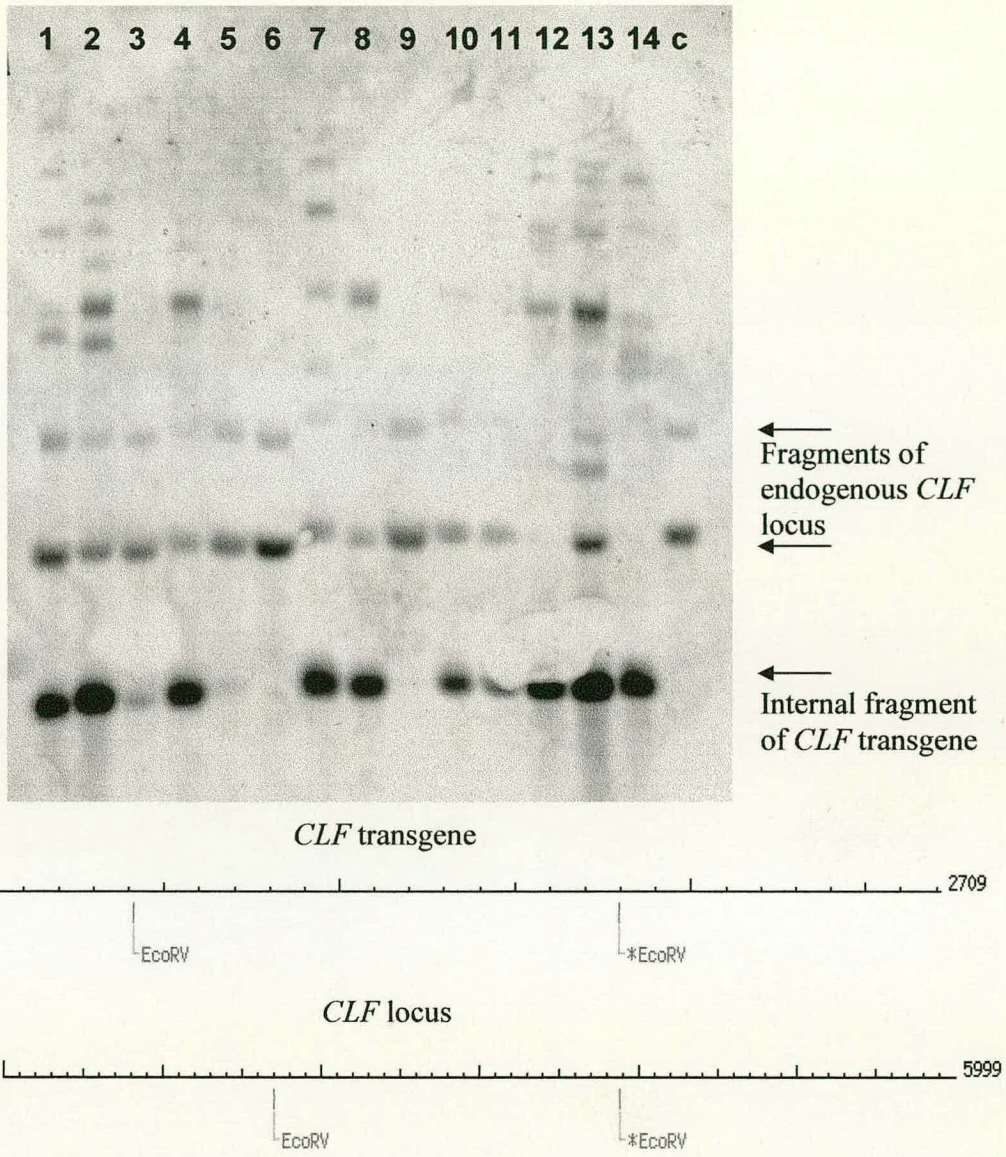


Figure 3.6 Southern blot analysis of 35S::*CLF* primary transformants
Genomic DNA was cut using *Eco*R V restriction endonuclease and separated by gel electrophoresis. The membrane was hybridised using the *CLF* cDNA as a probe. The endogenous *CLF* locus contains two *Eco*R V sites, so that the probe should detect a fragment of 2.2 kb deriving from the *CLF* locus. A second weaker band of approximately 3-4 kb is consistently detected, presumably due to fragment containing an *Eco*R V site outside the immediate *CLF* genomic locus. *clf-50* homozygotes lack the *CLF* locus, and so lack these fragments altogether. The *CLF* transgene contains 2 *Eco*R V sites 1.4 kb apart, so should detect a small internal band of 1.4 kb corresponding to the internal fragment of transgene. The intensity of this band is proportional to copy number. Variable larger bands are present corresponding to fragments containing the transgene and variable flanking locus DNA. On the right (c) is DNA from the wild-type progenitor background (Ws) showing genomic *CLF* bands. Lanes 12 and 14 show individuals carrying the *CLF* transgene but the genomic *CLF* locus is absent.

Figure 3.7 Complementation of *clf* mutants by promoter swap constructs
White lines represent scale bar. A. *CLF*$^{+}$ individual showing normal development. B. *clf-50* homozygous mutant, note early flowering phenotype, small curled leaves and small size. C, D and E are all homozygous *clf-50* mutants; C contains the 35S::*CLF* transgene and shows complete restoration of the *clf-50* phenotype, D and E contain the 35S::*MEA* and the 35S::*SWN* transgenes respectively and show no complementation of the *clf-50* phenotype. F, G and H all show a *CLF*$^{+}$ individual left next to and a homozygous *clf-50* mutants right carrying the p(*CLF*)::*CLF*, p(*CLF*)::*MEA* and p(*CLF*)::*SWN* transgenes respectively. The p(*CLF*)::*CLF* shows partial complementation of the *clf-50* phenotype; the plant is greater in stature, the leaves are larger but still curled and the cauline leaves are also significantly bigger. Individuals carrying the p(*CLF*)::*MEA* and p(*CLF*)::*SWN* transgenes show no complementation of the *clf-50* phenotype.

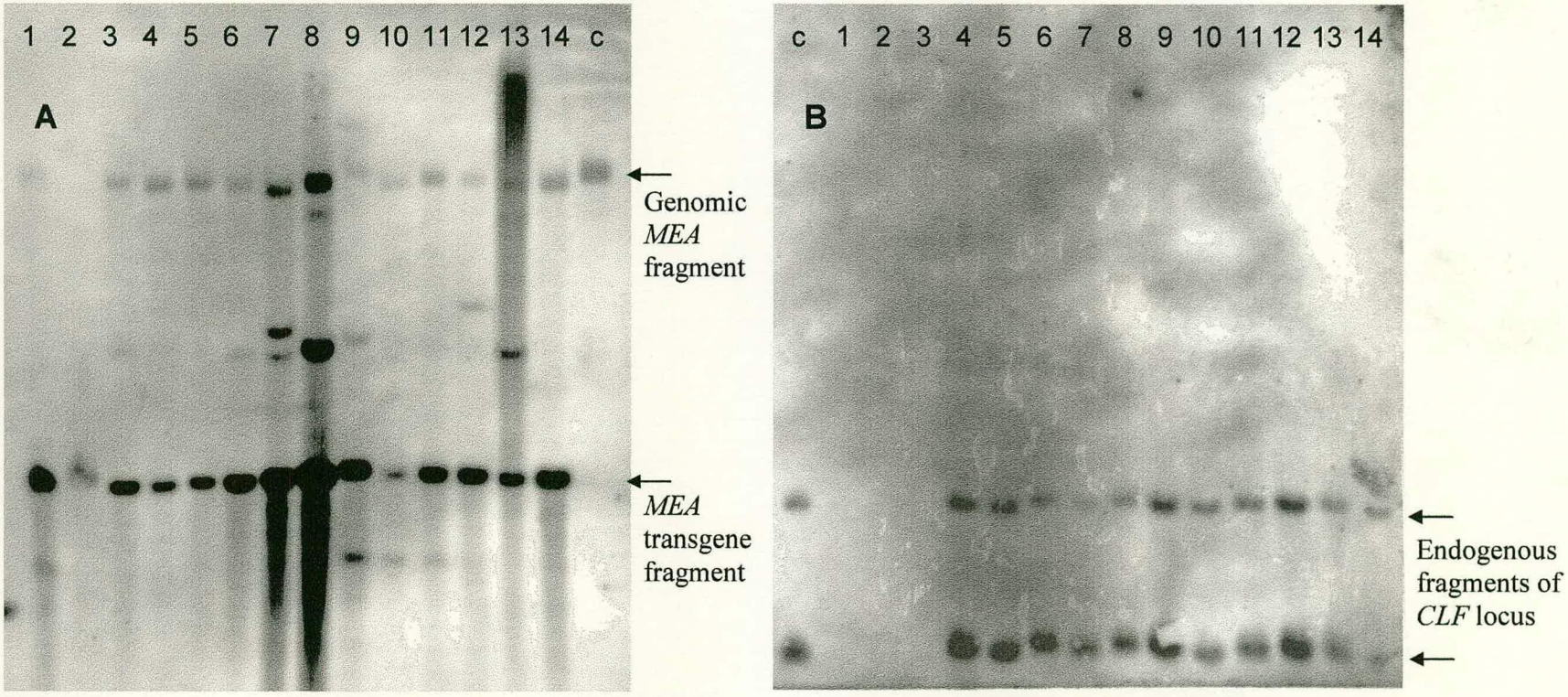


Figure 3.8 Southern blot analysis of 35S::*MEA* primary transformants

A. Fourteen independent transformants and a wild-type control were cut with *Eco*R I and *Bam*H I and the fragments separated by gel electrophoresis. The membrane was probed with *MEA* cDNA. The *Eco*R I and *Bam*H I sites cut the entire *MEA* cDNA from the transgene resulting in a 2.3 kb fragment. All 14 transformants contain this band and therefore the *MEA* transgene. There are also two *Bam*H I sites 6 kb apart in the genomic *MEA* locus. This produces a fainter band in each sample corresponding with genomic *MEA*. B. The same DNA preparations were cut with *Eco*R V and the resulting Southern blot hybridised with a *CLF* cDNA probe. The first three transformants are all *clf-50* homozygotes, as shown by absence of the 2.2 and 3.5 kb fragments of the *CLF* locus (see figure 3.6 legend for explanation).

To confirm that 35S::*SWN* complemented *clf*, I therefore analysed T2 families derived from these T1 plants with wild-type phenotypes. T2 lines were sown on kanamycin plates and plates lacking antibiotics. For 35S::*CLF* approximately one third of the T2 families segregated approximately 1 in 16 for plants with a *clf* mutant phenotype when grown without antibiotics. When seed from the same populations was sown onto plates containing kanamycin, plants with such a phenotype were never observed. When seed from the 35S::*MEA* line was sown on both antibiotic-free media and kanamycin plates, plants with a *clf* phenotype were seen with an equal frequency (approximately 1 in 4) in almost one third of the plates. When 35S::*SWN* lines were sown on plates lacking antibiotics few populations were observed which segregated for plants with a *clf* phenotype. This suggested that for some reason the occurrence of the *clf-50* allele was less frequent than expected in the T0 seed, perhaps due to a mixture of wild-type and *clf-50*/+ plants being transformed. 16 individuals from three independent T2 populations which segregated for *clf-50* were genotyped by PCR analysis at the *CLF* locus; it was found that not one plant with a wild-type phenotype was a *clf-50* homozygous mutant, whereas all those with a *clf* phenotype were homozygous *clf* mutants. Four individuals from each of five independent T2 families with a *clf* phenotype were genotyped by PCR for presence of the transgene; it was found that in a significant number of cases, these individuals carried the 35S::*SWN* transgene. This provided firm evidence that despite the absence of mutants with a *clf* phenotype in the primary transformants, 35S::*SWN* is unable to complement the *clf* phenotype.

This data confirms that the 35S::*CLF* transgene is able to rescue a *clf* mutant phenotype, however 35S::*MEA* and 35S::*SWN* are not. In order to ensure that the failure of 35S::*MEA* to complement a *clf-50* homozygote was not due to a lack of *MEA* expression, Northern blot analysis was performed in seven independent transformants to test levels of *MEA* RNA in vegetative tissue (figure 3.9). Although the levels of *MEA* RNA are variable, *MEA* transcript is easily detectable in all seven independent transformants, but not in wild-type control plants.

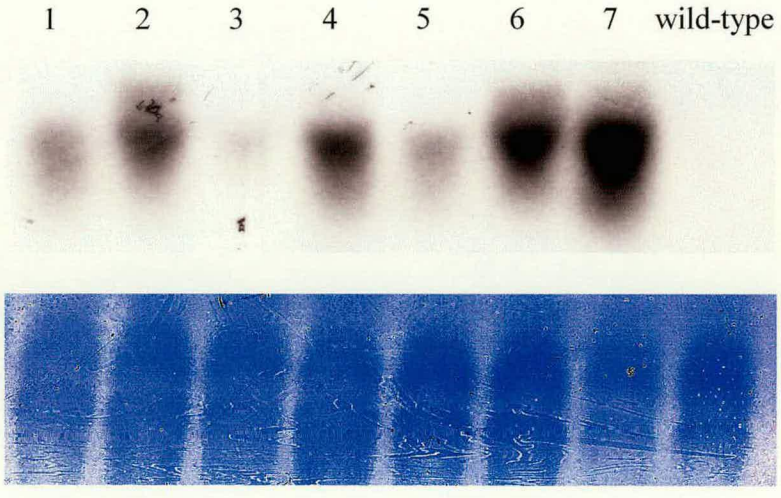


Figure 3.9 *MEA* is ectopically expressed in 35S::*MEA* lines
RNA was extracted from leaves of 7 independent 35S::*MEA* transformants and a wild-type control. RNA was separated by agarose gel electrophoresis, transferred to a nylon membrane and hybridised with a radiolabelled *MEA* probe.

Top panel shows *MEA* expression in all transformants at varying levels but none in the wild-type control. Bottom panel shows the same filter stained with methylene blue dye to demonstrate approximately equal loading of RNA. There is no detectable expression of *MEA* RNA in wild-type leaves.

Promoter fusion constructs were also made using the endogenous *CLF* promoter, to test whether *SWN* or *MEA* could complement *clf* when expressed in a manner similar to native *CLF* expression, in terms of both expression levels and expression patterns. Primary transformants were scored for *clf* mutant phenotype (table 3.3). However, even the control line p(*CLF*)::*CLF* was unable to fully complement *clf* homozygotes. 1 in 10 transformants showed a weak *clf* phenotype; the plants were significantly bigger, showed increased leaf size, greater numbers of lateral branches and reduced leaf curling when compared with null *clf* mutants (figure 3.7 B). Genotyping by Southern blot analysis confirmed that these plants with a weak *clf* phenotype carried the transgene and were homozygous *clf* mutants (data not shown).

The endogenous promoter used was therefore insufficient to drive *CLF* expression in a manner comparable to its native expression pattern. The fact that this promoter had previously been shown to complement *clf* when used in the context of a genomic fragment of *CLF*, suggests that important regulatory sequences must lie within the introns or 3'-sequences of *CLF*.

In the p(*CLF*)::*SWN* line primary transformants with a null *clf* phenotype were seen at a frequency of about 1 in 10 (table 3.3 and figure 3.7 E) and in the p(*CLF*)::*MEA* line approximately 1 in 7 transformants had a null *clf* mutant phenotype. Fourteen p(*CLF*)::*MEA* primary transformants were genotyped by Southern blot analysis for presence of the *MEA* transgene and genomic *CLF* locus. The plants with the *clf* phenotype were shown to *clf-50* homozygotes which also carried the p(*CLF*)::*MEA* transgene. These results support those from the promoter swap constructs using the 35S promoter, by showing that SWN and MEA are unable to complement a *clf* phenotype.

T2 progeny derived from T1 plants with CLF^{+} phenotype from all three lines were sown onto tissue culture media with kanamycin. In the case of p(*CLF*)::*MEA* and p(*CLF*)::*SWN* about one third of the families segregated for individuals with a *clf* phenotype with a frequency of about 1 in 4. In the cases of p(*CLF*)::*CLF* plants with a

weak *clf* phenotype were observed, however the severity of the phenotype was significantly reduced relative to the T1.

As an additional check to ensure the 35S::*MEA* and p(*CLF*)::*MEA* transgenes produced functional MEA protein, they were fertilised with pollen from a *mea-2* heterozygote. Unfortunately, the cross to 35S::*MEA* failed, and so only the F1 of the p(*CLF*)::*MEA* cross was analysed. Seed abortion in F1 siliques was scored, and individuals were identified showing ratios of seed abortion less than the 50 % observed in *mea* heterozygotes. If the p(*CLF*)::*MEA* transgene fully complemented a *mea* mutation, seed abortion would be expected to occur with a frequency of only 25 % (corresponding to the *mea*$^-$ transgene$^-$ ovules). However, the ratio of seed abortion was reduced to only 37 % (88:151). Statistical analysis through a chi squared test confirmed this ratio was significantly different from the 50 % abortion predicted if the transgene had no effect on seed viability. The ability of the p(*CLF*)::*MEA* transgene to partly complement a *mea* mutation provides further evidence that the *MEA* constructs are able to provide functional MEA protein. A cross between 35S::*SWN* and *clf*/+ *swn*/+ heterozygous double mutants was initiated to confirm that 35S::*SWN* could complement a *clf swn* double mutant, but due to time constraints the results of this cross are not yet available.

Taken together these results provide clear evidence that neither *SWN* nor *MEA* is able to regulate *AG*. Although *SWN* is able to act redundantly with *CLF* in controlling the expression of many genes, the reason why it does not do so for *AG* is unknown. However, these results support the phylogenetic analysis which suggests that the CLF protein is distinct from both SWN and MEA. The inability of both *MEA* and *SWN* to complement a *clf* mutant when expressed vegetatively must be due to divergence in protein sequence. Recent research has suggested an additional role for *CLF* in negatively regulating the expression of *MEA* in vegetative tissue; this can be seen by significant elevation of *MEA* transcript accumulation in the rosette leaves of *clf* mutant plants (Katz et al., 2004). This result supports the data that *MEA* is unable to complement *clf*, as it appears to be routinely mis-expressed in these mutants.

3.7 Identifying domains which distinguish the functions of CLF and MEA

The previous results indicated that MEA and CLF are not equivalent, which leads to the question of how these two proteins differ. This next section investigates this question using chimaeric fusions between CLF and MEA proteins to test for complementation of the *clf* phenotype.

Two different sets of chimaeric swaps were made. In the first set, the conserved SET domain and adjacent cysteine rich region were exchanged between MEA and CLF. This is the region of MEA and CLF which shows the greatest homology; the SET domain has been shown to be essential for CLF function *in-vivo*. The chimaeric gene fusions were made by Overlap Extension PCR (OE-PCR) designed to express fusions of MEA and CLF proteins in *Arabidopsis* under the control of the 35S constitutive promoter. 35S::*CLF*-[*MEA* SET] encoded the N-terminal half of CLF fused to a fragment of MEA containing the cysteine rich region, the SET domain and the C-terminal tail (see chapter 2). The construct 35S::*MEA*-[*CLF* SET] encoded the N-terminal half of MEA fused a fragment of CLF similar to that described previously for *MEA*.

These chimaeric swaps should indicate which half of MEA has diverged sufficiently for it to be unable to complement a *clf* mutant. Alternatively, it is possible that there are multiple regions of divergence between the two proteins separating their biological functions, and these may map to multiple locations across the proteins.

The second set of swaps exchanged the post SET domains CLF and MEA. The post SET domain was chosen because members of the CLF/SWN orthology group share homology in the C-terminal tail. However, MEA and E(Z) contain a shorter C-terminal tails and do not share these homologous regions. Constructs were named 35S::*CLF*-[*MEA* TAIL] and 35S::*MEA*[*CLF*-TAIL].

A T0 generation, comprising of the progeny of a *clf-50*/+ plant, was transformed using the four constructs and the T1 transformants scored for complementation of *clf* phenotype (table 3.4). In the 35S::*MEA*-[*CLF* SET] and 35S::*MEA*-[*CLF* TAIL] transgenic lines, individuals with a null *clf* phenotype occurred at frequencies of 1 in 5 and 1 in 6 respectively. This is not far different from the 1 in 6 frequency expected if the transgene fails to complement. Seven plants with a null *clf* phenotype for each of these constructs were genotyped by Southern blot analysis and were shown to be homozygous *clf* mutants (figure 3.10). Images of these plants can be seen in figure 3.11 A and B. This confirmed that these constructs did not complement. Also, the fact that all plants with a *clf* phenotype were *clf-50* homozygotes indicates that co-suppression was not a significant cause of the *clf* phenotype.

The inability of both 35S::*MEA*-[*CLF* SET] and 35S::*MEA*-[*CLF* TAIL] chimaeric transgenes to complement *clf*, is presumably due to a requirement for the N-terminal region of CLF. It is possible that differences in the SET and adjacent domains also specify the biological function of MEA and CLF, and this may be resolved by the analysis of the other chimaeras.

The ratio of primary transformants with a *clf* phenotype was significantly higher than expected in lines carrying the 35S::*CLF*-[*MEA* SET] and 35S::*CLF*-[*MEA* TAIL] transgenes (table 3.4). Furthermore, the degree of phenotypic change was extremely variable (figure 3.11 C-E), with some plants showing normal flowering time and only weak leaf curling, whereas others resembled null *clf* mutants and a third group showed an intermediate phenotype. For both lines the frequency of plants in the different classes was roughly consistent (figure 3.12), although it should be noted that the three classes of phenotype are subjective, and the phenotypic range was such that certain individuals could have been assigned to either of two classes.

Table 3.4 Phenotypes of primary transformants for chimaeric swaps

Construct	Number of primary transformants	Number with *clf* phenotype	Approximate ratio
35S::*MEA*-[*CLF* SET]	87	17	1 in 5
35S::*MEA*-[*CLF* TAIL]	61	9	1 in 7
35S::*CLF*-[*MEA* SET]	47	20	1 in 2
35S::*CLF*-[*MEA* TAIL]	73	33	1 in 2

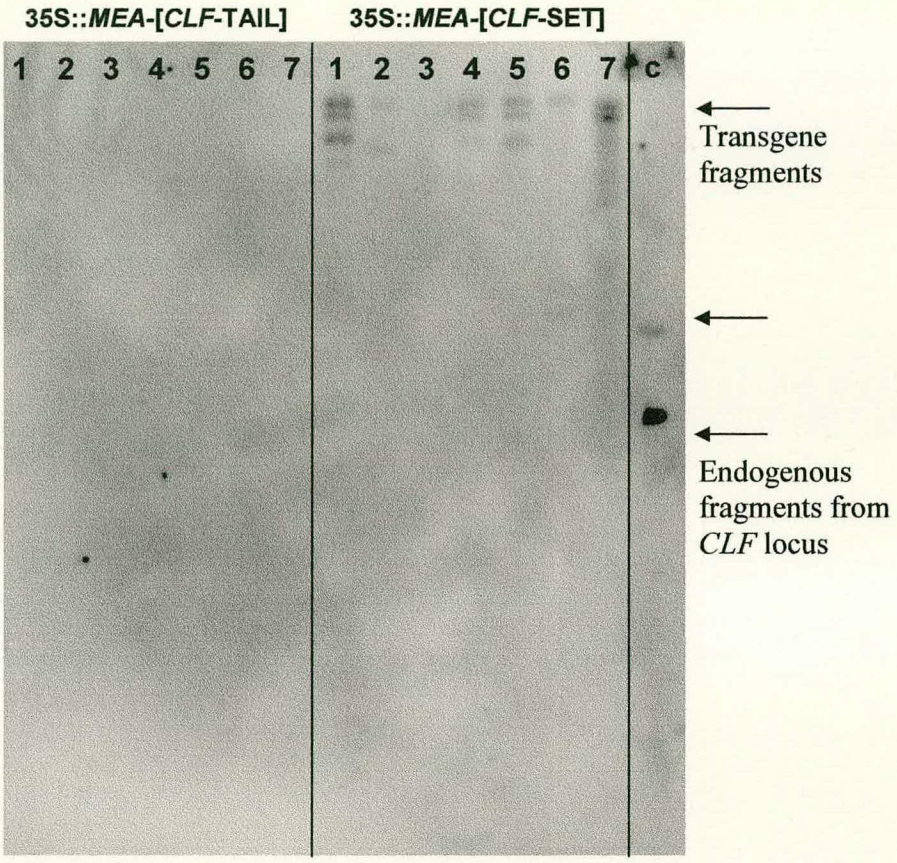


Figure 3.10 Southern blot analysis of primary transformants for 35S::*MEA*-[*CLF* SET] and 35S::*MEA*-[*CLF* TAIL]

Genomic DNA was extracted from 7 primary transformants with a *clf* phenotype for both experiments involving 35S::*MEA*-[*CLF* SET] and 35S::*MEA*-[*CLF* TAIL] tarnsgenes. This was digested with *Eco*R V and hybridised to a *CLF* cDNA probe. The endogenous fragments from the *CLF* locus were absent in all individuals present despite its occurrence in a wild-type control (c). This indicates that all 14 individuals are homozygous for the *clf-50* mutation. The large fragments in the 35S::*MEA*-[*CLF* SET] line are likely be due to hybridisation with the transgene and variable flanking sites. These are not seen in the 35S::*MEA*-[*CLF* TAIL] primary transformants, because the *CLF* cDNA probe did not include the 5'-end.

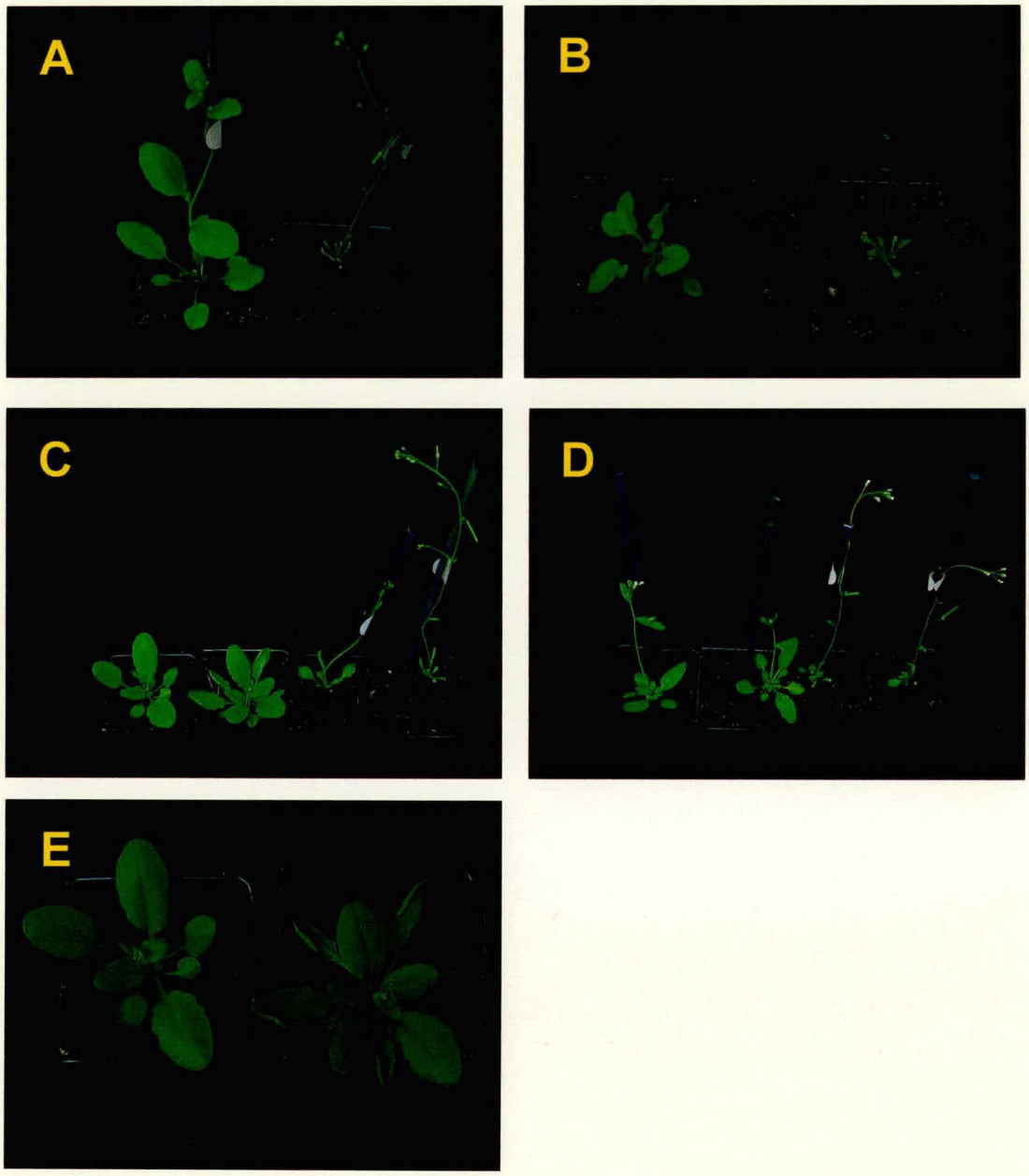


Figure 3.11 Phenotypes of plants with *MEA* and *CLF* chimaeric transgenes
A. 35S::*MEA*-[*CLF* SET] in *clf* background shown against wild-type. B. 35S::*MEA*-[*CLF* TAIL] in *clf* background shown against wild-type. C. & D. Range of phenotypes observed in the 35S::*CLF*-[*MEA* SET] and 35S::*CLF*-[*MEA* TAIL] lines. Degrees of severity vary from left to right, from individuals with a wild-type phenotype through to those with a strong *clf* phenotype. E. An individual carrying the 35S::*CLF*-[*MEA* SET] transgene and displaying only slight leaf curling is shown against an individual with a wild-type phenotype.

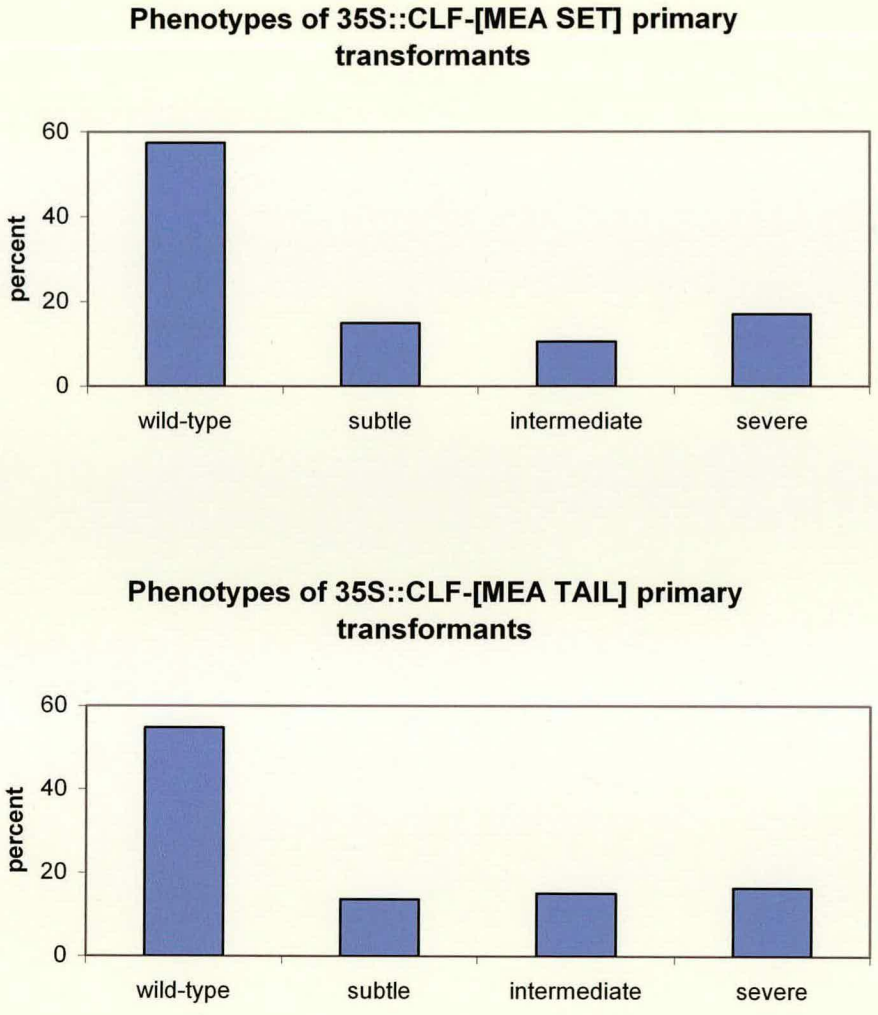


Figure 3.12 **The proportions of the different phenotype classes observed in plants carrying 35S::*CLF-MEA* chimaeric transgenes**

In order to test whether the strong and intermediate phenotypes were due to the plants being *clf-50* homozygotes rather than a direct result of presence of the transgene, three individuals with an intermediate phenotype and four with a severe phenotype were genotyped by Southern blot analysis (figure 3.13). In both lines, individuals with strong and intermediate *clf* phenotypes were identified which were CLF^{+}/CLF^{+} or $CLF^{+}/clf\text{-}50$ in genotype; this confirmed the suspicion that the high proportion of mutant phenotypes was caused by the transgene. This phenomena was more pronounced in the 35S::*CLF*-[*MEA* TAIL] line where all seven of the individuals tested carried one or two CLF^{+} loci.

There are two possible reasons why the presence of transgenes may confer dominant mutant phenotypes. The first is through co-suppression of the genomic *CLF* locus and the second is through displaying dominant-negative (antimorphic) effects of the chimaeric protein. In the first scenario, post-transcriptional gene silencing results in degradation of the RNA from the chimaeric fusion after transcription; however the endogenous *CLF* locus is also silenced due to presence of a related gene (the chimaeric fusion). In the second scenario, the chimaeric transgene produces functional protein which is actively expressed and produces a chimaeric protein which is able to enter the complex in which CLF is normally present. It is, however, unable to provide *CLF* activity and effectively poisons the complex. Furthermore, because of the high expression of the transgene under the control of the 35S promoter, it swamps the endogenous CLF causing a dominant negative effect, resulting in plants which carry the *CLF* genomic locus having a *clf* phenotype. Either mechanism can give a range of phenotypes due to variation in expression levels.

To attempt to distinguish between these two hypotheses the quantities of both endogenous *CLF* and total *CLF* transcripts (which will mostly originate from the transgene) were assayed by RT-PCR. If co-suppression was occurring in these lines, then less endogenous *CLF* RNA and less total *CLF* RNA may be observed in those primary transformants with severe phenotypes compared with those with only mild phenotypes. If the presence of the transgene caused dominant negative effects then

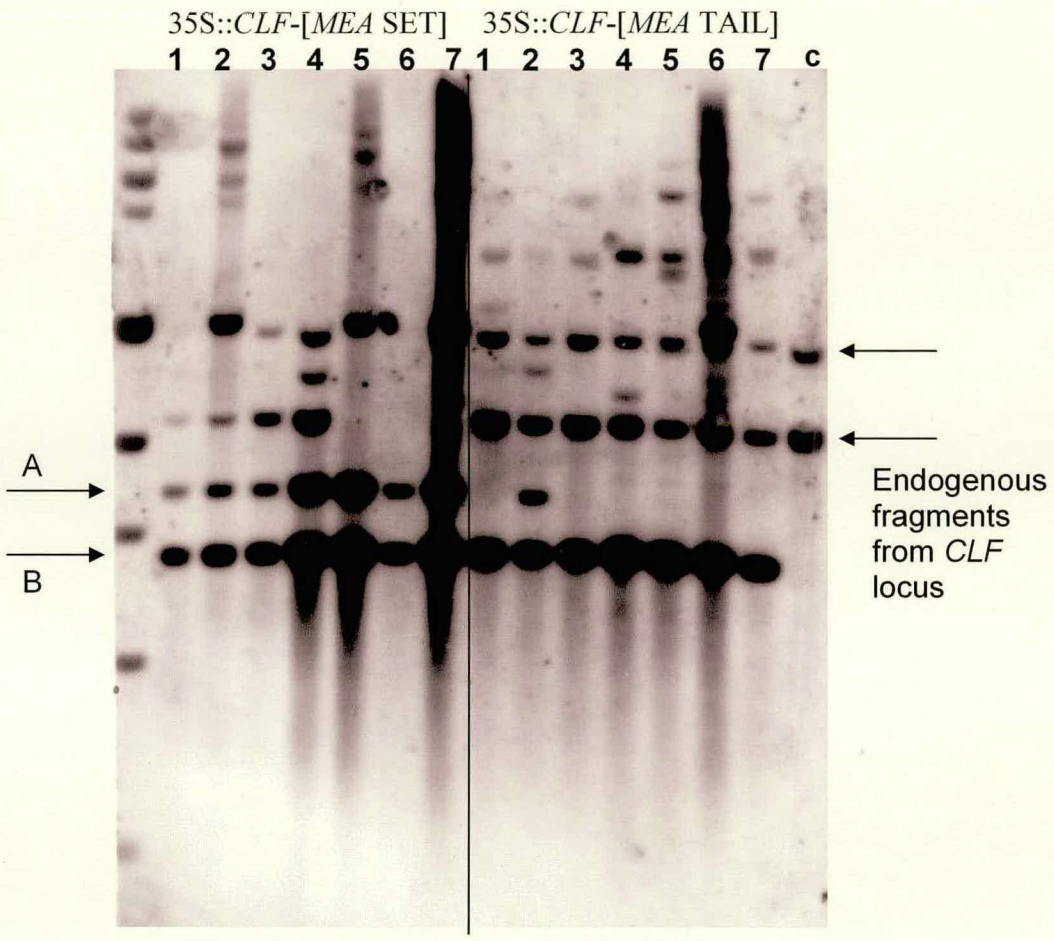


Figure 3.13 Southern blot analysis of primary transformants with severe and intermediate *clf* phenotypes

For each transgenic line individuals 1-3 have an intermediate *clf* phenotype and individuals 4-7 have a severe *clf* phenotype. Genomic DNA was cut with *Eco*R V and probed with DIG-labelled *CLF* cDNA. The band labelled B is present in all transformants although the second band labelled A is specific to those carrying the 35S::*CLF*-[*MEA* SET] transgene. For 35S::*CLF*-[*MEA* TAIL] all individuals carry one or more CLF^+ alleles, whereas in 35S::*CLF*-[*MEA* TAIL] only individuals 1-4 carry CLF^+.

greater levels of total *CLF* RNA may be observed in those individuals with a strong phenotype. If the phenotypes observed are due to co-suppression of the endogenous loci then it may be possible to observe lines where silencing does not occur, and these may give an insight into whether the transgene complements.

RNA was extracted from leaves of nine individuals in each line and used as a template in a reverse transcriptase reaction to synthesise cDNA. The endogenous *CLF* transcripts were amplified separately from the transgenic *CLF* transcripts using the primers G11 and AB6; AB6 was located in the 3'-untranslated region of *CLF* which was absent in the chimaeric constructs. Total *CLF* RNA was amplified using G11 and YC2 primers which amplify an internal *CLF* fragment common to both *CLF* and the transgenes. A third reaction was also performed to amplify *EIF-4a*, using the primers EIF-B22I and EIF-T22; this was used as a control to ensure that levels of template were even in all reactions (Metz et al., 1992). Because *CLF* under control of the 35S promoter is likely to be expressed at far higher levels than when under the control of the endogenous promoter the number of amplification cycles was altered accordingly. The amplification reaction of total *CLF* was cycled only 12 times, whereas the reactions for endogenous *CLF* and for *EIF-4a* were cycled 18 times. The reactions were then run on an agarose gel to separate fragments based on molecular weight, transferred to a nylon membrane and hybridised to a Dig-labelled probe for either *CLF* or *EIF-4a*. For total and endogenous *CLF* a slightly larger band can be seen in both lines; this is due to contamination of RNA preparations with genomic DNA, and represents amplification of a genomic fragment larger than the cDNA fragments due to presence of introns. This is not observed in *EIF-4a* because the genomic locus contains an intron so large it is not amplified under the PCR conditions used (figure 3.14).

The 35S::*CLF*-[*MEA* TAIL] shows RNA expression at a consistent level in all primary transformants (with the exception of number eight where the reactions have clearly failed). Endogenous *CLF* RNA is expressed in all these lines, indicating that none of these nine individuals are homozygous *clf* mutants. From these data it is not possible to

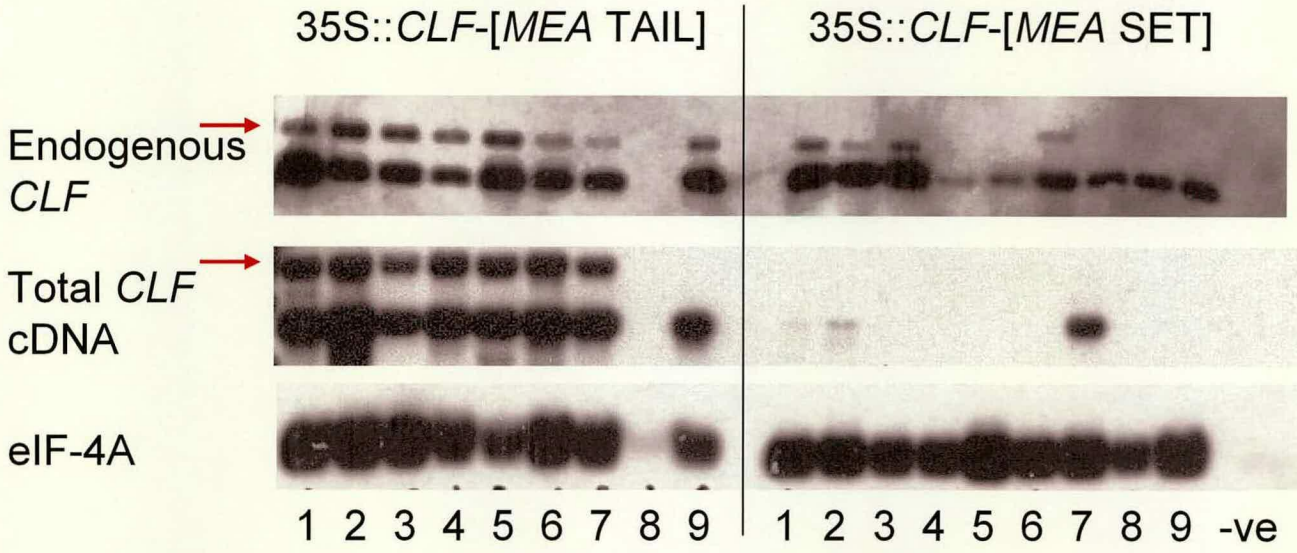


Figure 3.14 RT-PCR showing expression of endogenous and total *CLF* RNA in primary transformants for 35S::*CLF*-[*MEA* SET] and 35S::*CLF*-[*MEA* TAIL] transgenic lines.

Individuals 1-3 show weak leaf curling, 4-6 show an intermediate phenotype and individuals 7-9 show a severe *clf* phenotype. Reaction 8 for 35S::*CLF*-[*MEA* TAIL] failed for unknown reasons. Red arrows show larger bands corresponding to genomic DNA contamination. In 35S::*CLF*-[*MEA* TAIL] individuals there is fairly uniform *CLF* expression, perhaps more consistent with dominant-negative effects than co-suppression. In 35S::*CLF*-[*MEA* SET], the plants with weaker phenotypes show higher levels of *CLF* than those with intermediate and severe phenotypes, resembling co-suppression. The level of total *CLF* is low in almost all 35S::*CLF*-[*MEA* SET] lines, this may be due to post-transcriptional gene silencing. The phenomena is unlikely to be due to technical reasons because repeated experiments (using the same cDNA) gave similar results.

draw a conclusion concerning the dominant negative or co-suppressive nature of this transgenic line.

In the 35S::*CLF*-[*MEA* SET] line endogenous *CLF* expression can be observed in all individuals, therefore again no individual is a *clf-50* homozygote. However, the three plants with only weak leaf curling (1-3) have a higher level of expression of endogenous *CLF* than those with an intermediate or severe *clf* phenotype. Whilst this could be due to co-suppression of the genomic *CLF* locus in the other lines; it is important to note this trend is not continued and that those individuals with a severe *clf* phenotype have a higher level of endogenous *CLF* than individuals 4 and 5 which have only an intermediate phenotype. These differences could also be partly explained by differences in the number of copies of the genomic *CLF* locus. These data for the 35S::*CLF*-[*MEA* SET] line are more consistent with co-suppression than with dominant negative effects.

The total *CLF* expression is also very low in the 35S::*CLF*-[*MEA* SET] plants with weak expression in individuals one and two and strong expression only in individual seven. It is not clear why expression should be significantly lower in this line than 35S::*CLF*-[*MEA* TAIL], especially when there is no correlation between individuals expressing elevated levels of total *CLF* and those expressing high levels of endogenous *CLF* unless different degrees of co-suppression were observing in each line.

In order to attempt to further characterise these lines T2 seed was sown derived from individuals carrying severe, intermediate, weak *clf* phenotypes, and also those which appeared wild-type. It was found that the progeny of lines which previously gave severe and intermediate phenotypes, continued to give individuals with severe phenotypes in the T2 generation. However, a small number of lines were identified where all the T2 progeny appeared wild-type. Individuals from these lines were genotyped by PCR at the *CLF* locus, however, no *clf-50* homozygotes were identified.

It is not possible to conclude whether the CLF N-terminal chimaeras were able to complement a *clf-50* mutant. It is however, possible to conclude that the CLF and MEA proteins do not have different functions solely because of differences in the SET and adjacent cysteine-rich domain. It is highly likely that some, or all, of the factors which recruit CLF and MEA to there specific biological functions lie in the N-terminal portion of the proteins.

4 Characterisation of Protein-Protein Interactions between Pc-G Genes

4.1 Introduction

The approach taken in the previous chapter to identify regions of CLF and MEA that specify biological function, involved choosing domains based purely upon sequence similarity. Whilst data was presented suggesting that CLF likely acts as a histone methyltransferase and that this activity is mediated, at least in part, through the SET domain, little consideration was given to other domains and what their functions may be. An understanding of the function of these domains, and whether they have diverged between MEA and CLF, may show how CLF and MEA differ. These data would help explain why ectopic expression of *SWN* and *MEA* is unable to complement a *clf* phenotype.

In addition to choosing areas that are well conserved and expressing chimaerical protein fusions *in-vivo*, a second approach was taken to identify factors which may distinguish the biological functions of CLF, SWN and MEA. In *Drosophila*, although E(Z) acts as a histone lysine methyltransferase, it can only confer this activity as part of a multimeric protein complex (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). Therefore, the ability of E(Z) to interact with other proteins is vital to its biochemical function. It may be that in plants, where several E(Z) homologues have evolved, that they are recruited to different complexes, due to altered affinities for protein partners.

An E(Z) complex has been purified from *Drosophila* embryos containing the Pc-G proteins ESC and SU(Z)12 (Müller et al., 2002). Not only are these three proteins conserved in many organisms including mammals and nematode worms, but so are the interactions between them (Sewalt et al., 1998; Xu et al., 2001). There are homologues

of both ESC and SU(Z)12 in *Arabidopsis* and there is significant evidence for their involvement in Pc-G complexes similar to those purified from *Drosophila* embryos.

The aim of this chapter is to identify further functional domains of CLF, with the goal of comparing the corresponding domains of MEA. By investigating the protein-protein interactions of the *Arabidopsis* Pc-G proteins it is hoped to identify interactions unique to CLF. Differences in the protein-protein interactions of MEA and CLF may be responsible for the recruitment of these proteins into different functional complexes. Additionally, the domains which mediate these interactions may be defined, and this in turn may help to identify small changes within these proteins that have profound effects of their roles *in-vivo*. This information could then be used to design new chimaeric constructs in which these domains are swapped.

4.2 Interaction of CLF, SWN and MEA with FIE

The *Drosophila* ESC-E(Z) partnership was shown to be conserved in plants through the interaction of FIE with MEA (Spillane et al., 2000). FIE and MEA have a shared role in repressing endosperm proliferation, and the closeness of their relationship has not only been shown by the physical interaction of the two proteins in yeast, but also through overlapping expression patterns and similar mutant phenotypes (Spillane et al., 2000; Luo et al., 1999).

The completion of the *Arabidopsis* genome sequence has confirmed that *FIE* is a single copy gene in *Arabidopsis* (Riechmann et al., 2000). *FIE* also differs from the other *FIS* genes, because it is expressed in vegetative tissues as well as endosperm and embryo tissues (Ohad et al., 1999) suggesting that *FIE* might have a role beyond seed development. This was confirmed when transgenic lines were created which supplied *FIE* activity in the seed enabling a rescue of *fie* ovules during seed development (Kinoshita et al., 2001). However, in these lines there was no detectable *FIE* activity during vegetative growth enabling the recovery of homozygous *fie*$^{-}$ plants. These

homozygous plants flower extremely prematurely, and show developmental abnormalities. Although the phenotypes of the rescued *fie*⁻ mutants are variable, they are similar to *clf swn* double mutants, for example giving minute early flowering plants with poorly differentiated organs. Both *fie*⁻ and *clf* mutants show mis-expression of floral homeotic genes. In addition to these similar mutant phenotypes, FIE has been shown to interact with CLF and SWN in yeast two-hybrid assays (Chanvivattana, 2002; MacDougall pers. comm.). Together, these observations suggest that FIE may act with MEA during seed development and in a different complex with CLF and SWN during vegetative growth. This would mean that the ESC-E(Z) partnership is also conserved in the repression of flowering.

Previous studies of the interaction of E(Z) with ESC have implicated a 33 amino acid region at the N-terminal region of E(Z) as essential for this interaction (Jones et al.,1998; Tie et al., 1998). The localisation of this domain is also conserved in *Arabidopsis*. In MEA and SWN the first 110 N-terminal amino acids are sufficient to interact with FIE (Spillane et al., 2000; Chanvivattanna, 2002).

To test whether CLF and FIE can also interact through the corresponding N-terminal region of CLF, yeast two-hybrid assays were performed. A portion of the *CLF* cDNA encoding amino acids 1-108 was cloned in frame as a 'prey' fusion to the yeast Gal4 transcriptional activation domain (TA) (see methods). This was introduced into HF7c cells, together with a 'bait' of full length FIE fused to the yeast Gal4 DNA-binding domain (DB). N-terminal truncations of MEA (Spillane et al., 2000) and SWN (Chanvivattana, 2002) were also introduced with the FIE bait for comparison. Yeast transformants were selected on plates lacking the essential amino acids leucine and tryptophan. The plasmid pBI-881, used for the TA fusion, contained the *LEU2* gene and pBI-880, used for the DB fusion, contained the *TRP1* gene. Only cells which contained both these plasmids were able to synthesise leucine and tryptophan and hence grow on media lacking these essential amino acids. A detailed description of the principles by which the yeast two-hybrid system operates is included in chapter 2. Physical

interaction between the two protein products was assayed by the expression of two reporter genes *HIS3* and *LacZ*. Expression of *HIS3* allows growth on plates lacking the essential amino acid histidine, and expression of *LacZ* can be quantitatively assayed by measuring β-galactosidase activity.

Yeast strains which carried DB-FIE with GAL4-TA (empty prey vector) were unable to grow on plates lacking histidine, despite control colonies growing well on plates lacking leucine and tryptophan (figure 4.1). Neither did they express significant levels of β-galactosidase. This indicates that the DB-FIE construct did not activate either reporter gene, and that there was no interaction between DB-FIE and GAL4-TA. However, cells containing DB-FIE with TA-CLF, TA-SWN or TA-MEA expressed the *HIS3* and *LacZ* reporter genes, (figure 4.1). This shows that FIE can interact with the N-terminal regions of all three *Arabidopsis* E(Z) homologues.

To investigate the degree of conservation between N-terminal regions of CLF, SWN, MEA and E(Z), the first 110 amino acids of each protein were aligned using the software MegAlign 5.03 (figure 4.2a). The N-terminal regions showed surprisingly little conservation with only 2 % of residues being conserved in all four proteins and approximately 10 % being conserved in three out of four proteins. As expected, the three *Arabidopsis* proteins showed most similarity. Not only has there been significant sequence divergence between the plant and animal lineages, but also within *Arabidopsis*.

Blast searches identified one other *Arabidopsis* protein, AtMYB5, which showed homology with the N-terminal 110 amino acids of CLF. Interestingly, the region of AtMYB5 showing homology with the CLF N-terminus was also located at the N-terminus, and although the first 20 amino acids showed little similarity, the spacing between the conserved residues and the start of the protein matched identically. However, the two residues previously identified as being conserved in E(Z) and the three *Arabidopsis* homologues are different in AtMYB5 (figure 4.2b). Overall AtMYB5 showed comparable similarity to the N-terminus of CLF (23 %) as SWN did (22 %).

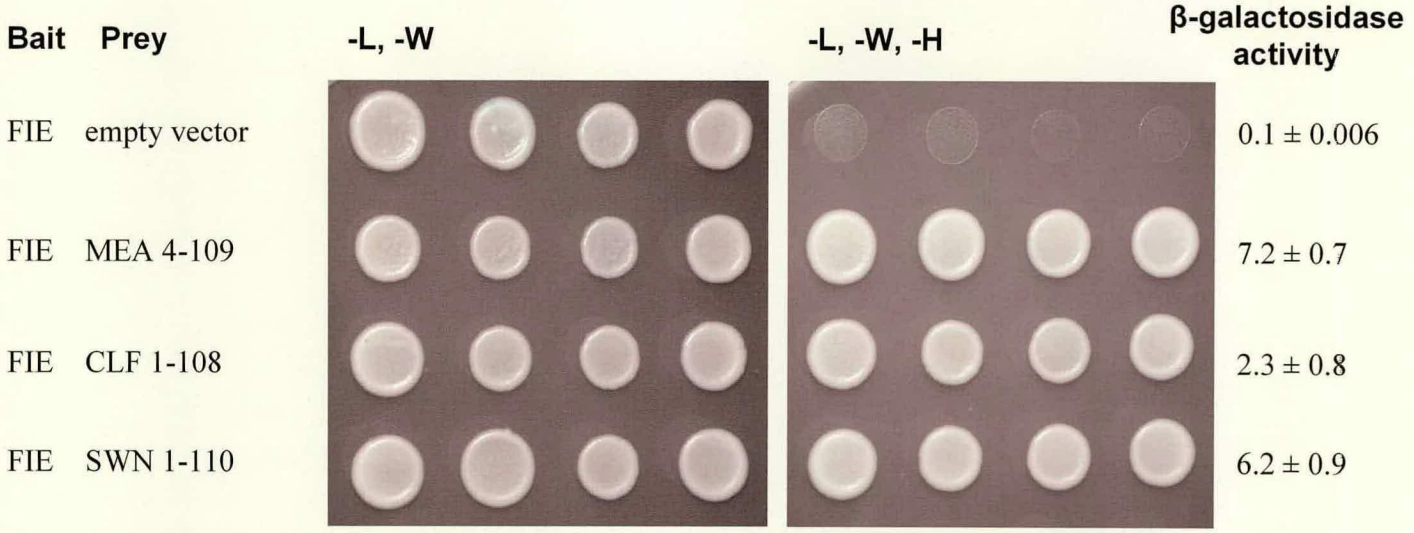


Figure 4.1 Interaction of FIE with *Arabidopsis* E(Z) homologues
Four independent HF7c transformants were plated on media lacking leucine and tryptophan (-L, -W) and on media lacking leucine, tryptophan and histidine (-L, -W, -H). All yeast colonies contain both bait and prey constructs, this can be seen by growth on –L, -W plates. However, only the transformants containing MEA, CLF or SWN bait fusions are able to grow on media also lacking histidine. This indicates that in these colonies the *HIS3* reporter gene has been activated by physical interaction of the two protein products. The average data for β-galactosidase activity is also shown. This is based on triplicate assays of independent transformants and the standard error is shown. Again, activity of the reporter gene is only observed in transformants containing MEA, CLF or SWN bait fusions.

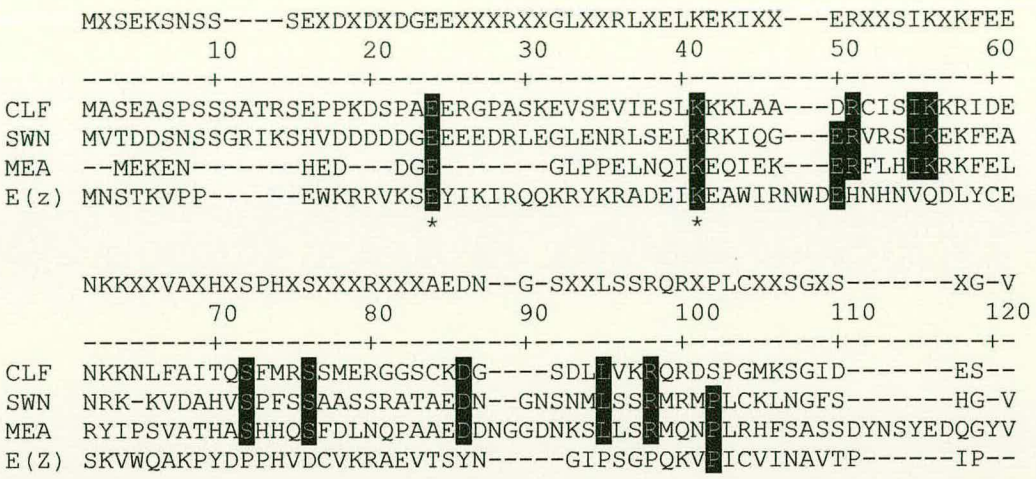


Figure 4.2a Alignment of N terminal regions of CLF, SWN, MEA and E(Z)
A consensus sequence is shown above the alignment. The black boxes show residues conserved in at least 3 out of 4 sequences, and the asterisks mark residues conserved in all 4 sequences.

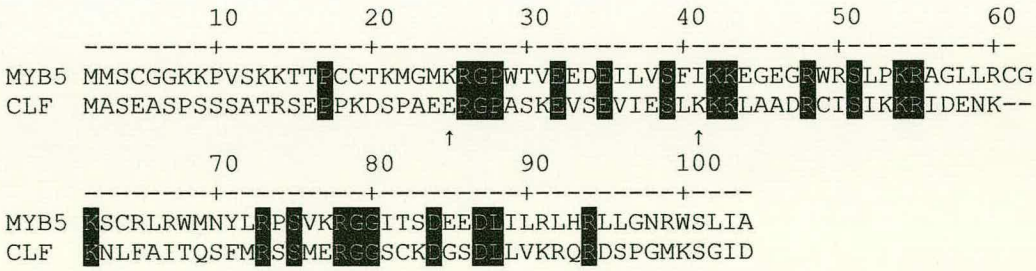


Figure 4.2b Alignment of CLF N-terminus with AtMYB5 N-terminus
The black boxes indicate residues conserved both sequences. The two residues which were previously shown to be identical in CLF, SWN, MEA and E(Z) are indicated with small arrows. In ATMYB5 neither of these residues is conserved.

Although there was little secondary conservation between the N-terminal regions of CLF, SWN, MEA and E(Z), the three proteins may have similar tertiary structures. To test this, the three protein sequences were submitted to Predict Protein (an Internet-based programme predicting protein structure) (figure 4.3). Where CLF and SWN appeared to be similar (an initial loop, followed by a large helix, then a much smaller loop-helix-loop), the structure of MEA and E(Z) appeared less similar. The structure proposed for MEA is difficult to interpret because there is little certainty over the structure taken by many of the residues. However, there is some structural homology between these four proteins. The N-terminal 100 amino acids of AtMYB5 were also submitted to Predict Protein (data not shown) however the predicted structure was different and contained more numerous smaller helix-loop-helix sheets.

4.3 Interaction between EMF2 and CLF

The previous experiments suggested that FIE may participate not only in a complex with MEA, but also in additional complexes involving CLF and SWN. This partnership between FIE and the three *Arabidopsis* E(Z) homologues is mediated through the same domain in each case, and clearly not something which separates biological function. It is unlikely that the interaction with FIE distinguishes CLF and MEA, as both can interact with FIE.

The PRC2 complex isolated by Müller et al. (2002) from *Drosophila* embryos also contained the Pc-G protein SU(Z)12. In *Arabidopsis* there are three homologues of *SU(Z)12*; *EMF2*, *FIS2* and *VRN2*, each have different biological functions. Although no physical interaction has been shown between these proteins and other *Arabidopsis* Polycomb proteins, genetic evidence suggested that these genes act together in a complex.

Genetic evidence also suggested an interaction between EMF2 and CLF. For example, weak *emf2* alleles phenocopy *clf* mutants and show strong genetic interactions with *clf* mutants. To test whether those genetic interactions might reflect physical interactions

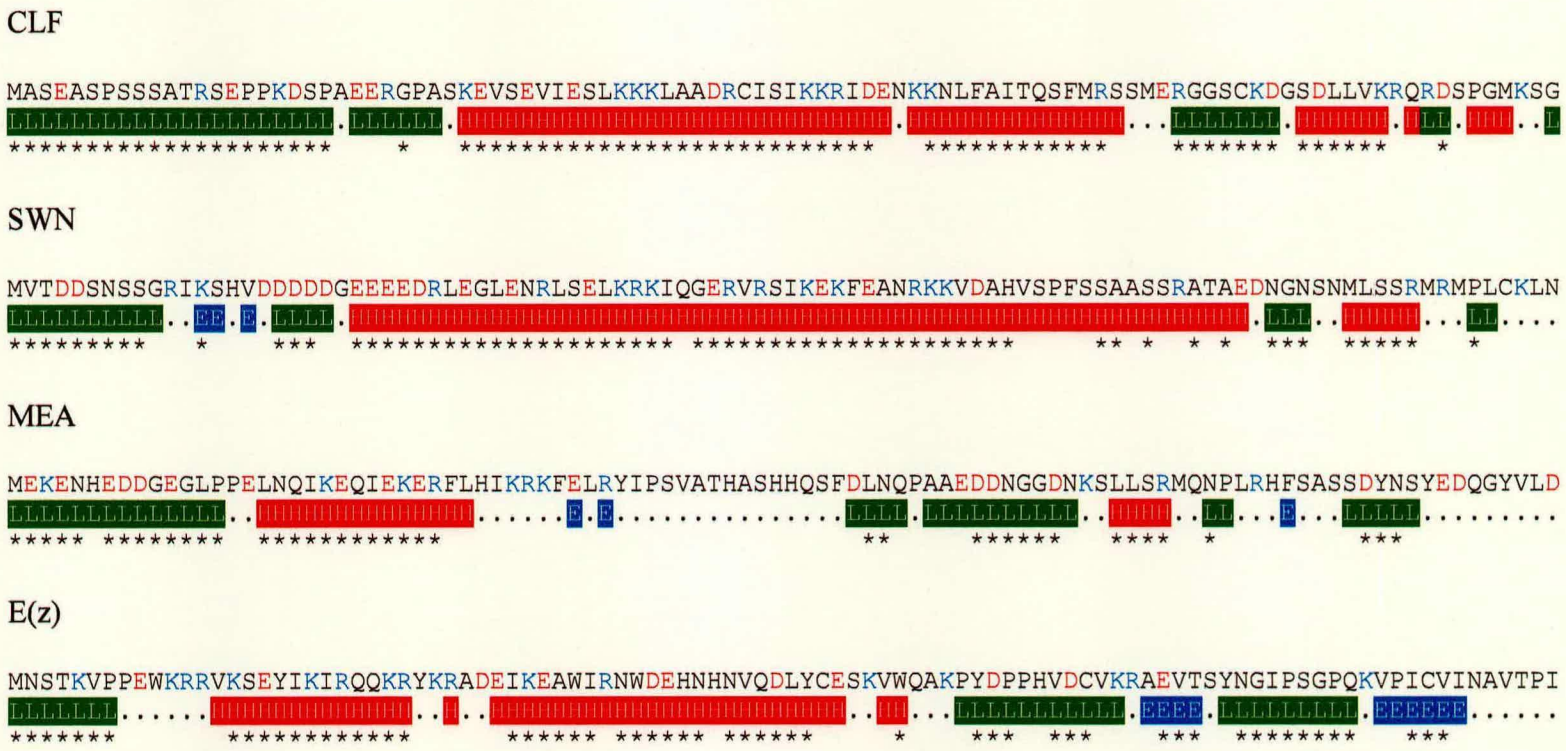


Figure 4.3 The N-terminal regions of E(Z) homologues show some structural conservation
HH=Helix EE=Extended sheet LL=Loop **=Residues with strong predictions

between the protein products, the yeast two-hybrid system was used to test the ability of CLF and EMF2 to interact.

Previous work (Colin MacDougall pers. comm.) suggested that the SET domain of CLF was often inhibitory for yeast two-hybrid assays. I therefore used a truncated CLF fusion protein which lacked the SET domain (pCMGB2, kindly supplied by Colin MacDougall). HF7c yeast cells were transformed with pCMGB2 and a series of prey constructs containing various truncations of EMF2 (pEMF2 1-5, see chapter 2); colonies were selected on plates lacking both leucine and tryptophan. Interactions between the two proteins were tested by examining growth on plates lacking histidine and measuring the quantity of β-galactosidase produced (figure 4.4). There was no growth of control colonies (pCGMB2 with pGAD424) on plates lacking histidine and only minimal levels of β-galactosidase, indicating that there was no auto-activation of reporter genes or interaction with empty vector. Full length EMF2 showed no interaction with pCMGB2, nor did truncations 2 and 3. However, truncations 4 and 5 (which contained the VEFS box) interacted with CLF in both assays, albeit with rather weak expression of β-galactosidase. Because of this, the *LacZ* assays were repeated using several time points and the weak *LacZ* activation was reproducible. It was puzzling that the construct expressing full length EMF2, including the VEFS domain, was unable to interact with CLF when the VEFS domain alone did. One possibility is that in the full length construct the interaction site may have been masked preventing access to the VEFS domain. Another possibility is that the full length EMF2 protein did not express due to instability or toxicity.

The smallest region of EMF2 which interacts with CLF corresponds to the last 113 amino acids. This region contains the VEFS box followed by a small adjacent terminal region. No function has previously been ascribed to this domain, although within this family it is the most highly conserved region. Figure 4.5 shows an alignment of the VEFS domain, with EMF2 and VRN2 being the most similar.

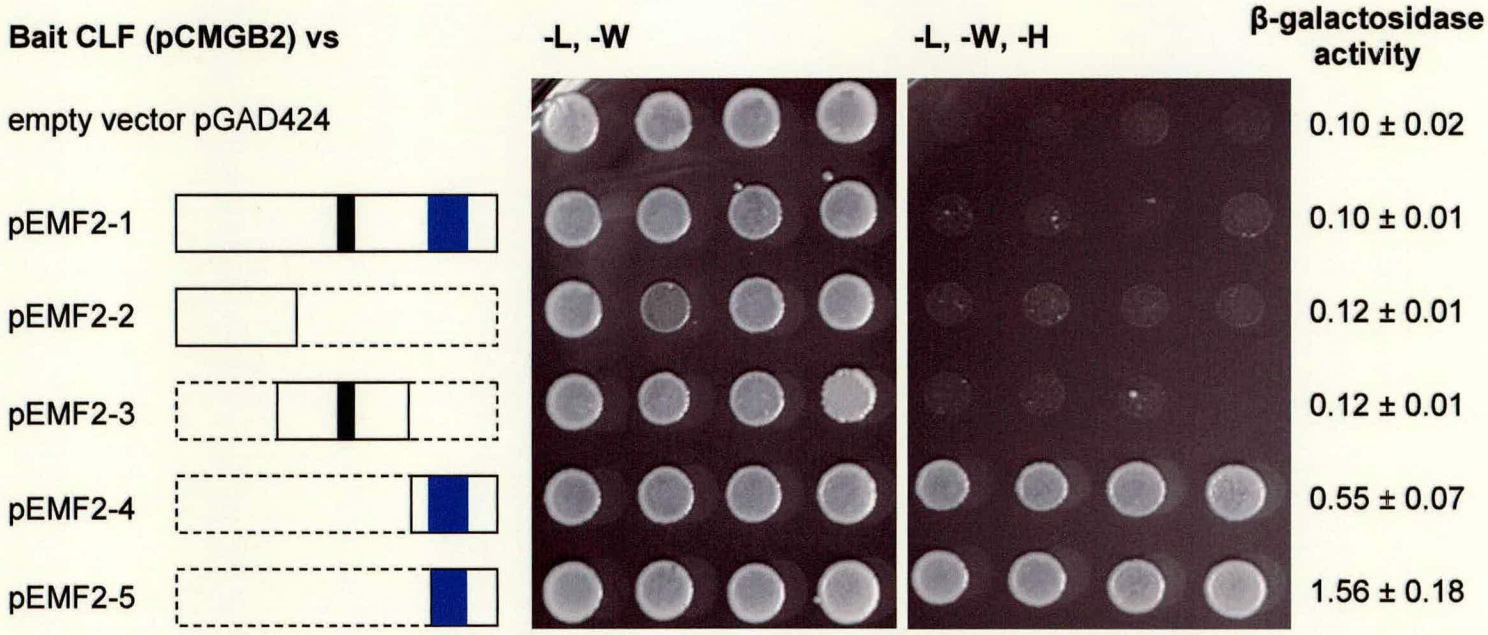


Figure 4.4 The interaction between EMF2 and CLF mediates to the VEFS domain of EMF2
Four independent HF7c transformants were tested for activation of the *HIS3* growth on media lacking histidine, and three independent transformants tested for activation of the *LacZ* reporter gene by measuring β-galactosidase activity. Only the prey constructs pEMF2-4 and pEMF2-5 showed an interaction with CLF. The schematic diagram shows approximate composition of these EMF2 truncations; the black box represents the zinc finger and the blue box the VEFS box.

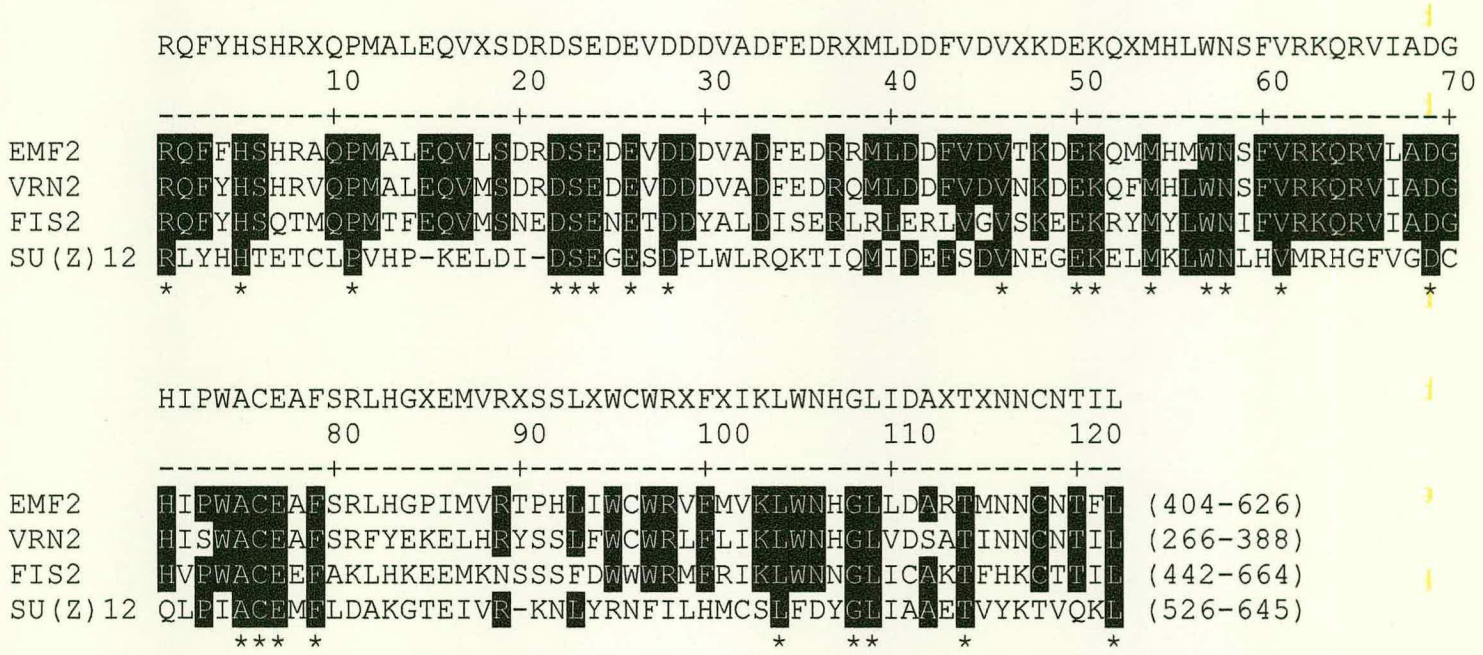


Figure 4.5 Alignment of VEFS domains in *Arabidopsis* SU(Z)12 homologues
Black boxes show residues conserved in 3 out of 4 sequences, residues conserved in all four sequences are shown with an asterisk. Diagram adapted from Yoshida et al. (2001).

In order to map the interacting region of CLF, a number of 'bait' fusion proteins with truncated CLF inserts were created and the tested for an interaction with EMF2. Interactions were tested in the same way by looking for growth on histidine deficient plates and measuring β-galactosidase levels in HF7c yeast cells. Control transformations containing either pCMGB2 with pGAD424 or pEMF2-4 with pGBT9 showed no activation of either *HIS3* or *LacZ* reporter genes (see figure 4.6). The first two C-terminal deletions tested were pCMGB2 and 3. pCMGB2 lacks the SET domain and pCMGB3 lacks the SET domain and adjacent cysteine rich domain. These both showed an interaction with EMF2, demonstrated by growth on plates lacking histidine and through the *LacZ* assay. The next two constructs tested, pCMGB4 and 5, had more substantial deletions. In pCMGB4 the protein was terminated midway through the C5 region and pCMGB5 contained only the first 139 amino acids. Neither of these constructs showed any interaction with EMF2 in either of the assays performed. To test whether disruption of the C5 region caused the failure of the interaction, a further six truncated CLF constructs were made. Constructs pCMGB6-8 encoded a new series of C-terminal truncations, spanning the region between the C5 domain and the pCMGB3 truncation. All of these interacted with EMF2 in both assays. A series of N-terminal truncations were also produced (pCMGB9-11). Each of these was progressively smaller until the last (pCMGB11) was a mere 74 amino acids spanning the C5 region. All of these constructs showed interaction with EMF2 through growth on histidine deficient plates and production of β-galactosidase. β-galactosidase activity suggested that the small truncations containing the C5 region interacted more strongly than the larger truncations did.

I also tested whether the reciprocal fusions could interact, i.e. GAL4 DB-EMF2 was tested with GAL4 TA-CLF. This interaction was once again tested in HF7c cells using both *HIS3* and *LacZ* reporter genes. In both cases no reciprocal interaction was detected (data not shown). However, there are many cases where *bona fide* interactions can be seen when one partner is fused to the bait construct but can not be detected if that partner is fused to the prey. A good example is the interaction between MEA and FIE which

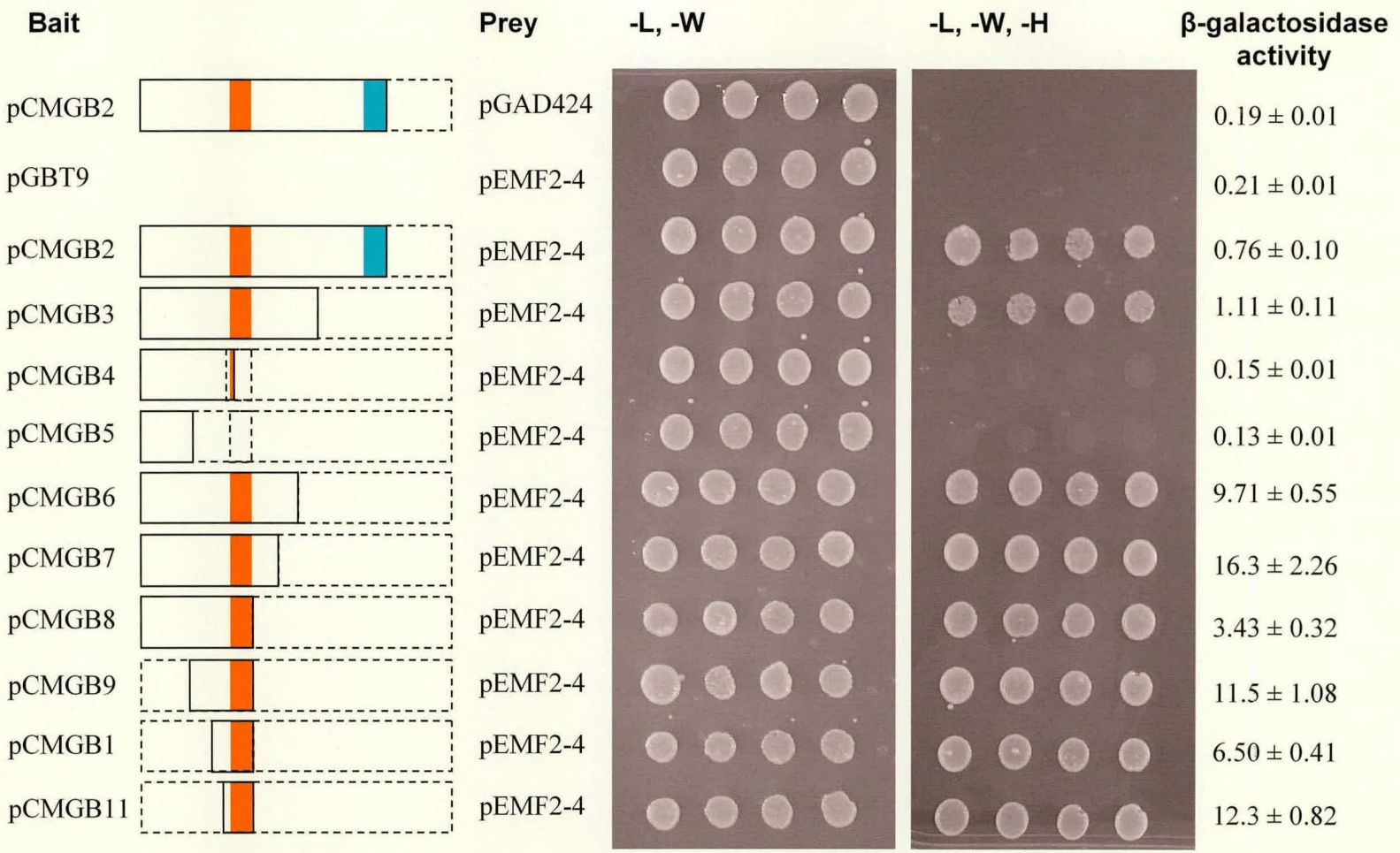


Figure 4.6 The interaction between CLF and EMF2 maps to the C5 region of CLF
The histidine assay is based on four replicates and the *LacZ* assay is an average of three replicates, also shown is standard deviation. On the schematic diagrams of CLF truncations, the C5 region is shown in orange, and the CXC region in blue.

can only be seen when FIE is used as the bait and MEA used as the prey (Spillane et al., 2000).

The C5 region was originally identified as a region rich in cysteine residues, that is well conserved between CLF, E(Z), MEA and SWN (Goodrich et al., 1997; Grossniklaus et al., 1998). This region is characteristic of E(Z) and its homologues, and does not occur in other SET domain proteins, or any other *Arabidopsis* proteins. The smallest fragment of CLF which can interact with EMF2 corresponds to amino acids 257-331, spanning the C5 region. Within this there are seven cysteine residues, of which five are conserved between SWN, MEA and E(Z) (Figure 4.7). In addition to these cysteine residues, there is also a small area of similarity at the N-terminus of the C5 region. There are also arginine, asparatic acid, histidine and tyrosine residues conserved in all four proteins. Despite this region being described in previous papers no known function has been assigned to it. These results show, that at least in the case of CLF, it has a role in mediating the physical interaction with another Polycomb protein, EMF2.

In order to examine the significance of the C5 region *in-vivo* a literature search was performed to identify any alleles of *E(Z), CLF* or *MEA* that specifically affected the C5 region. There was one allele, *e(z)-28*, which contained a single nucleotide mutation causing the fifth conserved cysteine to be substituted by a tyrosine in the protein product (Carrington and Jones, 1996). In *Drosophila*, this mutation causes a temperature sensitive phenotype. If the C5 region is generally required to mediate an interaction with a VEFS domain containing partner, it is possible that this mutation affects the ability of E(Z) to bind to SU(Z)12.

To test whether a mutation of the fifth conserved cysteine residue in the CLF C5 region (C317) would have a similar effect in disturbing the interaction with EMF2, the pCMGB11 construct was modified by site-directed mutagenesis. This introduced a mis-sense mutation that substituted the C317 for a tyrosine, and the resulting construct was designated pCLF-C317Y. This construct was then transformed into yeast cells, and its

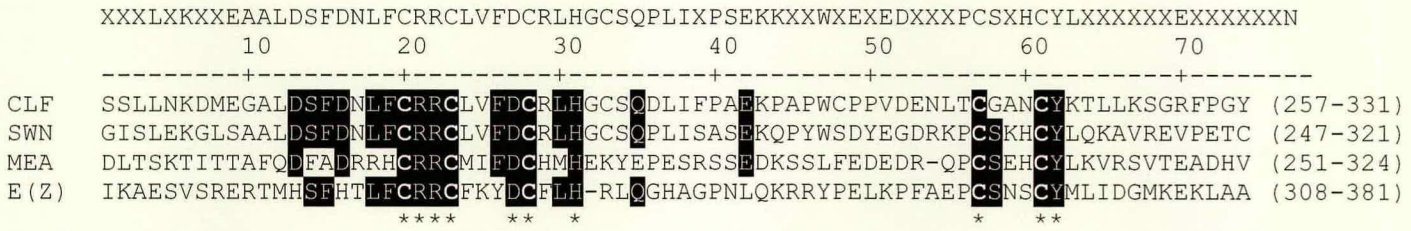


Figure 4.7 Alignment of C5 region of CLF, SWN, MEA and E(Z)
A consensus sequence is shown above the alignment. Black boxes indicate residues conserved in at least 3 out of 4 sequences, and asterisks mark residues conserved in all 4 sequences.

interaction with EMF2 was compared with that of CLF. However, HF7c cells transformed with pCMGB11 and pGAD4242 showed some autoactivation of the *HIS3* reporter gene, although activation of *LacZ* within acceptable background limits. In AH109 cells, which carry a stringent *ADE2* reporter gene, there was no growth of pCMGB11 and pGAD424 transformants on plates lacking adenine. As a second marker the *LacZ* assay was performed as before using HF7c cells.

For both *ADE2* and *LacZ* reporter genes, pCLF-C317Y showed an interaction with EMF2, comparable to that of pCMGB11 (figures 4.8). In the *LacZ* assay pCLF-C317Y gave somewhat lower β-galactosidase activity. However, the reduction was small, and it is questionable whether this is significant. In *Drosophila* the *e(z)-28* mutation is not null, but causes a temperature sensitive phenotype. With hindsight it may have been of more value to have changed this cysteine to a more drastically different residue (i.e. a charged one).

4.4 Specificity of Pc-G complexes

The data from the previous assays has provided evidence of a complex through which CLF most likely acts. Unpublished work by Christine Stock has shown a physical interaction in yeast two-hybrid between MEA and the VEFS box of FIS2 (figure 4.9). From this new information, two Pc-G complexes can be constructed, one involved in controlling aspects of seed development, the other repressing flowering (figure 4.10). The single copy Pc-G protein FIE is common to both complexes. They differ through the function of CLF and EMF2 in floral repression and MEA and FIS2 in seed development. Data presented in the third chapter showed that despite the similarities between *MEA* and *CLF*, *MEA* was not equivalent to *CLF* and was unable to rescue a *clf* mutant when ectopically expressed. In contrast, *EMF2*, *FIS2* and *VRN2* show virtually no sequence conservation outside the VEFS box and zinc finger motifs. Could it be that the interaction between the VEFS and C5 domains specifically recruits CLF and MEA to

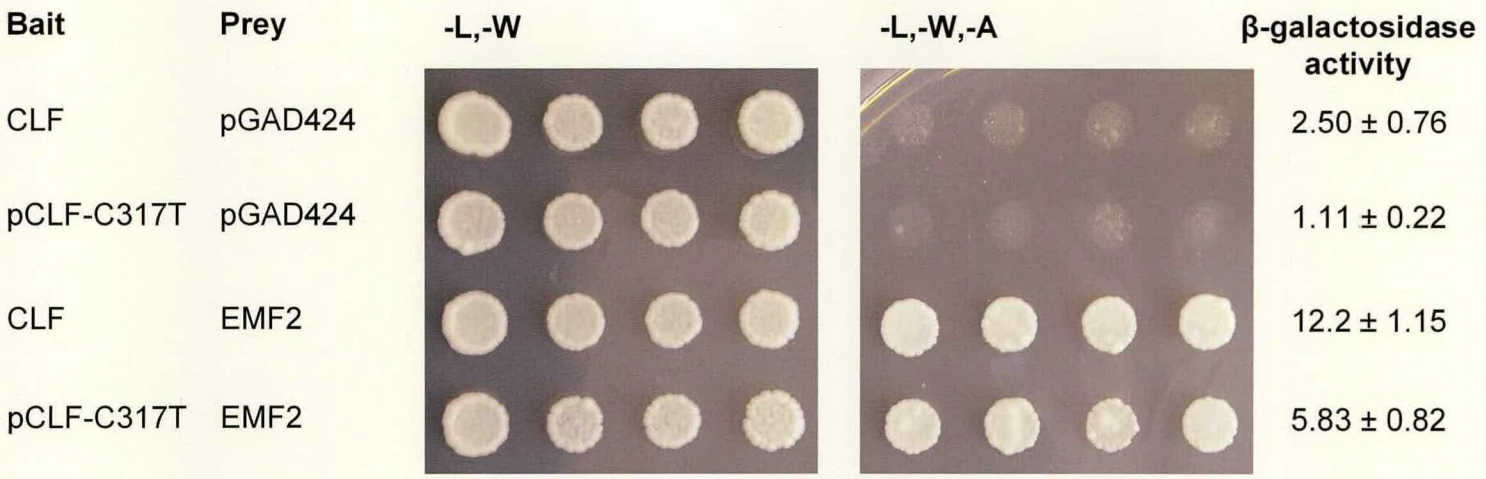


Figure 4.8 Effect of CLF C317T mutation on interaction with EMF2
Four independent transformants (in AH109 cells) were plated on –L, -W and –L, -W, -A media to assay activation of the *ADE2* reporter gene. β-galactosidase activity was assayed in three independent transformants (in HF7c cells). Average β-galactosidase activity is shown with standard deviation. The C317T mutation has little effect on activation of the *ADE2* reporter gene. Although there is less β-galactosidase activity associated with the C317T mutation, this is not likely to be significant (compare also auto-activation of CLF bait and pCLF-C317T bait).

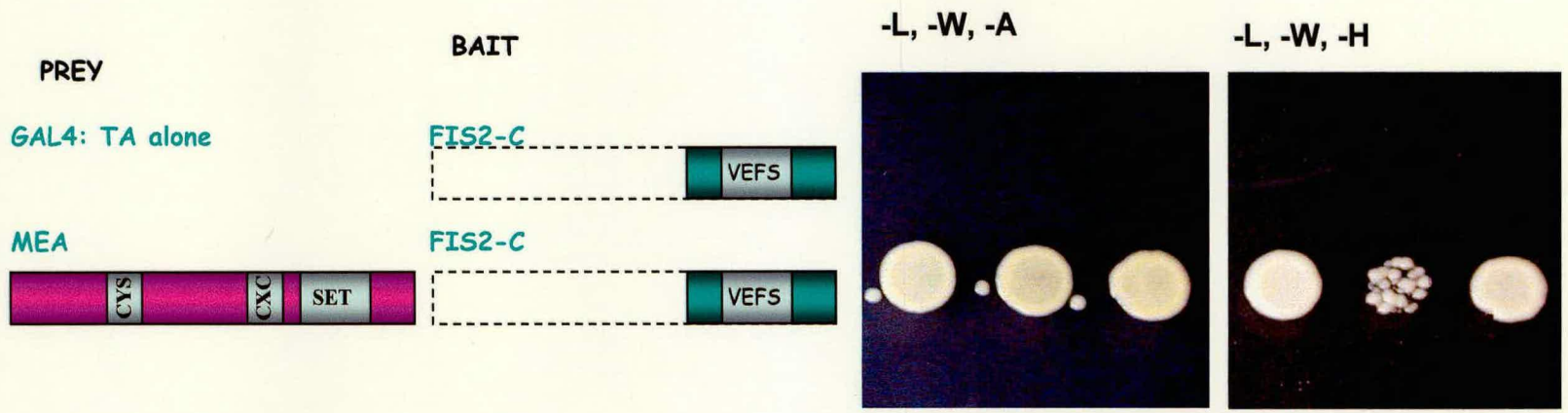


Figure 4.9 **MEA Interacts with the FIS2 C-terminus in yeast.**

Figure and data by Christine Stock

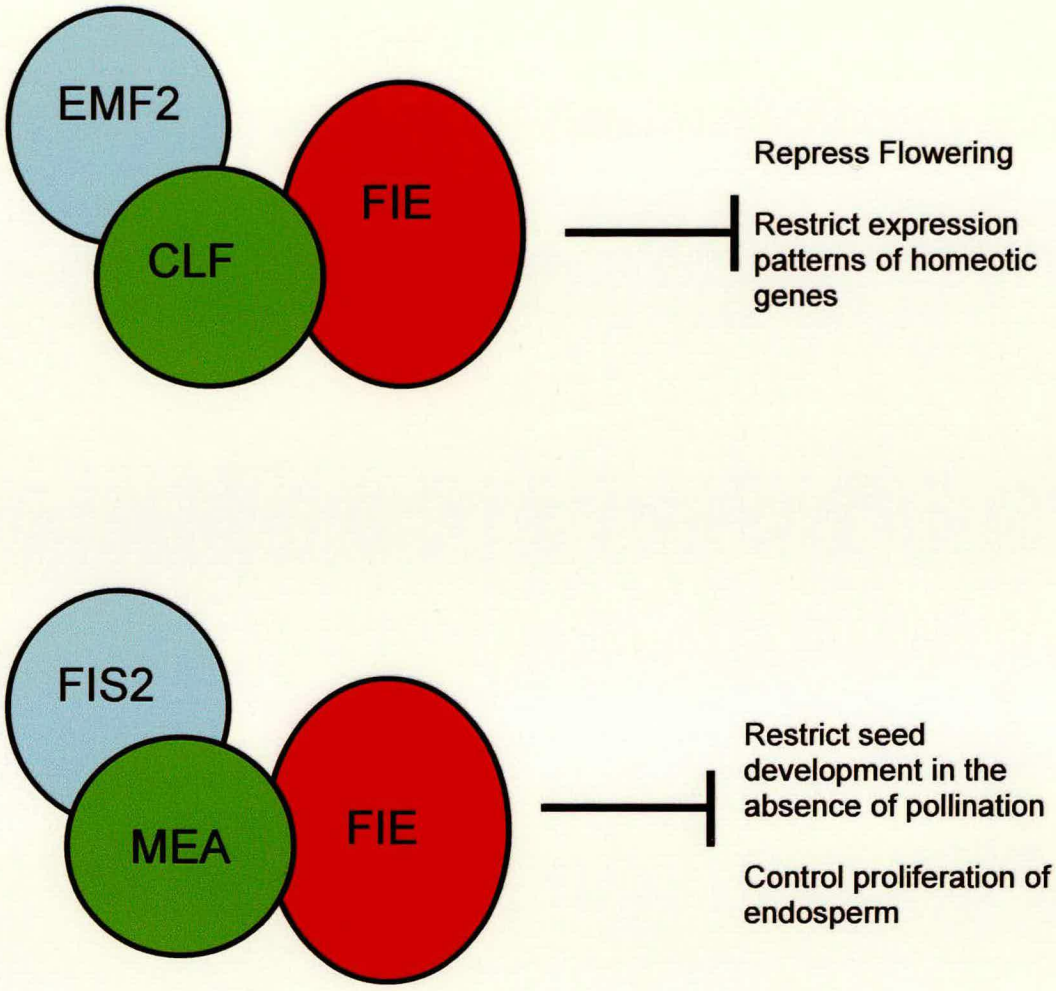


Figure 4.10 Proposed Polycomb complexes involved in seed and flower development
The colours indicate homology; green represents E(Z) homologues which contain a SET domain, adjacent cysteine rich regions and a C5 domain, blue indicates SU(Z)12 homologues which contain a zinc finger and VEFS box, and red represents the ESC homologue, FIE, which has 7 WD40 repeats. Areas of contact represent physical interaction of protein products.

the two separate complexes? This would mean that MEA was unable to complement *clf* due to an inability to interact with EMF2.

To investigate this further, yeast two-hybrid assays were performed to test the ability of the different *Arabidopsis* Pc-G proteins to interact with one another. In addition to establishing whether CLF and MEA differed in their ability to interact with different VEFS domain proteins, this also served to identify potential Pc-G partners for VRN2.

Initially, constructs similar to pCMGB11 were engineered to express the C5 domains of SWN and MEA in pGBT9, and constructs similar to pEMF2-4 were engineered to express the VEFS domains of FIS2 and VRN2 in pGAD424. Interactions were tested using the adenine reporter gene in AH109 and the *LacZ* reporter gene in HF7c. It was decided not to use the *HIS3* reporter gene in either yeast strain because both CLF and SWN C5 domains showed some autoactivation on histidine deficient plates.

Both CLF and SWN showed similar interactions with all three *Arabidopsis* VEFS proteins in yeast two-hybrid assays, demonstrated by both growth in the absence of adenine and production of high levels of β-galactosidase (figures 4.11 and 4.12).

MEA showed no interaction with any of the VEFS proteins in either assay (data not shown). To test whether there was a problem with the MEA C5 construct, it was sequenced and shown to be free of PCR-induced mutations. Two more constructs, MEA C5 II and III, were produced; although these still contained the C5 region, they were slightly larger containing additional residues both N and C-terminal of the original construct. Again, these constructs showed no interaction with any of the VEFS proteins (data not shown). When Christine Stock showed an interaction between MEA and FIS2, she used a different vector system (pBI-880 and pBI-88I) and a full length MEA construct. Two new constructs were produced, one with full length MEA and the other with a MEA truncation missing the SET domain. Neither of these showed a convincing

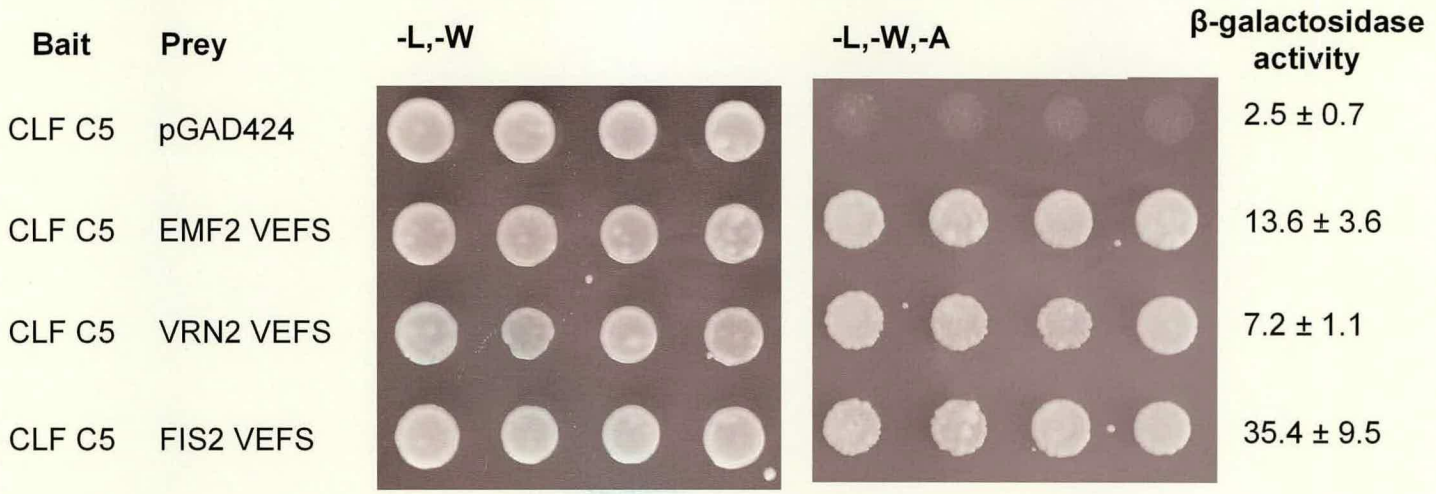


Figure 4.11 CLF interacts with all 3 *Arabidopsis* VEFS proteins
Interactions between the C5 region of CLF and the VEFS boxes of EMF2, VRN2 and FIS2 are shown using two independent markers. Activation of the *ADE2* reporter gene is shown for four independent transformants in AH109 cells. Activation of the *LacZ* reporter gene was quantified by measuring β-galactosidase activity in HF7c cells.

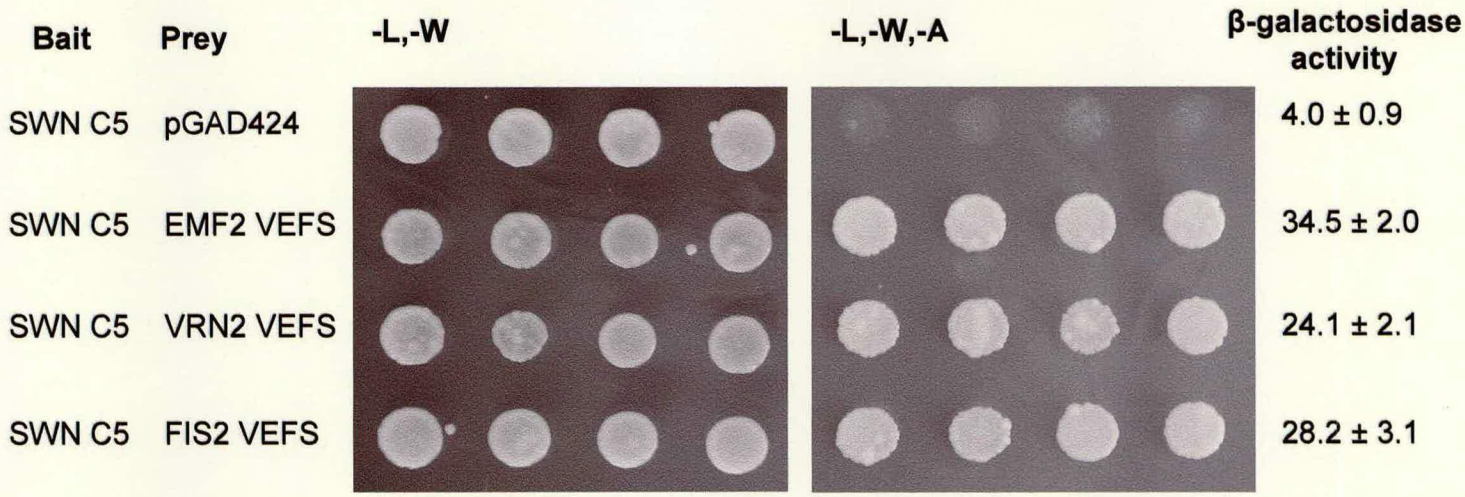


Figure 4.12 SWN interacts with all 3 *Arabidopsis* VEFS proteins
Interactions between the C5 region of SWN and the VEFS boxes of EMF2, VRN2 and FIS2 are shown using two independent markers. Activation of the *ADE2* reporter gene is shown for four independent transformants in AH109 cells. Activation of the *LacZ* reporter gene was quantified by measuring β-galactosidase activity in HF7c cells.

repeatable interaction on plates lacking either histidine or adenine. The production of β-galactosidase was not measured.

There are a number of reasons why the MEA constructs may have failed to interact with FIS2 in my assays. These could involve, incorrect folding leading to the masking of interaction site, general instability of the protein fusion in yeast or the protein may have been toxic to the yeast cells. It has been noted earlier that proteins can behave differently in slightly different fusions. This may explain why the interaction was apparent using the pBI-880 and pBI-881 vectors (Kolhami, 1998) and not with the pGBT9 and pGAD424 vectors (Clontech). An additional difference between Christine Stock's and my own experiments was that Christine Stock used MEA as an AD fusion, whereas I used it as a DB fusion. It has previously been noted that some yeast two-hybrid interactions can only be observed when a partner is in a particular bait or prey fusion, and not in the reciprocal fusion.

Both CLF and SWN showed no preference towards any of the three VEFS proteins as potential partners in yeast two-hybrid assays. It can not be said whether MEA has the ability to interact with EMF2 or VRN2, but based upon the results for CLF and SWN it is possible that it too would also show no specificity. It appears that the interactions mediated by the C5 and VEFS domains are not sufficient to recruit MEA and CLF into different complexes, and therefore other domains are most likely responsible for the inability of MEA to complement *clf* homozygotes. Such factors could be interactions with other, as yet unidentified, members of the multimeric protein complexes or, different interactions with transcription factors or different histone methyltransferase activity.

This experiment has, however, shown that both CLF and SWN are potential partners for VRN2. In addition to having overlapping expression patterns, both CLF and SWN have shown the ability to interact with VRN2 in yeast two-hybrid assays. This suggests that

there is a third Pc-G complex operational in *Arabidopsis* to control vernalization, with VRN2, FIE and either CLF or SWN providing the histone methyltransferase activity.

5 The Requirement for CLF and SWN in Epigenetic Control of the Vernalization Response

5.1 Introduction: a role for CLF and SWN in the vernalization response?

Exposure of *Arabidopsis* seedlings to a period of cold (vernalization) accelerates flowering subsequently when the plants are provided ambient growth conditions. The molecular mechanism for this has been shown to be through the repression of *FLC* mRNA in a mitotically stable manner, suggesting that this is achieved through an epigenetic mechanism (Sheldon et al., 1999; Gendall et al., 2001). More recent identification of proteins involved in the vernalization response suggested that this is mediated through a Pc-G complex. The VRN2 protein is required for mitotic stability of *FLC* repression and shows significant homology with the *Arabidopsis* Pc-G proteins FIS2 and EMF2 and the *Drosophila* protein SU(Z)12. The similarities between these proteins suggests that VRN2 mediates the response to vernalization as part of a larger Pc-G complex contributing to the maintenance of transcriptional regulation of *FLC* through the modification of chromatin (Gendall et al., 2001).

It has now been shown that vernalization causes changes in histone methylation at discrete sites within the *FLC* locus by increasing the quantity of K9 and K27 residues of histone H3 which are dimethylated (Bastow et al., 2004; Sung and Amasino, 2004). Furthermore the K9 methylation mark is lost in *vrn1* mutants and the K9 and K27 mark lost in *vrn2* mutants (Bastow et al., 2004). These data provide further evidence that VRN2 functions in a PRC2-like manner, i.e. not only is it structurally similar to a PRC2 member, it also has a similar biochemical activity towards K27 on histone H3.

The data presented in the previous chapter demonstrated the potential of both CLF and SWN to interact with VRN2. To test whether this interaction in yeast was biologically relevant in plants, the vernalization response of wild-type and *clf* and *swn* mutants were

compared, to see if either mutant impaired the vernalization response in a similar manner to *vrn2* mutants.

5.2 Testing the vernalization response in *clf* and *swn* mutants

Despite the fact that both *clf* and *swn* mutants have been well characterised (Goodrich et al., 1997; Chanvivattana, 2002), the role of *CLF* and *SWN* in the response to vernalization has not been investigated. This is because the *clf* and *swn* mutants which have been studied, have been in the Landsberg *erecta* (L*er*), Columbia (Col) and Wassileskija (Ws) backgrounds which have only a small vernalization requirement. To enhance the response, *clf* and *swn* mutants were crossed into the *fca* and FRI^{+} backgrounds, which are late flowering and confer a strong vernalization requirement.

The transition from vegetative development to flowering in *Arabidopsis* is controlled by several pathways which converge on the *FLC* gene. This acts as a strong floral repressor by negatively regulating the expression of genes that promote the floral transition, including *SOC1* (*AGL20*) and *FT* (Michaels and Amasino, 1999). In many laboratory ecotypes of *Arabidopsis*, the *FRI* or *FLC* gene is inactive so that levels of *FLC* mRNA are low, creating early flowering lines and removing the requirement for vernalization. Levels of *FLC* mRNA can be elevated to those seen in naturally occurring over-wintering accessions by the presence of a dominant FRI^{+} allele or a recessive mutation at the *FCA* locus (Johanson et al., 2000; Michaels and Amasino, 1999). *FRI* promotes the expression of *FLC*, and these elevated expression levels can only be removed by a period of vernalization. Rapid cycling ecotypes such as Columbia and Landsberg *erecta* have non-functional *FRI* alleles, and this removes the need for vernalization (Johanson et al., 2000). Recessive mutations in genes in the autonomous pathway have a similar effect in increasing *FLC* RNA levels. Autonomous genes such as *FCA* negatively regulate *FLC* expression; in *fca* mutants levels of *FLC* mRNA are able to rise and consequently flowering is delayed if no vernalization treatment is provided (Michaels and Amasino, 1999).

Plants homozygous for the *swn-3* mutation (Columbia) were crossed with a transgenic line carrying a functional *FRI*$^{+}$ transgene introduced in the Columbia background. The *swn-3* mutation is likely to be null as it contains a T-DNA insert in an intron in the middle of the gene, which should truncate the CXC and SET domains. The F2 seed were harvested, sown directly on to soil, and vernalized for six weeks at 4°C or stratified for two days at 4°C before being transferred to ambient growth conditions.

A homozygous *clf-2* mutant (L*er* background) was pollinated with pollen from an *fca-1* mutant (also L*er*). Individuals were identified in the F2 generation which were homozygous for both the *fca-1* and *clf-2* loci and their seed collected. This was sown directly on to soil and vernalized or stratified as before.

fca-1 clf-2 double mutants were grown under both long day and short day conditions and the flowering time measured by counting the number of rosette leaves and the age of the plant in days when the inflorescence meristem bolted to 1cm. The double mutants show a similar response to vernalization as *fca* mutants (figure 5.1 and table 5.1). Vernalized *clf fca* plants flower earlier and with fewer leaves than non-vernalized plants in both long and short day conditions. The *clf fca* doubles have slightly fewer leaves than *fca* mutants at flowering in both long and short days; this is probably due to the early flowering nature of *clf*, for example caused by mis-expression of *AG*. Due to these pleiotropic effects of *clf* on flowering time, it was not possible to exclude subtle effects of *clf* on vernalization response. Nonetheless, the data from short day particularly show that *clf fca* mutants retain a strong vernalization response.

For the *swn-3 FRI*$^{+}$ cross, 3/16 of the F2 population would be expected to be homozygous for the *swn* mutation and carry a functional copy of the *FRI*$^{+}$ gene. If *SWN* has a role in vernalization, then these individuals would likely show a reduced vernalization response. These data show many individuals flowering late (with more than 20 leaves in long day conditions) even after a period of vernalization (the data for

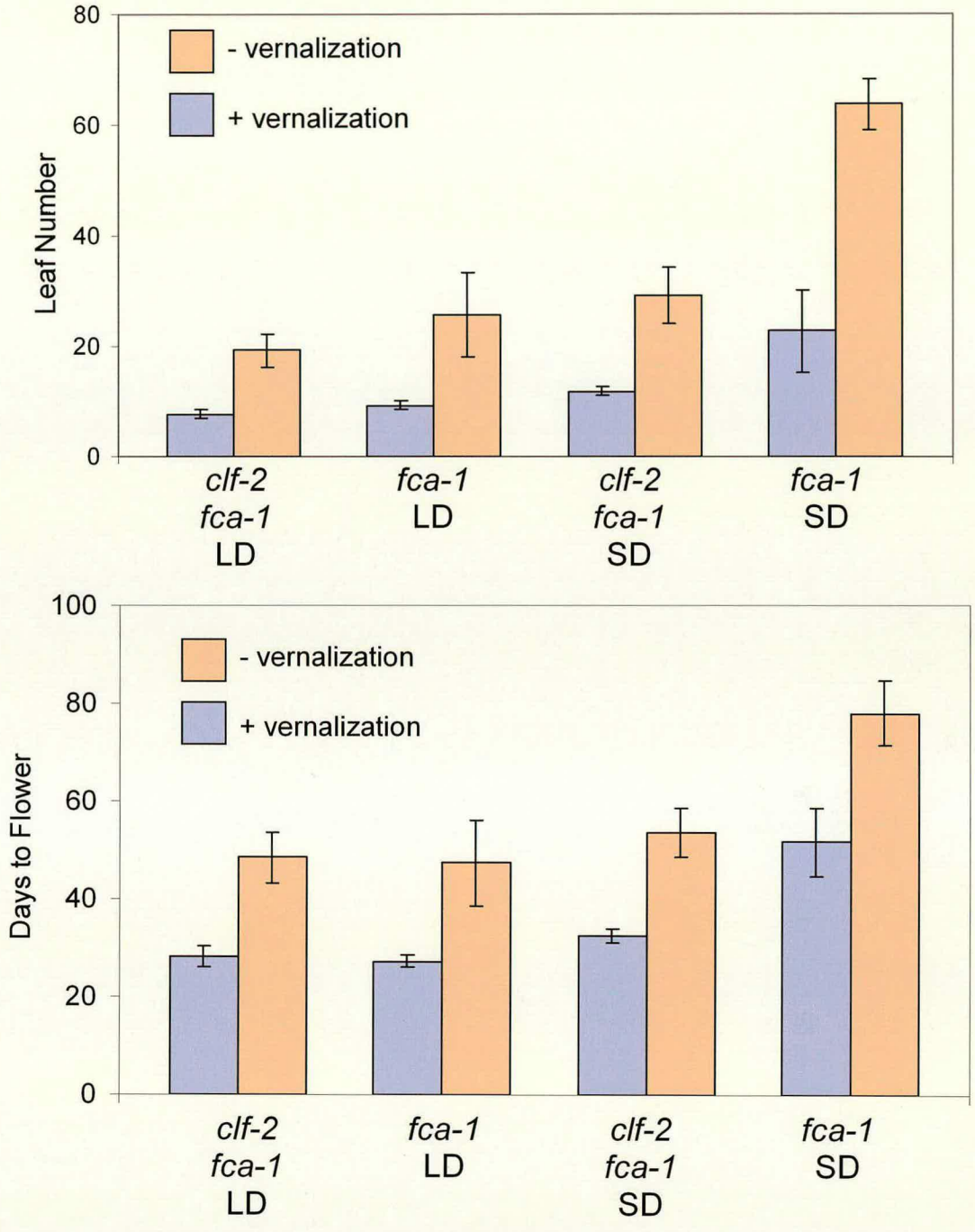


Figure 5.1 ***clf*** **mutants respond to vernalization**
Seed were sown directly onto soil and either treated at 4°C for six weeks or two days. The mean leaf number and age of the plant in days are shown when the inflorescence meristem bolts to 1 cm. Error bars show standard deviations based on about 40 plants. SD = short day LD = long day

Table 5.1 Flowering times for *clf fca* double mutants

	vern	Day Length	Age in days*	Number of rosette leaves*
clf-2 fca-1	+	Long	28 ± 2	8 ± 1
clf-2 fca-1	-	Long	48 ± 5	19 ± 3
fca-1	+	Long	27 ± 1	9 ± 1
fca-1	-	Long	47 ± 9	26 ± 8
clf-2 fca-1	+	Short	33 ± 1	12 ± 1
clf-2 fca-1	-	Short	54 ± 5	29 ± 5
fca-1	+	Short	52 ± 7	23 ± 8
fca-1	-	Short	78 ± 7	64 ± 5

* when inflorescence meristem has bolted to 1 cm
Errors shown are standard deviations based on approximately 40 individuals

leaf number is shown in figure 5.2; the data for age at flowering is similar but not shown). This frequency of late flowering individuals was considerably higher than the 3/16 which would be predicted if *swn* mutants showed a reduced vernalization response. To identify whether these plants flowered late as a result of *swn*, they were genotyped at the *SWN* locus by Southern blot analysis (data not shown). Plants homozygous for the *swn* mutation were observed with equal frequency among individuals flowering late and those flowering early. The homozygous *swn* plants were also genotyped for the presence of a functional FRI^{+} locus by PCR using the primers UJ24 and UJ34 (Johanson et al., 2000). The genomic *FRI* locus in the Columbia ecotype contains a 16 bp deletion; presence of an active FRI^{+} allele can be determined by the presence of a second larger band in the PCR reaction. Individuals were identified as swn^{-} FRI^{+} which flowered as early as the FRI^{+} parental line (14 leaves), indicating that SWN^{+} activity was not required for vernalization response.

In order to test for subtle differences in vernalization response the two experiments were repeated using a shorter vernalization treatment. The theory was that a marginal cold treatment may mean that the vernalization is suboptimal and subtle changes in the vernalization response may become apparent. To this end, the vernalization response in *clf fca* mutants was measured after a period of just three weeks treatment at 4°C. Under short day conditions both *fca* and *clf fca* double mutants flowered earlier after a three week vernalization treatment than when sown with no prior vernalization treatment (figure 5.3 and table 5.2). In both *fca* and *clf fca* double mutants the response to a three week period of vernalization is less than that seen after a six week vernalization. Statistical analysis through paired samples *t*-tests (using Analyse-it software) confirmed that the acceleration in flowering time caused by a three week vernalization was significantly different (95 % confidence interval) from non-vernalized plants in both *fca* and *clf fca* double mutants. However, in *fca* there is still a fairly big response to vernalization after three weeks cold treatment; plants flower with on average of 48 % of the leaves of non-vernalized individuals. In comparison, the response of *clf fca* double mutants to a three week vernalization treatment is more marginal; plants flower with

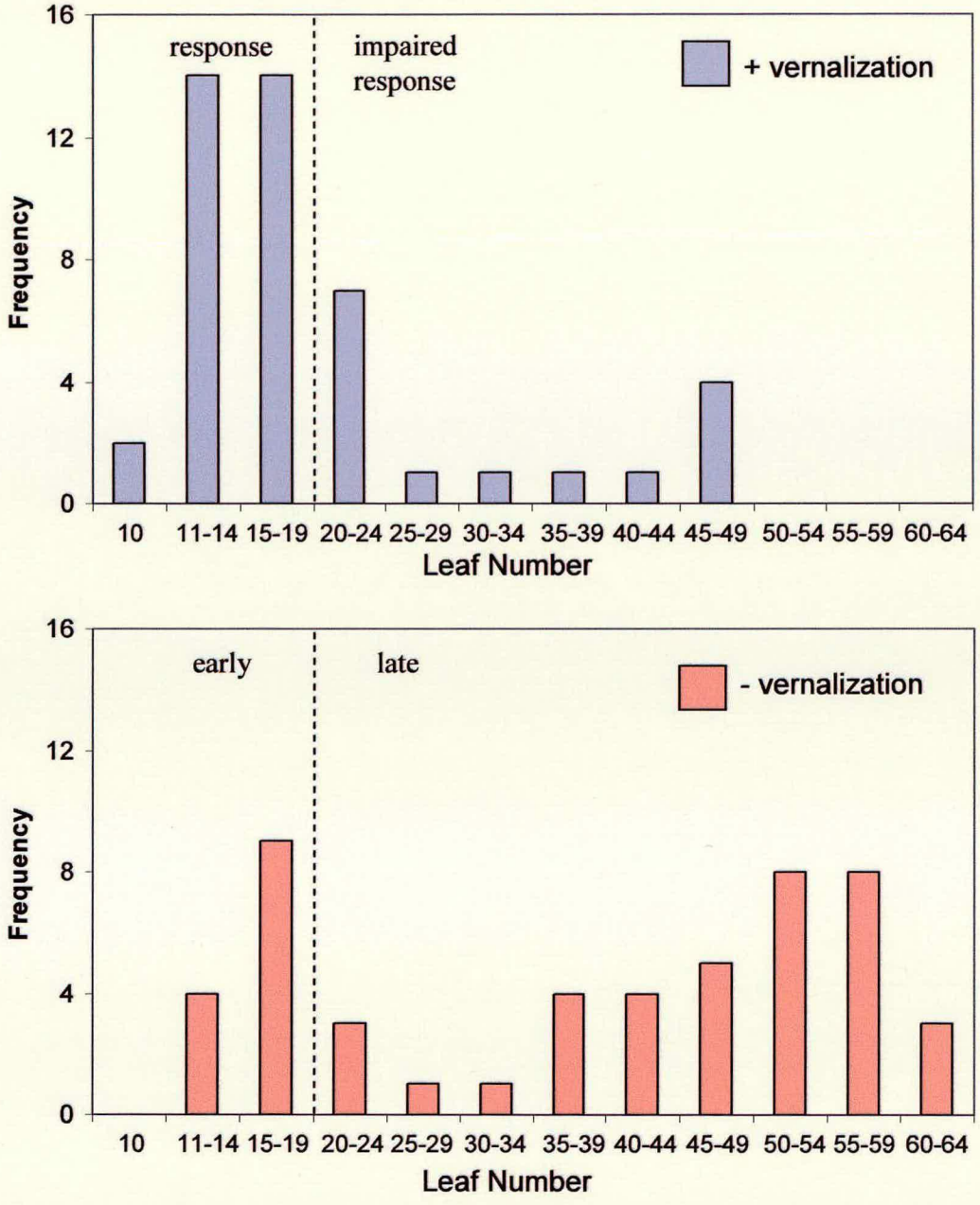


Figure 5.2 Flowering time in an F2 population segregating for *swn*-3 and *FRI*$^{+}$
Plants were supplied with either a six week cold treatment (vernalized) or a two day cold treatment (non-vernalized) and grown under long day conditions. The number of rosette leaves when the inflorescence meristem bolts to 1 cm is shown on the x axis, and the frequency based upon a population of approximately 40 individuals is shown on the y axis. In the non-vernalized plants, approximately one 1:3 (13:38) flower early, consistent with the expected number of *fri*$^{-}$ individuals. In the vernalized plants the ratio of plants showing an impaired vernalization response is higher (1:2) than the expected value of 3:13 if *swn* mutants affect vernalization.

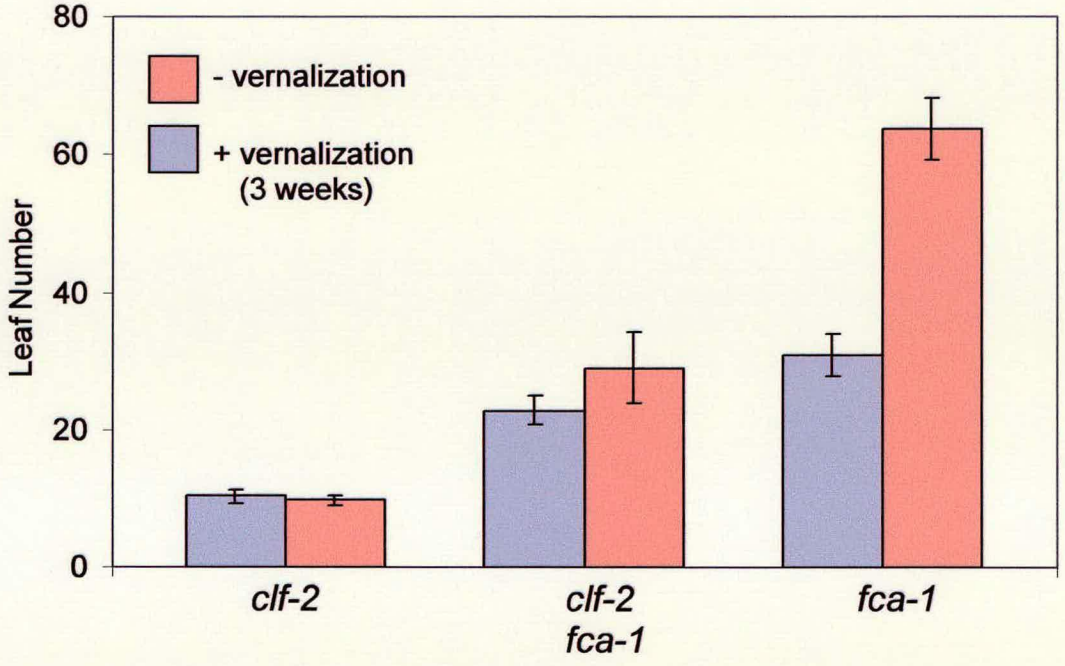


Figure 5.3 The vernalization response is reduced in *clf fca* mutants when a marginal (3 weeks) exposure to cold is provided

Plants were grown under short day conditions with either three weeks or two days cold treatment. Numbers of rosette leaves were counted when individuals bolted to 1cm. Bars represent the mean leaf number based on approximately 20 individuals. Error bars show the standard deviation. *clf fca* double mutants do show a statistically significant response to a 3 week period of vernalization. However, it is much smaller than either the response seen in *fca* plants to a 3 week vernalization period, or the response seen in *clf fca* double mutants to a longer (six week) cold treatment.

Table 5.2 Flowering times for *clf fca* double mutants with 3 week vernalization

	vernalization	Number of rosette leaves*
clf-2	3 weeks	10 ± 1
clf-2	-	10 ± 1
clf-2 fca-1	3 weeks	23 ± 2
clf-2 fca-1	-	29 ± 5
fca-1	3 weeks	31 ± 3
fca-1	-	64 ± 5

* when inflorescence meristem has bolted to 1 cm
Error values show standard deviation based on about 40 individuals

79% of the leaves of non-vernalized individuals. Although these data need to be confirmed, they suggest a role for *CLF* in the response to vernalization, and confirm the yeast two-hybrid data.

To test the vernalization response of *SWN*, it was decided to use two new alleles *swn-5* and *swn-7*, because of the possibility that the *swn-3* line was segregating for an additional mutation affecting the vernalization response. Due to time restraints it was not possible to analyse F2 crosses with an individual carrying a dominant FRI^{+} allele. Analysis was therefore performed on the progeny of a heterozygote under short day conditions.

It was first necessary to ascertain that under short day conditions the Columbia ecotype did show a measurable response to a three week period of vernalization. Leaf number was compared between populations of 40 Columbia plants which had been vernalized for three weeks or merely stratified for two days. This period of vernalization accelerated flowering from non-vernalized individuals having a mean of 42 leaves at flowering, whilst vernalized plants flowered with a mean of 38 leaves (figure 5.4). There was however overlap between the two datasets with some non-vernalized plants flowering with fewer leaves than some vernalized individuals. To test whether the differences between these mean values were significant a paired samples *t*-test was performed on the two datasets using the Analyse-it 1.71 software for Microsoft Excel. This test confirmed that the mean values were significantly different with a confidence interval of 95 %.

For both populations segregating *swn-5* and *swn-7* a slight shift can be seen in flowering time as a response to vernalization (figure 5.5). Again, there is significant overlap in flowering times between the vernalized and non-vernalized populations. To test if *SWN* had any effect on the response after this marginal vernalization period, both early and late flowering plants were genotyped at the *SWN* locus by PCR. It was found that in both *swn-5* and *swn-7* lines, homozygous *swn* mutants were found with approximately

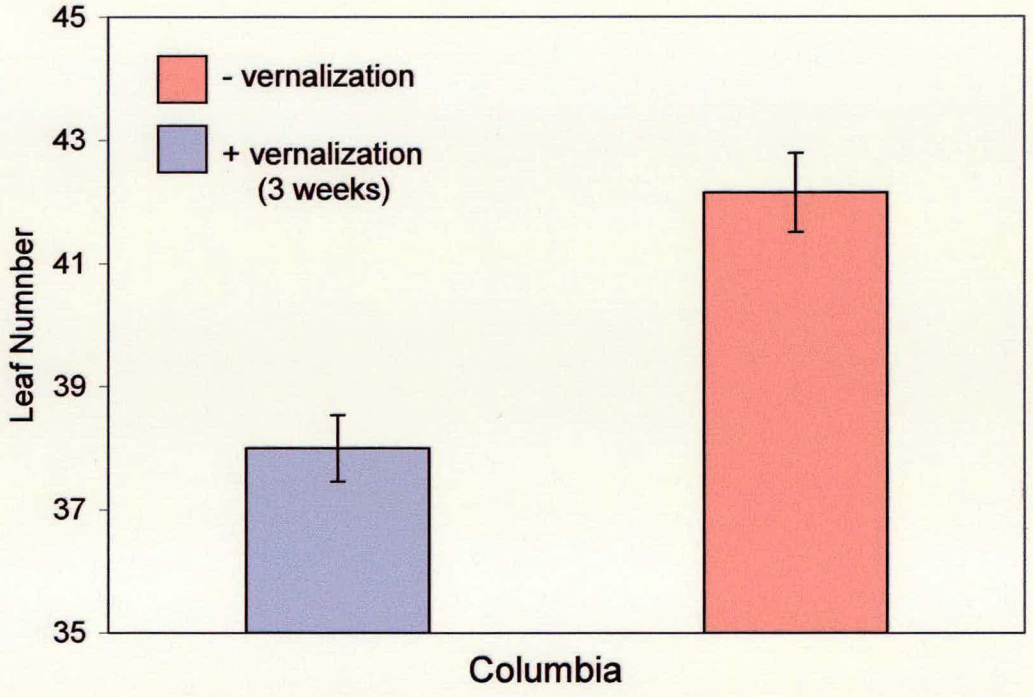


Figure 5.4 The Columbia ecotype shows a small vernalization response
Mean leaf number for approximately 40 individuals grown under short day conditions with and without a 3 week period of vernalization. Error bars show standard error.

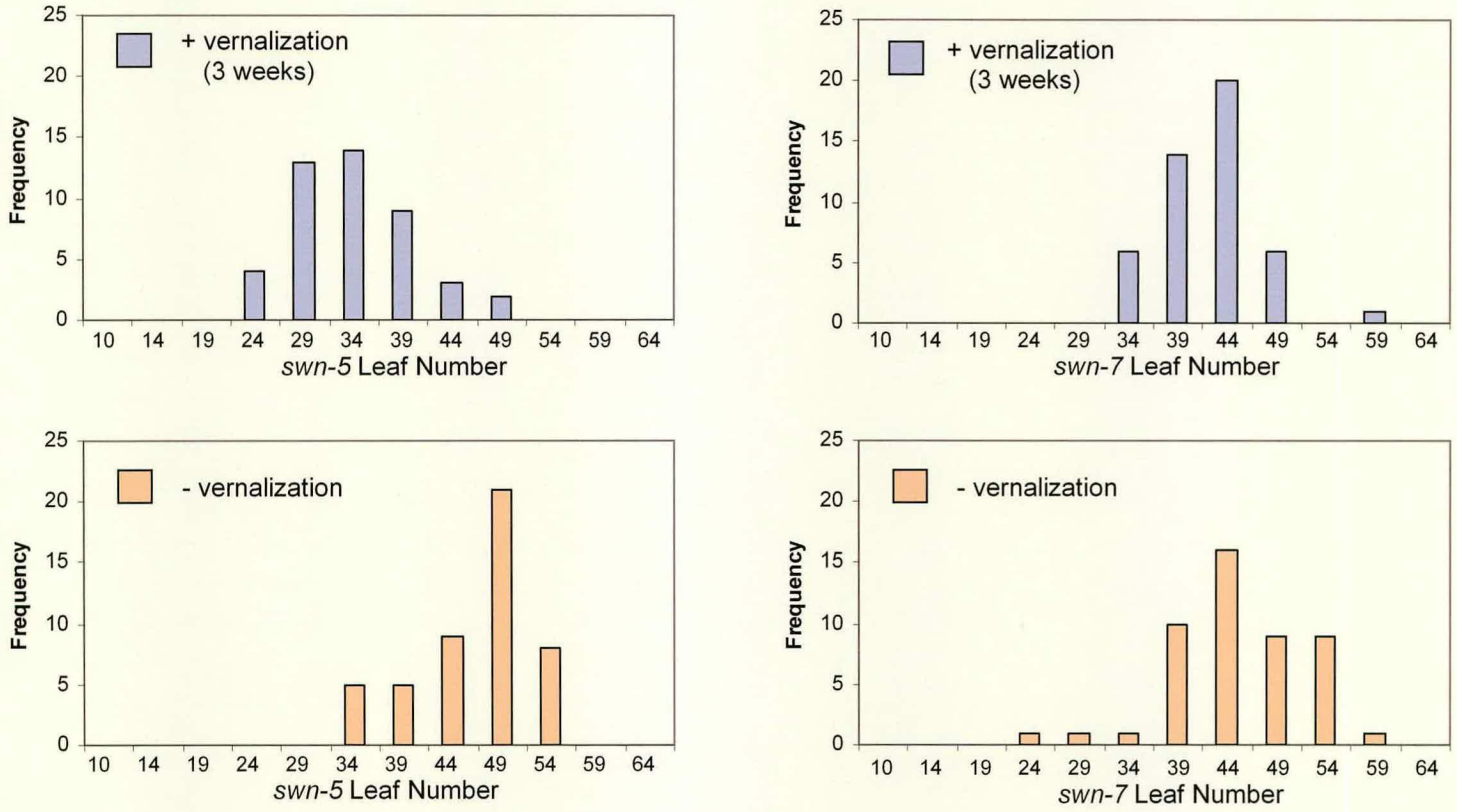


Figure 5.5 The vernalization response in populations segregating for *swn-5* and *swn-7*
The number of rosette leaves when the inflorescence meristem bolts to 1 cm is shown on the x axis, and the frequency based on a population of approximately 50 individuals is shown on the y axis. All plants were grown under short day conditions.

equal frequency among individuals flowering early and those flowering late, indicating that loss of *SWN* activity had no effect on vernalization response. The results would be clearer if these *swn* lines were crossed with a line carrying FRI^{+} and the experiment repeated after a short vernalization period. The difference between vernalized and non-vernalized plants would be greater and any individuals showing an altered response to vernalization would, therefore, be easier to identify.

5.3 Which E(Z) homologue(s) give VRN2 its histone methyltransferase activity?

The data show that loss of *CLF* function has minor effects on the response to vernalization. In the light of the findings that VRN2 is required *in-vivo* for histone methylation at the *FLC* locus, and my findings that it can interact with the *Arabidopsis* E(Z) homologues, it seems likely that VRN2 acts in a complex with one or more of CLF, SWN and MEA. It is possible that *MEA* has a role in vernalization, but this is unlikely because it is expressed at such low levels in vegetative organs. To test this would be problematic due to the difficulty in producing viable *mea* homozygotes.

A more likely hypothesis would be that CLF and SWN act redundantly in the vernalization pathway. This hypothesis is attractive because the two genes clearly show redundancy with respect to many of their other functions. In the previous chapter both CLF and SWN proteins were shown to be capable of physically interacting with VRN2 in yeast two-hybrid assays, and both proteins are expressed vegetatively. If the two proteins show redundancy with respect to repressing *FLC* it would explain why no, or only minor changes to the vernalization response were observed in the single mutants.

Testing the vernalization response in *clf swn* double mutants can not be performed by measuring flowering time because the phenotype is so severe that the null double mutants fail to produce recognisable flowers and the weaker double mutants flower extremely early. If the double mutants were crossed with an accession carrying a

dominant *FRI*$^{+}$ allele, the levels of *FLC* mRNA could be compared between vernalized and non-vernalized individuals by either Northern blot analysis or by the use of an *FLC* reporter gene. The difficulty would be that many genes are mis-regulated in the *clf swn* doubles and it would not be possible to ascertain whether *FLC* was a direct target. If weak double mutants were used the reduction in *FLC* levels could be compared with that of an *emf2 FRI*$^{+}$ cross; the problem here would be that because the loss of CLF and SWN function is incomplete, there may still be response to vernalization. A better method may be to use ChIP analysis of the *FLC* locus of plants with and without a vernalization treatment, using antibodies against CLF and SWN it would be possible to determine whether CLF and SWN proteins are bound to the *FLC* locus. To demonstrate a direct role in vernalization, antibodies against methylated K27 histone H3 could be used, and enrichment of the *FLC* locus compared in *clf*, *swn* and wild-type plants, both with and without vernalization. The problem here would be that if the CLF and SWN show redundancy, it would be likely that there would be little or no loss of K27 methylation in either of the single mutants.

6 Role of *FIE* Post Seed Development

6.1 Introduction

The evidence for the inclusion of FIE in a multimeric protein complex comprising key Pc-G members has been discussed in detail in chapter 4. Not only is *FIE* the only *ESC* homologue in *Arabidopsis*, but it is expressed beyond seed development suggesting additional roles involving other Pc-G genes, such as *CLF* and *EMF2*. In addition to the role of *FIE* in seed development, it is likely to be involved in repression of floral induction genes in embryos and seedlings, repression of homeotic and meristem identity genes in seedlings and maintenance of an epigenetic memory of vernalization.

However, because embryos with maternally inherited *fie* mutations abort, it has not been possible to generate homozygous *fie* mutant plants, and this has prevented the investigation of post-embryonic phenotypes. Recently, this problem has been addressed by Kinoshita et al. (2001), who used a modified *FIE* transgene which specifically provides *FIE* expression during seed development but not during any later stages of plant development. This was sufficient to rescue the embryo lethality of *fie* homozygotes, but not the subsequent effects on late embryo and vegetative development. The shoot apical meristem of these rescued *fie* mutant seedlings bypasses vegetative development and produces flowers and floral organs. Floral organs can also be seen on hypocotyls and roots. Furthermore, floral induction and floral homeotic genes (such as *LFY*, *AG* and *AP3*) are mis-expressed at hugely elevated levels within these seedlings.

Whilst these transgenic rescue mutants have demonstrated additional roles of *FIE*, they have been unable to demonstrate involvement in the vernalization pathway. Characterising the involvement of *FIE* in vernalization remains a challenging problem, as homozygous *fie* mutants flower so prematurely that the vernalization pathway is almost certainly bypassed.

This chapter aims to further characterise the roles of *FIE* outside of seed development. Two approaches have been taken to address this issue. The first involved the analysis and characterisation of novel *fie* alleles identified by TILLING. It was hoped that these mutations could split the functions of *FIE* and affect some pathways but not others; this would be especially useful if the vernalization response was affected whereas other aspects such as seed development and flowering time were not. The second approach involves the creation of a transgenic line where *FIE* activity is placed under the control of a steroid-inducible receptor. This can then be used to switch the expression of *FIE* on and off. This could provide useful information about the temporal requirements for FIE^{+} activity (and might be informative for understanding the vernalization requirement if introduced into a FRI^{+} *fie/fie* p(*FLC*)::GUS background).

6.2 Identification of novel *fie* alleles

Nine *fie* mutations have been discovered and characterised to date. These were identified in two genetic screens based on silique elongation in the absence of fertilisation (Chaudhury et al., 1997 and Ohad et al., 1996). All nine of these mutations show similar phenotypes; namely endosperm development in the absence of fertilisation and precocious endosperm development leading to embryo abortion in fertilised seed. The nature of both genetic screens allowed only mutations showing strong fertilisation-independent development to be identified and consequently all nine *fie* alleles are likely null. The severity of these mutants, in particular the maternal effect lethality, may have masked additional roles of *FIE* outside of seed development. Cloning of *FIE* has confirmed the null status of these *fie* alleles (Ohad et al., 1999). For the *fie-1*, *fie-2* and *fie-3* alleles, a G residue was changed to an A residue at the 3'-border of the first, third and eighth introns, respectively; mutations in this region have been shown to disrupt the splicing reaction (Brown 1996). The *fie-4*, *fie-5*, *fie-6* and *fie-9* alleles have single-base changes which result in the formation of stop codons in the ninth, seventh, thirteenth and fifth exons, respectively. The *fie-7* allele has a 27-bp deletion in the tenth exon. There

is a single-base change coupled with a two-base deletion in seventh exon of *fie-8* which alters the reading frame.

The first method of investigating the role of *FIE* post seed development was through the identification and analysis of novel *fie* alleles. The previous screens for autonomous endosperm development, which had a low penetrance and preselected for severe/null *fie* alleles. Here, I used a reverse genetics approach, based on TILLING (see chapter 3), so that a range of weak and strong alleles might be generated, possibly revealing new aspects of *FIE* function.

The original TILLING population was mutagenised with EMS, in this method single-base changes of G:C to A:T predominate. Transitions such as these lead to a variety of changes in the DNA including nonsense, missense and silent mutations. As the aim is to identify alleles which confer either a subtle phenotype or affect only one aspect of *FIE* function, missense mutations will be the most relevant. These mutations change only one amino acid, and this may affect how *FIE* functions without the lethal effects caused by null alleles. Although both the original screens, through which *FIE* was identified, used EMS to mutagenise seed and discovered only null alleles, it is likely that a series of silent and missense mutations were also produced, but were not identified in screens based upon phenotype.

A variety of *esc* mutations have been discovered, some of which do not confer a null phenotype. The *EXTRA SEX COMBS* locus was originally defined by a single recessive allele esc^{1}. This mutation was described as being variable in terms of both penetrance and expressivity, suggesting that this mutation might reduce but not eliminate the gene product (Struhl 1981; Slifer 1942; Tokunaga and Stern 1965). These variations in phenotype were also due to distinct maternal and zygotic contributions, for example esc^{-}/esc^{-} embryos could develop more or less normally (with the exception of sex combs) as the ESC^{+}/esc^{-} mother pumps ESC^{+} into the esc^{-} egg. Sequencing this allele revealed two point mutations, the second and more likely to cause the phenotype, substituted a leucine

for an arginine in the fourth WD repeat (Gutjahr et al., 1995). Since this mutant is no longer available it is not possible to ascertain which substitution causes the mutant phenotype. Another allele *esc*^{17} is temperature sensitive, so not a null. When the embryos are maintained at the permissive temperature 17°C, embryos develop normally and develop into first-instar larvae with normal or almost normal segmentation patterns; however, when the embryos are maintained at the non-permissive temperature 29°C they develop into first-instar larvae showing the characteristic *esc* phenotype (Struhl and Brower 1982). The fact that weak or non-lethal mutations do occur in *ESC* suggests that it may be possible to identify mutations showing only weak phenotypes in other genes with similar structures. The task is to identify similar mutations in *FIE*.

6.3 Choosing which domain to TILL

Like ESC, the FIE protein contains seven WD repeats. A predicted tertiary structure of ESC protein has been generated by homology mapping to the G_β protein, where the seven WD repeats form a doughnut-shaped β-propellor structure (See introduction and Ng et al., 1997). The nature of this structure means that it is hard to select conserved domains which may affect function, because the modules are similar and disrupting any one part can block the whole protein from folding correctly as the blades are interdependent. In the case of *CLF*, the SET domain was an obvious place to target. Firstly, because the high level of conservation with both plant and animal homologues suggested that this region was functionally important, and secondly, because previous studies had shown the SET domain to confer histone methyltransferase activity in similar proteins. Although FIE shows a high level of conservation with ESC, this is not restricted to certain domains but rather spreads the length of all seven WD repeats.

In order to identify areas of *FIE* which show the highest conservation, the *Arabidopsis* FIE protein was compared with homologues from three divergent plant species. These sequences were aligned using the MegAlign programme and are displayed in figure 6.1. All seven WD repeats show remarkable similarity throughout the protein (see also table

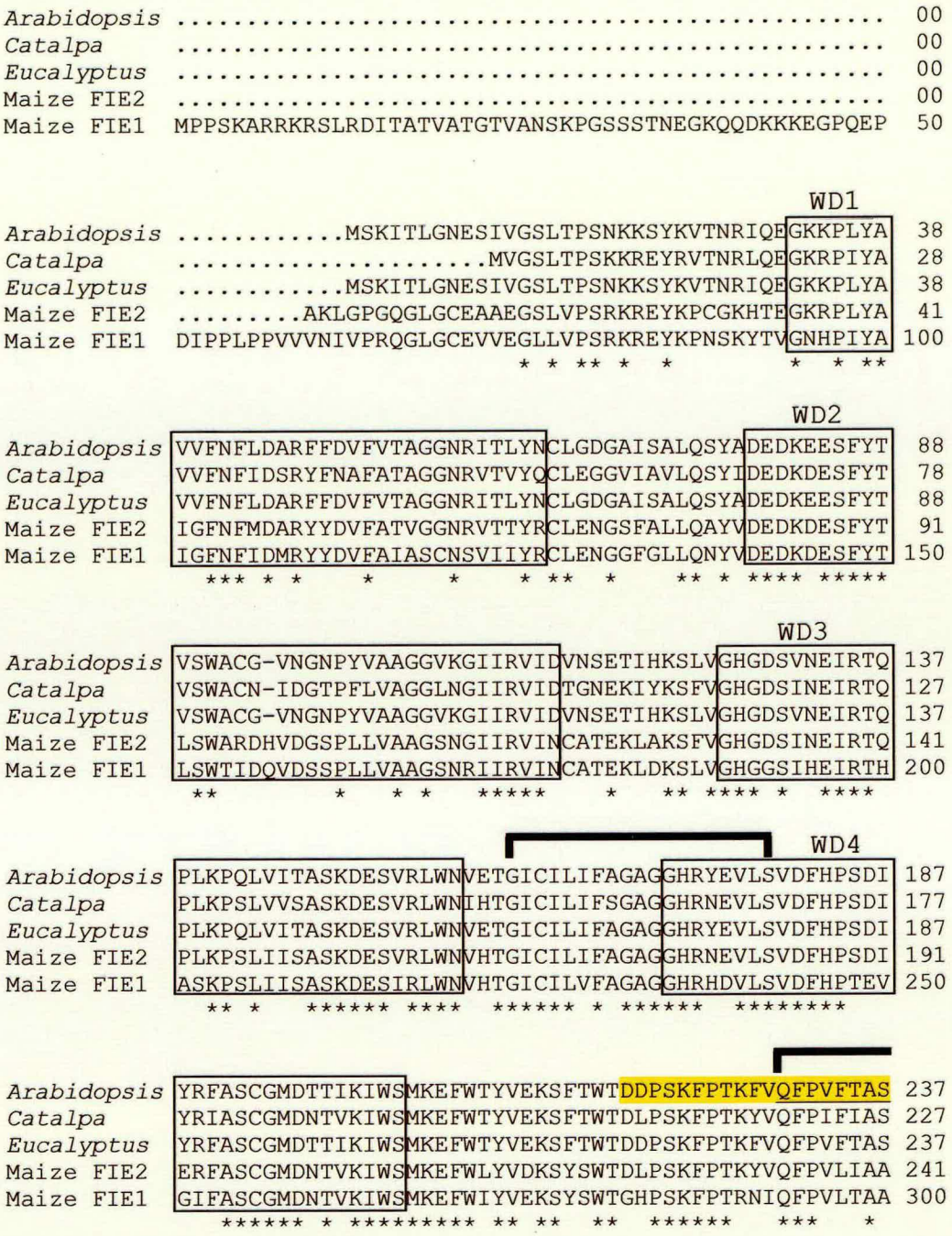


Figure 6.1 Alignment of *Arabidopsis* FIE with plant homologues
Residues conserved in all 5 sequences are indicated with a star. Black boxes denote WD repeats (defined in Ng et al., 1997). Region submitted to *Arabidopsis* TILLING service is highlighted. Surface-accessible regions predicted for ESC (Ng et al., 1997) are bracketed. Figure continues on next page.

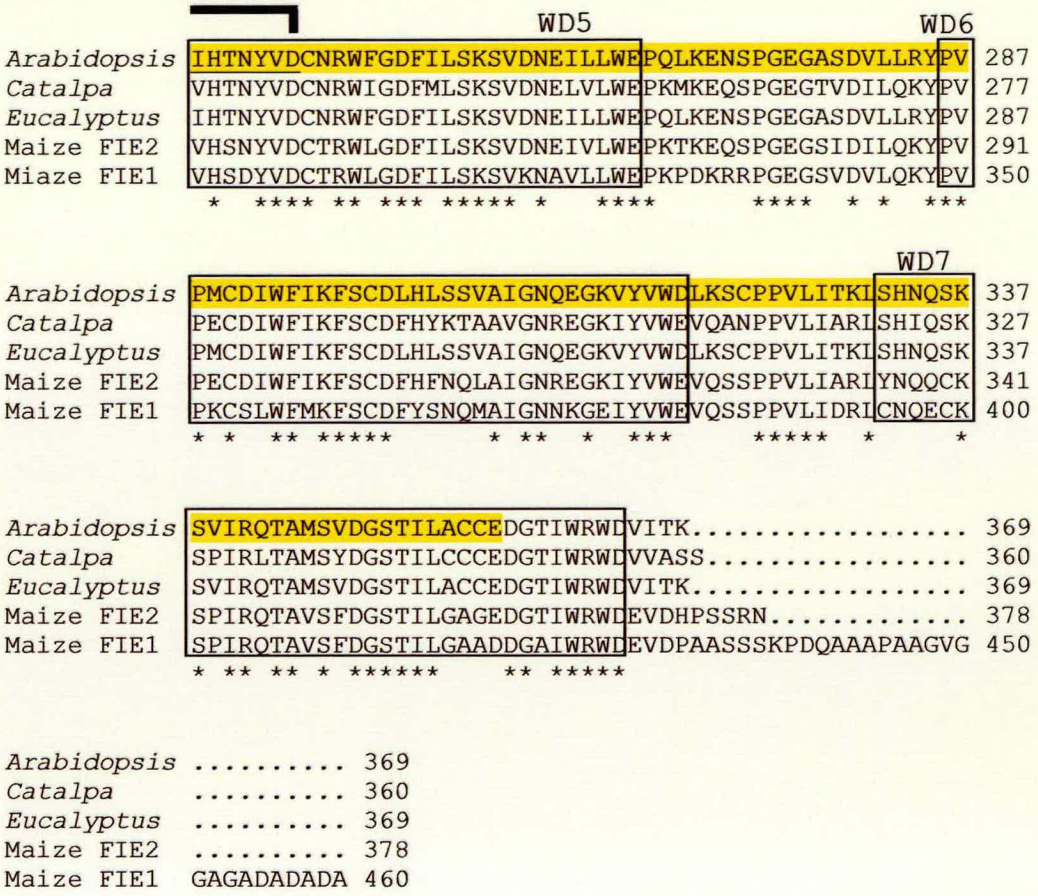


Figure 6.1 continued from previous page

Table 6.1 Evolutionary conservation in the different WD repeats of FIE

Portion of FIE	% of residues identical in all four proteins
N-terminal	19%
WD1	35%
WD 2	51%
WD 3	67%
WD 4	69%
WD 5	66%
WD 6	52%
WD 7	59%

6.1), although the strongest similarities are found within WD repeats 3 and 4. One potential problem with identifying important domains by conservation is that it is not known whether the sequence specific similarities between these proteins represent functional conservation. In the case of maize, it is likely that both *FIE* homologues have evolved different *FIE*-like functions; *FIE1* is expressed specifically in kernels whereas *FIE2* expression is found throughout the plant (Springer et al., 2002).

The first choice to submit to TILL was a 1 kb region encoding WD repeats 3 and 4. As well as spanning the two most highly conserved repeats this region had several other advantages. Two extensive surface-accessible regions of ESC that exhibit near absolute conservation between insect *ESC* homologues were identified, the first of these lies between WD repeats 3 and 4, and the second between WD repeats 4 and 5 (Ng et al., 1997). The only two *esc* mutations known to be missense lie within the first of these surface-accessible regions, one of these produces a null and the other a variable phenotype (Gutjahr et al., 1995; Sathe and Harte 1995). mRNA carrying site-directed mutations in a number of residues in these regions were tested for *ESC* function *in-vivo* by injection into esc^- eggs. It was found that mutations within both these areas produced a variety of strong loss-of-*ESC*-function as well as more modest loss-of-function phenotypes (Ng et al., 1997).

For these reasons, a 1 kb region encoding WD repeats 3 and 4 was submitted for TILLING. This region was, unfortunately, unavailable because it overlapped extensively with a TILLING request submitted by another party. A second area was selected spanning WD repeats 5, 6 and most of 7. This contained the second surface-accessible region identified by Ng. This request was optimised using CODDLe and submitted to the *Arabidopsis* TILLING project.

6.4 Mutations identified in TILLING screen

In total 18 polymorphisms were identified within the 993 bp of *FIE* submitted, these are shown in table 6.2. Of the 18 polymorphisms 15 were unique and three were duplicate alleles. Presumably, this is either because of contamination of the seed lines at the TILLING service, or more likely a preference of EMS mutagen to make particular mutations. Of the 15 novel polymorphisms a further six were located in introns and, therefore, bore no effect on the protein product. Of the remaining nine polymorphisms located in the coding region, two made silent changes. This left seven novel *fie* alleles, each making a missense mutation and changing the amino acid encoded. These were numbered *fie-10-14*, *fie-20* and *fie-21* (figure 6.2). None of these seven mutations were positioned in the surface-accessible region that proved effective at producing *esc* alleles of varying penetrance. However, both *fie-13* and *fie-20* altered residues which were identical in FIE, ESC and EED. The *fie-20* mutation is of particular interest because a nonpolar proline is substituted by a polar serine.

6.5 Analysis of novel *fie* alleles

Seed for all seven TILLING lines with missense mutations were ordered from the *Arabidopsis* Stock Centre. These lines were grown on GM plates before being moved to soil. They were genotyped using either CAP or dCAP markers; a list of methods employed to genotype these lines is shown in table 6.3. In the case of *fie10-14* and *fie20* homozygous mutant plants were identified. Superficially, these plants looked wild-type, were of normal size, leaf shape and flowered at a time comparable to wild-type. Closer examination of siliques showed normal development in all lines with virtually 100 % viable seed. A typical dCAPs gel used to genotype *fie-11* is shown in figure 6.3, along with a segregating population showing no phenotypic differences. Unfortunately, no mutant plants (either heterozygous or homozygous) were identified in the seed supplied for *fie21*.

Table 6.2 Mutations identified in *FIE* locus through TILLING screen

#	Nucleotide Change[a]	Effect[b]	Genotype of parent	Allele designation
1	G285→ A	Intron (silent)	Hetero	
2	C344→ T	P274→ S	Hetero	*fie-20*
3	A406→ T	Intron (silent)	Homo	
4	G447→ A	D280→ N	Homo	*fie-10*
5	G447→ A	D280→ N	Hetero	
6	G460→ A	R284→ K	Hetero	*fie-11*
7	G460→ A	R284→ K	Hetero	
8	G467→ A	(silent)	Hetero	
9	C471→ T	P288→ S	Hetero	*fie-12*
10	G534→ A	G309→ S	Homo	*fie-14*
11	G554→ A	Intron (silent)	Hetero	
12	G554→ A	Intron (silent)	Hetero	
13	G574→ A	Intron (silent)	Hetero	
14	G595→ T	Intron (silent)	Hetero	
15	C618→ T	Intron (silent)	Hetero	
16	G645→ A	E312→ K	Hetero	*fie-13*
17	G653→ A	(silent)	Hetero	
18	G828→ A	M345→ I	Hetero	*fie-21*

[a] Nucleotide number relates to position within area TILLED (1-993 bp).
[b] Changes within coding region are shown by their effect on the protein. Residues are numbered normally (1-369).

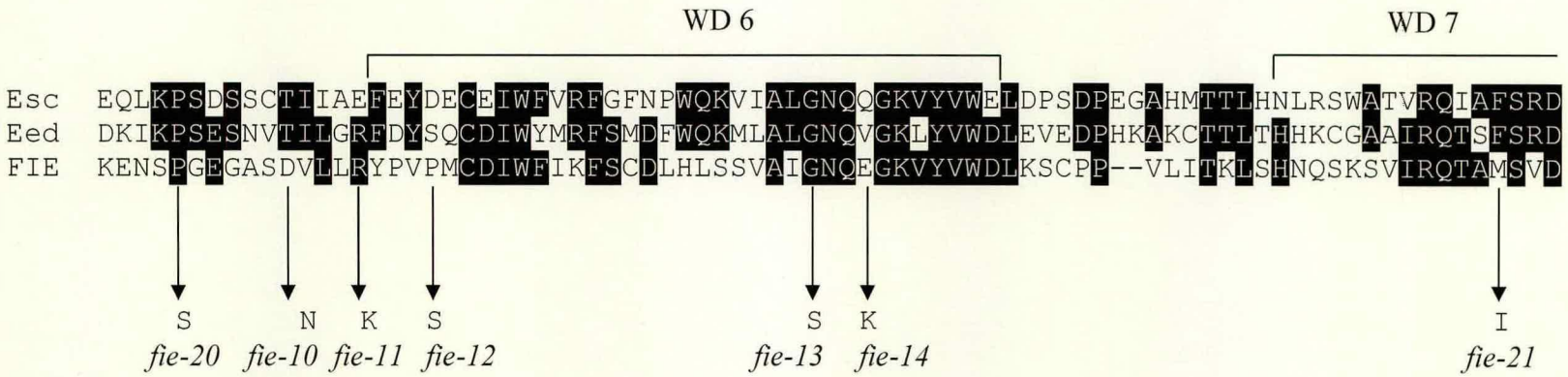


Figure 6.2 Location of novel *fie* mutations identified through TILLING

Table 6.3 Methods employed to genotype *fie* alleles

fie allele	Method of analysis	Amplification with primers	PCR product cleaved by	Cleaved band corresponds to
fie-10	dCAP	AB17 and AB18	*Dde* I	wild-type
fie-11	dCAP	AB15 and AB19	*Eco*R V	wild-type
fie-12	CAP	AB17 and AB16	*Nla* IV	wild-type
fie-13	CAP	AB17 and AB16	*Hpy*178 III	wild-type
fie-14	dCAP	AB20 and AB15	*Sca* I	mutant
fie-20	CAP	fie20f and AB20	*Mbo* II	mutant
fie-21	dCAP	fie21f and fie21r	*Eco*R V	mutant

For definitions of CAP and dCAP markers see chapter 3.

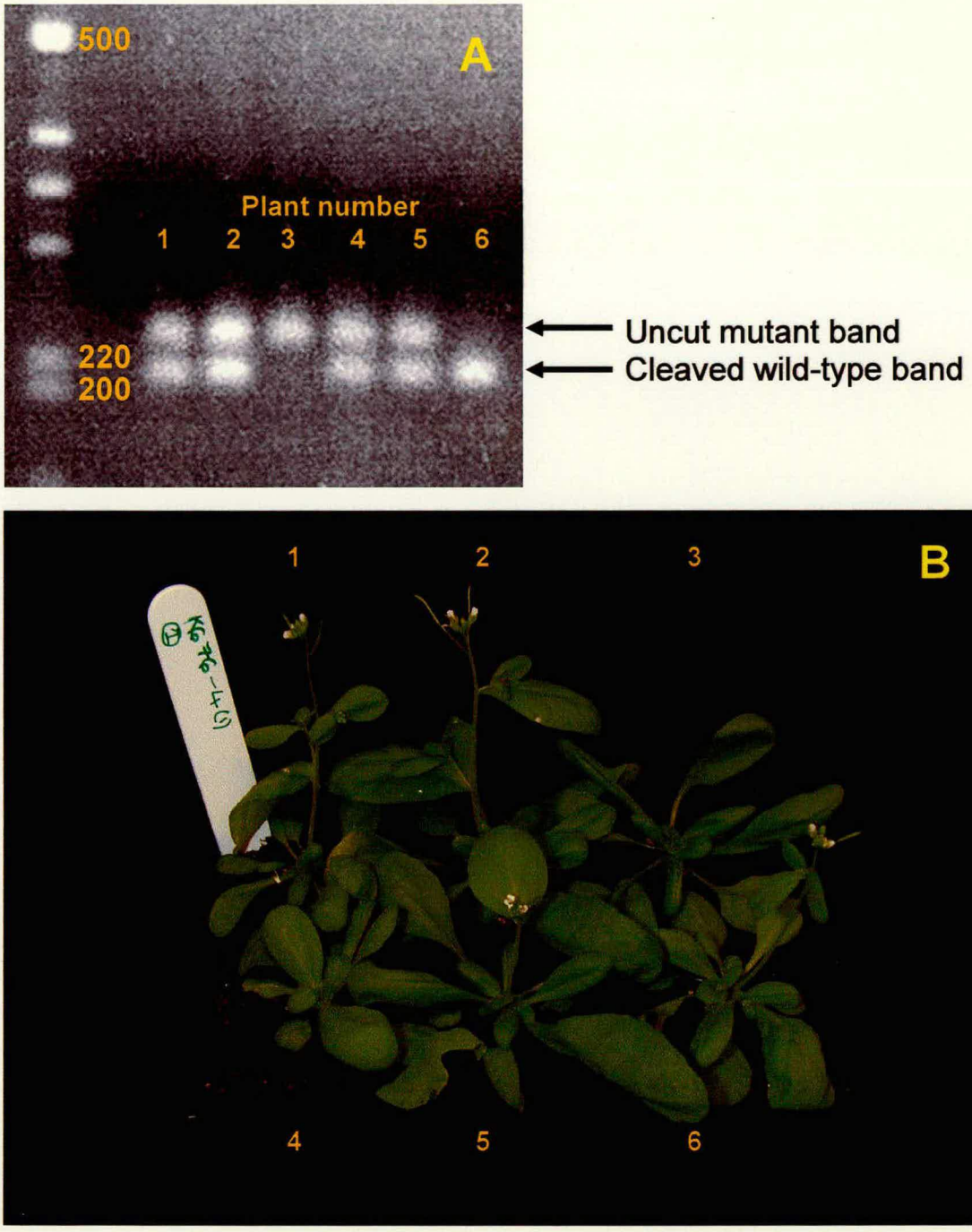


Figure 6.3 Analysis of *fie-11* mutant.
A. The wild-type band is cleaved with *Eco*R V. Plants 1, 2, 4 and 5 are heterozygous for *fie-11*, 3 is homozygous *fie-11* and 6 is wild-type. B. All of the 6 plants show no phenotype associated with the *fie-11* mutation.

A crossing programme was initiated to identify subtle phenotypes which may not have been identifiable in a weak *fie* background. The first cross was made with lines homozygous for the new TILLING alleles to the null allele *fie-2*. In a background in which *FIE* activity is further reduced, subtle changes associated with the new *fie* alleles may become more pronounced. Pollen from *fie-2*/+ plants was used to fertilise homozygotes for the novel alleles. Half of the resulting F1 plants were expected to be *fie*/fie-2* (where *fie** indicates the novel tilling allele). Examination of developing siliques indicated that 50 % of plants showed 50 % seed abortion. These corresponded to the *fie-2/fie** plants. In all other aspects they appeared morphologically normal. The novel alleles therefore appeared aphenotypic.

A second cross was made between *fie10-14* and the strong *clf-2* allele to test whether these alleles enhance the *clf* phenotype in double mutant combinations. Here *clf-2*/+ pollen was used to pollinate plants homozygous for the various TILLING alleles. The F1 population was then grown on GM plates containing kanamycin, which selects for the *clf-2* allele (due to a linked T-DNA carrying kan^R). The surviving plants were transferred to soil, allowed to self-pollinate and the seed collected. The F2 progeny were then sown directly on soil and plants with a *clf* phenotype were examined closely. There appeared to be no plants with an enhanced *clf* phenotype, however, to confirm that some of these plants were *fie* homozygotes they were genotyped. In each case homozygous *fie* mutants were identified. This shows that even in a strong *clf* background these novel *fie* alleles did not enhance the *clf* phenotype.

The final cross was to a transgenic line of Landsberg *erecta* carrying a functional FRI^+ transgene (kindly supplied by Prof. Caroline Dean). Pollen from this line was crossed to plants homozygous for *fie10*, *fie11*, *fie12* and *fie13*. The F1 plants were vernalized, grown on GM plates containing kanamycin to select for the FRI^+ transgene, and seed were collected. The F2 progeny was also sown on plates containing kanamycin in two batches, one batch was vernalized at 4°C for six weeks prior to being transferred to long day growth facilities, and the other batch was given a cold treatment for just two days.

After 10 days further growth (at 23°C) on kanamycin, 40 resistant seedlings were transferred to soil. The flowering time was measured and the number of rosette leaves at flowering recorded. Flowering time was defined as the point at which the floral inflorescence meristem bolted to 1cm. In the *FRI*$^{+}$ background, all the F2 families flowered with fewer than 12 leaves when vernalized (figure 6.4), whereas they produced more than 30 leaves if sown without vernalization (data not shown). Therefore, they showed a vernalization response. However, in F2 families from crosses to *fie10*, *fie12* and *fie13*, some of the plants flowered fairly late even after vernalization (18 or more leaves). To see if these corresponded to *fie* homozygotes showing a reduced vernalization response, eight late flowering plants from each family were genotyped at the *FIE* locus. In each case, these did not contain a significant number of individuals homozygous for the *fie* mutation. This delayed flowering in some individuals may be due to other mutations within the TILLING populations, which could be expected as the parent generation was heavily mutagenised such that above 30 lesions per haploid genome are predicted. In conclusion, no affect on the vernalization response or any other aspect of seed or vegetative development was observed in these lines.

6.6 Steroid-inducible rescue of *fie* ovules

One method of investigating the role of *FIE* post-seed development would be through the creation of transgenic plants which supply *FIE* function during seed development but not during any stage of later growth. Such plants could then be crossed with *fie*/+. In the next generation the *FIE* transgene will rescue those seed inheriting a maternal *fie* allele, allowing viable homozygous *fie* mutant seed to develop. The *FIE* transgene will not be expressed during germination and seedling growth, and hence have no subsequent effect on the phenotype of *fie*$^{-}$ seedlings (if the epigenetic effects of *FIE*$^{+}$ are not maintained in the absence of *FIE*$^{+}$ activity). This theory is supported in the data shown in Kinoshita et al. (2001).

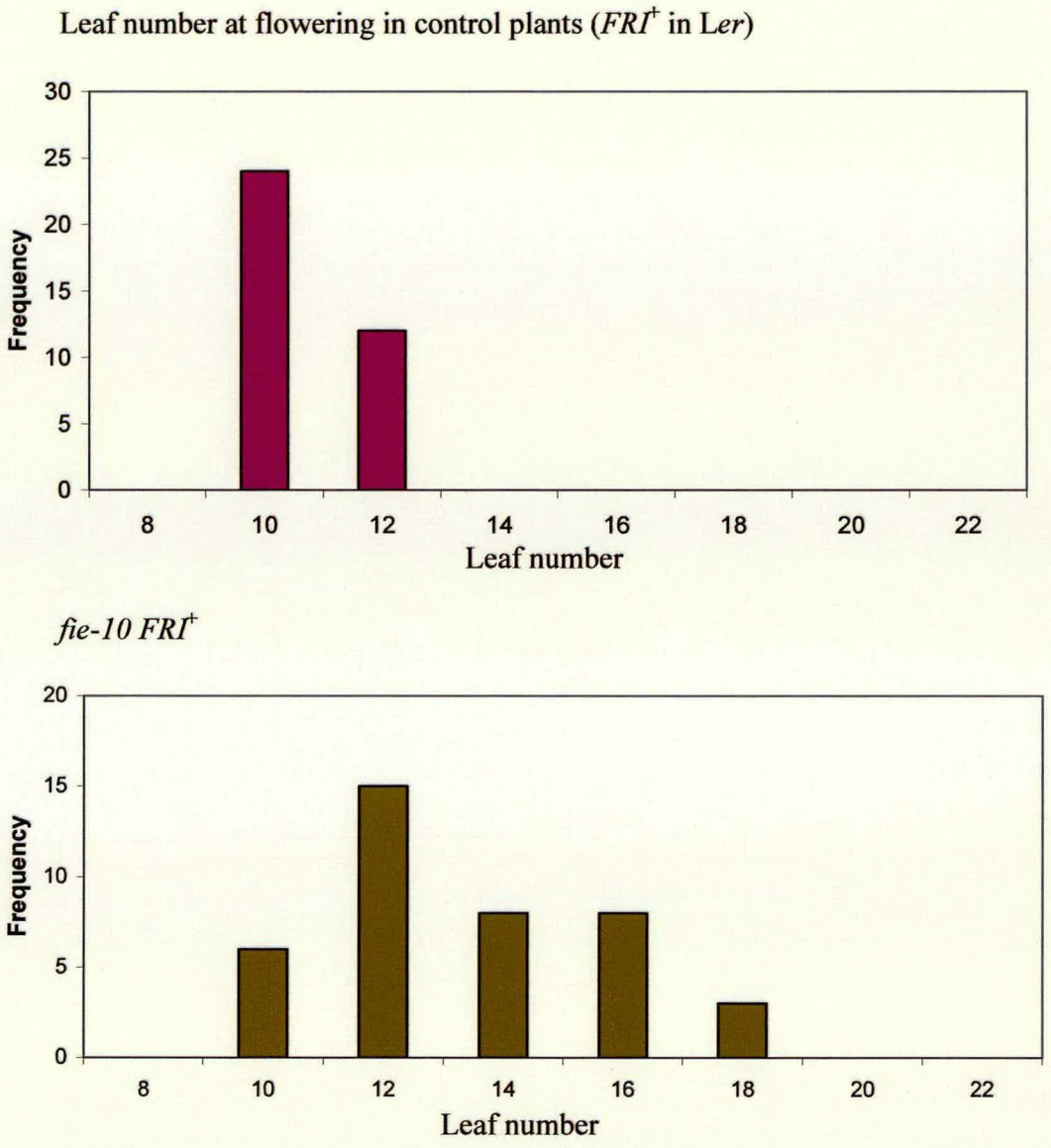
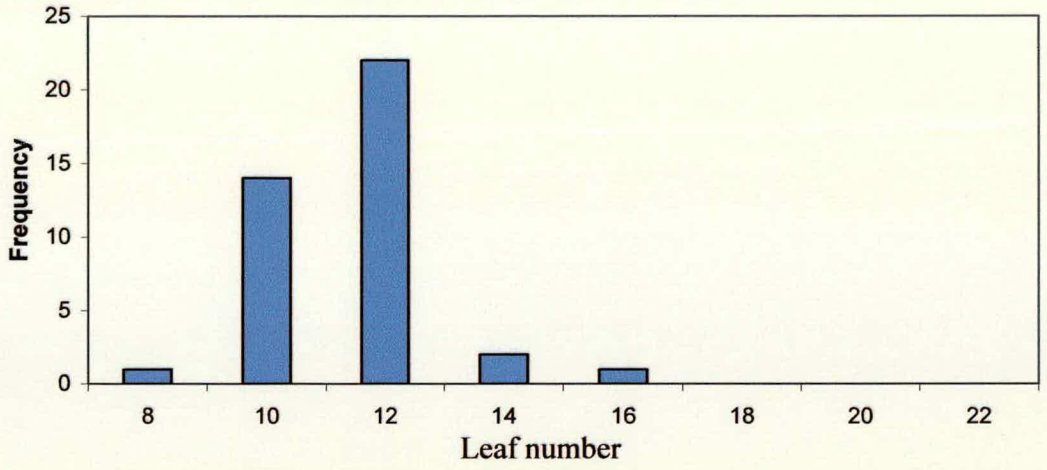
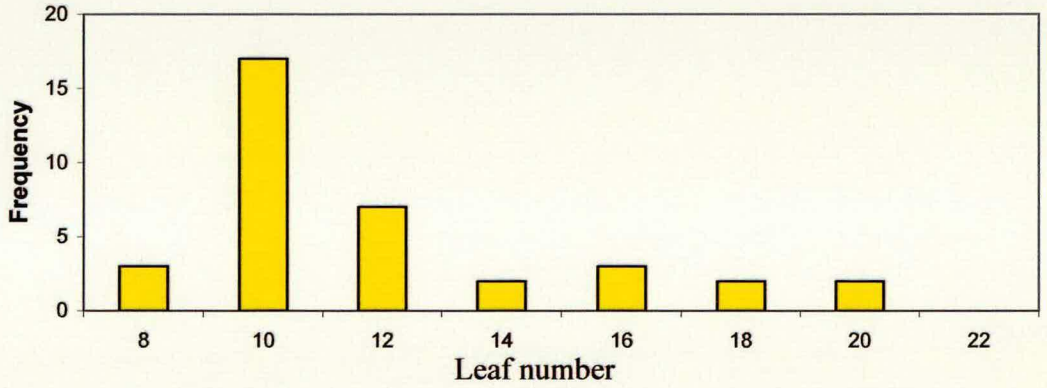


Figure 6.4 Vernalization response in new *fie* alleles
Flowering time is measured by number of rosette leaves when the plant has bolted 1 cm. Frequency is shown on the y axis and is based on 40 individuals. All lines have been vernalized for 6 weeks. Figure is continued on following page.

fie-11 FRI^{+}



fie-12 FRI^{+}



fie-13 FRI^{+}

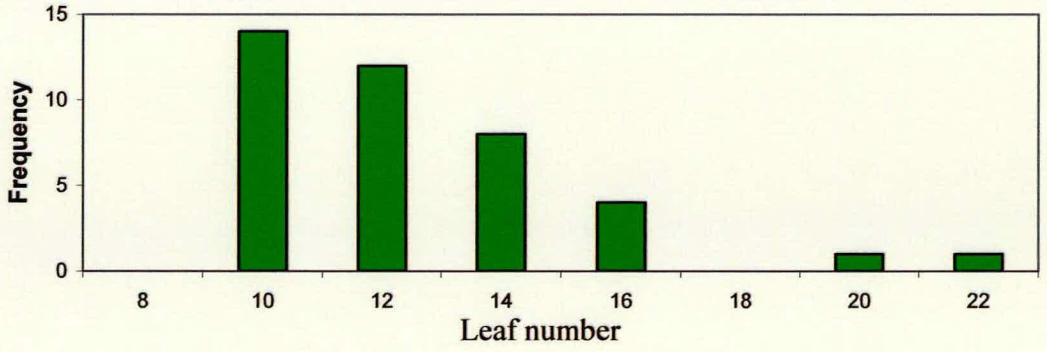


Figure 6.4 continued from previous page

There are two ways such a rescue could be performed. Firstly, *FIE* could be placed under the control of a seed-specific promoter such as that of the *ABI3* gene (Devic et al., 1996). Disadvantages with this method are that it might be difficult to find a promoter which matched the complicated pattern in which *FIE* is expressed. If *FIE* was over-expressed in the seed it might have dramatic affects upon seed development. Another problem is ensuring that there is no expression outside of seed development associated with the promoter driving *FIE* expression, as low levels of *FIE* expression are likely sufficient for FIE^{+} activity.

A second approach would be to place *FIE* under the inducible control of an external factor in such a way that expression can be turned on and off by the presence or absence of this factor. The rat glucocorticoid receptor is a protein that binds to and activates a class of hormone-dependent transcriptional enhancers. It has been used in a variety of heterologous systems, including plants, to provide activity of a fused protein which is dependent on external application of steroids. This technique has been previously been used successfully to study a range of genes in *Arabidopsis* (*CONSTANS* – Simon et al., 1996b; *APETALA 3* – Sablowski and Meyerowitz 1998; *KNAt2* – Pautot et al., 2001). In these studies a translational fusion was made between the gene of interest and amino acids 519-795 corresponding to the ligand-binding domain of the rat glucocorticoid receptor (GR) (Miesfeld et al., 1986; Schnea et al., 1991; Picard 2000). The glucocorticoid receptor inactivates proteins to which it is fused, it is thought because they are held in a complex with the Hsp90 chaperone. However, in the presence of the mammalian steroid DEXAMETHASONE (DEX), the receptor undergoes allosteric change in conformation and is released from the Hsp90 complex. This induced activation is both rapid and reversible. Furthermore, seedlings grown to maturity in high concentrations of DEX are undistinguishable from untreated plants (Schnea et al., 1991).

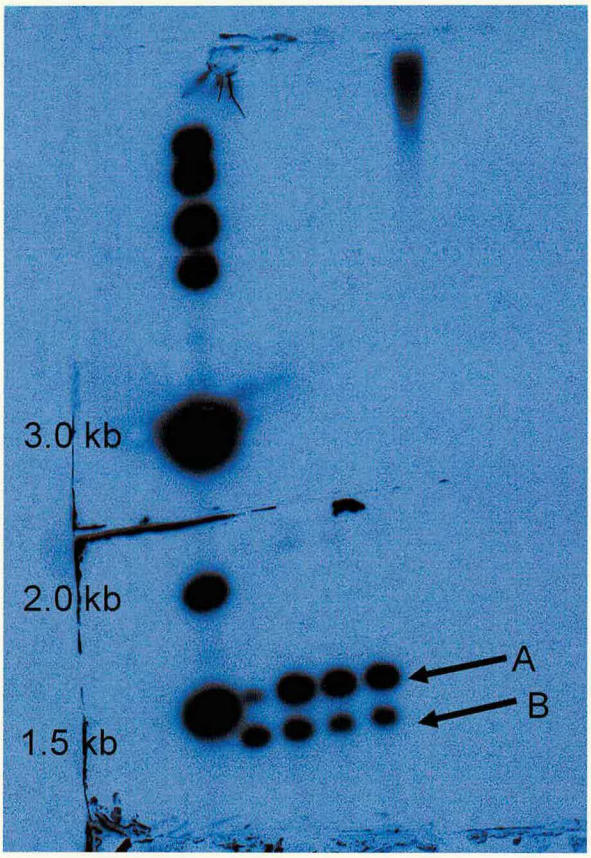
An N-terminal translational fusion protein was made between the rat GR and *FIE* (see methods chapter). This was made in a plasmid containing a 10 kb genomic clone spanning the entire *FIE* locus and containing 4 kb of DNA upstream of the

transcriptional start codon. This fusion was then cloned into the binary vector pBIB (hyg) and introduced into *fie-2*/+ plants by floral dip transformation.

The transformation efficiency was extremely low for unknown reasons, perhaps due to reduced fertility of *fie-2*/+ plants. However, four primary transformants were identified. Of these four primary transformants, two carried *fie-2* and two were FIE^{+} homozygotes; the *fie-2*/+ plants were easily identifiable as they showed 50 % seed abortion. The T2 progeny of both these plants were grown on media containing hygromycin to select for the p(*FIE*)::*GR-FIE* transgene and DEX to provide inducible *FIE* activity. One family grew poorly on hygromycin plates, producing stunted plants with negligible seed set. The other family produced healthy plants and all subsequent analyses used this family.

To ensure that this family contained the p(*FIE*)::*GR-FIE* transgene four hygromycin resistant T2 plants were genotyped by Southern blot analysis (figure 6.5). The genomic DNA was cut with *Hind* III endonuclease. A fragment of genomic *FIE* was isolated from the plasmid pBSgen*FIE*, radiolabelled for use as a probe; this showed homology with the both the 5'-end of the gene and the promoter sequence upstream. In all four individuals two bands were observed, a native *FIE* band at 1.4 kb and a transgenic band at 1.5 kb containing part of the *FIE* regulatory sequence and part of the glucocorticoid receptor. There would also have been a transgenic band at 0.8 kb but this has "run off" the gel. Plasmid DNA from the pBIB p(*FIE*)::GR-*FIE* clone was used as a control to size the bands but this did not cut and can be seen at top right of image.

These T2 plants were grown on DEX throughout, so that fie^{-} ovules might be rescued. Initially the plants were grown in tissue culture on plates containing 10 μM DEX for two weeks, they were then transferred to soil and watered with 2 μM DEX and sprayed with 10 μM DEX every three days. When the first siliques neared maturity the spraying with DEX was stopped and seed from the first siliques collected separately. The plants were allowed to continue setting seed; two weeks later the plants were examined for seed abortion and the *fie2*/+ individuals identified.



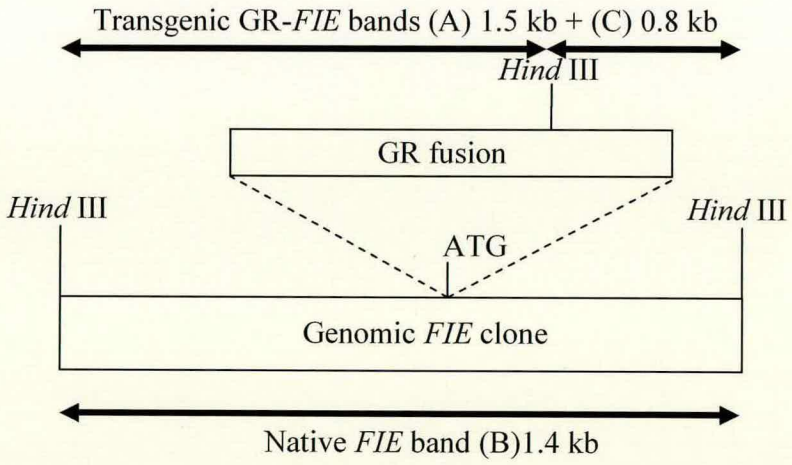


Figure 6.5 Southern Blot showing *FIE*-GR in transgenic lines
Genomic DNA from four progeny of the p(*FIE*)::GR-*FIE* transgenic line used for rescue of *fie* mutants was cut with *Hind* III restriction endonuclease and separated by gel electrophoresis. It was transferred to a nylon membrane and probed with a radiolabelled probe showing homology to the 5'-end of the *FIE* gene and upstream sequence. Band A (1.5 kb) is the larger fragment of the transgene. Band B (1.4 kb) is the fragment from the native *FIE* genomic locus. The smaller fragment (0.8 kb) of the transgene has migrated beyond the end of the gel. All four plants contain the transgene.

T3 seed from the first siliques of T2 *fie2*/+ parents were grown in tissue culture plates with and without DEX (10μM). After 10 days growth on plates lacking DEX abnormal plants were observed (Figure 6.6). These individuals were considerably smaller than wild-type and showed a range of developmental abnormalities. These included termination of the shoot apical meristem and the appearance of poorly differentiated organs. In the most severe cases an unidentifiable mass of organs originated from the seed and the plants measured just over 1 mm. One mutant was identified which appeared to have a homeotic change producing roots from what appeared to be vegetative organs (Figure 6.6E).

Similar plants were never observed when the same seed was sown on to DEX plates or when seed of the progenitor line was sown onto plates with or without DEX. If these abnormal plants corresponded to *fie-2* homozygotes, a 1:3 or 1:5 segregation ratio would be expected. If the parent is *fie-2*/+ and homozygous for the transgene, 1:3 progeny would be expected to be *fie-2*/*fie-2* and FIE^{+}. If the parent is hemizygous for the transgene, and the transgene must be inherited through the mother for rescue, 1:5 progeny would be expected to be *fie-2*/*fie-2* and FIE^{+}. The actual segregation rates were 10:45, 6:46 and 1:27. These differences could most likely be explained by uneven distribution of DEX, particularly through spraying, but also by dosage of the p(*FIE*)::GR-*FIE* transgene.

The abnormal plants were genotyped using a CAP marker. Initially a 1.6 kb fragment spanning the *fie2* mutation was amplified by PCR using the primers FIS3-1special and FIS3-1r. FIS3-1special was designed to amplify the endogenous *FIE* locus and not the *FIE* transgene. This was achieved because the FIS3-1special primer was based upon sequence spanning the site of the GR insertion. The PCR product was then cleaved with *Bsg* I restriction endonuclease which cut wild-type sequence but not the mutant sequence. This confirmed that the abnormal plants were *fie-2* homozygotes.
These *fie-2* homozygotes could be partially rescued by translocation to tissue culture media containing DEX. In these cases roots developed normally and there was

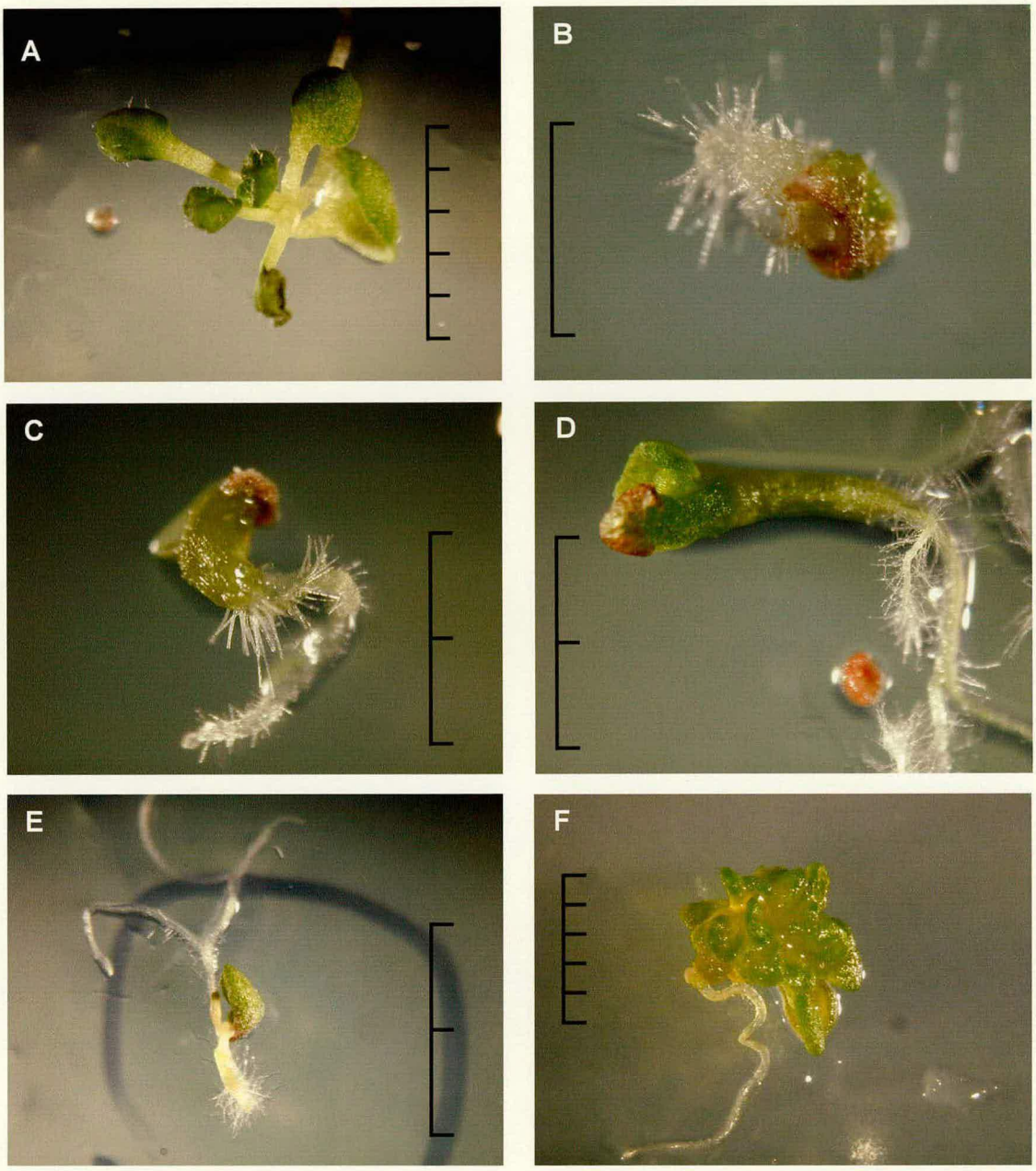


Figure 6.6 ***fie-2*** **homozygous mutants rescued by the GR-*FIE* line**
Divisions on black bars are 1mm. A shows a wild-type seedling 10 days old. B to E show 10 day old homozygous *fie* mutants grown on media lacking DEX, left on DEX-deficient plates these individuals do not develop much further and die prematurely. F shows a partial rescue of these mutants if transferred to media containing DEX; the root appears normal and numerous leaf like organs grow from the central meristem.

increased meristematic activity at the shoot apical meristem producing numerous leaves. A rescued individual is shown in figure 6.6F. These individuals were subsequently confirmed as *fie* homozygotes using CAP markers and direct sequencing of PCR products. If these individuals were left on DEX-deficient media they failed to develop further and died prematurely.

Together these results support those of Kinoshita et al. (2001) and Katz et al. (2004) demonstrating a role for FIE post seed development. The similarities between the homozygous *fie* mutants and the *clf swn* double mutants is consistent with FIE, CLF and SWN acting in a common complex. The construction of a steroid-inducible *FIE* line may be beneficial to future studies. It may be used to generate a large number of *fie* homozygotes, which could be used for histone preparations to investigate the global levels of K9 and K27 methylation. It would also be interesting to see if the same targets are deregulated in *fie* mutants as *clf swn* double mutants.

7 Discussion

I have shown that the *Arabidopsis* Pc-G proteins have the potential to form multimeric complexes similar to the *Drosophila* PRC2 complex. In *Arabidopsis* at least three of these complexes exist with at least partly discrete functions. I have shown that neither MEA nor SWN are able to provide CLF function, indicating that the specificity of these complexes is not due purely to differences in expression patterns. In yeast two-hybrid assays neither CLF nor SWN show any preference towards different Pc-G partners, suggesting that other factors may be responsible for recruiting the different complexes to their targets. However, this data demonstrated that CLF and SWN have the potential to interact with VRN2, genetic analysis has been unable to confirm this and I therefore conclude that the two proteins are likely to act redundantly to mediate the vernalization response.

7.1 CLF is likely to act as a histone methyltransferse

In animals, the core members of the PRC2 complex are ESC, E(Z), SU(Z)12 and p55 (Cao et al., 2002; Müller et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002). In *Arabidopsis*, these members are represented predominantly by small gene families. The single copy gene *FIE* is the only *ESC* homologue. However, *E(Z)* and *SU(Z)12* are represented by small gene families each with three members and p55 by a small family with five members (Goodrich and Tweedie, 2002).

The histone methyltransferase activity of the PRC2 complex has been shown to be mediated through the SET domain of E(Z) (Cao et al., 2002; Müller et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002). Characterisation of a novel mutation at the *CLF* locus has identified a change which is predicted to affect the ability of the CLF protein to bind histone tails. If histone binding and modification is an essential element of CLF activity, it would be predicted that by preventing the binding of histone tails, this

mutation would render CLF non-functional causing a strong loss-of-function phenotype. Similar to the prediction, this mutation causes a null phenotype suggesting that CLF does act as a histone methyltransferase *in-vivo*. The characterisation of the other mutations within the SET domain that I obtained through TILLING has also shown that this domain is vital to CLF function, however, the effects that these mutations cause is hard to predict as they do not fall into regions with strong structural predictions. It will be important to follow this up with *in-vitro* studies to determine if the mutations affect histone binding or methyltransferase activity. The two features could be distinguished by testing first for binding and then for histone methyltransferase activity. If the mutated proteins do not bind histones they will not show methyltransferase activity, however some may be capable of binding histones but unable to confer methyltransferase activity.

Unpublished work by Daniel Schubert has shown a loss of di-methylation of K9 and di and tri-methylation of K27 on the histone H3 tails at the *AG* locus in *clf* mutants. This provides further evidence of the role of CLF as a histone methyltransferase, although it does not show that the effects are direct, for example, it can not be ruled out that CLF might regulate another histone methyltransferase. This could be confirmed with new assays to test whether CLF can modify histones *in-vitro*. This approach would have the additional advantage of confirming whether these residues are mono, di, or tri-methylated (the *in-vivo* results are sometimes difficult to interpret as the antibodies can cross react, particularly those for di and tri-methylated residues).

7.2 The three *E(Z)* homologues have distinct functions

Bioinformatic analysis has revealed that multiple E(Z) homologues exist in diverse plant species including *Arabidopsis,* rice, maize, *Petunia* and *Antirrhinum*. Only those in *Arabidopsis* have been well characterised. MEA and CLF have distinct functions, whilst CLF and SWN show partial redundancy. More detailed characterisation of *swn* single mutants may identify distinct cryptic functions or subtle phenotypes.

Some lines of evidence suggest that these proteins may show overlapping specificity with respect to target genes. Firstly, the proteins show a high degree of conservation and most likely have a similar biochemical activity towards histones. Secondly, data from rice and maize suggest there are no *MEA* homologues in monocotyledonous plants, and expression studies in these plants suggest that *MEA* activity is provided by *SWN* homologues (Springer et al., 2002; Thakur et al., 2003). On rare occasions in *Arabidopsis*, the maternal effects of null *mea* alleles are leaky and homozygous mutants escape (Vielle-Calzada et al., 1999). This suggests that either the requirement for MEA function is not absolute or that another protein has the potential to supply E(Z)-like function in the embryo and endosperm. As the requirement for FIE and MSI1 is absolute (Luo et al., 2000; Köhler et al., 2003a) the presence of a functional FIS complex is almost certainly required, and therefore another protein must have the potential to supply E(Z)-like function to the FIS complex, most likely CLF or SWN. Taken with the partial redundancy between CLF and SWN it seems likely that there is some overlap between the functions of these proteins.

To test the potential of MEA and SWN to provide CLF function, I engineered a series of transgenic lines which mis-expressed SWN and MEA vegetatively at high levels under the control of the 35S promoter and also under the control of a fragment of DNA believed to constitute the *CLF* endogenous promoter. In both experiments I observed that the neither mis-expression of MEA or SWN was able to complement *clf* mutants, whereas CLF was able to provide full complementation when expressed under the control of the 35S promoter, and partial complementation under its native promoter. This result demonstrates, that at least in *Arabidopsis*, the three E(Z) homologues are not equivalent with respect to regulation of *AG*. It also confirmed that although the expression pattern of MEA is different to that of CLF and SWN, it is not the only factor in determining the specificity of these proteins. One possibility is that differences in target gene specificity could be mediated by specific protein-protein interactions between members of the PRC2 complex, and these may be sufficient to recruit CLF,

MEA and SWN to different targets. Another possibility is that these specificities are controlled by interactions with different sequence specific transcription factors and these recruit the PRC2 to different targets.

7.3 At least three distinct PRC2 complexes exist in *Arabidopsis*

Strong evidence has previously been presented for a PRC2 complex involved in the control of endosperm development during seed development which includes the ESC homologue FIE, the SU(Z)12 homologue FIS2, the E(Z) homologue MEA and the p55 homologue MSI1. This complex was identified based on physical interactions between the protein products, overlapping expression patterns and strong genetic interactions (Spillane et al., 2000; Luo et al., 2000; C. Stock pers. comm.; Köhler et al., 2003a; Chaudhury et al., 1996; Ohad et al., 1996).

There is also evidence suggesting a second PRC2 complex exists which regulates expression of floral homeotic genes in embryos and seedlings, as a genetic interaction has been demonstrated between *CLF* and *EMF2* (Chanvivattana, 2002). The identification of VRN2 demonstrated a novel function of Pc-G genes in the epigenetic control of vernalization. As the memory of vernalization has been shown to be maintained through a PRC2-like mechanism by methylation of lysine residues 9 and 27 on histone H3, it is likely that a third PRC2 complex exists with VRN2 as a key player (Bastow et al., 2004; Sung and Amasino, 2004).

The Pc-G protein FIE has been shown to interact physically with all three *Arabidopsis* E(Z) homologues (Spillane et al., 2000; Chanvivattana, 2001; Luo et al., 2000; Katz et al., 2004). As it is a single copy gene and common to all complexes, it is unlikely to provide any specificity to the different complexes. I have shown that in all three cases the region of this interaction maps to the N-termini of the E(Z) homologues. Although this region shows poor sequence conservation it has a conserved function through its physical interaction with FIE. There is also genetic evidence to support these

interactions, as *fie* ovules resemble *mea* ovules, and transgenic lines with reduced *FIE* activity show phenotypes similar to *clf* or *clf swn* double mutants (Ohad et al., 1999; Chaudhury et al., 1999; Kinoshita et al., 2000; Katz et al., 2004). The construction of a steroid-inducible line has enabled the recovery of homozygous *fie* mutants and this confirms a role for *FIE* post seed development. These mutants resemble the *clf swn* double mutants, suggesting a common role for these proteins in embryo and seedling development.

I have also demonstrated the potential of the two Pc-G proteins CLF and EMF2 to interact in yeast two-hybrid assays. This supports genetic evidence of an interaction between CLF and EMF2 where, firstly, weak *emf2* mutants phenocopy *clf* mutants, and secondly, weak *clf* and *emf2* allele show a strong interaction so that the double mutant resembles severe *emf2* mutants (Chanvivattana, 2002). Recently, the physical interaction has also been independently confirmed using yeast split-ubiquitin and *in-vitro* binding assays (D. Schubert pers. comm.). The yeast split-ubiquitin assay differs from the yeast two-hybrid assay as the candidate proteins are fused to portions of the ubiquitin protein and the fusions are expressed in the cytoplasm rather than the nucleus (Johnsson and Varshavsky, 1994; Kim et al., 2002; Stagljar et al., 1998). The *in-vitro* binding assays provides independent confirmation that CLF and EMF2 proteins bind to one another directly (i.e. few other proteins are involved).

I have defined discrete regions of CLF and EMF2 which mediate the interaction: the conserved C5 domain of CLF and the VEFS box of EMF2. Subsequently, mammalian S(U)Z12 has been shown to associate with EZH2 (one of the mammalian E(Z) homologues) by coimmunoprecipitation. The region of SU(Z)12 to which this interaction mapped was identified by *in-vitro* binding assays as the VEFS box (Yamamota et al., 2003).

The C5 region encodes a novel domain which is conserved in all E(Z) homologues but is not found in any other SET domain proteins. This domain is well conserved in E(Z) and

its plant and mammalian homologues, but only poorly conserved in MES2 from *C. elegans*. Of the five cysteine residues the first three have conserved spacing and can be aligned identically in all E(Z) homologues. The last two cysteine residues in the C5 region of MES2, like in E(Z) are separated by three residues, however, the distance between the third and fourth cysteine residues is conserved between the *Arabidopsis*, *Drosophila* and mammalian homologues but not with MES2 (probably a result of a 21 bp insertion in the C5 region of MES2). More strikingly, other conserved residues close to the key cysteine residues are different in MES2. The importance of these residues is unknown; it is also not known if the difference in spacing between cysteine residues 3 and 4 reflects significant differences in their overall locations within the protein product. For example, the insertion could produce a loop such as the one shown in figure 7.1. In both these cases the five cysteine and one histidine residues are conserved and may both have the same biochemical properties (e.g. the binding of a metal ion). The other conserved residues in close proximity may have a specific role in interacting with SU(Z)12 homologues. As there are no SU(Z)12 homologues in *C. elegans*, the C5 region may have evolved to interact with a new class of proteins or maybe a relict of a C5 domain which used to mediate interactions with an ancestral VEFS containing protein. The importance of these residues could be tested by using S-D IVM to create mutations at these residues and the ability of the mutagenised proteins to interact with EMF2 tested in yeast two-hybrid assays or *in-vitro* pull down assays. The absence of SU(Z)12 homologues in *C. elegans* demonstrates that there may not be an absolute requirement for SU(Z)12 in the biochemical activity of the PRC2 complex, and the interaction with SU(Z)12 may therefore play a role in specifying the targets of the different plant Pc-G complexes.

A physical interaction between MEA and FIS2 has recently been demonstrated using the yeast two-hybrid assay (C. Stock pers. comm.). This supports previous genetic evidence that the two proteins interact based on the similarities between *mea* and *fis2* phenotypes and demonstrates that the E(Z) – SU(Z)12 partnership is conserved in the FIS complex (Chaudhury et al., 1997; Ohad et al., 1997).

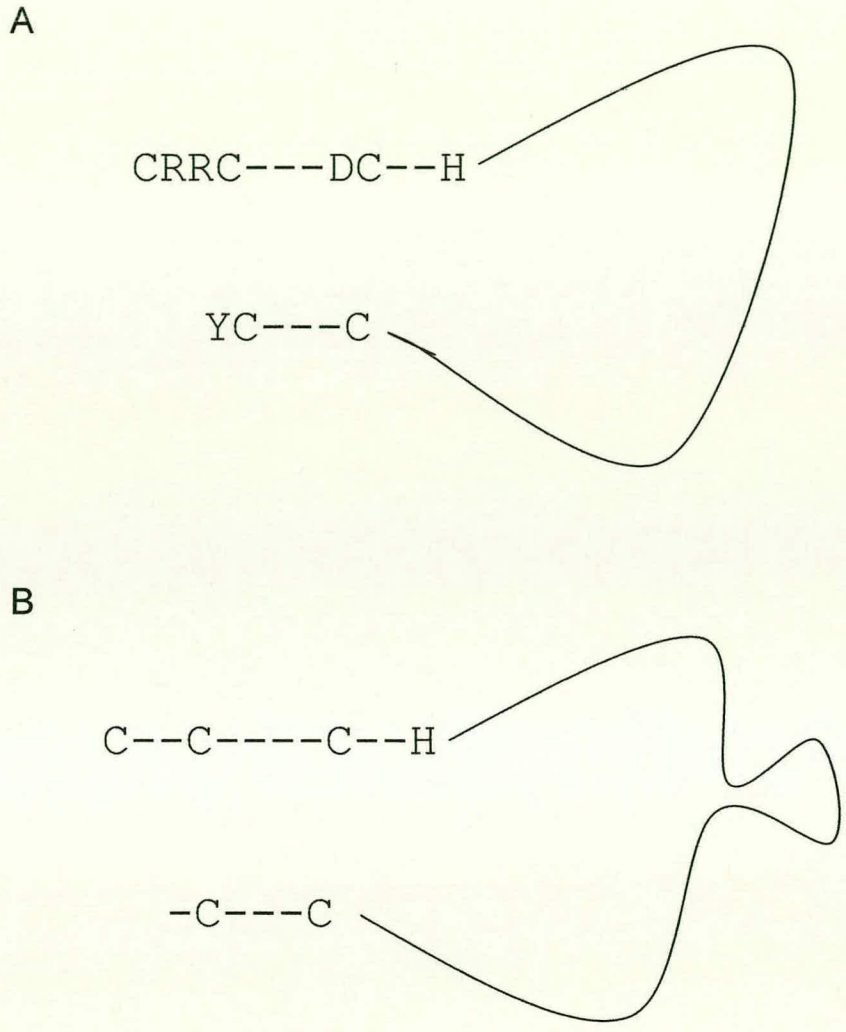


Figure 7.1 Hypothetical structures for the C5 domains
Figure A shows a consensus of the C5 domain from E(Z), CLF, SWN and MEA. Identical residues are shown, non identical residues are shown with a dash, the black loop represents 25 non-conserved residues. Figure B shows the C5 domain for MES2. Here the sequence between the conserved cysteine residues contains a 7 residue insertion (shown by an additional loop). Taken together these diagrams illustrate one possible scenario where although the sequence of the C5 region is different between E(Z) homologues, the overall spacing of the five cysteine residues may not be.

As previously noted CLF, MEA and SWN are not equivalent and fail to cross complement a *clf* mutant when mis-expressed. Clearly, all three show a similar interaction with FIE and this is not something which separates the biological function. I proposed that this may be due to CLF, SWN and MEA showing different affinities for EMF2, VRN2 and FIS2. I did not observe these differences in yeast assays; CLF and SWN showed similar potential to interact with each of the EMF2, VRN2 and FIS2 proteins. However, the interactions in yeast may not accurately reflect subtle differences in affinity *in planta*. Although the results from the chimaeric fusion were unclear, they suggest that differences between MEA and CLF most likely map to the N-terminal half of the protein; this region includes the C5. One possible way of elucidating whether differences in affinities for specific SU(Z)12 homologues recruit MEA and CLF to specific targets would be by making new chimaeric fusion proteins swapping the C5 domains of CLF and MEA.

Another core component of the PRC2 complex is the p55 protein. One of the *Arabidopsis* p55 homologues, MSI1, has been shown to interact with FIE *in-vitro* and *in-planta* (Köhler et al., 2003a). This interaction is again supported by genetic evidence as null *msi1* mutants have a *fis* phenotype, and partial loss of *MSI1* activity produces plants resembling *clf* mutants and lines with reduced *FIE* activity (Hennig et al., 2003; Köhler et al., 2003a; Katz et al., 2004; Goodrich et al., 1997).

In *Arabidopsis*, MSI is represented by a family of five proteins; however, it is unlikely to provide specificity observed in CLF, SWN and MEA. This is because MSI1 is involved in multiple aspects of plant development, and because its interaction with the PRC2 complexes is through FIE only, which is common to all three PRC2 complexes. However, as p55 is represented by a small family of five genes in *Arabidopsis*, if these were expressed in different regions they could recruit the PRC2 complexes to different targets. Future studies may confirm this by studying the expression pattern of these genes, and also the phenotypes associated with mutants and double mutants affecting

these loci. Whilst the expression patterns of the p55 family will be unable to explain the specificity of CLF, SWN and MEA, they may provide an extra mechanism by which the PRC2 complexes can be assigned to different targets.

7.4 A role for CLF and SWN in the vernalization response

The result that CLF and SWN both show the potential to interact with VRN2 in yeast two-hybrid assays suggests that these two proteins may function with VRN2 in a third PRC2 complex. It has been previously shown that VRN2 is required for vernalization-dependent mK27 H3 methylation of chromatin at the locus of the *FLC* gene (Bastow et al., 2004; Sung and Amasino et al., 2004). In animals, mK27 H3 methylation by the PRC2 complex requires the presence of E(Z) (Cao et al., 2002; Müller et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002). Taken together these data suggest that like SU(Z)12, VRN2 will require an E(Z) homologue to make the methylation marks on chromatin which confer an epigenetic memory of vernalization. Of the three candidates CLF and SWN are the more likely to mediate the vernalization response because they show strong expression in the shoot apical meristem, the site of the vernalization response (Goodrich et al., 1997; Chanvivattana, 2002; Sheldon et al., 2002). Although no gross effect on vernalization, similar to that of *vrn2*, was observed in either *clf* or *swn* mutants, a subtle effect was noticed in *clf* mutants exposed to cold for a short (3 week) period. The most likely scenario is that CLF and SWN act redundantly with respect to vernalization. This is consistent with the observation that *clf swn* double mutants have far greater morphological effects than either of the single mutants, for example no effect on root development is observed in either single mutant but double mutants show ectopic shoot formation from the primary root.

To further test whether *clf* mutants show an impaired response when subjected to a suboptimal vernalization, it will be necessary to repeat the experiment with a variety of periods of cold exposures (two to six weeks). If the reduced vernalization response after a three week period of cold treatment in *clf* mutants is reproducible, it would be

interesting to see the effect of altering *SWN* levels. This could be achieved by analysing firstly homozygous *clf* mutants which are heterozygous for *swn*, and secondly in *clf* mutants carrying the 35S::*SWN* transgene, where the level of SWN would be higher. This data may indicate whether there is an absolute need for CLF in the response to a suboptimal period of vernalization, or more likely, if small differences in combined levels of CLF and SWN have more profound effects when the period of vernalization is marginal. It is also important to reanalyse the response of *swn* mutants to a short period of vernalization. If *swn-5* and *swn-7* lines were crossed with FRI^{+}, the vernalization response would be more pronounced and individuals conferring altered responses more apparent.

Taken with other results these data strongly suggest that unlike in animals, at least three distinct PRC2 complexes exist in *Arabidopsis*. FIE is a common partner to all three complexes. MEA and FIS2 act as the key Pc-G proteins in a complex specific to regulation of endosperm development. EMF2, CLF and SWN act together in a complex responsible for the repression of floral homeotic and other genes in embryos and seedlings. VRN2 acts most likely with CLF and SWN in a complex stably repressing *FLC* in a vernalization dependent manner. These complexes can be seen schematically in figure 7.2

7.5 Diversification of plant PRC2 complexes

Whilst differences in expression patterns may account for the distinct roles of the three complexes, they are not the only factor. MEA and FIS2 have seed-specific expression patterns, whereas CLF, SWN, EMF2 and VRN2 all have much wider expression patterns, in vegetative tissue and seedlings (Luo et al., 2000; Vielle-Calzada et al., 1999; Goodrich et al., 1997; Chanvivattana, 2001; Gendall et al., 2001; Yoshida et al., 2001). However, these changes are insufficient to explain the altered roles of the proteins as only ectopic mis-expression of CLF is able to complement a homozygous *clf* mutant. In addition, there appears to be overlap between VRN2 and EMF2 expression, as the site of

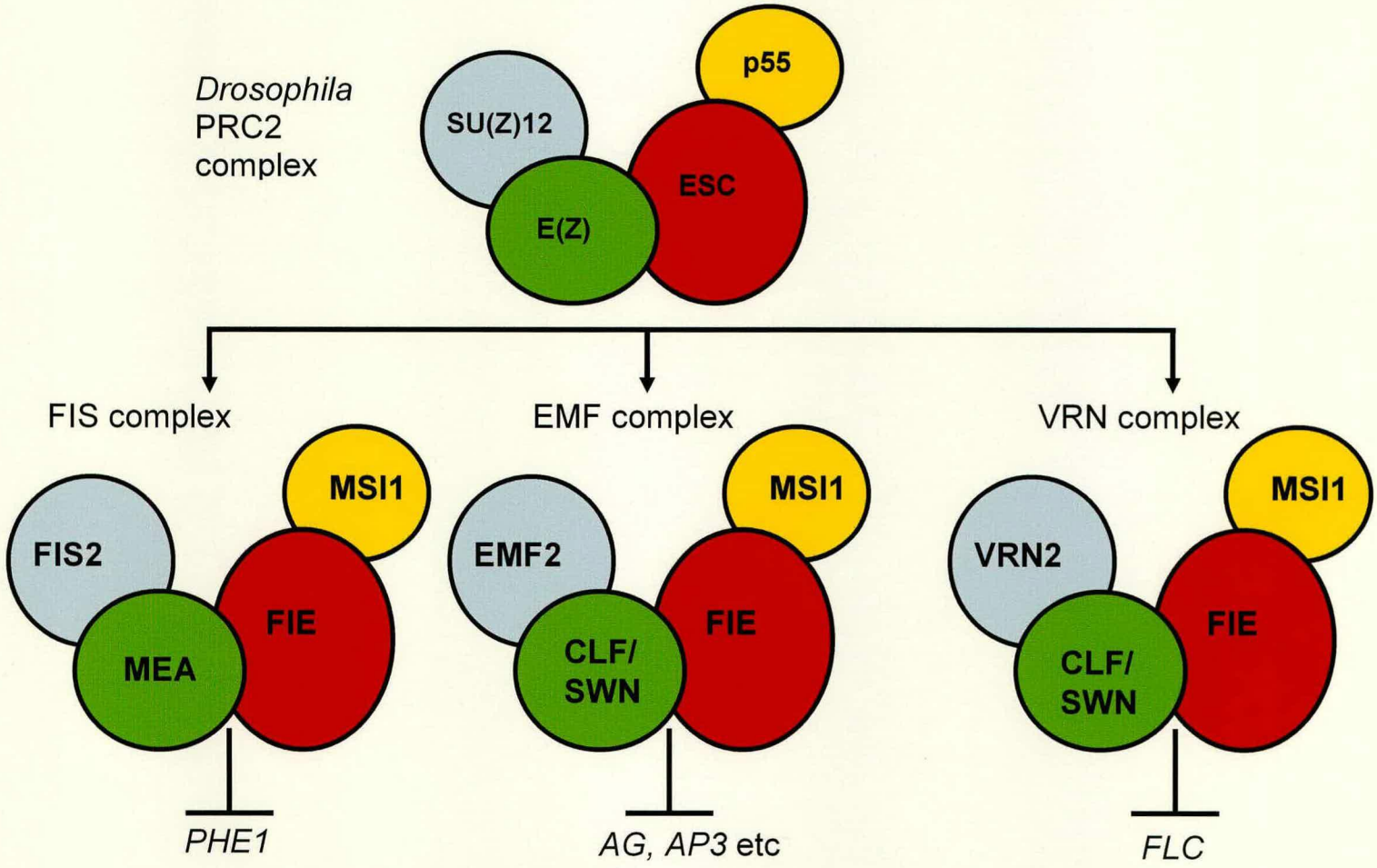


Figure 7.2 Three PRC2 complexes have been identified in *Arabidopsis* controlling distinct targets
The core components of the *Drosophila* PRC2 complex are shown above. In *Arabidopsis*, an ancestral complex is proposed to have diverged into three complexes with at least partially discreet functions. The same colour indicates homologous proteins, and regions of contact indicate physical interactions between proteins. The list of target genes is not comprehensive, and with time it is likely that new target genes will be identified.

the vernalization response (the shoot apical meristem) corresponds to a site where EMF2 is strongly expressed, but despite this both *emf2* and *vrn2* have distinct mutant phenotypes (Sheldon et al., 2002; Yoshida et al., 2001; Gendall et al., 2001).

One modern theory which may partly explain why in plants E(Z) and SU(Z)12 are represented by small gene families with distinct functions is the Duplicate-Degeneration-Complementation (DDC) model (Force et al., 1999). Previously, classic models used to describe the evolution of duplicate genes have stated that the most common outcome of gene duplications is that one member of the duplicate pair degenerates by accumulation of deleterious mutations, whilst the other duplicate retains the original function. Only on rare occasions, one duplicate may acquire a new adaptive function, resulting in the preservation of both duplicates, one with the classic function and one with the new. Force and colleagues (1999) present empirical data showing that in eukaryotes a much greater proportion of gene duplications is preserved than can be explained using the classic models. They present evidence that in a variety of plant and animal systems following polyploidization events, the number of genes avoiding nonfunctionalization (i.e. retaining some or all ancestral function) is considerable higher than can be predicted with classic models. For example, following allopolyploidization in maize 11 million years ago, an estimated 70 % of genes have retained some functionality (White and Doebly, 1998). Force and colleagues also note observations made in *Xenopus laevis* that nucleotide substitution patterns of ancient duplicated genes are consistent with both copies being subject to purifying selection (i.e. a significant proportion of nucleotide polymorphisms do not alter the protein sequence), and following more recent duplications, few null alleles can be observed in extant populations.

To combat these discrepancies between classic theory and observed data the DDC model was proposed. This model proposed that degenerative mutations (particularly those in regulatory elements) could increase the probability of gene preservation and that the usual mechanism of duplicate gene preservation is through the partitioning of functions

between ancestral genes. If duplicate chromosomes lose different genes, then for the organism to remain viable, the two chromosomes must complement each other by jointly retaining the full set of subfunctions on the original ancestral chromosome. Equally, if duplicate genes lose different regulatory subfunctions, the two genes must complement each other by jointly retaining the full set of subfunctions of the original ancestral gene. The advantage that this model has over the classic model is that more frequent deleterious mutations can help preserve duplicate genes. The DDC model proposes that instead of rare occasions when a mutation in one copy of a duplicate gene gives rise to a new and beneficial function being the main reason for fixation, a more common fate would be fixation through each duplicate experiencing a loss or reduction of expression necessary for a particular subfunction. In these cases, the combined action of both genes is needed to fulfil the requirements of the ancestral locus.

This hypothesis may help explain the evolution of *CLF* and *SWN*. My phylogenetic analysis showed that the divergence of *CLF* and *SWN* is an ancient one, and because homologues to both genes exist in diverse taxa, both genes are clearly important and have been strongly selected. The DDC model would suggest that the combined function of CLF and SWN is equivalent to that of the ancestral locus and that the two duplicate genes have been preserved because they each fulfil different complementary ancestral subfunctions. *In Arabidopsis*, exactly what these subfunctions are is not readily apparent. *CLF* and *SWN* have similar expression patterns and similar functions. However, SWN is not able to completely complement CLF function (e.g. with respect to regulation of *AG*), this can be observed in *clf* SWN^{+} plants and also in transgenic lines over-expressing SWN. Whilst this suggests a defined subfunction for CLF, a subfunction for SWN is less clear, especially as homozygous *swn* mutants have no discernable phenotype. A more holistic approach considering all flowering plants rather than just *Arabidopsis* suggests that *SWN* homologues may have a distinct role in seed development. For example, in rice and maize despite the existence of extensive EST databases and the screening of cDNA libraries no *MEA* homologue has been identified (Springer et al., 2002; Thakur et al., 2003). However, *SWN* homologues have been

identified with seed expression; furthermore one of the maize *SWN* homologues, *MEZ2*, shows tissue specific alternative splicing, with one spliced variant showing kernel-specific expression. This raises the possibility that distinct functions could arise from the same gene in different cells and that MEZ2 may provide MEA-like function in monocotyledonous plants. A role for SWN in seed development may then constitute a specific subfunction which CLF does not share.

Future studies could employ a number of methods to test these hypotheses. Firstly, reverse genetics could be employed in maize to investigate the exact functions of MEZ1, MEZ2 and MEZ3. The DDC theory would suggest that the sum of subfunctions of these genes would be broadly equivalent to the sum of subfunctions of CLF, SWN and MEA. It should be noted that the evolutionary time frames are so great that these genes may have acquired some new functions. Secondly, the phylogenetic analysis could be improved by the inclusion of more data; in particular the relationship with MEA needs to be more carefully studied. If *MEA* is absent from rice and maize it would be interesting to know if *MEA* is widely represented in dicotyledonous plants. My analysis shows that the divergence of CLF and SWN predated the divergence of monocotyledonous and dicotyledonous plants, as MEA is very different to CLF/SWN the divergence of MEA from the CLF/SWN common ancestor will be basal to this. The most parsimonious explanation would be that MEA has been lost in an evolutionary line containing the monocotyledonous plants maize and rice, but maintained in other plants. This could be investigated by confirming whether MEA homologues exist in other plant taxa, for example *Antirrhinum* and *Petunia*. It would also be interesting to note if discernable MEA and CLF/SWN homologues exist in more distant plant taxa, such as in those believed to be basal to the angiosperms. This may allow a timeframe to be placed on the divergence of MEA from the CLF/SWN common ancestor. The theory could also be tested by examining more closely related duplications such as those in *Antirrhinum* CLF homologues. Evidence suggests that evolution through subpartitioning is actively occurring in maize, as it was recently shown that there are two *FIE* homologues with more restricted expression patterns than the *Arabidopsis FIE*

gene, so that one is restricted to expression in seed and the other to vegetative and floral tissues (Danilevskaya et al., 2003; Springer et al., 2002).

It is likely that despite subpartitioning of gene function, considerable overlap exists between the *Arabidopsis* Pc-G. For example, the phenotype of *emf2* mutants is far less severe than either *clf swn* doubles or *fie* mutants. This suggests that there might be some partial redundancy between *VRN2* and *EMF2*: the construction of new double mutants would test this. In *Arabidopsis*, some lines of evidence suggest that some FIS2 and MEA-like function can be provided by other proteins. Firstly, the *fie* phenotype in seed is more severe than that of *fis2* or *mea*; in *mea* and *fis2* ovules the embryo arrests at the torpedo stage with no obvious morphological defects, whereas, in *fie* mutants severe morphological defects are observed (Sørenson et al., 2001). Secondly, the maternal lethal effects of *fis2* and *mea* can be leaky and viable homozygous seed produced (Vielle-Calzada et al., 1999; Luo et al., 2000). This suggests that activity of an E(Z) or SU(Z)12 homologue is not an absolute requirement, or more likely that similar activity can be provided by homologous proteins. These leaky maternal effects are never observed in *fie* mutants (Luo et al., 2000).

This hypothesis is supported by the theory that in *Arabidopsis* MEA-like function has the potential to be provided by other proteins. The observation that transgenic lines carrying the p(*CLF*)::*MEA* transgene were able to rescue *mea* ovules suggests that there is some *CLF* expression in the seed, and further supposes that CLF and MEA proteins have distinct functions. As the endogenous promoter failed to recreate the exact expression pattern of *CLF*, this interpretation should be viewed as provisional. Analysis of gametophytic *CLF* and *SWN* expression by *in-situ* hybridisation will be needed to determine whether they are expressed in the developing seed, and new double mutant combinations with *mea* may show whether CLF and SWN have a role in seed development, for example by looking for an enhancement of *mea*$^-$ phenotype.

7.6 Directions for future studies

Similar PRC2 complexes regulate diverse developmental processes in plants and animals and these likely have a similar biochemical activity towards histones. However, these differ because in *Arabidopsis* at least three separate complexes exist and each likely targets at least some different target genes. Further use of chimaeric fusion proteins may identify how proteins within these complexes confer this target specificity *in-vivo*. For example, new motifs such as the C5 region may be exchanged between CLF and MEA. Another approach would be to identify new motifs within CLF/MEA and EMF2/VRN2 which define new interactions or functions. This work is in progress already through yeast two-hybrid screens carried out for new proteins which have the potential to interact with CLF. New motifs defined by these interactions could be used in further chimaeric fusion proteins to test the significance of these regions on defining the specificity *in-vivo*. However, the possibility of future transgenic plants expressing chimaeric proteins exhibiting a dominant phenotype is a genuine concern. One possible way of elucidating whether this effect is due to co-suppression would be using Northern blot hybridisation to test for the presence of short interfering RNA raised against CLF. If co-suppression is the cause of these problems, it may be that other constructs do not cause these effects or that the effects can be lessened by using carefully selected lines with single insertions or by crossing to mutants which are defective in post-transcriptional gene silencing such as *sde3*, *sde1*, *sgs2* or *sgs3* (Dalmay et al., 2001; Dalmay et al., 2000; Mourrin et al., 2000).

Recovery of *fie* homozygotes through the construction of a steroid-inducible line may provide a valuable tool for analysing the biochemical nature of the PRC2 complex. This may allow histone preparations to be made from homozygous *fie* plants, so that levels of global K9 and K27 histone methylation can be compared between wild-type, *fie* mutants and *clf swn* double mutants. The advantage of the steroid-inducible approach is that once completely rescued *fie* homozygotes are identified they may produce large quantities of *fie*$^{-}$ seed, meaning that the large amounts of tissue necessary for histone preparations can be harvested. ChIP analysis could also be used to determine the histone

methylation status at particular targets, and this data could again be compared with the *clf swn* double mutants.

At the moment it is not certain whether the steroid-inducible construct is able to provide complete rescue of *fie* homozygotes to such a degree where they are able to set viable seed. Clearly, there is some rescue but it still needs to be confirmed how DEX-dependent or leaky this rescue is. One problem is that so far all analysis has been performed on the progeny of just one transformant. Analysis of other independent lines is necessary and may identify individuals were the DEX-dependent rescue is more complete, and these may assist in further analyses.

Comparisons between the expression levels of target genes can be made between *fie* and *clf swn* doubles by either RT-PCR or by microarray analysis. Although, these have similar phenotypes it would be useful to ascertain whether the same genes are de-regulated in both mutants. The steroid-inducible line may also be used to investigate the timing of FIE requirement. Although preliminary experiments suggest that the DEX-mediated rescue is incomplete, these experiments may be possible if new lines are identified were the DEX-dependent FIE activity complements *fie* homozygous mutants completely.

Many components necessary for Polycomb-mediated silencing in *Drosophila* are absent from *Arabidopsis*. There are no *PLEIOHOMEOTIC*, *ZESTE* and *GAGA* homologues in *Arabidopsis* suggesting that new DNA-binding proteins and *cis*-acting sequence motifs have been acquired to recruit the PRC2 complex to its target genes (Goodrich and Tweedie, 2002). It is also likely that plants use different factors than POLYCOMB to bind the mK27 mark as there is no PC homologue in *Arabidopsis*. This suggests that new mechanisms to interpret, to maintain and to re-set epigenetic information conveyed by the PRC2 complex may have evolved independently within the plant kingdom, or that factors involved in resetting epigenetic marks may be expressed more, i.e. beyond germline/meiosis resetting). The recruitment of an independent complex providing

PRC1-like function or the wider expression of these may account for the greater flexibility in plant development than that observed in animals. Future genetic screens to identify modifiers of the plant PRC2 mutant phenotypes may resolve this issue. For example, genetic screens for modifiers of *clf* may identify mutations with enhanced phenotypes and these may be due to a failure to interpret the K9 and K27 marks. Alternatively, these factors could also be identified biochemically by identifying proteins which bind specifically to methylated histones *in-vitro*.

8 References Cited

Aho, S., Arffman, A., Pummi, T., and Uitto, J., (1997). A novel reporter gene *MEL1* for the yeast two-hybrid system. **Anal. Biochem.** 253, 270-272.

Akam, M., (1983)., The location of *Ultrabithorax* transcripts in *Drosophila* tissue sections. **EMBO J.** 2:2075-84.

Akam, M., (1989). Hox and HOM: Homologous gene clusters in insects and invertebrates. **Cell** 57:347-9.

Akam, M. and Martinez-Arias, A., (1985). The distribution of *Ultrabithorax* transcripts in *Drosophila* embryos. **EMBO J.** 4:1689-1700.

Aubert, D., Chen, L., Moon, Y.-H., Castle, L., Yang, C.-H. and Sung, Z., (2001). EMF1, a novel protein involved in the control of shhot architecture and flowering in *Arabidopsis*. **Plant Cell** 13:1865-75.

Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidman, J., Smith, J. and Struhl, K., (1996). Current Protocols in Molecular Biology: Volume 1. John Wiley and Sons, Harvard.

Bannister, A., Zegerman, P., Partridge, J., Miska, E., Thomas, J., Allshire, R. and Kouzarides, T., (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. **Nature** 410:120-4.

Bartel, P., Chien, C. T., Sternglanz, R., and Fields, S., (1993). Elimination of false positives that arise in using the two-hybrid system. **Biotechniques** 14, 920-924.

Bastow, R., Mylne, J., Lister, C., Lippman, Z., Martienssen, R. and Dean, C. (2004). Vernalization requires epigenetic silencing of *FLC* by histone methylation. **Nature** 427:164-167.

Baumbusch, L., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I., Reuter, G. and Aalen, R., (2001). The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. **Nucleic Acids Res.** 29:4319-4333.

Becker, D., (1990). Binary vectors which allow the exchange of plant selectable markers and reporter genes. **Nucleic Acids Res.** 11;18(1), 203-206.

Birve, A., Sengupta, A., Beuchle, D., Larsson, J., Kennison J., Rasmuson-Lestander, A. and Muller, J., (2001). *Su(z)12*, a novel *Drosophila* Polycomb group gene that is conserved in vertebrates and plants. **Development** 128:3371-9.

Blázquez, M., and Weigel, D., (2000). Integration of floral inductive signals in *Arabidopsis*. **Nature** 404:889-92.

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

Bowman, J., Drews, G. and Meyerowitz, E., (1991b). Expression of the floral homeotic gene *AGAMOUS* is restricted to specific cell types late in development. **Plant Cell** 3:749-58.

Brown, J.W.S., (1996). *Arabidopsis*, introns, mutations and pre-mRNA splicing. **Plant J.** 10:771-780.

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

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. and Zhang, Y., (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. **Science** 298:1039-1043.

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

Chandler, J., Wilson, A. and Dean, C., (1996). *Arabidopsis* mutants showing an altered response to vernalization. **Plant J.** 10:637-44.

Chanvivattana, Y., (2002). Interactions amongst Polycomb-group genes regulating flowering in *Arabidopsis*. PhD thesis. University of Edinburgh.

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

Chen, L., Cheng, J., Castle, L. and Sung, Z., (1997). *EMF* genes regulate *Arabidopsis* inflorescence development. **Plant Cell** 9:2011-24

Coen, E. and Meyerowitz, E., (1991). The war of the whorls: genetic interactions controlling flower development. **Nature** 353:31-7.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A. and Pirrotta, V., (2002). *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. **Cell** 111:185-196.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S.and Baulcombe, D., (2000). An RNA dependent RNA polymerase gene in *Arabidopsis* is required for post-transcriptional gene silencing by a transgene not a virus. **Cell** 101:543-53.

Dalmay, T., Horsefield, R., Hartig Braustein, T. and Baulcombe, D., (2001). *SDE3* encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis*. **EMBO J.** 8:2069-77.

Devic, M., Albert, S. and Delseny M., (1996). Induction and expression of seed-specific promoters in *Arabidopsis* embryo-defective mutants. Plant J. 9:205-215.

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

Feng, Q., Wang, H., Ng, H., Erdjument-Bromage, H., Tempst, P., Struhl, K. and Zhang, Y., (2002). Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. **Curr. Biol.** 12:1052-1058.

Fields, S. and Song, O., (1989). A novel genetic system to detect protein-protein interactions. **Nature** 340, 245-246.

Fischle, W., Wang, Y., Jacobs, S., Kim, Y., Allis, C. and Khorasanizadeh, S., (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. **Genes Dev.** ;17:1870-81.

Force, A., Lynch, M., Pickett, F., Amores, A., Yan, Y. and Postlethwait, J., (1999). Preservation of duplicate genes by complementary, degenerative mutations. **Genetics** 151:1531-45.

Franks, R., Wang, C., Lerevin, Z. and Lui, Z., (2002). *SEUSS*, a member of a novel family of plant regulatory proteins, represses floral homeotic gene expression with *LEUNIG*. **Development** 129:253-63.

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

Gleave, A., (1992). A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. **Plant Mol. Biol.** 20, 1203-1207.

Goodrich, J., Puangsomlee, 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.

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

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

Gutjahr, T., Frei, E., Spicer, C., Baumgartner, S., White, R. and Noll, M., (1995). The Polycomb group gene, *extra sex combs*, encodes a member of the WD-40 repeat family. **EMBO J.** 14:4296-4306.

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

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

Holdeman, R., Nehrt, S. and Strome, S., (1998). MES-2, a maternal protein essential for the viability of the germline in *Caenorhabditis elegans* is homologous to a *Drosophila* Polycomb protein. **Development** 125:2457-67.

Hur, M., Laney, J., Jeon, S., Ali, J. and Biggin, M., (2002). *Zeste* maintains repression of *Ubx* transgenes: support for a new model of Polycomb repression. **Development** 129:1339-43.

Jack, T., Fox, G. and Meyerowitz, E., (1994). *Arabidopsis* homeotic gene *APETALA3* ectopic expression: transcriptional and post-transcriptional regulation determine floral organ identity. **Cell** 76:703-16.

Jacobs, S., Harp, J., Devarakonda, S., Kim, Y., Rastinejad, F. and Khorasanizadeh, S., (2002). The active site of the SET domain is constructed on a knot. **Nat. Struct. Biol.** 9:833-838.

Jenuwein, T., (2001). Re-SET-ting heterochromatin by histone methyltransferases. **Trends Cell Biol.** 11:266-73.

Johnsson, N. and Varshavsky, A., (1994). Split ubiquitin as a sensor of protein interactions *in vivo*. **Proc. Natl. Acad. Sci. U. S. A.** 91:10340-10344.

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R. and Dean, C., (2000). Molecular analysis of *FRIGIDA* a major determinant of natural variation in *Arabidopsis* flowering time. **Science** 290:344-7.

Jones, R. and Gelbart, W., (1990). Genetic analysis of the *Enhancer of zeste* locus and its role in gene regulation in *Drosophila melanogaster*. **Genetics** 126:185-99.

Jones, R. and Gelbart, W., (1993). The *Drosophila* Polycomb group gene *Enhancer of zeste* contains a region with sequence similarity to trithorax. **Mol. Cell. Biol.** 13:6357-66.

Jones, C., Ng, J., Peterson, A., Morgan, K., Simon, J. and Jones, R., (1998). The *Drosophila* esc and E(z). proteins are direct partners in polycomb group-mediated repression. **Mol. Cell Biol.** 18:2825-2834.

Katz, A., Oliva, M., Mosquna, A., Hakim, O. and Ohad, N., (2004). FIE and CURLY LEAF polycomb proteins interact in the regulation of homeobox gene expression during sporophyte development. **Plant J.** 37:707-719.

Kim, M., Panstruga, R., Elliot, C., Muller, J., Devoto, A., Yoon, H., Park, H., Cho, M. and Schulze-Lefert, P., (2002). Calmodulin interacts with MLO protein to regulate defence against mildew in barley. **Nature** 416:447-51.

Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL. Imprinting of the *MEDEA* polycomb gene in the *Arabidopsis* endosperm. **Plant Cell.** 1999 Oct;11(10):1945-52.

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

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J., Goldberg, R. and Fischer, R., (1999). Control of fertilization-independent endosperm development by the *MEDEA* polycomb gene in *Arabidopsis.* **Proc Natl Acad Sci U S A.** 96:4186-91.

Köhler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U. and Gruissem, W., (2003a). *Arabidopsis* MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. **EMBO J.** 22:4804-14.

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

Kolhami, S., (1998). Identification and characterisation of proteins using yeast two-hybrid analysis. Plant Molecular Biology Manual. M1:1-30.

Koncs, C. and Schell, J., (1986). The promoter of the TL-DNA gene 5 controls the tissue specific expression of chimeric gene carrier O by a novel type of *Agrobacterium* binary vector. **Mol. Gen. Genet.** 204:283-396.

Koornneef, M., Hanhart,C. and van der Veen,J., (1991). A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. **Mol. Gen. Genet.** 229:57-66.

Korf, I., Fan, Y. and Strome, S., (1998). The Polycomb group in *Caenorhabditis elegans* and maternal control of development. **Development** 125:2469-78.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. and Reinberg, D., (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. **Genes Dev.** 16:2893-2905.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. and Jenuwein, T., (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. **Nature.** 410:116-20.

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.

Lawrence, P., (1992). The making of a fly: the genetics of animal design. Blackwell Scientific Publications, Oxford.

Lawrence, P. and Morata, G., (1994). Homeobox genes: their function in *Drosophila* segmentation and pattern formation. **Cell** 78:181-9.

Levy, Y., Mesnage, S., Mylne, J., Gendall, A. and Dean, C., Multiple roles of *Arabidopsis* VRN1 in vernalization and flowering time control. **Science** 297:243-6.

Lewis, E., (1978). A gene controlling segmentation in Drosophila. **Nature** 276:565-70

Lim, M., Kim, J., Kim, H., Chung, K., Seo, Y., Lee, I., Kim, J., Hong, C., Kim, H. and Park, C., (2004). A new *Arabidopsis* gene *FLK* encodes an RNA binding protein with K homology motifs and regulates flowering time via FLOWERING LOCUS C. **Plant Cell** 16:731-40.

Liu, Z. and Meyerowitz, E., (1995). *LEUNIG* regulates *AGAMOUS* expression in *Arabidopsis* flowers. **Development** 121:975-91.

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.

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

Luo, M., Bilodeau, P., Dennis, E., 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* seed. **Proc. Natl. Acad. Sci. U. S. A.** 97, 10637-10642.

Lyko, F. and Paro, R., (1999). Chromosomal elements conferring epigenetic inheritance. Bioessays. 21:824-32.

Ma, J. and Ptashne, M., (1987). Deletion analysis of GAL4 defines two transcriptional activating segments. **Cell** 48, 847-853.

Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C. and Dean,C., (1997). *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. **Cell** 89:737-745.

Mayama, T., Ohtsubo, E., and Tsuchimoto, S., (2003). Isolation and expression analysis of petunia *CURLY LEAF*-like genes. **Plant Cell Physiol.**44, 811-819.

Metz, A., Timmer, R. and Browning, K.S., (1992). Sequences for two cDNAs encoding *Arabidopsis thaliana* eukaryotic protein synthesis initiation factor 4A. **Gene** 120, 313-314.

Michaels, S. and Amasino, R., (1999). *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. **Plant Cell** 11:949-56.

Michaels, S. and Amasino, R., (2001). Loss of *FLOWERING LOCUS C* eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutants but not responsiveness to vernalization. **Plant Cell** 13:935-41.

Miesfield, R., Rusconi, S., Godowski, P., Maler, B., Okret, S., Wikström, A-C., Gustafsson, J-A. and Yamamoto, K., (1986). Genetic complementation of a glucocorticoid receptor deficiency by expression of cloned receptor cDNA. **Cell** 46:389-399.

Miller, J., (1972). Experiments in Molecular Genetics. Cold Spring Harbour Laboratory, Cold Spring Harbour, NY).

Min, J., Zhang, X., Cheng, X., Grewal, S. and Xu, R.M., (2002). Structure of the SET domain histone lysine methyltransferase Clr4. **Nat. Struct. Biol.** 9:828-832.

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

Moon, Y.-H., Chen, L., Pan, R., Chang, H.-S., Zhu, T., Maffeo, D. and Sung, Z., (2003). *EMF* genes maintain vegetative development by repressing the floral program in *Arabidopsis*. **Plant Cell** 15:681-693.

Mourrain, P., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.-B., Jouette, D., Lacombe, A.-M., Nikic, S., Picault, N., Rémoué, K., Sanial, M., Vo, T.-A. and Vaucheret, H., (2000). *Arabidopsis SGS2* and *SGS3* genes are required for post-transcriptional gene silencing and natural virus resistance. **Cell** 101:533-542.

Müller, J., Hart, C., Francis, N., Vargas, M., Sengupta, A., Wild, B., Miller, E., O'Connor, M., Kingston, R. and Simon, J.A., (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. **Cell** 111, 197-208.

Nakayama, J., Rice, J., Strahl, B., Allis, C. and Grewal, S., (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. **Science** 292:110-3.

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

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

Ohad, N., Margossian, L., Hsu, Y., Williams, C., Repetti, P. and Fischer, R., (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., Goldberg, R., and Fischer, R., (1999). Mutations in *FIE*, a WD polycomb group gene, allow endosperm development without fertilization. **Plant Cell** 11, 407-416.

Orlando, V., (2003). Polycomb, epigenomes, and control of cell identity. **Cell** 112:599-606.

Pautot, V., Dockx, J., Hamant, O., Kronenberger, J., Grandjean, O., Jublot, D. and Traas J., (2001). *KNAT2*: evidence for a link between Knotted-like genes and carpel development. **Plant Cell** 13:1719:1734.

Picard, D., (2000). Post-translational regulation of proteins by fusions to steroid-binding domains. **Mech. Enzymol.** 327:385-401.

Poethig, R., (1990). Phase change and the regulation of shoot morphogenesis in plants. **Science** 250:923-30.

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

Putterill, J., Laurie, R. and Macknight, R., (2004). It's time to flower: the genetic control of flowering time. **Bioessays** 26:363-73.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B., Sun, Z., Schmid, M., Opravil, S., Mechtler, K., Ponting, C., Allis, C. and Jenuwein, T., (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. **Nature** 406:593-9.

Reichmann, J., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda O., Ratcliffe, O., Samaha, R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J., Ghandehari, D., Sherman, B. and Yu, G., (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. **Science** 290:2105-10.

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

Reiser, L. and Fischer, R., (1993). The ovule and the embryo sac. **Plant Cell** 5:1291-1301.

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

Samach, A. and Coupland, G., (2000). Time measurement and the control of flowering in plants. **BioEssays** 22:38-47.

Samach, A., Onouchi, H., Gold, S., Ditta, G., Schwartz-Sommer, Z., Yanofsky, M. and Coupland, G., (2000). Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. **Science** 288:1613-6.

Sathe, S. and Harte, P., (1995). The *Drosophila* extra sex combs protein contains WD motifs essential for its function as a repressor of homeotic genes. **Mech. Dev.** 52:77-87.

Saurin, A., Shao, Z., Erdjument-Bromage, H., Tempst, P. and Kingston, R., (2001). A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins. **Nature** 412:655-60

Schena, M., Lloyd, A. and Davis W., (1991). A steroid-inducible gene expression system for plant cells. **Proc. Natl. Acad. Sci. U. S. A.** 88:10421-10425.

Schumacher, A., Lichtarge, O., Schwartz, S. and Magnuson, T., (1998). The murine Polycomb-group gene *eed* and its human orthologue: functional implications of evolutionary conservation. **Genomics** 54:79-88.

Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M., Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., and Shinozaki, K., (2002). Functional annotation of a full-length *Arabidopsis* cDNA collection. **Science** 296, 141-145.

Sewalt, R., Gunster, M., Hamer, K., den Blaauwen, J., Satijn, D., Hendrix, T., van Driel, R. and Otte, A., (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., Wu, C., Bender, W. and Kingston, R., (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. **Cell** 98:37-46.

Sheldon, C., Burn, J., Perez, P., Metzger, J., Edwards, J., Peacock, W., and Dennis, E., (1999). The *FLF* MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. **Plant Cell** 11, 445-458.

Sheldon, C., Rouse, D., Finnegan, E., Peacock, W. and Dennis, E., (2000). The molecular basis of vernalization: the central role of *FLOWERING LOCUS C* (*FLC*). **Proc. Natl. Acad. Sci. U. S. A.** 97:3753-3758.

Simon, J., Bornemann, D., Lunde, K. and Schwartz, C., (1996a). The extra sex combs protein contains WD40 repeats and its time of action implies a role distinct from other Polycomb group products. **Mech. Dev.** 53:197-208.

Simon, R., Igeño, M. and Coupland, G., (1996b). Activation of floral meristem identity genes in *Arabidopsis*. **Nature** 384:59:62.

Simpson, G. and Dean C., (2002). *Arabidopsis*, the Rosetta stone of flowering? **Science** 296:285-9.

Slifer, E., (1942). A mutant stock of *Drosophila* with extra sex-combs. **J. Exp.Zool.** 30:31-40.

Sørensen, M., Chaudhury, A., Robert, H., Bancharel, E. and Berger, F., (2001).. Polycomb group genes control pattern formation in plant seed. **Curr. Biol.** 20:277-81.

Spillane, C., MacDougall, C., Stock, C., Kohler, C., Vielle-Calzada, J.-P., Nunes, S., 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.

Springer, N., Danilevskaya, O., Hermon, P., Helentjaris, T., Phillips, R., Kaeppler, H., and Kaeppler, S., (2002). Sequence relationships, conserved domains, and expression patterns for maize homologs of the polycomb group genes *E(z), esc*, and *E(Pc)*. **Plant Physiol.** 128, 1332-1345.

Stagljar, I., Korostensky, C., Johnsson, N. and te, H., (1998). A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins *in vivo*. **Proc. Natl. Acad. Sci. U. S. A.** 95:5187-5192.

Struhl, G., (1981). A gene product required for correct initiation of segmental determination in *Drosophila*. **Nature** 293:36-41.

Struhl, G., (1982). Genes controlling segmental specification in the *Drosophila* thorax. **Proc. Nat. Acad. Sci. U. S. A.** 79:7380-4.

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

Struhl, G. and Akam, M., (1985). Altered distribution of *Ultrabithorax* in *extra sex combs* mutant embryos of *Drosophila*. **EMBO J.** 4:3259-64.

Sung, Z., Belachew, A., Bai, S., Bertrand-Garcia, R., (1992). *EMF*, an *Arabidopsis* gene required for vegetative shoot development. **Science** 258:1645-7.

Sung, S. and Amasino, R., (2004). Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. **Nature** 427:159-164.

Tautz, D., (1988). Regulation of the *Drosophila* segmentation gene *hunchback* by two morphogenetic centres. **Nature** 332:281-4.

Thakur, J., Malik, M., Bhatt, V., Reddy, M., Sopory, S., Tyagi, A. and Khurana, J.P., (2003). A POLYCOMB group gene of rice (*Oryza sativa* L. subspecies indica), *OsiEZ1*, codes for a nuclear-localized protein expressed preferentially in young seedlings and during reproductive development. **Gene** 314:1-13.

Thali, M., Müller, M., DeLorenzi, M., Matthias, P. and Bienz, M., (1988). *Drosophila* homeotic genes encode transcriptional activators similar to mammalian OTF-2. **Nature** 336:588-600.

Tie, F., Furuyama, T., and Harte, P., (1998). The *Drosophila* Polycomb Group proteins ESC and E(Z). bind directly to each other and co-localize at multiple chromosomal sites. **Development** 125:3483-3496.

Till, B., Colbert, T., Tompa, R., Enns, L., Codomo, C., Johnson, J., Reynolds, S., Henikoff, J., Greene, E., Steine, M., Comai, L. and Henikoff, S., (2003a). High-Throughput TILLING for Functional Genomics. **Methods Mol. Biol.** 236, 205-220.

Till, B., Reynolds, S., Greene, E., Codomo, C., Enns, L., Johnson, J., Burtner, C., Odden, A., Young, K., Taylor, N., Henikoff, J., Comai, L. and Henikoff, S. (2003b). Large-scale discovery of induced point mutations with high-throughput TILLING. **Genome Res.** 13, 524-530.

Tokunaga, C. and Stern, C., (1965). The development autonomy of extra sex combs in *Drosophila melanogaster*. **Dev. Biol.** 11:51-81.

Trievel, R., Beach, B., Dirk, L., Houtz, R., and Hurley, J., (2002). Structure and catalytic mechanism of a SET domain protein methyltransferase. **Cell** 111:91-103.

Tschiersch, B., Hofmann, 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.** 13:3822-31.

van Leeuwen, F., Gafken, P. and Gottschling, D., (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. **Cell** 109:745-56.

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

Warren, R., Nagy, L., Selegue, J., Gates, J. and Carroll, S., (1994). Evolution of homeotic gene regulation and function in flies and butterflies. **Nature** 372:458-61

Weigel, D. and Meyerozitz, E., (1994). The ABCs of floral homeotic genes. **Cell** 78:203-9.

White, R. and Wilcox, M., (1984). Protein products of the Bithorax complex in *Drosophila*. **Cell** 39:163-71.

White, R. and Wilcox, M., (1985). Distribution of Ultrabithorax proteins in *Drosophila*. **EMBO J.** 4:2035-43.

White, R. and Lehmann, R., (1986). A Gap gene, *hunchback*, regulates the spatial expression of *Ultrabithorax*. **Cell** 47:311-21.

White, S. and Doebly, J., (1998). Of genes and genomes and the origin of maize. **Trends Genet.** 14:327-332.

Wilson, R., Heckman, J. and Somerville, C., (1992). Gibberellin is required for flowering in *Arabidopsis thaliana* under short days. **Plant Physiol.** 100:403-8.

Wilson, C., (2001). Characterisation of Polycomb-group genes from *Antirrhinum majus* and *Arabidopsis thaliana*. PhD thesis. University of Edinburgh.

Wilson, J., Jing, C., Walker, P., Martin, S., Howell, S., Blackburn, G., Gamblin, S. and Xiao, B., (2002). Crystal structure and functional analysis of the histone methyltransferase SET7/9. **Cell** 111:105-115.

Xiao, B., Wilson, J.R. and Gamblin, S., (2003). SET domains and histone methylation. **Curr. Opin. Struct. Biol.** 13:699-705.

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., Cohen, G., Katz, A., Choi, Y., Katz, A., Nakashima, K., Harada, J., Goldberg, R., Fischer, R. 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** 12:2367-2382.

Yamamoto, K., (1985). Steroid receptor regulated transcription of specific genes and gene networks. **Annu. Rev. Genet.** 19 209-252.

Yang, C.-H., Chen, L.-J. and Sung, Z., (1995). Genetic regulation of shoot development in *Arabidopsis*: role of the *EMF* genes. **Devel. Biol.** 169:421-435.

Yoshida, N., Yanai, Y., Chen, L., Kato, Y., Hiratsuka, J., Miwa, T., Sung, Z. and Takahashi, S., (2001). EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates shoot development and flowering in *Arabidopsis*. **Plant Cell** 13:2471-81.

Zachgo, S., Silva Ede, A., Motte, P., Trobner. W., Saedler, H. and Schwarz-Sommer, Z., (1995). Functional analysis of the Antirrhinum floral homeotic *DEFICIENS* gene *in-vivo* and *in-vitro* by using a temperature-sensitive mutant. **Development** 121:2861-75.

Zhang, X., Tamaru, H., Khan, S., Horton, J., Keefe, L., Selker, E. and Cheng, X., (2002). Structure of the *Neurospora* SET domain protein DIM-5, a histone H3 lysine methyltransferase. **Cell** 111:117-127.

Zhang, H., Azevedo, R., Lints, R., Doyle, C., Teng, Y., Haber, D. and Emmons, S., (2003). Global regulation of Hox gene expression in *C. elegans* by a SAM domain protein. **Dev. Cell.** 4:903-15.

9 Appendices

Primers used to insert *Eco*R I site in SWN cDNA to make 35S::*SWN*

AB-EcoRIswn5f	cgacggccagtgaattctaatacgact
AB-EcoRIswn5r	agtcgtattagaattcactggccgtcg

Primers used in the cloning of the p(*FIE*)::GR-*FIE* construct

AB-GRBAM1F	CTGTTTCATCAAAAAgGAtccAGCTAGTTCTAGAGC
AB-GRBAM1R	GCTCTAGAACTAGCTcctTCcTTTTTCATGAAACAG
AB-GRBAM1F	GCTTCTGTTTCATCAAAAggatCcGAGCTAGTTCTAGAGCGGCCGC
AB-GRBAM2F	GCGGCCGCTCTAGAACTAGCTCgGatccTTTTGATGAAACAGAAGC
AB-KOBam1	ctgcagcccgggggaaccactagttctaga
AB-KOBam2	tctagaactagtggttcccccgggctgcag
AB-FIEBam1F	cgaatatcgaatgtcgaaggatcccttagggaacgagtcaatag
AB-FIEBam1R	ctattgactcgttccctaagggatccttcgacattcgatattcg

Primers used to construct MEA and CLF fusion proteins

AB3	cttgatatcgaattcgcccttatg	Amplifies fragment for MCS and MCT
AB4	gcttctcctcggatccaaaactaacg	Amplifies fragment for CMS and CMT
AB5	cgcgacccggatcttaa	Amplifies fragment for CMS and CMT
AB6	tcaatgtgttcttccgtgtg	Amplifies fragment for MCS and MCT
AB7	gctaagttttattgccgtcagtttaatccatgt	Amplifies fragment for MCS
AB8	aaactgacggcaataaaacttagcttctccact	Amplifies fragment for MCS
AB9	aaagaccagcccaagcactacacaccatgcact	Amplifies fragment for CMS
AB10	tgtgtagtgcttgggctggtctttcttctcagt	Amplifies fragment for CMS
AB11	ccagaacatgcgccagcttgggccaaaaaacct	Amplifies fragment for MCT
AB12	ggcccaagctggcgcatgttctggtccatagca	Amplifies fragment for MCT
AB13	ccagatcgagctgattggtcgcgtggtcgagaa	Amplifies fragment for CMT
AB14	acgcgaccaatcagctcgatctggctcataccg	Amplifies fragment for CMT

Primers used to sequence *CLF* cDNA clones

JG1	gaaccattcctgaactggcaggtc
YC24	tgcctgatatcaaacgtttgtc
G2r	agcagtttgcaccacaggtta
G6	cctactcgtctcccactacct
G11	agaggcgataattatgagtgc
G11r	gcactcataattatcgcctct
H2	tggctcataccggtaatcgta
LP8	ggaactcgctcagcaac
LP7	ctccagcaaacttaatg
YC25	acaggaacgtcatccttcaca
YC2	aactttaacccgagtctcag

Primers used to sequence *MEA* cDNA clones

measeq	aatggatcgtggtagcttct
AB-seqmea15	ttaccaagtgtcaagcttccaa
AB-seqmea16	ggcaatgacgtctatcagca
AB-seqmea17	gctggtgaggcttctgattt
AB-seqmea18	ttggtaacctgattgctttg

AB-seqmea19	tacatgcgcgaacaagatca	
AB-seqmea20	tgcatgttcttgcattggat	
AB-seqmea21	tggcactcttggtgagacac	
AB-seqmea22	gacgggctttagacctt	

Primers used to create constructs pCMGB6-11

jg43	TTCCGGGATATGtCgaCATTGAAGGTAA	Introduces *Sal* I site
jg44	TTACCTTCAATGtcGaCATATCCCGGAA	Introduces *Sal* I site
jg45	CCGATAAAGTTTCATCgTCGaCAAAGGTGAAAGG	Introduces *Sal* I site
jg46	CCTTTCACCTTTGtCGAcGATGAAACTTTATCGG	Introduces *Sal* I site
jg47	taatgaagatgctacGtcGActtctcagaagcatg	Introduces *Sal* I site
jg48	catgcttctgagaagTCgaCgtagcatcttcatta	Introduces *Sal* I site
AB-CLFtrunc1f	ggacctgccagttcagaattcgttcaaggatctagtgtcc	*Eco*R I site
AB-CLFtrunc1r	ggacactagatccttgaacgaattctgaactggcaggtcc	*Eco*R I site
AB-CLFtrunc2f	gcaactaggtctttcagacgaattcctggcggaactagc	*Eco*R I site
AB-CLFtrunc2r	gctagttccgccaggaattcgtctgaaagacctagttgc	*Eco*R I site
AB-CLFtrunc3f	aaagatatggaaggagaattcgattcttcgataacctgt	*Eco*R I site
AB-CLFtrunc3r	acaggttatcgaagaatcgaattctccttccatatcttt	*Eco*R I site

Primers used to create pCLFC137Y

AB-CLF-C317Y	cctgtggtgcaaactactataaaacgcttctc	mimics E(z)28
AB-CLF-C317Yr	gagaagcgttttatagtagtttgcaccacagg	mimics E(z)28

Primers used to modify pBI-880 CLF

AB-CLF108f	gatatgtagaagatggagctcccagttcaggaatgg	inserts *Sac* I
AB-CLF108r	ccattcctgaactgggagctccatcttctacatatc	inserts *Sac* I

Primers used to amplify C5 region of SWN for pGBT9 construct

AB-SWNC5f	gctgggaatatctctggaattcggcttggtgcagctc	*Eco*R I site
AB-SWNC5r	gctttctcttctgctgtcgacgcaaaattactgcatg	*Sal* I site

Primers used to amplify fragments used for MEA C5 constructs

AB-MEAC5f	ggtgaggcttctgaattcacatccaagaca	MEAC5 (*Eco*R I)
AB-MEAC5r	gatatagagttatcattgtcgaccacatcatcagc	MEAC5 (*Sal* I)
MEAC52f	agatacaatgaattcaagcttaag	MEAC5II (*Eco*R I)
MEAC52r	gtgttgtttggagcgtcgaccacaatc	MEAC5II (*Sal* I)
MEAC53f	aagtacctcgaattcgatgtttcggac	MEAC5III (*Eco*R I)
MEAC53r	ctctacaggcgtcgacatagtgttgc	MEAC5III (*Sal* I)

Primers used to create EcoR I site at MEA 5' to clone full length construct into pGBT9

MEA-Eco-5'f	gaaggaaaaccatgaattcgatggtgagggtttgc
MEA-Eco-5'r	gcaaaccctcaccatcgaattcatggttttccttc

Primers used to amplify FIS2 to clone VEFS box into pGAD424

FIS2VEFS	cctaaggtgcgtgaattcaatgatgataat	contains *Eco*R I
FIS2VEFSr	agccttgttgatccctag	

Primers used to repair PCR-induced error in FIS2 constructs

Rep-FIS2r	ggttcagaaggctgagcttttgc
Rep-FIS2f	gcaaaagctcagccttctgaacc

Primers used to genotype clf TILLING alleles

AB32	gttctcgcgatcgtatatcttcgcg	Score *clf-57* and *clf-58*
AB33	gatgtttctggttggggagct	dCAP *clf-57 Hha* I
AB6	tcaatgtgttcttccgtgtg	Score *clf-58* and *clf-55*
H1	ctggttggggagctttcttaa	Score *clf-55*
AB34	cacataaagaggcagataaccgcg	Score *clf-54*
AB35	cctttcctgtaagcatctag	dCAP *clf-54 Sac* II
AB-G799E	agtgtaagtaagcatgaataccatg	dCAP *clf-56 Nco* I

Primers used to amplify eIF as a RT-PCR control

eIF4a-T22	ttcgctcttctctttgctctc	(See Metz et al., 1992)
eIFB221	gaactcatcttgtccctcaagta	

Primers used to genotype *FRI*

UJ26	agatttgctggatttgataag	(See Johansen et al.,2002)
UJ34	atatttgatgtgctctcc	

Primers used to genotype *fie* TILLING alleles

AB15	aggagaattctctttcagttg	Score *fie-14* and *fie-11*
AB16	ccatcgacagacatggctgtttg	Score of *fie-12* and *fie-13*
AB17	ctttttgtgtttgcagggagcctca	Score of *fie-12* and *fie-13*
AB18	cctgataacaaaccagcgag	dCAPS *fie-14 Dde* I
AB19	ccaaatatcacacattggaaccgggtatc	dCAP *fie-11 Eco*R V
AB20	atctaacgagctctctgattac	dCAP *fie-14 Sca* I
Fie20f	ggtgattttatcctctcaaag	Score *fie-20*
fie21f	ctgtaatcaggcaaacagt	dCAP *fie-21 Eco*R V
fie21r	gtcctcgcagcaagcaagaat	Scoring *fie-21*

Media

All recipes are for liquid media. To make plates for *E. coli*, *Agrobacteria* or yeast 20g/L of agar was added.

LB Agar (1L)

10g NaCl
10g bacto-tryptone
5g bacto-yeast extract
20g agar
pH to 7.0 with NaOH
Autoclave

YEP (1L)

10 g bacto yeast extract
10 g bacto-peptone
5 g NaCl
pH to 7.5 with NaOH

SOC

20g bacto-tryptone

5g bacto yeast extract
0.5g NaCl
10ml 250mM KCl
pH to 7 and make up to 975ml
autoclave
add 5ml 2M $MgCl_2$ and 20 ml 1M glucose

YPDA
20g Difco peptone
10g yeast extract
15ml 0.2 % adenenine hemisulphate
pH to 6.5 make up to 950ml
autoclave
cool to 60°C and add 50ml 40 % glucose

YMM (to make selective plates)
6.7g yeast nitrogen base (without amino acids)
100ml 10x appropriate dropout (eg. –L,-W)
pH to 5.7 make up to 950ml
autoclave
cool to 60°C and add 50ml 40 % glucose

Dropout 10x stock

L-Adenine hemisulfate salt	200 mg/L
L-Arginine HCl	200 mg/L
L-Histidine HCl monohydrate	200 mg/L
L-Isoleucine	300 mg/L
L-Leucine	1000 mg/L
L-Lysine HCl	300 mg/L
L-Methionine	200 mg/L
L-Phenylalanine	500 mg/L
L-Threonine	2000 mg/L
L-Tryptophan	200 mg/L
L-Tyrosine	300 mg/L
L-Uracil	200 mg/L
L-Valine	1500 mg/L

To make –L,-W dropout leucine and tryptophan were omitted.