THE UNIVERSITY
of EDINBURGH

Characterization of a novel trithorax group gene candidate in *Arabidopsis*

Shih Chieh Liang

Degree of Doctor of Philosophy

School of Biological Science

The University of Edinburgh

2013

Table of contents

Declaration

This scientific work described in this thesis was carried out in the School of Biological Science in University of Edinburgh between October 2009 and March 2013. Unless otherwise stated, it is the work of the author and has not been submitted in whole or in support of an application for another degree or qualification of this or any other University or institute of learning.

Signed:

Date:

Acknowledgements

It seemed not long ago that I landed in Edinburgh Airport. I still remember when the plane took off from Taiwan, I was quite excited that the new life was going to begin! Looking back for all those years, I am so pleased that I have come here for studying. I met nice people, experienced different cultures, and indeed have gained a lot of knowledge and have learned the ways of doing research. There are tons and tons of people whom I would like to thank to. First of all is my supervisor, Dr. Justin Goodrich. He offered great project to me. Moreover, he is very kind, accommodating, knowledgeable and helpful. He is no doubt my model of being a scientist. Also, I would like to give my big thanks to Dr. Noreen Murray and Dr. Kenneth Murray who offer overseas students Darwin Trust Scholarship. It is almost impossible for me to come to UK without the support of this scholarship. Sadly they have passed away during my PhD. I am sure their love will last forever, and I will remember them with gratitude. Also thank Prof. David Finnegan and Dr. Myriam Calonje for helping me accomplish the last step of my PhD.

I must thank my parents for their support of my basic study before PhD. Especially my father, Chin Sheng Liang, who gives me space, freedom and his trust. The great thanks go to Erica de Leau, and her family members/Churchill-laan crew. They are like my family in Edinburgh/Amsterdam. Particularly thank Erica---- no matter what happens, in private or at work, she is always there to help or consult.

This thesis cannot be done without the help from many people. Justin's cooperators, Dr. Daniel Schubert, Dr. Franziska Turck, and people in their lab, especially Dr. Mareike Hohenstatt and Dr. Benjamin Hartwig, have provided technique supports and inspiring idea. Their contributions are addressed in the texts. Lots of people from the Big lab, Swann kindly offer their help as well. Prof. Andrew Hudson shares a lot of things with us and also helped me with phylogenetic analysis. Dr. Steven Spoel helped me enormously with immunoprecipitation when I had absolutely no clue how to do it. In addition, he also gave me useful suggestions about future career. Prof. Juri Rasperi and Flavia Alves in his lab are experts in Mass-Spectrometry and are the good hands of my interactome analysis. Dr. Ralf Müller gave technique and knowledge support. Dr. Anne Moore voluntarily offered her help for proof-reading of my thesis although my time was limited so in the end my horrible writing went to Justin straight away.....Andrew Waters, and Dr. Xin (Eric) Tian are good company. We discussed about both science and life. Two short-term students, Marianne Loong and Patrick Dickinson helped with transgenic line and yeast two hybrid screening, respectively. George McMillan, Patrica Watson and Sophie Haupt are always there to make sure our equipments are functional.

Last, I would like to thank my friends, Dr. Jhenyi Wu, Yi-Shuan (Delphine) Yu, Wei-Ting Chen, Yi-Ching (Claire) Ku, Chia-Chen Chang, Chia-yi Lin, Tina Tsai for bringing me the beautiful life in Edinburgh.

This thesis is dedicated to everyone who ever offered help in my life.

Abstract

The Polycomb group (Pc-G) and trithorax group (trx-G) genes play crucial roles in development by regulating expression of homeotic and other genes that control cell fate. Both groups catalyse modifications in chromatin, including histone methylation, leading to epigenetic changes in gene activity. The trx-G antagonises the function of Pc-G genes by activating Pc-G target genes, and consequently trx-G mutants suppress Pc-G mutants. The trx-G genes are relatively poorly characterised in plants. We identified a novel trx-G candidate *SUPRESSOR OF POLYCOMB 12* (*SOP12*) by a genetic screen for suppressors of mutants for the *Arabidopsis* Pc-G gene *CURLY LEAF (CLF)*. Thus *sop12* mutations have no discernible phenotype in wild type backgrounds but partially suppress the leaf curling and early flowering phenotypes of *clf* mutants. Molecular cloning shows that *SOP12* encodes a Harbinger transposase nuclease-like protein which is conserved in green plants, although key residues required for the catalytic activity of the nuclease domain are not conserved. In *sop12 clf* double mutants, many *CLF* target genes are down-regulated relative to *clf* mutant, which suggests SOP12 is a general activator of Pc-G target genes instead of a target of CLF or a late

flowering suppressor. The *CLF* gene encodes an H3K27me3 histone methyltransferase, however chromatin immunoprecipitation (ChIP) analysis indicates that *SOP12* does not antagonise Pc-G by removing H3K27me3 methylation, which is consistent with the fact that *sop12* suppresses mutants for another Pc-G gene, *LIKE HETEROCHROMATIN PROTEIN 1* (*LHP1)*, which is not involved in H3K27me3 deposition. . Rather, genetic analysis shows that *sop12* enhances the phenotype of mutants of *EARLY FLOWERING IN SHORT DAYS* (*EFS*), a trx-G gene involved in deposition of H3K36me3, and of *ULTRAPETALA* 1 (*ULT1*), a plant specific trx-G gene. The enhancement indicates SOP12 may act together with ULT or EFS proteins, or at least regulate the same targets in synergistic ways. For example, SOP12 activates *AP3* expression, a role which overlaps with EFS. Yeast two hybrid screening and imunoprecipitation followed by Mass spectrometry were performed to identify numerous potential SOP12 interacting proteins but await further validation. One protein (SUP1) identified through yeast two hybrid screens was independently identified by another group as a Pc-G suppressor, suggesting that SOP12 and SUP1 may act in a common complex to regulate Pc-G targets.Collectively, my data suggests

that *SOP12* represents a domestic transposase that has acquired a role as a novel, plant specific trx-G members.

Abbreviation

A	Alanine
AB	Abdominal segment
ABRC	Arabidopsis Biological Resource Center
ANOVA	Analysis of Variance
AP	Adaptor primers
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bp	Base pairs
BX-C	Bithorax complex
C. elegans	*Caenorhabditis elegans*
CDD	Conserved Domain Database
Cdna	Complementary DNA
CDS	Coding sequence
co-IP	Co-immunoprecipitation
Col	*Columbia-0* ecotype
D	Aspartic acid
DDO media	Double drop-out media
DEX	Dexamethasone
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide Triphosphate
DTT	Dithiothretitol
E	Elutamic acid
E. Coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EMS	Ethyl methane sulphonate
EtBr	Ethidium bromide
EtOH	Ethanol
GFP	Green fluorescent protein
GSP	Gene specific primers
GST	Glutathione S-transferase
GUS	β-glucuronidase
H	Histidine
HDAC	Histone deacetylase
HLH	Helix-loop-helix

HMTase	Histone methyltransferase
IP	Immunoprecipitation
K	Lysine
L	Leucine
MS	Mass spectrometry
MS (medium)	Murashige and Skoog
NCBI	National center for biotechnology information
NEB	New England Biolabs
NLS	Nuclear localisation signal
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
Pc-G	Polycomb group
PCR	Polymerase chain reaction
PEG	Polyethylene glyco
PI	Propidium iodide
PRE	*Pc* group response element
QDO media	Quadruple drop-out media
RNA	Ribonucleic acid
RNA pol II	RNA polymerase II
rpm	Revolutions per minute
RT-PCR	Reverse-transcriptase polymerase chain reaction
SDS	Sodium dodecyl sulfate
TAIR	The Arabidopsis information resource
T-DNA	Transfer DNA
TDO media	Triple drop-out media
TE	Transposable element
TE (reagent)	Tris-HCl Ethylenediaminetetraacetic acid
TEMED	N,N,N',N'-Tetramethylethylenediamine
TIR	Terminal inverted repeat
Tris	Tris(hydroxymethyl)aminomethane
trx-G	Trithorax group
TSA	Trichostatin A
W	Tryptophen
Ws	*Wassilewskija* ecotype
Y2H	Yeast two hybrid
YEP	Yeast extract peptone
YPDA	Yeast peptone dextrose adenine

Chapter 1.Introduction: Cell fate and epigenetic control of gene expression

The body plans of multicellular organisms are largely specified by gene regulation networks: although all cells in an organism generally have the same DNA content, the different cell types are distinguished by different patterns of gene activity, particularly of key transcription factors such as homeotic genes. The body plans are typically set up early in development in small groups of cells, such as in embryos or organ primordia. As cells undergo further proliferation to give rise to a complete organism, developmental patterning is progressively elaborated and refined but patterns of homeotic gene expression established early in development have to be maintained to ensure proper and correct development (Ingham, 1988; Lawrence and Morata, 1994). There are two mechanisms by which the identity of cells can be maintained through cell division: first, gene activity states may be inherited through mitosis, so that fate is intrinsic to a cell and its progeny; alternatively, cells may adopt fates according to signals from their neighbours, in which case fate is controlled by cell extrinsic factors often termed positional information. Cell fates which are maintained by environmental cues are more flexible. Cells may communicate with the neighbouring cells through signals such as morphogens to determine their cell identities. If a cell is destroyed in early stage of development, other cells near the damaged position may take the part of the destroyed cell and replace it. Or when the cells are transplanted, they tend to differentiate to the same

cells as their new neighbours are rather than the original source. Most organisms use both mechanisms. For example, even in the nematode *Caenorhabditis elegans* (*C. elegans*), whose body normally develops from an invariant pattern of cell lineage, laser ablation studies indicate that the identity of cells may change when their neighbours are destroyed.

Epigenetic changes in gene activity likely provide the main mechanism by which cell fates may be inherited through cell division. In this context the term epigenetic refers to changes in gene activities which are heritable through cell cycles yet are not caused by changes in DNA sequences. There are two well characterised systems mediating epigenetic changes. The first involves modification of DNA by cytosine methylation. DNA methylation is heavily associated with guarding genome integrity and defence against exogenous genetic material, for example in silencing transgenes and transposons in plants (Moazed, 2009; Chen, 2012). The second system is controlled by two antagonistic groups of genes, termed the Polycomb-group (Pc-G) and the trithorax-group (trx-G) and is strongly implicated in controlling developmental processes both in plants and animals. For example, most of these genes have been identified in mutant screens due to their effects on developmental patterning (see section 1.1 and 1.2). The mechanism by which Pc-G and trx-G proteins cause stable changes in gene activity has been obscure, but an important advance has been the finding that many members are involved in modifying the histone proteins that package DNA into chromatin.

In plants, most studies suggest that cell fate is controlled by cell extrinsic factors and relatively easy to alter – for example, cloning of new individuals from adult tissues is relatively easy compared to mammalian systems. Despite this, several Pc-G and trx-G genes are conserved in plants and regulate expression of homeotic genes and other key developmental regulators. (Bemer and Grossniklaus, 2012). This suggests either that the Pc-G and trx-G function differently in plants from animals, or that the changes in chromatin they impart are more easily reversed. To find out more about the plant Pc-G and trx-G genes a genetic screen was made to identify modifiers of Pc-G mutant phenotypes. This led to the isolation of *SOP12*, a novel gene which may correspond to a new plant trx-G member. Here I first review the key features of Pc-G and trx-G genes and how they act, and then explain the use of genetic screens to identify novel members.

1.1 Polycomb group (Pc-G) genes

Pc-G genes were initially identified from genetic screens in *Drosophila melanogaster* (fruit fly). Their developmental defect phenotypes attracted interests and led to massive research. Later, homologues in a wide variety of groups, including vertebrates, nematodes, and plants, were identified and also found to regulate developmental patterning (Schwartz and Pirrotta, 2007; Bemer and Grossniklaus, 2012). The key features of Pc-G genes are 1) they repress transcription. However, they maintain repression rather than establish the inactivation state of their targets, 2) they regulate

multiple homeotic and other developmental target genes, and 3) they confer changes in chromatin, often via histone modification.

Pc-G genes in *Drosophila*

The Pc-G member, *Polycomb* (*Pc*), was the first gene functionally characterised as a repressor of homeotic genes (Lewis, 1978). In *Drosophila*, the identity of segments along the anterior posterior axis is specified by the activity of eight homeotic genes. In severe *pc* mutants, the anterior segments, such as thoracic and initial abdominal segments, are transformed into the most posterior abdominal segment (abdominal 8, AB8) (Lewis, 1978) due to derepression of multiple homeotic genes during embryogenesis. Later, mutants which also show segment transformation toward the posterior identities were categorised into Pc-G. For example, the identification of *Posterior sex combs* (*Psc*) and *Sex comb on midleg* (*Scm*) was based on the similar phenotypes of their mutants with *pc* mutants (Jurgens, 1985).

In Pc-G mutants, Pc-G targets are expressed normally during early embryogenesis in *Drosophila*, but ectopically expressed later on. For example, when embryos are 3-6 hour-old, the expression pattern of *Ultrabithorax* (*Ubx*), a homeotic gene that specifies identity of posterior thoracic and abdominal segments, is indistinguishable in embryos of *extra sex combs* (*esc*) mutants and wild type ones. Later on *Ubx* is mis-expressed in all segments in *esc* rather than restricted to the posterior as in wild type (Struhl and Akam, 1985). This characteristic distinguishes the maintenance role

Pc-G genes from the activity of the transcription factors that establish gene inactivation.

Various Pc-G genes were cloned and molecularly characterised, but initially this was uninformative as their protein products were mostly novel and there were seldom common features between different Pc-G members despite their common genetic properties. However, several lines of evidence strongly suggested Pc-G components work together in complexes regardless of their protein sequence diversity. Immunolocalisation of Pc-G proteins on polytene chromosomes showed that they co-localised at the homeobox gene cluster bithorax complex (BX-C) and at other targets (Orlando and Paro, 1993). Biochemical analyses such as co-immunoprecipitation (co-IP) and yeast two hybrid studies also supported direct interaction between Pc-G proteins. Moreover, when Pc protein was fused to lexA DNA binding domain, the fusion protein bound to reporter genes carrying lexA target sites, and it recruited other Pc-G members to silence the downstream reporter gene (Poux et al., 2001). The most robust evidence is the purification of *in vivo*. protein complexes containing Pc-G components. Shao et al. (1999) expressed epitope-tagged Pc in *Drosophila*, and used affinity chromatography to purify the first Pc-G complex, call Polycomb Repressive Complex 1 (PRC1) *in vivo.* They showed that Pc, Polyhomeotic (Ph) and Psc are included in PRC1. Later Levine et al. identified other missing components, making PRC1 core parts, Pc, Ph, Psc, Sex comb extra (Sce) and Scm (Shao et al., 1999; Levine et al., 2002).

Not long after the purification of PRC1, another Pc-G complex, PRC2 was identified via similar strategies. Ng et al. identified parts of the PRC2 {Esc and Enhancer of zeste [E(z)]} by gel-filtration of HA-tagged Esc (Ng et al., 2000). Succeeding researches indicated PRC2 contains four core members—E(z), Esc, Suppressor of zeste 12 [Su(z)12] and nucleosome-remodelling factor 55 (Nurf 55) (Müller et al., 2002).

i. Biochemical activities of PRC2 complexes

The purification of Pc-G protein complexes led to a major advance, namely the demonstration of a biochemical activity of Pc-G towards chromatin. Thus, the PRC2 was found to have histone methyltransferase activity, specifically methylating the 27th lysine residues on histone H3 (H3K27) in *in vitro* methylation assays, and the catalytic subunit within PRC2 was E(z) (Czermin et al., 2002; Müller et al., 2002). E(z) bears a SET (Suvar3-9, Enhancer of zeste, trithorax) domain, a conserved motif first shown to have histone methyltransferase (HMTase) activity by characterisation of the Su(var)3-9 product (Jenuwein et al., 1998). The methylation at H3K27 (H3K27me) catalysed by PRC2 is important for gene silencing *in vivo*. When embryos carrying the temperature sensitive $e(z)^{61}$ mutation where shifted to restrictive temperatures, Western analysis of total histone extracts using an antibody against H3K27me2 revealed that H3K27me2 was rapidly depleted (Cao et al., 2002), which is correlated with de-repression of *Ubx* gene (LaJeunesse and Shearn, 1996). Although E(z) is the component possessing catalytic functional domain, when one of

the three core components in addition to E(z) is mutated, the complex is still intact but cannot execute proper methyltransferase activity (Ketel et al., 2005). This illustrates that all four core proteins are essential for this activity. Subsequent studies have identified variants of the PRC2 complex with subtle differences in activity. For example, Pcl-PRC2, comprising the core PRC2 with an additional Pc-G protein called Polycomb-like (Pcl) is required for full H3K27me3 levels at target genes. H3K27me3 is reduced, but not abolished, in *pcl* mutants, and homeobox genes in BX-C are mis-expressed in embryos (Nekrasov et al., 2007).

ii. Biochemical activities of PRC1 complex

The isolation of Pc-G members in structurally distinct PRC1 and PRC2 complexes raised the question of how to explain their common role in repression. Surprisingly, chromatin immunoprecipitation experiments revealed that not only PRC2 components are co-localised with H3K27me2 at the *Ubx* target gene, but also Pc, a PRC1 component (Cao et al., 2002). A potential clue came from the fact that Pc contained a motif, the chromodomain, which was shared with the Heterochromatin Protein 1 (HP1) protein. The HP1 chromodomain was shown to bind histone H3 methylated at K9 (H3K9), a modification catalysed by the Su(var)3-9 (Czermin et al., 2001), suggesting that Pc might play a similar role in binding H3K27me3. Indeed, *in vitro*. binding assays indicated that Pc binds to H3K27me3 via its chromodomain (Fischle et al., 2003) and transgenic experiments swapping the chromodomains of

HP1 and Pc confirmed the importance of the Pc chromodomain in localising Pc *in vivo*.

What does PRC1 do after binding to H3K27me3? King et al. found the biochemically purified native PRC1 is able to repress transcription *in vitro*. (King et al., 2002). This may result from the PRC1 activity in chromatin compacting (Francis et al., 2004). In the presence of PRC1, the chromatins fold under the observation of electron microscopy. The key component of these activities is Psc. Psc alone is sufficient to inhibit transcription *in vitro*, but not any one of the other three core components of PRC1 (Pc, dRing, and Ph) is (King et al., 2002). PRC1 lacking Ph, or Psc alone are both still able to compact chromatin yet PRC1 with a truncated Psc lacking amino acids 572 to 872 is not (Francis et al., 2004), although the core components still assemble with the mutated form of Psc (King et al., 2005).

Another activity of PRC1 may be to ubiquitylate H2A. A biochemical approach to look for a H2A E3 ligase in human HeLa cell nuclear extract showed the fraction which display the activity contained Ring1 and Ring2, the two proteins that are homologous to a *Drosophila* PRC1 component Sce (Wang et al., 2004). Furthermore, knockdown of Sce by RNAi indicated that H2A ubiquitylation (H2Au) is required for repressing *Ubx* in tissue culture cells. In addition, H2Au is downstream of H3K27me3, since H2Au is affected by knocking down Sce (*Drosophila* Ring protein, also known as dRing) but H3K27me3 is not. (Wang et al., 2004). The PRC1 activity in human is associated with H2Au (de Napoles et al., 2004), and therefore, it seemed

PRC1 recognise H3K27me3 and then add H2Au at targets. However, Lagarou et al. (2008) identified a dRing-associated factors (dRAF) complex, which is responsible for H2A ubiquitylation and is distinct from PRC1. The dRAF complex shares common components Sce/dRing and Psc with PRC1, but contains extra factors like dKdm2 and excludes Pc (Lagarou et al., 2008). They reconstituted dRAF complex and demonstrated the activity to ubiquitylate H2A is generated by dRAF rather than PRC1. Nevertheless, the *in vitro*. test cannot rule out the possibility that *Drosophila* PRC1 still possess H2A ubiquitylation activity. How H2Au conveys Pc-G repression is not clear yet. It was proposed that H2Au probably regulates gene expression by inhibiting RNA polymerase II (RNA pol II) from transcription elongation in mouse ES cells and human cells (Stock et al., 2007; Zhou et al., 2008). Since H2Au is abundant on inactivated X chromosome (Xi) and PRC1 co-localizes with this mark in mice (de Napoles et al., 2004), H2Au may silence gene through a pathway similar to Xi chromosome inhibition (Simon and Kingston, 2009).

iii. The role of H2Au in Pc-G repression mechanism

The requirement of H2Au for Pc-G repression is rather complicated. H2Au is not absolutely downstream of H3K27me3. Genome-wide expression profile showed a portion of Pc-G targets were de-repressed in mutants of *Psc*, *Pc* and *Ph* but not in mutants of *Sce*, which implies some of the repression is probably independent from H2Au (Gutiérrez et al., 2012). Moreover, a mutagenesis screen identified a novel Pc-G gene which unexpectedly encodes a deubiquitinating enzyme named *Calypso* (de

Ayala Alonso et al., 2007). Mutants of *Calypso* demonstrate the abdominal segment transformation phenotype like other Pc-G mutants. The AB5 to AB7 segments transform to segments resembling AB8 (de Ayala Alonso et al., 2007). When a Calypso-containing complex was purified by tandem affinity purification (TAP)-tagging and identified by MS, it includes a known Pc-G component, Additional sex comb (Asx). Reconstituted Polycomb repressive deubiquitinase (PR-DUB) complex displays a H2Au-specific de-ubiquitylation activity (Scheuermann et al., 2010). Both unbiquitylation and de-unbiquitylation are required for Pc-G repression, which suggests possibly the dynamic balance of H2Au is important rather than the steady-state amount of H2Au.

iv. Recruitment of Pc-G complexes to target genes

Prior to PRC1 and PRC2 exhibiting their enzymatic activities, Pc-G components have to be located to target genes. A complex cis-element called the Polycomb response element (PREs), is present at several Pc-G target promoters (Mihaly et al., 1998), and is bound by Pc-G proteins in ChIP assays (Cao et al., 2002). Deletion of PRE in the promoter region of Pc-G targets leads to gene de-repression (Simon et al., 1993; Chan et al., 1994). However, most of the Pc-G components do not possess DNA binding motifs. Brown et al. used radioactively labelled PRE of the *Drosophila Engrailed* gene as a bait to screen binding proteins and identified Pleiohomeotic (Pho) as the possible protein responsible for Pc-G protein recruitment (Brown et al., 1998). The binding between Pho and PRC1 components, Pc, Psc and Ph, is also supported

by co-IP, which is an *in vivo*. evidence (Mohd-Sarip et al., 2002). By using TAP-tagging, Pho-containing complexes were biochemically purified from nuclear extracts of *Drosophila* embryos. A Pho-containing complex called Pho repressive complex (PhoRC) is comprised of Pho and Sfmbt (Scm-like with four MBT domain-containing protein) which potentially binds to histones (Klymenko et al., 2006). Even though no stable interactions were observed between PhoRC and PRC1 or PRC2 (Klymenko et al., 2006), PhoRC co-localises with PRC1 and PRC2 at target genes (Kwong et al., 2008; Oktaba et al., 2008), indicating the possible function of PhoRC as the anchor of PRC1 and PRC2 to targets. In addition to PRE, it was proposed recently that Pc-G might be recruited by stalled RNA pol II and short RNA transcripts in *Drosophila* (Enderle et al., 2010). Yet this needs further investigation to clarify, for example, whether the stalled RNA pol II and non-coding RNA recruits PRC or the recruitment of PRC results in stalled RNA pol II and short transcripts.

The Pc-G repression context in *Drosophila* is summarised in Figure 1-1, which is reproduced from Bantignies and Cavalli (2011). A cis-regulatory element in Pc-G targets, PRE is bound by PhoRC in a sequence-specific manner. PhoRC may tether PRC1 and PRC2 to target sites. PRC2 generates H3K27me3, with Pcl-PRC2 required for high levels. PRC1 is able to recognise H3K27me3 and possibly to ubiquitylate H2A synergistically with dRAF. The dynamic balance of H2Au is controlled by the ubiquitylation by PRC1 and dRAF, and the de-ubiquitylation by PR-DUB.

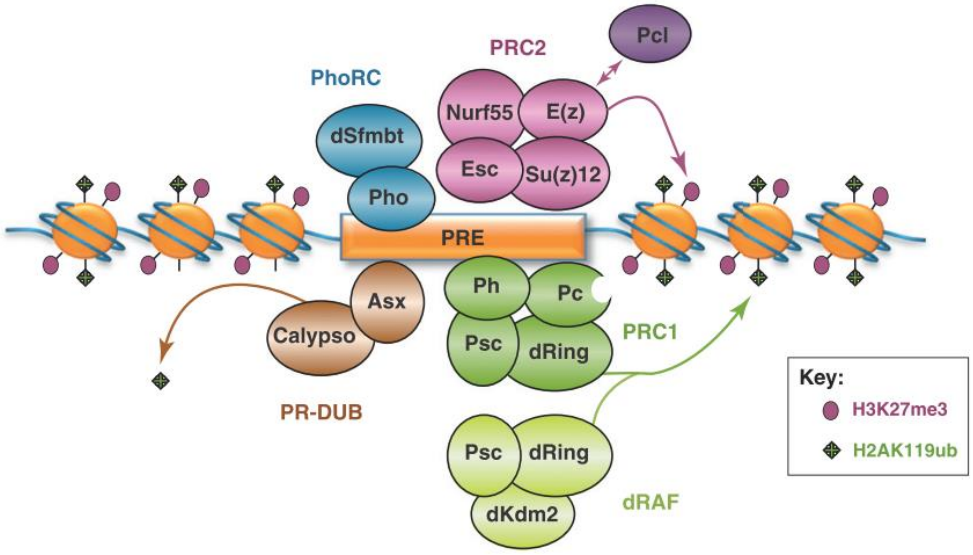


Figure 1-1. The schematic summary of Pc-G repression in *Drosophila*.(Bantignies and Cavalli, 2011)

Five Pc-G complexes in *Drosophila* are gathered at PRE of Pc-G target promoters. PhoRC is responsible for sequence specific binding and recruiting PRC1 and PRC2. PRC2 methylates H3K27, and H3K27me3 is extensively spread with the help of Pcl. PRC1 and dRAF generate the second repressive mark, H2A119ub. PR-DUB removes this repressive mark, but somehow the deubiquitylation also results in repression. The orange circles represent histone octamers, with DNA (blue strings) wrapping on it.

Pc-G genes in *Arabidopsis* development

Although plants do not have a segmental body plan like *Drosophila*, they are built from reiterated units, for instance, the whorls of floral organs whose identity is specified by homeotic genes. Pc-G genes have been identified mostly from forward genetic screens for developmental phenotypes in the plant model organism

Arabidopsis thaliana. Here I summarised the information of Pc-G genes discovered in *Arabidopsis*.

i. PRC2 components in *Arabidopsis*

The four core components of PRC2 are all conserved and are represented by multiple homologues with the exception of *Esc*, which is represented by a single copy gene. The details of the multiple homologues are summarised in Table1-1. Chanvivattana et al. proposed the multiplicate homologues form three different PRC2s with universal Esc and Nurf55-related proteins, MSI1 and FIE, but distinct Su(z) and E(z) homologues (Chanvivattana et al., 2004) (Figure 1-2). The three distinct PRC2s are named by their specific Su(z)12 components: VRN complex, EMF complex, and FIS complex. Succeeding biochemical purification has confirmed the presence of FIS and VRN complexes in *Arabidopsis* (Köhler et al., 2003b; Wood et al., 2006; De Lucia et al., 2008), but the purification of the EMF complex has not been reported yet.

The three plant PRC2s are partially but not fully redundant. They are expressed in different tissues in some cases and participate in different developmental pathways. The FIS complex is pivotal for seed development, whereas the VRN complex regulates flowering in response to cold treatments (vernalisation) and the EMF complex regulates floral organ development and flowering time (reviewed in (Bemer and Grossniklaus, 2012). Only the EMF-PRC2 will be considered here, as it is most relevant for my work.

Table 1-1. The homologues of PRC2 in *Arabidopsis*. This table was adapted from (Hennig and Derkacheva, 2009).

Drosophila PRC2 component	*Arabidopsis* homologue	Protein domain	Refs.
E(z)	CLF	SET	(Goodrich et al., 1997)
E(z)	SWN	SET	(Chanvivattana et al., 2004)
E(z)	MEA	SET	(Grossniklaus et al., 1998)
			(Kiyosue et al., 1999)
Su(z)12	EMF2	Zn finger	(Yoshida et al., 2001)
Su(z)12	VRN2	Zn finger	(Gendall et al., 2001)
Su(z)12	FIS2	Zn finger	(Luo et al., 1999)
Nurf55	MSI1	WD40	(Ach et al., 1997)
Esc	FIE	WD40	(Ohad et al., 1999)
Pcl*	VRN5 (VIL1)	PHD finger	(Köhler and Hennig, 2010)
		FN III	(Zheng and Chen, 2011)
Pcl*	VIN3	PHD finger	(Sung and Amasino, 2004)
		FN III	
Pcl*	VEL1	PHD finger	(Köhler and Hennig, 2010)
		FN III	(Zheng and Chen, 2011)

* Functionally similar instead of sequence homology (De Lucia et al., 2008).

The full name of *Arabidopsis* PRC2 components are described as follows: CLF, CURLY LEAF; SWN, SWINGER; MEA, MEDEA; VRN, VERNALIZATION; EMF, EMBRYONIC FLOWERING; FIE, FERTILIZATION-INDEPENDENT ENDOSPERM; MSI, MULTICOPY SUPPRESSOR OF IRA.

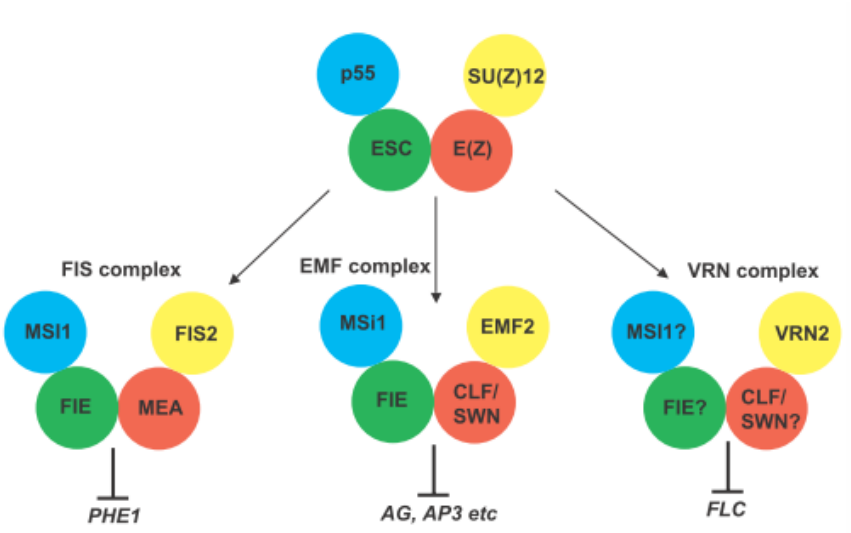


Figure 1-2. A schematic summary of three distinct PRC2 in *Arabidopsis*.

Multicopies of PRC2 homologues were found in *Arabidopsis*. The complex on the top represents PRC2 core components in *Drosophila*. p55 is equal to Nurf55. Three complexes at the bottom are PRC2 in *Arabidopsis*. Homologues are indicated by the same colour. For example, FIS2, EMF2, and VRN2 are homologues of Su(z)12 (yellow). The target genes of each complex are diverged, suggesting three complexes are not redundant.

The EMF-PRC2 complex delays flowering and prevents the shoot meristem from prematurely developing into a floral meristem (FM). Thus *emf2* mutants omit vegetative growth and produce seedlings in which a terminal flower develops immediately after cotyledons with no intervening rosette leaves (Sung et al., 1992; Yoshida et al., 2001). The alteration of the meristem identity results from de-repression of the FM identity genes *LEAFY* (*LFY*) and *APETALA1* (*AP1*), and thus the meristem is unable to produce vegetative tissues like rosette leaves. Weak *emf2* mutants display vegetative growth but are early flowering with narrow curled leaves.

This phenotypes resembles that of mutants in *CLF*, which encodes the catalytic component of the EMF-PRC2, and it is due to mis-expression of the floral organ identity genes *AGAMOUS* (*AG*) and *APELATA3* (*AP3*) (Chanvivattana et al., 2004). In addition to floral developmental genes, genome-wide transcriptomic analysis showed that *EMF2* has many other targets involved in aspects like hormone and stress responses (Kim et al., 2009). By contrast, the effects of *emf2* on seed developmental genes are less drastic. For instance, the expression of the seed storage protein *Arabidopsis 2S Albumin3* (*AT2S3*), a pivotal gene that is expressed during seed development and serves as nitrogen resources for seed germination, remains indistinguishable between *emf2* mutants and wild type plants.

In addition to sequence conservation, the role of the PRC2 in histone methylation is well conserved. Thus chromatin immunoprecipitation (ChIP) experiments reveal that the H3K27me3 is present at about 20% of *Arabidopsis* genes and is lost in severe Pc-G mutants (Schubert et al., 2006; Bouyer et al., 2011). In addition, the EMF-PRC2 complex has been reconstituted *in vitro.* and show to catalyse H3K27 methylation (Schmitges et al., 2011).

ii. The plant PRC1 and PRC1-like components

In contrast to the well-conserved PRC2, fewer *Arabidopsis* proteins share sequence homology with PRC1 components of animal models. In *Drosophila* PRC1, the core components are Pc, Ph, Sce/dRing, Psc and Scm (Shao et al., 1999; Levine et al., 2002), whereas in mammalian PRC1, an additional H2A E3 ligase, human B

lymphoma Mo-MLV insertion region 1 homolog (Bmi1), is also included (Cao et al., 2005). Both the homologues of Ring and Bmi1 proteins are found in *Arabidopsis*: ATRING1a, ATRING1b (Xu and Shen, 2008), ATBMI1a, ATBMI1b and ATBMI1c (Sanchez-Pulido et al., 2008). All five E3 ligases were shown to have the activity to ubiquitylate H2A *in vitro*. (Bratzel et al., 2010; Li et al., 2011). Although uH2A has been detected *in vivo*, it is unclear how it is distributed relative to H3K27me3, and only a sub-set of Pc-G targets are derepressed in the *ring* mutants (Li et al., 2011). The proteins homologous to Pc, Psc, Ph, and Scm have not been found yet. Nevertheless, a few novel, plant-specific proteins are classified as plant PRC1 because they functionally mimic PRC1 components. For example, EMBRYONIC FLOWER1 (EMF1) was proposed to play a role similar to Psc because EMF1 is able to repress transcription *in vitro* (Calonje et al., 2008) and has similar charge properties and predicted disordered structure (Beh et al., 2012). EMF1 very likely participates in the same pathway as EMF2 does since *emf1* demonstrates a nearly identical phenotype to *emf2*. Both mutants skip the vegetative phase and produce terminal flowers immediately after germination (Aubert et al., 2001). In transciptome analysis, a high percentage of target genes overlap, including genes involved in floral organ identity, and stress response, etc. (Kim et al., 2009). Genetic analysis of the *emf1 emf2* double mutant indicated that *emf1* is epistatic to *emf2* (Calonje et al., 2008), which suggests that the two act in a common pathway. Another two plant specific, PRC1-like components have been discovered so far are VERNALIZATION 1 (VRN1) (Levy et al., 2002), and LIKE-HETEROCHROMATIN PROTEIN 1/ TERMINAL FLOWER 2 (LHP1/TFL2) (Gaudin et al., 2001). VRN1 was isolated as a gene required for

responses to vernalisation (Levy et al., 2002). In the *vrn1* mutant, H3K27me3 is retained (Bastow et al., 2004), but its target *FLOWERING LOCUS C* (*FLC*) is mis-expressed after plants return to warm conditions (Sung and Amasino, 2004). This implies VRN1 is downstream of H3K27me3, the repressive mark generated by PRC2, in *FLC* regulation. LHP1 shares homology with HP1 in animals especially the chromodomain (Gaudin et al., 2001). HP1 interact with heterochromatin mark H3K9me3. In contrast, instead of binding H3K9me3 or heterochromatin, LHP1 associates with H3K27me3 and euchromatin (Libault et al., 2005; Nakahigashi et al., 2005; Turck et al., 2007; Zhang et al., 2007), and therefore LHP1 is reckoned functionally similar to Pc in *Drosophila*. In *lhp1*, H3K27me3 remains in the compatible level with wild type plants (Turck et al., 2007), yet the target genes are de-repressed. The PRC1-like proteins in *Arabidopsis* and their functional or sequence homologues are list in Table 1-2.

Table 1-2. The homologues of PRC1 components in *Arabidopsis*.

Arabidopsis protein	Protein domain	Homologue	Refs.
AtRING1a	RING	Sce	(Sanchez-Pulido et al., 2008; Xu and Shen, 2008)
AtRING1B	RING	Sce	(Sanchez-Pulido et al., 2008; Xu and Shen, 2008)
LHP1	Chromodomain	Pc*	(Libault et al., 2005; Nakahigashi et al., 2005; Turck et al., 2007; Zhang et al., 2007)
EMF1		Psc*	(Aubert et al., 2001)
VRN1	B3		(Levy et al., 2002)
AtBMI1a	RING	BMI1	(Sanchez-Pulido et al., 2008)
AtBMI1b		BMI1	(Sanchez-Pulido et al., 2008)
AtBMI1c		BMI1	(Sanchez-Pulido et al., 2008)

*Functionally homologues rather than sequence homologues.

Although in plants the Pc-G genes are not fully conserved, the Pc-G repression mechanism probably still remains similar (Bemer and Grossniklaus, 2012). The proposed mechanism is summarised in Figure 1-3. The core PRC2 establishes the repression by generating H3K27me3 at target loci. H3K27me3 is subsequently recognised by chromodomain of LHP1. LHP1 may participate in a PRC1-like complex consisting of EMF1, one or more ATRING proteins and ATBMI1 proteins. The binding of LHP1 to H3K27me3 possibly tethers the PRC1-like complex to target loci, as Pc does. EMF1 promotes chromatin compacting and inhibiting transcription, whereas ATRING and ATBMI1 proteins generate H2Au to conduct stable repression.

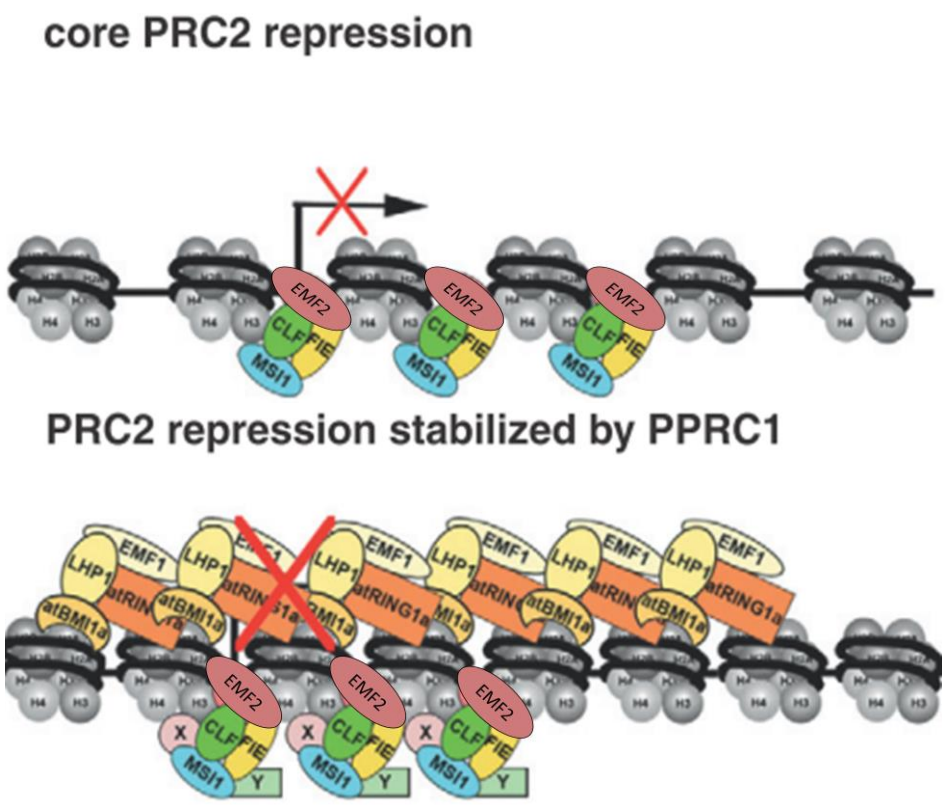


Figure 1-3. Pc-G repression of Pc-G target genes.

This figure is adapted from the review published by Bemer and Grossniklaus (2012). PRC2 establishes primary repression by adding H3K27me3. H3K27me3 is then bound by LHP1 through the chromodomain and then brought in the plant PRC1 (PPRC1) to stabilise the repression by generating H2Au. Chromatin becomes more condensed and the transcription is inhibited because of EMF1. The black string represents DNA, wrapping around histone octamers. The arrow indicates the gene expression and in the presence of PRC2, it is repressed (the red crosses).

1.2 The Trithorax group (trx-G) genes

Trx-G genes also work as developmental regulators but have opposite effects on target genes to Pc-G genes. In general, trx-G genes antagonize Pc-G functions by promoting or maintaining the expression of Pc-G targets, and therefore trx-G mutants are usually suppressors of Pc-G mutants (Kennison and Tamkun, 1992). In addition to the maintenance role of trx-G genes toward homeotic gene expression, trx-G genes may also be directly involved in transcription activation (Armstrong et al., 2002; Srinivasan et al., 2008). Trx-G genes are conserved among yeast, *Drosophila*, human, and partially in *Arabidopsis* (Schuettengruber et al., 2011). The existence of distinct trx-G in plants implies plants may have evolved diverse trx-G components after separating from the animal kingdom in evolution (Carles and Fletcher, 2009). Trx-G genes could be classified into three categories according to their mechanisms: histone-modifying, chromatin-remodelling, and DNA-binding trxG proteins. The last one includes genes that cannot be categorised into any of the three classes (Schuettengruber et al., 2011).

Trx-G components are also epigenetic regulators. They generate histone marks such as H3K4me3, H3K36me3, which are associated with actively transcribed gene (Shilatifard, 2008).

Genetic identification of trx-G genes in *Drosophila*

The landmark gene in this group, *Trithorax* (*Trx*), was identified as a spontaneous mutation in *Drosophila* (Ingham and Whittle, 1980). In *trx* mutants, the body segments are transformed towards an anterior identity. For instance the abdominal segments AB2 to AB7 are all transformed to AB1 (Ingham and Whittle, 1980; Ingham, 1981). The segment transformation phenotype in *trx* mutants results from decrease in expression of the homeobox genes, such as *Ubx*, in the BX-C (Breen and Harte, 1991). By contrast, in a Pc-G mutant *esc*, the abdominal segments AB1 to AB7 are transformed in to AB8, which is due to ectopic expression of BX-C genes (Struhl, 1981). The antagonistic effect of *trx* and *esc* on BX-C expression led to the investigation of the genetic interaction between these two genes. The *trx esc* double mutant partially suppresses the esc phenotype so that although wild type segment identities are not fully recovered, their morphology is closer to AB6 or AB7 instead of AB8 (Ingham, 1983). After discovery of *Trx*, genes are categorized into trx-G if their mutant phenotypes resemble loss of function of homeobox genes, such as anterior transformation of segments as in *trx*, or suppress phenotypes of known Pc-G mutations. Several trx-G genes are characterised by the later criterion, such as *Brahma* (*Brm*), *Moira* (*Mor*), *Osa* and *Kismet* (*Kis*) (Kennison and Tamkun, 1988). In addition to

being suppressors of Pc-G genes, trx-G mutants may enhance each other. Taking *Absent small or homeotic 1* (*Ash1*) and *Ash2* as examples, it was shown that their mutations caused homeotic transformation as was observed in *trx*, and that they suppress *pc* mutants. Therefore they are characterised as trx-G genes. A higher percentage of *ash1 trx* or *ash2 trx* double, or *ash1 ash2 trx* triple mutant flies showed segment transformation than in single mutants of each gene (Shearn, 1989).

Similar to Pc-G genes, trx-G genes are necessary both during embryogenesis and later in development. The analysis of mitotic clones showed that after embryogenesis is completed, removing Trx activity in imaginal discs resulted in altering identity, for example transforming cells in the anterior halteres towards wings (Ingham, 1981). Also, trx-G genes are responsible for maintenance of homeobox gene expression. In *trx* mutants, the expression patterns of BX-C genes are not altered up to 4 hour-old embryos, although expression levels are weaker (Sedkov et al., 1994). Unlike Pc-G, trx-G proteins may directly be involved in transcription activation. For example, a trx-G gene, *Brm*, participated in transcription activation as a co-activator of Zeste (Kal et al., 2000).

By using similar approaches as for identifying Pc-G complexes, many trx-G complexes have been biochemically purified so far, but it is not clear whether those complexes function sequentially or independently and their biochemical functions are more diverse. The following is not comprehensive, but provides examples of the main classes found in animals

i. Trx complexes involved in histone modification

While purifying native complexes in which Trx is involved in *Drosophila*, Trithorax acetylation complex 1 (TAC1) was found (Petruk et al., 2001). TAC1 includes Trx and *Drosophila* CREB-Binding Protein (dCBP). The dCBP subunit can acetylate H3K27 (H3K27ac) and subsequently prevent H3K27 from being mehtylated by PRC2 (Tie et al., 2009). In addition to TAC1, dCBP potentially forms a complex with Ash1. Ash1 has a SET domain and was shown generate two active histone marks H3K4me3 (Beisel et al., 2002) and H3K36me3 (Tanaka et al., 2007). The interaction between Ash1 and dCBP was shown by *in vitro*. pull-down assay (Bantignies et al., 2000). The direct evidence has not been provided yet whether the TAC1 complex or the Ash1-dCBP complex possesses the activity to acetylate H3K27, yet H3K27me3 is increased both in *trx* and *ash1* (Srinivasan et al., 2008), which implies possibly Trx and Ash1 prohibit H3K27 from being methylated by acetylating H3K27.

Trx is also involved in a Complex Proteins Associated with Set1 (COMPASS)-like complex in *Drosophila*. The COMPASS complex was initially discovered in yeast by isolating the Set1-containing complex via chromatography (Miller et al., 2001). Set1 is the yeast homologue of *Trx* (Nislow et al., 1997). Recently, the COMPASS complex and two COMPASS-like complexes were purified in *Drosophila* as well (Mohan et al., 2011). The core components of COMPASS are dSet1 (the *Drosophila* Set1 homologue), Ash2, Dumpy 30 (Dpy30), Retinoblastoma binding protein 5 (Rbbp5), and Will die slowly (Wds) (Takahashi et al., 2011). The COMPASS complex is an

H3K4 methyltransferase. It generates H3K4me3, a histone mark that correlates with gene activation. The catalytic subunit is dSet1 (Ardehali et al., 2011; Hallson et al., 2011; Mohan et al., 2011). The COMPASS-like complexes share the core components except for dSet1 with the COMPASS complex, and contain subunits specific for each complex. In COMPASS-like complexes, the dSet1 is substituted by Trx or Trx-related (Trr). The COMPASS-like complexes also have the H3K4 methyltransferase activity. Their catalytic subunits are Trx and Trr. The COMPASS complex is the major H3K4 methylatransferase, since H3K4me3 decreases drastically in cells knocking down *dSet1*, whereas in cells with reduced *Trx* or *Trr* expression, the degree of decrease is less strong (Mohan et al., 2011).

ii. Trx complexes involved in chromatin remodelling

In addition to histone modifying complexes, some trx-G proteins are involved in histone remodelling complexes which are able to slide or exchange nucleosomes, or alter chromatin conformation, such as looping and unwinding, so-called chromatin remodelling (Ho and Crabtree, 2010). The best studied is the chromatin-remodelling complex containing Brm. The Brm-associated proteins (BAP) complex comprises of Brm, Osa and Mor, Snf5-related protein 1 (SNR1) and Supporter of activation of yellow protein (SAYP) (Dingwall et al., 1995; Papoulas et al., 1998; Crosby et al., 1999). Brm is the homologue of Switch 2 (SWI2) in yeast (Tamkun et al., 1992), which belongs to the switch/sucrose nonfermentable (SWI/SNF) family. Proteins in this family harbour the activity to hydrolyse ATP and induce the conformation

change of chromatins (Ho and Crabtree, 2010). In the yeast system, SWI/SNF proteins function as multimeric complexes and are required for transcription activation (Laurent et al., 1993). BAP contains counterparts of yeast SWI/SNF complexes, for example SNR1 and Mor are similar to yeast SNF5 and SWI3, respectively (Dingwall et al., 1995; Crosby et al., 1999), so it possibly activates homeobox gene expression to antagonise the Pc-G repression via the similar mechanism.

In addition to Brm, Imitation switch (Iswi) is another *Drosophila* homologue of yeast SWI2 (Elfring et al., 1994). Iswi differs from Brm that it lacks of the bromodomain, the domain recognises acetylated histone (Zeng and Zhou, 2002). Iwsi has a PHD domain and a SWI3, ADA2, N-CoR and TFIIIB (SANT) domain for binding to H3K4me3 and unmodified histones, respectively (Wysocka et al., 2006). Iwsi is included in the complex called nucleosome-remodelling factor (NURF). NURF is also able to hydrolyse ATP and activate the expression of *Heat shock protein 70* (*HSP70*), which suggests the function of Iswi may be close to BAP which is participated in activation of transcription (Tsukiyama et al., 1995). It was shown to be required for H1 incorporation and wide range of gene activation (Corona et al., 2007).

The trx-G gene *Kis* belongs to the chromodomain-helicase-DNA binding (CHD) family, which is a sub-family of SWI/SNF distinct from *Brm* and *Iswi*, (Daubresse et al., 1999). This family is characterised by the common feature of tandem chromodomains (Marfella and Imbalzano, 2007). Most of the members are involved

in gene activation. For instance, CHD1 is important for histone H3 variant H3.3 exchange (Konev et al., 2007). Substitution of H3 to H3.3 leads to gene activation (Ahmad and Henikoff, 2002). Moreover, it was suggested that CHD3 is associated with active transcription because it co-localises with RNA pol II when their distribution on polytene is revealed by immunostaining (Murawska et al., 2008).

iii. Trx complexes with DNA binding or other activity

Some trx-G proteins do not have enzymatic activities, but they directly bind to DNA and are required for homeobox gene expression, such as Zeste and GAGA (also known as Trithorax-like, Trl) (Kaufman et al., 1973; Farkas et al., 1994). Zeste and GAGA were both shown to activate *Ubx* expression, and the cis-elements they bind to were also found in *Ubx* regulatory regions (Biggin et al., 1988; Biggin and Tjian, 1988; Laney and Biggin, 1992). Their targets might not be limited to homeobox genes since GAGA is an activator of heat shock genes (Wilkins and Lis, 1997). Zeste and GAGA may be responsible for localising trx-G complexes to target genes. They bind DNA in a sequence specific manner, and meanwhile they bind to trx-G proteins as well. Zeste can be pulled down by BAP, or to be more precise, the leucine-zipper of Mor (Kal et al., 2000). In addition, NURF was purified when Tsukiyama et al. tried to identify the chromatin remodelling complex which cooperates with GAGA (Tsukiyama et al., 1995; Tsukiyama and Wu, 1995). Surprisingly, both Zeste and GAGA are involved in Pc-G repression apart from their roles as trx-G proteins. Instead of being a suppressor of Pc-G mutants, *zeste* displayed both enhancement and

suppression interaction with Pc-G genes E(z) (Phillips and Shearn, 1990; Pelegri and Lehmann, 1994). GAGA was found to interact with Pc, and is required for silencing of a homeobox gene *Frontabdominal-7* (*Fab-7*), which implies that GAGA is essential for Pc-G repression (Hagstrom et al., 1997; Horard et al., 2000). Furthermore, the consensus sequence bound by GAGA factor can displace the binding of PRE with Pc (Horard et al., 2000). The cis-element where trx-G complexes are tethered, named *Trithorax* response element (TRE) are closely located with PRE (Tillib et al., 1999). These findings suggest that *in vivo*., trx-G and Pc-G proteins may compete to bind to their common targets and regulate them in antagonistic ways.

Trx-G genes in *Arabidopsis* and their roles in development

In plants, trx-G genes are less well understood compared to Pc-G genes. In the past decade, more information about trx-G genes in *Arabidopsis* has been disclosed partly as a result of the completion of the *Arabidopsis* genome sequence. Trx-G genes are conserved in *Arabidopsis* since several trx-G genes have been identified based on the sequence similarity to trx-G genes in animal models. Although various sequence homologues have been found, only a portion of them are genetically or functionally characterised as trx-G genes. In addition, no intact trx-G complex has been biochemically purified yet in *Arabidopsis*. Here, I concentrate on those members for which there are reasonable functional genetic data to support a role as trx-G genes.

i. Plant trx-G genes which are involved in histone modification

By searching for proteins which share sequence homology to Trx and Ash1, many potential trx-G genes in *Arabidopsis* were identified. *ARABIDOPSIS TRITHORAX 1* (*ATX1*) to *ATX5* (Alvarez-Venegas and Avramova, 2001; Baumbusch et al., 2001) are homologous to *Trx*, whereas *Ash1 HOMOLOG 1* (*ASHH1*) to *ASHH4* and are homologues of *Ash1* (Baumbusch et al., 2001). In addition, there are proteins that only share similarity to the SET domain, or part of the SET domain of Trx. They are designated as ARABIDOPSIS TRITHORAX-RELATED 1 (ATXR1) to ATXR7. According to the same rules, ASH1-RELATED 1 (ASHR1) to ASHR3 are proteins that harbour SET domains close to the one of Ash1. ATX1, ATX2, ATXR3 and ATXR7 are H3K4 methytransferases (Alvarez-Venegas et al., 2003; Saleh et al., 2008; Yun et al., 2012), whereas ASHH2 methylates both H3K4 and H3K36 (Kim et al., 2005; Xu et al., 2008). However, ATXR5 and ATXR6 display H3K27 methyltransferase activities and are required for transposon silencing (Jacob et al., 2009).

Genetically, the best evidence for a trx-G role is for *ATX1*. *atx1* mutants show weak floral homeotic transformations correlated with reduced expression of floral homeotic genes in flowers (Alvarez-Venegas et al., 2003). Moreover, *atx1* was shown to suppress the Pc-G mutant *clf* in double mutant combinations (Saleh et al., 2007). *ATX2* is closely related in sequence to *ATX1* and shares some target genes although it has subtle differences in functions (Saleh et al., 2008). Both ATX1 and ATX2 have been

expressed and confirmed *in vitro.* that they have H3K4 methyltransferase activity (Alvarez-Venegas et al., 2003; Saleh et al., 2008). It has been shown recently ATX1 and ATX2 add tri- and di-methylation, respectively on histone H3K4 and have their own particular roles in regulation of gene expression (Saleh et al., 2008). ATX1 is responsible for adding H3K4me3 on *FLC* chromatin and up-regulating *FLC* expression. *FLC* is one of the major flowering time genes which delays flower by repressing flowering promoting genes such as *FLOWERING LOCUS T* (*FT*) and *SUPPRESOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) (Searle et al., 2006). Loss of ATX1 reduces *FLC* expression and H3K4me3 on *FLC* chromatin (Pien et al., 2008).

Several other histone modifying enzymes have been identified through their effects on *FLC* expression and flowering time. In particular, mutation in *EARLY FLOWERING IN SHORT DAYS* (*EFS*, also known as *ASHH2* or *SET-DOMAIN GROUP PROTEIN 8*, *SDG8*), is early flowering under short day conditions (Soppe et al., 1999). The early flowering phenotype is associated with reduced activity of *FLC* (Kim et al., 2005). *EFS* was reported as the main methyltransferase for H3K36me2 and H3K36me3. In *efs* mutants, H3K36me3 reduces globally, which was shown by Western blotting (Xu et al., 2008). H3K36me3 is necessary for activation of *FLC* (Zhao et al., 2005) and *APETALA3* (*AP3*), a floral homeotic gene. In *efs*, *AP3* is down-regulated resulting in a partially sterile phenotype, and this is correlated with reduction of H3K36me3 at *AP3* chromatin (Grini et al., 2009).

ii. Plant trx-G components with chromatin remodelling functions

In addition to homologues of chromatin modifying group, homologues of chromatin remodelling were also identified. *BRAHMA* (*BRM*) is the homologue of *Drosophila Brm* (Hurtado et al., 2006); *SPLAYED* (*SYD*) shares homology to yeast *Snf2* (Wagner and Meyerowitz, 2002b). The *ATSWI3* family in *Arabidopsis*, *ATSWI3a*, *ATSWI3b*, and *ATSWI3c* are homologous to yeast *Swi3* (Sarnowski et al., 2002). *SYD* was identified in a forward genetic screen and has trx-G properties. The *syd* mutant was found as an enhancer of weak *lfy* mutants due to a role of *SYD* in activation of floral homeotic genes such as *AP3* (Wagner and Meyerowitz, 2002a). *BRM* and *SYD* have redundant roles. Therefore the single mutants of *brm* or *syd* shows minor phenotypes, but *brm syd* double mutants are embryo lethal, and knockdown of *BRM* in *syd* backgrounds strongly enhances the floral phenotypes (Wu et al., 2012). *SYD/BRM* and *CLF* have antagonistic effects on floral homeotic gene activity, such that *clf* mutants partly suppress the defects of *syd* and *BRM* knockdown in flowers, whilst *syd* mutants suppress the effects of *clf* on leaf curling. *SYD* and *BRM* thus have both structural and functional characteristics of trx-G.

PHOTOPERIOD INDEPENDENT EARLY FLOWERING 1 (*PIE1*) is similar to yeast *Swr1* in the *Iswi* family (Noh and Amasino, 2003). *PICKLE* (*PKL*) and its redundant gene *PKL-RELATED 1* (*PKR1*) are CHD3 homologues, whereas another *PKL* redundant gene *PKL-RELATED 2* (*PKR2*) is closer to *CHD1* (Eshed et al., 1999; Ogas et al., 1999). PKL and PIE1 have been implicated as trx-G genes by their

genetic interactions with Pc-G genes. *pkl* was also shown to suppress the clf leave curling phenotype (Aichinger et al., 2009), as does *pie1* (Noh and Amasino, 2003).

iii. ULT---- a novel trx-G component

A few years ago, a plant-specific trx-G gene, *ULTRAPETALA1* (*ULT1*), was identified (Carles and Fletcher, 2009). *ult1* mutants were identified from an extra floral organ phenotype which likely reflects increased floral stem cell number due to impaired activity of the floral homeotic gene *AG (Carles et al., 2005)*, which is required to suppress stem cell activity in the centre of the flower (Liu et al., 2011). Intriguingly, when *ULT1* is over-expressed by *CaMV 35S* promoter, the transgenic plants exhibit an up-curled leaf phenotype, which resembles the phenotype in plants harbouring *35S::AG* transgene, or in the Pc-G mutant *clf*. Subsequently, the genetic test shows *ult1* suppresses *clf*. At the molecular level, over-expressed *ULT1* is able to activate *AG*, and removing *ULT1* from *clf* background down-regulates mis-expressed *AG*. ChIP assays showed that in the *AG* locus in *ult1* mutants, H3K27me3 is increased and H3K4me3 is decreased in seedlings (Carles and Fletcher, 2009). ULT1 possesses DNA-binding SAND (SP100, AIRE-1, NUCP41/75, DEAF-1) domain (Carles et al., 2005), which is not found in trx-G proteins in animal. The identification of *ULT1* as a plant-unique trx-G gene implies that, the trx-G genes in *Arabidopsis* may be diverged than in animal models. Thus, it is not sufficient to identify trx-G in *Arabidopsis* depending on sequence homology. The mechanism for *ULT1* action is unclear. ULT1 may interact with ATX1 protein as the two proteins interact in a split YFP assay when

expressed in onion epidermis but robust biochemical evidence is lacking (Carles and Fletcher, 2009).

1.3 The antagonism of trx-G and Polycomb group (Pc-G) genes

Pc-G and trx-G members share part of their target genes but generate antagonistic effect to each other, which brings about the genetic interaction that trx-G mutants suppress Pc-G mutants, or vice versa. The scenario behind this phenomenon is that, in Pc-G mutants, the Pc-G targets are mis-expressed resulting from losing Pc-G repression. In the double mutants of a Pc-G gene and the trx-G gene which up-regulates the same targets as the Pc-G gene, the targets fail to mis-express and the phenotype recovers towards wild type. This genetic interaction provides a good strategy for forward genetics to discover more novel players in trx-G or Pc-G in plants.

Pc-G and trx-G complexes add histone modifications that result in antagonistic effects on gene expression. Interestingly, the complexes contain subunits that actively remove the opposing histone marks. For instance, the Pc-G complexes in human and *Drosophila* carry histone deacetylases (HDACs) to remove histone acetylation, a mark associated with active expressed genes, at targets (van der Vlag and Otte, 1999; Tie et al., 2001). One of the components in the *Drosophila* H2A E3 ligase complex dRAF, dKdm2 is a demethylase for removing H3K36me2 (Lagarou et al., 2008). In trx-G TAC1 and ASH1 complexes, dCBP deposits H3K27ac which blocks methylation at H3K27 (Tie et al., 2009). This process is reversible. By removing H3K27ac, PRC2 is

recruited to generate H3K27me3 and represses gene expression (Reynolds et al., 2012). Furthermore, via an unclear mechanism, H3K36me3 inhibits methylation at H3K27me3, resulting in discrimination of one from the other (Yuan et al., 2011). The clearance or inhibition of the opposite marks may assist the interactions between histones and histone modifying complexes. Recently it was shown that H3K4me3 inhibits the binding of PRC2 to histone H3, and subsequently H3K27me3 cannot be added (Schmitges et al., 2011).

Because Pc-G and trx-G affect gene expression oppositely, one would expect genes that are marked with active codes by trx-G to not have repressive marks generated by Pc-G. However, in mouse embryonic stem cells (ES cells), chromatin with both active and repressive marks were found (Vastenhouw and Schier, 2012). Possibly because the cell fates of ES cells are not determined yet, the existence of both marks facilitates quick switch between ON and OFF of status gene expression. Bivalent marks were also found at *AG* locus in *Arabidopsis* (Saleh et al., 2007). Yet it is ambiguous if bivalent marks do exist in *Arabidopsis*, since the experiment was done in a mixture of types of cells.

1.4 Suppressor screening of *clf-50* and identification of *SUPPRESSOR OF POLYCOMB 12* (*SOP12*)

Compared to Pc-G, less trx-G members have been discovered in *Arabidopsis*. In order to identify novel plant trx-G genes which have not been revealed by looking for

orthologues of trx-G genes from other species, a suppressor screening of Pc-G mutant *clf* has been established in Dr. J. Goodrich's lab. CLF is one of the three homologues of E(z) in *Arabidopsis* (Goodrich et al., 1997). It plays the role of methyltrasferase in both EMF and VRN complexes (De Lucia et al., 2008). The SET domain is essential for CLF function since a mis-sense mutation in the SET domain resembles a *clf* null mutant (Schubert et al., 2006). Mutations in *CLF* causes reduction in H3K27me3 at target genes, resulting in mis-expression of many Pc-G target genes, including *FLC*, *FT*, *AG* and *SEP3*, etc. (Goodrich et al., 1997; Chanvivattana et al., 2004; Jiang et al., 2008). These changes contribute to the pleiotropic phenotypes in *clf*, such as curled leaves and early flowering.

Based on researches in *Drosophila*, genetic screening for modifiers is an efficient strategy to discover more Pc-G or trx-G genes (Jurgens, 1985). A screening which utilizes a large scale mutagenesis was performed earlier in Goodrich Laboratory, involving T-DNA mutagenesis of a conditional *clf* mutant in *Wassilewskija* **(***Ws*) background. *clf* mutants are partial sterile due to the defect in flower structures, which affects the efficiency of transformation used to generate random T-DNA insertions. To improve fertility, a conditional CLF transgene was constructed by fusing *CLF* genomic sequence in frame with the ligand binding domain of the rat glucocorticoid receptor (GR) and the resulting *CLF-GR* transgene was introduced into *clf-50* background. In the presence of the glucocorticoid steroid analogue dexamethasone (DEX), the CLF::GR fusion translocates into nucleus to fulfil CLF activity, and therefore the clf phenotypes could be rescued. Whereas in the absence of DEX,

CLF::GR stays in cytosol, so the transgenic plants *clf-50 CLF-GR* show clf phenotypes (Schubert et al., 2006; Lopez-Vernaza et al., 2012). The mutagenesis was performed in transgenic plants *clf-50 CLF-GR* with DEX treatment to produce large number of T1 seeds. In the subsequent screen, the plants showing suppressed phenotypes when growing without DEX were selected as candidates for further characterisation.

In the *clf* mutant background, Pc-G activity is reduced but not eliminated, so the plants are more sensitive to subtle changes resulting from mutations. The partial loss of Pc-G activity is due to redundancy between *CLF* and *SWN*. For example, SWN is one of the core components in VRN2 and EMF complexes. Mutation of *swn* in wild type background does not give obvious phenotypes, but drastically enhances *clf* phenotype: *swn clf* completely loses cell identity and develops into callus on tissue culture (Chanvivattana et al., 2004). In other words, mutating the "modifier genes" in wild type background might not produce obvious phenotype because of the buffering Pc-G activity. When *CLF* is lost, Pc-G activity reduces dramatically, resulting in loss of buffering. These modifier mutants might demonstrate big differences with *clf* single mutant. The up-curled leaf phenotype *clf* (Figure1-4), caused by mis-expression of *SEP3* and *AG* (Goodrich et al., 1997; Lopez-Vernaza et al., 2012) is utilized as an index of the suppression. Moreover, expansion of plant size and delaying the flowering time are also indicators of suppression. Those suppressors are designated as *SUPRESSOR OF POLYCOMB* (*SOPs*). A few *SOP*s have been characterised as late flowering mutants or Pc-G target mutants, like *FT* and *SEP3* (Lopez-Vernaza et al.,

2012). Among other *SOP*s uncovered, *sop12-1* was revealed in the M2 population as a recessive suppressor with less curled leaves compared to *clf-50* (Figure 1-4). Cloning of a T-DNA insertion from *sop12-1* plants suggested that *SOP12* might correspond to a novel gene *AT3G63270*.

Aims

Unlike many of the other sop mutants, *sop12* had fairly normal flowering time and few pleiotropic effects, suggests that it might interact rather specifically with Pc-G. It therefore represented a good candidate for a plant trx-G gene. The aims of my research can be separated into three parts. First of all, to confirm *AT3G63270* corresponds to the *SOP12* locus. The second is to further characterise the role of SOP12 by genetic and molecular analysis to test whether it acts like a novel plant trx-G member. The third aim is to investigate the mechanism underneath the suppression phenotypes.

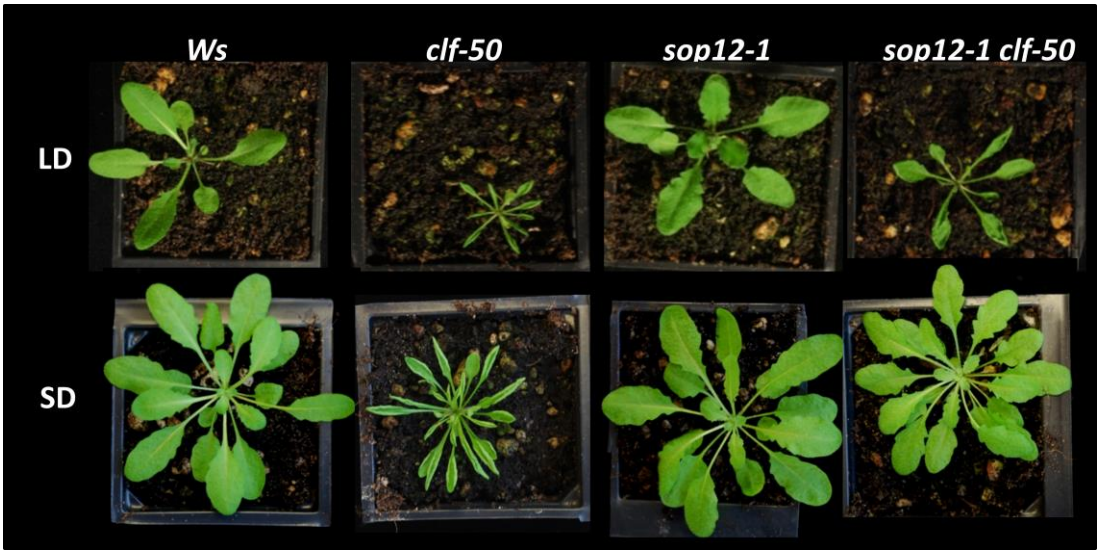


Figure 1-4. The *SOP* mutant *sop12-1* allele in Ws backgound also suppresses the clf phenotype

sop12-1 was selected from the mutagenesis screening of *clf* suppressors. In *sop12-1 clf-50*, leaves are less curled and the morphology is more similar to wild type (*Ws*). Photos were taken when plants were grown under long day (LD) for 26 days or short day (SD) for 51 days in 5 cm x 5 cm square pots. The inflorescence stems were removed from all the plants.

Chapter 2. Material and methods

2.1 Plant material

Growth conditions

Plants grown on soil were grown in growth chambers under fluorescent light. The day/night temperature was 22^{o}C/20^{o}C. Long day conditions (LD) were 16 hours light and 8 hours dark cycles, whereas short day (SD) conditions were 8 hours of light and 16 hours of dark.

For the plants sown on Murashige-Skoog (MS) agar plate (0.64 g MS salts (Duchefa) in 1 L of water containing 0.9% w/v agar), the seeds were sterilized before plating to avoid fungi contamination. Seed were soaked in more than 10 times volume of sterilization solution (70% v/v ethanol and 0.1% v/v Triton X-100) and agitated at the speed of 20 rpm for 10 minutes (min), and then the tubes were left on a rack for a while until seeds settled at the bottom of the tubes. After the supernatant was discarded, seeds were washed 3 times with 95% ethanol and left dry on filter paper. Addiotional washes in 95% ethanol were included if needed in order to remove dried plant debris.

Plant materials

Most of the ChIP and RNA analysis experiments were done using 2-week-old seedlings unless clarified. In tissue specificity experiments, juvenile leaves are round

in shape and possess fewer trichomes (the first to fourth true leaves), and adult leaves are oval-shaped leaves with dense trichomes.

Mutant lines

The origin and genetic backgrounds for the genetic materials used are summarised in Table 2-1 below.

Table 2-1. The mutant lines used.

Name	Locus	Mutation type	Background	Source or reference
sop12-1	*AT3G63270*	T-DNA insertion	Ws	Dr. J. Goodrich
sop12-3	*AT3G63270*	Ds insertion	Ler	Enhancer trap line (CSHL_ET1398)
clf-2	*AT2G23380*	Ds insertion	Ler	(Long et al., 1993)
clf-50	*AT2G23380*	Deletion	Ws	Dr. J. Goodrich
atx1-1	*AT2G31650*	T-DNA insertion	Ws	Dr. Z. Avramova (Alvarez-Venegas et al., 2003)
atx2-2	*AT1G05830*	T-DNA insertion	Col	ABRC (SALK_117262)
ult1-1	*AT4G28190*	Mis-sense mutation	Ler	Dr. Cristel C. Carles (Fletcher, 2001)
ult2-2	*AT2G20825*	Nonsense mutation	Ler	Dr. Cristel C. Carles
efs	*AT1G77300*		Ler	(Soppe, et al., 1999)
lhp1-2	*AT5G17690*	Nonsense mutation	Ws	(Gaudin, et al., 2001)
as1	*AT2G37630*	Mis-sense mutation	Ler	Dr. B. Hwang (Sun et al., 2002)
sop12l1-2	*AT3G55350*	T-DNA insertion	Col	ABRC (SALK_122829)
emf1-1	*AT5G11530*	Nonsense mutation	Col	Dr. R. Sung (Sung et al., 1992)

emf2-3	*AT5G51230*	35-bps deletion	Col	Dr. R. Sung (Yang et al., 1995)
emf2-10	*AT5G51230*	Deletion	Ws	(Chanvivattana et al., 2004)

2.2 RNA analysis

Total RNA extraction

Around 0.1 g of plant material was collected and quickly frozen and ground thoroughly in liquid nitrogen. 1 ml Trizol (Invitrogen) was added to the tissue powder and mixed by vortexing. The mixture was centrifuged for 10 minutes (min) at 16,000 g at 4^{o}C. The supernatant was transferred to new tube and 200 µl of chloroform was added. After mixing by vortexing, the milky solution was centrifuged for 10 min at 16,000 g and the clear upper liquid was transferred into a new tube. 0.5 volumes of a mixure of 1.2M sodium chloride and 0.8M sodium citrate (pH 5.5), and 0.7 volumes of iso-propanol were added and mixed well. Then the solution was stored at -20^{o}C for 20 min to help RNA precipitation. A white pellet of RNA was formed by centrifugation for 10 min at 16,000 g at 4^{o}C. The pellet was washed with 70% ethanol in DEPC treated H_2O and dried in air. When the pellet was dried, it was dissolved in 50 µl DEPC treated H_2O and stored at -80^{o}C

Preparation of cDNA by reverse transcription (RT)

1 µg of RNA was mixed with 1 µl of 100 µM anchored oligo dT primer (5'-TTTTTTTTTTTTTTTTTTTTV-3') and DEPC-treated H_2O to make the volume

up to 10 µl. The mixture was heated to 65^{o}C for 10 min and immediately cooled on ice. The reverse transcription reaction was performed in a 20 µl volume containing 10 U RNase inhibitor (Promega), 100 µM DTT, 10 mM dNTP, 1× MMLV buffer (Promega) and 100 U of MMLV Reverse Transcriptase (Promega). The heating cycle was, initially at 25^{o}C for 30 min, followed by 42^{o}C for 90 min, and finally at 70^{o}C for 10 min. The final cDNA product was diluted to 200 µl in DEPC-treated H_2O and stored at -20^{o}C.

Realtime PCR (qPCR) for quantification of gene expression and chromatin immunoprecipitation (ChIP) experiments

LightCycler480 SYBR® Green Master I qPCR Kit（Roche） was used as a 2X SYBR master premix. SYBR® green is a dye which binds into double stranded DNA and gives fluorescence at 530 nm. A 15 µl reaction mixture consisted of 7.5 µl of the 2X SYBR master premix, 0.5 µl of 10 mM forward and reverse primer mix, and 7 µl template DNA. The details of the amplification program was: an initial 3 minutes at 95℃ followed by 50 cycles of denaturation at 95℃ for 15 seconds (sec), annealing at 55℃ for 15 sec, and finally extension/detection at 72℃ for 15 sec. Following the amplification procedure, the PCR products were heated up to 95℃ to denature the products and then gradually cooled down to 37℃ to allow reannealing. The fluorescence was recorded through the whole cooling step. This information was used for melting curve observation to make sure that only one product was amplified and that there was no primer dimer formed (visible as a second fluorescence peak caused

by the smaller product reannealing at a lower temperature than the longer, gene specific product). The data analysis was performed using the LightCycler480 (Roche) software. First, the melting curves of the PCR products were calculated. The single, gene specific PCR products should give a single peak in a plot of [–d(intensity of fluorescence)/d(temperature)] v.s. temperature. Second, the crossing point (Cp) values of each well were calculated using the 2nd Derivative Maximum Method in the Roche Software suite. For each primer set, the amplification efficiency was calculated by analysis of a calibration curve made using a 1 – 10^{-4} dilution series of cDNA or genomic DNA templates, again using the supplied software. The Cp values were converted to relative expression levels manually in Excel by calculating the value (Primer efficiency$^{)(\text{-Cp value})}$.

RNA sequencing

RNA was isolated as described in subsection "total RNA extraction". The RNA samples were processed by the next generation sequencing facility at the Max Plunck Institute, Cologne, Germany as part of the collaboration with Dr F. Turck, including the purification of poly adenine (polyA) tail-containing mRNA, reversed transcription, reads mapping and coverage, library generation, and statistics analysis.

2.3 DNA analysis and cloning

Crude genomic DNA extraction (Edward's prep)

For genotyping plants using PCR, a crude DNA extraction was made using the method of Edwards (Edwards et al., 1991; Kasajima et al., 2004). A piece of tissues was collected from single plant and ground in 400 µl Edward's prep buffer (200 mM Tris, pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5 % w/v SDS) using a homogenizer (ball miller) with 30 Hz frequency for 30 sec. The plant extract was centrifuged at 16,000 g for 10 min. 300 µl of the supernatant was transferred to a new tube and equal amount of isopropanol was added. The solution was mixed thoroughly and left at -20^{o}C for 30 min to help DNA precipitation. The DNA precipitant was pelleted by centrifugation in micro-centrifuge at 16,000 g for 10 min. The supernatant was discarded and the pellet was air-dried. Once all the supernatant had completely evaporated, 400 µl of T.1E buffer (10 mM Tris, pH 8.0, 0.1 mM EDTA) was used to dissolve the DNA. 5 µl of the DNA solution was used in a 20 µl PCR reaction.

High quality genomic DNA extraction

A higher quality extraction method was used to prepare genomic DNA for use in molecular cloning by genome walker PCR. About 0.1 g of plant tissue was harvested and ground in liquid nitrogen. 500 µl of extraction buffer (0.1M Tris, pH 8.0), 50 mM EDTA, 0.1 M NaCl, 1% w/v SDS) was applied to the ground powder and the mixture was vortexed. The extract was heated to 65^{o}C for 30 min and cooled down before 500

µl of phenol/chloroform was added. The two phases were separated by centrifugation for 10 min at 16,000 g and the upper aqueous phase was transferred into a new tube. 0.1 volumes of 3M sodium acetate (pH 5.5) and 0.7 volumes of iso-propanol were added and solution was inverted to mix well. Then the solution was stored at -20^{o}C for 20 min to precipitate genomic DNA. After precipitation, the genomic DNA was pelleted by centrifugation for 10 min at 16,000 g. The pellet was washed in 70% ethanol and dried in air. Once dried, the pellet was dissolved in 50 µl T.1E (10 mM Tris, pH 8.0, 0.1 mM EDTA) buffer.

Polymerase chain reaction (PCR)

20 µl reaction reagent of PCR was prepared as a mixture of 1X reaction buffer (50 mM Tris HCl, pH 8.3, 0.5% Ficoll, 500 µg/ml BSA, 1% sucrose, 30 mM KCl, 3 mM $MgCl_2$, and 1 mM tatrazine), 200 µM dNTP, 200 µM of DNA forward and reverse primers, 1 to 5 µl of the template, and 1 U of Taq DNA polymerase. The amplification conditions were an initial 10 min at 95℃ followed by 22~35 cycles of denaturation at 95℃ for 30 sec, annealing at 55 ℃ for 30 sec, and extension at 72℃ for 30 sec. An additional final extension at 72℃ for 10 min was included at the end of the reaction. The sequences of primers used are given in Table 2. The size of PCR products were analyzed by DNA electrophoresis (see the section of "DNA electrophoresis").

Table 2-2. Primers used in this thesis

Category	Primer name	Sequence	Locus
RT- and qRT-PCR	eLFT22	TTCGCTCTTCTCTTTGCTCTC	internal control (eIF4A)
	eLFT221	GAACTCATCTTGTCCCTCAAGTA	
	RM_FLC_1	CCTGGTTCTCTTCTTTCAGCATTTTCT	FLC
	RM_FLC_2	CGATGCTCTTGTTCAACTGGAGGA	
	SEP3 Q PCR F	TATGACGCCTTACAGAGAACC	SEP3
	SEP3 Q PCR R	ATACCCATCAGCTAACCTTAGTC	
	FT Q RT F	CCTCAGGAACTTCTATACTTTGGTTATGG	FT
	FT Q RT R	CTGTTTGCCTGCCAAGCTGTC	
	AGQPCR1F	TCCGAGTATAAGTCTAATGCC	AG
	AGQPCR1R	GCCTATATTACACTAACTGGAGAG	
	SOP12 real F	TGTTTCTCAGGTTACATGGAG	SOP12
	SOP12 real R	GTCATCTGATGCTTGCACAG	
	3260 real F	ATTGATGTGAATGTCGGAACTG	AT3G63260
	3260 real R	GATGAATAGGCTGTGGATTTGAG	
	3280 real F	CATTCTTACACATCTCCAGTTCC	AT3G63280
	3280 real R	GCAACCAATAAAGCGATCCA	
SOP12 cloining	Primer name	Sequence	other information
	SOP12 gDNA net	CACCGAAGAAGGAAAAAGAAGCAGAC	
	SOP12 gDNA R1	GTGGTGGTGGTCTCCAC	
	SOP12 cDNA F	CACCATGGCTCCGGTGAAGC	
	SOP12 cDNA R	TCACCTAAGCAAATGCTCAG	
	SOP12 cDNA R-S	CCTAAGCAAATGCTCAGCTA	used for C-terminal fusion constructs, like SOP12-GFP and SOP12-GUS
ChIP	Primer name	Sequence	Locus
	5S1	GGATGCGATCATACCAG	5S ribosomal DNA (control of ChIP against H3K27me3)
	5S2	CGAAAAGGTATCACATGCC	
	actin F	CGTTTCGCTTTCCTTAGTGTTAGCT	Actin (control of ChIP experiment against H3K4me3 and H3K36me3)
	actin R	AGCGAACGGATCTAGAGACTCACCTTG	
	SEP3 E3 F	TGACGCCTTACAGAGAACCCA	SEP3
	SEP3 E4 R	ACCCTGAGAGCTCTGATCTGC	

	5'ChIP AG UTR	CCCAAAGATTTTAGTGCCTCA	AG
	3'ChIP AG UTR	GGTTCAAGTTGGGCAATCAC	

Genotyping primers	Forward primer	Reverse primer	T-DNA/Ds insertion primer
atx1-1	CTTGTGCTCTTTCTTTCTC TTGG	AGATGGATTAGGTGGTGG GA	GATGCACTCGAAATC AGCCAATTTTAGAC
atx2-2	ATTGTTGGCGGATACTCTC G	CGTCTATGACCCGCTCATT ATC	GCTGTTGCCCGTCTC ACTGGTG
emf1-1	AGAGAGACCTTAGAAAAT GTTCA	CTAACATTAGAAGCACCCA TA	-
EMF1-1	AGAGAGACCTTAGAAAAT GTTCG		-
emf2-10	GCC AGG CAT TCC TCT TGT TA	TTG TAA GCA ACC CCA CAA CA	-
emf2-3	CGGACCGGGATAGTGAAG ATGAAG	AGTCAGAGAACATGGATG CGTATG	-
sop12-1	CAACCTCTACTCAGCTCCC T	ACATAACTTTGCTCAGAAT CCTCC	ATATTGACCATCATA CTCATTGC
sop12L-2	CTGTGTAAATGCGCTTGTA GG	GTCTTTAAGCTTCGATAGT GCC	GCTGTTGCCCGTCTC ACTGGTG
swn-3	CTG GAT GGG GTG CTT TTC TA	GCG TAG CAA TTG GGT TTA GC	GCTGTTGCCCGTCTC ACTGGTG
ult1-1	GGGTGATCCAAACTGGAA AGA	AAGTACACTCAGAGAAAC GG	-
ULT1-1	GGGTGATCCAAACTGGAA AGG		-
sop12-3	CAACCTCTACTCAGCTCCC T	ACATAACTTTGCTCAGAAT CCTCC	TTCCCATCCTACTTT CATCCCTG

Genome walk PCR

This method is modified from the method developed by Siebert et al. (Siebert et al., 1995). The purpose of genome walker PCR is to amplify the unknown DNA sequence flanking a known sequence such as a T-DNA or transposon insert. After PCR, the amplified fragments were sequenced to identify the flanking sequence. The genomic DNA was digested by restriction enzymes (using *Eco*RI and *Hin*dIII together, or *Eco*RV alone) and ligated with corresponding adaptors

(5’-CTAATACGACTCACTATAGGGCTCGAGCGGCCCGGGCAGGT-3’, plus restriction enzyme site at 3’ end). Ligation was done at 16^{o}C overnight in a 20 µl reaction containing 5 µM adaptor, 50 mM Tris-HCl,, pH 7.6, 10 mM $MgCl_2$, 0.5 mM rATP, 10 mM dithiothretitol [DTT]) The target fragments were amplified by two rounds of nested PCR using gene specific primers (GSPs) and adaptor primers (APs), where gene specific primers were based on known sequences such as the Left border of the pJG52 activation tagging T-DNA construct or the SOP12 gene. In the first round of PCR, 1 µl of ligation product was used as the template in a 25 µl reaction volume. *AP1* (5’-GATCCTAATACGACTCACTATAGGGC-3’) and *GSP1* were the primers and the reaction program is described as following: 96℃ initial denaturation for 3 min, 20 cycles of 96℃ for 15 sec, and 68℃ for 2 min 20 sec. Next, 25 cycles of 96℃ for 20 sec, 58℃ for 20 sec, and 72℃ for 2 min 20 sec. at the end of the program, the 72℃ extension was prolonged to 7 min. The PCR product from the first round was diluted 100 fold and 1 µl of diluted product was used as the template for the second round of PCR. The primers used in the second round were *AP2* (5'-AATAGGGCTCGAGCGGC-3’) and *GSP2* (the nested primer). The amplification program was: 96℃ initial denaturation for 3 min, 5 cycles of 94℃ for 20 sec, and 68℃ for 2 min 20 sec. Next, 23 cycles of 94℃ for 20 sec, 58℃ for 20 sec, and 72℃ for 2 min 20 sec. Finally, an extension step at 72℃ for 7 min was included. The products of both rounds were loaded on to 1% w/v agarose gel for electrophoresis (see the section of “DNA electrophoresis”). In cases where nested PCR bands were observed, i.e. a product from the first round of PCR was reamplified to produce a

slightly smaller product, the product of the second round PCR was then sequenced using *GSP2* primer.

DNA electrophoresis

DNA electrophoresis was performed to check the size of DNA fragments and whether PCR reactions were successful. When a voltage is applied, negatively-charged DNA fragments migrate towards the anode in agarose gels with different speed according to their molecular weight/size of the fragments. 0.7%~1.5% agarose gels containing 0.1 μg/ml ethidium bromide (EtBr) was used according to the size of DNA fragments. The higher percentage of agarose gel was applied for smaller fragments less than 0.5 kb in size. DNA solution was mixed 5:1 with 6X DNA dye (0.25% w/v bromophenol blue, 30% w/v glycerol) and loaded into wells in agarose gel which had been immersed in 0.5X TBE buffer (44.5 mM Tris,, pH 8.0, 44.5 mM Boric acid, 1 mM EDTA,, pH 8.0). As well as the DNA samples, 0.1 μg of 100 bp or 1 kb DNA ladders (Fermentas) were loaded as a size reference. 80~100 V was applied until the size of desired fragments could be distinguished.

Molecular cloning using Gateway technology

i. Generation of Gateway entry clones using the pENTR™/D-TOPO® entry system

The pENTR™ Directional TOPO® Cloning Kits (Invitrogen) was used to generate clones in the pENTR™/D-TOPO® plasmid vector (Invitrogen). PCR products are inserted into the linearised vector using topoisomerase I to join the two DNA molecules. In addition, the pENTR™/D-TOPO® vector contains *att*L1 and *att*L2 sites flanking the insertion sites such that the resulting clones can be used as Gateway entry clones. In the kit, pENTR™/D-TOPO® vector is provided as a linearized vector with topoisomerase I covalently bound to the 3'-phosphate at both DNA ends. Four nucleotides, GTGG are present at one end to produce a 5'-overhang that facilitates directional cloning of blunt-ended PCR products. The desired insert fragment was amplified by PCR using a forward primer with the extra four nucleotide sequence CACC at the 5'-end to promote recovery of recombinant clones with inserts in the desired orientation – the GTGG sequence from the vector invades the end of the PCR product and anneals to the complementary CACC sequence to stabilise the PCR insert in one orientation. The PCR reactions were performed using high fidelity Phusion taq (New England Biolab) which generates blunt-ended products compatible with the topoisomerase I cloning system. The size of the PCR product was checked by agarose gel electrophoresis. The gel contained 0.1 µg/ml ethidium bromide (EtBr) and DNA bands were detected by exposure under UV light. The intensity of the PCR product

band was quantified using Image J software and compared to the fragment of corresponding size within a 1 kb DNA ladder to estimate the concentration. About 5-10 ng of the PCR product was mixed with 0.5 µl of pENTR™/D-TOPO® vector. The reaction was carried out at room temperature for 30 min and then introduced into competent cells of *E. coli* strain DH5α by heat shock method (see "heat shock transformation of *Escherichia Coli*.").

ii. LR reaction

LR Clonase (Invitrogen) mediated recombination was used to transfer the cloned fragments from the entry clones into destination vectors. 50~150 ng of entry clone and 150 ng of destination vector were mixed and then H_2O was added to make the volume up to 4 µl. 1 µl of the LR Clonase II mix (Invitrogen) was supplied prior to the mixture being incubated at 25°C overnight. 1 µl of the LR reaction product was used for heat shock transformation of DH5α competent cells (see the next section "heat shock transformation of *Escherichia Coli*.").

Heat shock transformation of *Escherichia Coli*. (*E. Coli*.)

Competent cells (strain: DH5α for most plasmids, except for non-recombinant Gateway vectors which were propagated in strain DB3.1) were made by the method of Inoue el al. (Inoue et al., 1990) and stored at -80°C. Before the transformation, the competent cells were thawed on ice. 5µl of the ligation or gateway reaction, containing about 10 ng of plasmid was added to 100 µl of competent cells and the mixture was

kept on ice for 2 min before it was heated to 42^{o}C for 90 sec. After the heat shock, 1 ml of LB (1% w/v tryptone, 0.5% w/v yeast extract, and 1% w/v NaCl) was added to the cells to allow the cells to recover and express selection markers (for example, the plasmid-borne antibiotic resistance gene) at 37^{o}C for 1 hr with shaking at 100 rpm. The transformed cells were harvested by centrifugation at 7000 rpm in a micorcentrifuge for 3 min and plated onto LB agar plates containing antibiotics. The functional concentration for each antibiotic was: 100 µg/ml ampicilin, 25 µg/ml kanamycin, and 100 µg/ml spectinomycin.

Plasmid DNA mini-prep from *E. Coli*.

i. Alkaline lysis

1.5 ml of overnight culture was centrifuged at 7,000 rpm for 3 min in a microcentrifuge to harvest the cells. After discarding the supernatant, the pellet was re-suspended in 100 µl of solution I (50 mM glucose, 25 mM Tris, pH 8.0, 10 mM EDTA, pH 8.0) and then cells were lysed and DNA was denatured in 200 µl freshly prepared solution II (0.2 N NaOH, 1 % w/v SDS), followed by the addition of 150 µl of solution III (3 M potassium acetate [KOAc] and 11.5% v/v acetic acid). The acid neutralizes the pH in the solution, allowing the supercoiled plasmid DNA to reanneal, whereas the bacterial chromosomal DNA fails to reanneal as it is a longer sequence and is not supercoiled. . The viscous solution was gently inverted several times and then centrifuged at 16,000 g for 10 min to separate plasmid DNA (aqueous phase) from bacterial chromosomal DNA and proteins (precipitate). The supernatant was

transferred to a new tube and mixed thoroughly with 1 ml 95 % ethanol. The precipitated DNA was pelleted by centrifugation at 16,000 g for 10 min. The liquid phase was removed and the pellet was dried until no residual ethanol remained and then dissolved in 30 µl of T.1E (10 mM Tris, pH 8.0, 0.1 mM EDTA, pH 8.0).

ii. QIApure miniprep (plasmid mini-prep with Qiagen kit)

When high quality plasmid was needed, for example for DNA sequencing or for Gateway reactions, DNA was further purified on silica columns (Qiagen) according to the methods provided with the Qiagen miniprep kit.

Restriction enzyme digestion

Restriction enzyme digests were used to analyse the products of gateway cloning reactions, and also when cloning DNA flanking T DNA insertions using the Genome Walker procedure (see section 2.3). In the 20 µl reaction, 0.1-0.5 µg of DNA, 50 µg/ml RNAse, 10 U of restriction enzymes (NEB) and its supplied buffer (1X) were applied. In the case of double digestions, the buffer suggested by NEB catalog was used. The mixture was incubated at 37 ^{o}C for 1 hr and then the DNA fragments were separated by electrophoresis on 0.7~1.2% w/v agarose gel containing EtBr to check their size .

2.4 Floral dipping (transformation of Arabidopsis)

The floral dip transformation method (Clough and Bent, 1998) was modified according to Davis et al. (2009) (Davis et al., 2009) to simplify the procedure. In this method, the developing flowers of *Arabidopsis thaliana* are immersed into *Agrobacterium*-containing solution. *Agrobacteria* (*Agrobacterium tumefaciens*) can enter developing carpels in young flowers, where the carpels are not fused at their stylar ends, and insert T-DNA into the egg cells within the ovules (Desfeux et al., 2000). When the rare, transformed ovules are fertilized they developed into seeds harbouring transgenic embryos. The rare, transgenic seed can be selected for using the selectable markers introduced by the T-DNA, typically antibiotic or herbicide resistance.

Preparation of Agrobacteria competent cells

Competent cells were made by Tian Xin. Two single colonies of *Agrobacteria* (strain *GV3101* harbouring the mp90 helper plasmid conferring virulence and other functions for T DNA transfer) were firstly incubated in 3 ml YEP (1% w/v yeast extract, 1% w/v bacto peptone, 0.5% w/v NaCl) at 28^{o}C overnight, and then diluted into 20 ml YEP and grown again overnight. 500 ml of YEP were inoculated with the 20 ml culture and grown at 28 ^{o}C with 150 rpm of shaking until the optical absorbance at 600nm was about 0.5. Cells were harvested by centrifugation at 5000 rpm for 5 min. The supernatant was discarded and the pellet was resuspended in 100 ml 0.15 M NaCl.

Subsequently, the cells were harvested by centrifugation 6,500 rpm for 5 min. The supernatant was again disposed of and the pellet was resuspended in 10ml of 20 mM ice-cold $CaCl_2$. The cells were split into 200 µl aliquots and stored at -80 ^oC until needed.

Heat shock transformation of *Agrobacterium GV3101*

100 µl GV3101 competent cells were thawed on ice before 0.2 µg plasmid DNA were added. The cells and DNA were mixed by gently tapping the tube. The cells were frozen in liquid nitrogen for 1 min and then left in a 37 ^oC oven until melted. 1 ml of YEP was added and the cells were incubated at 28^oC for 1 hr with shaking at 100 rpm to help the transformed cells recover and express selection markers (antibiotic resistance genes). The cells were collected by spinning at 7000 rpm on a bench-top centrifuge for 3 min and plated on selective YEP plates.

Preparation of dipping solution

A single colony of the transformed *Agrobacteria* was picked and cultured overnight in 5 ml YEP medium containing 80 µg/ml gentamycin to select for the mp90 binary helper plasmid, together with the antibiotic appropriate for for the binary vector, in most cases kanamycin. 2 ml of overnight culture was diluted into 500 ml fresh YEP and cultured overnight. 100 µl of silwet L-77 was added into the overnight culture just before floral dipping was conducted. Silwet L-77 is a detergent which helps wetting of

plant tissue, which is hydrophobic due to the presence of a waxy cuticle, and facilitates *Agrobacteria* penetration into carpels.

Plant material for dipping

Plants were grown in SD conditions for four weeks and then transferred to LD conditions to induce flowering. Once bolting occurred, the primary inflorescence stem was cut to induce development of multiple, synchronised lateral inflorescences from axillary buds in the rosette. Plants were ready to dip when the inflorescence stems were about 15~20 cm-high.

Floral dipping

Flower buds were immersed in the dipping solution for 1 min and then rested for 1 min. This step was repeated at least 6 times. While immersed, the flower buds were gently pressed to help the dipping solution enter the buds. After dipping, the plants were covered with plastic bags for 8~12 hrs in the growth room. Then the bags were removed and the T0 plants were grown in LD conditions to set T1 seeds. Seeds were collected and sown on selection media to select primary (T1) transformants. The concentration of antibiotics for selection of transformants was 20 μg/ml for hygromycin B or 50 μg/ml for kanamycin.

2.5 Protein analysis

β-glucuronidase (GUS) staining

Plant material was harvested and briefly incubated for 30 min in 90% v/v acetone on ice in order to permeabilise tissue. Before the tissue was stained for activity of the *GUS* reporter gene, acetone was washed out by rinsing in distilled water (dH_2O). Then tissue was immersed in GUS staining solution (0.05 M $NaHPO_4$, pH 7.2, 0.5 mM $K_3Fe(CN)_6$, 0.5 mM $K_4Fe(CN)_6$, 1% Triton X-100, 2 mM 5-Bromo-4-chloro-3-indolyl β-D-glucuronide [X-GlucA]), a vacuum was applied for 10 min twice to facilitate incorporation of the substrate for the GUS gene product and the tissue was subsequently incubated at 37°C overnight. After the staining solution was removed, the stained tissue was washed with 70% ethanol (EtOH) for 1 hr to remove chlorophyll, followed by 85% EtOH for 1 hr and 95% EtOH for 1 hr. The stained tissue was stored in 95% EtOH at room temperature.

Confocal microscopy

To analyse GFP fluorescence in transgenic plants carrying GFP reporter gene constructs, freshly harvested roots of 10-day-old seedlings from MS plates were counterstained with 5 μg/ml propidium iodide (PI) by immersion in the solution for 5 minutes. PI is a fluorescence dye with the peak emission at 617 nm, and is commonly used to stain cell walls and nuclei. The latter are only stained in dead cells as the

plasma membrane of living cells excludes the dye, except at very high concentrations. The roots were detached from hypocotyls, placed on microscope slides and covered with cover slides, and observed on an Olympus Fluoview confocal microscope, using filter sets for texas red and GFP fluorescence.

Protein extraction

i. Crude protein extraction

0.1 g of plant material was harvested and ground in liquid nitrogen. 2X Laemmli buffer (0.125 mM Tris, pH 6.8, 4% w/v sodium dodecyl sulfate [SDS], 18% glycerol, 0.024% w/v bromphenol-blue, 1.43 M β–mercaptoethanol, 0.2% protease inhibitor cocktail [Sigma]) was pre-heated to 95°C before 200 µl hot buffer was added to the ground tissuepowder. The mixture was vortexed for 10 sec, immediately refrozen in liquid nitrogen, and then transferred to a heat block to boil at 95℃ for 10 min. Plant debris was separated from protein extracts by centrifuging the mixture at 16,000 g in a table top centrifuge for 10 min. The supernatant was transferred to a new tube and centrifuged again at 16,000 g for 10 min. The second supernatant was the crude protein extract. 10 or 15 µl of the extracts was analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting.

ii. Histone-enriched protein extraction

1 g of plant tissue was collected and ground in liquid nitrogen and then resuspended in 30 ml of extraction buffer (0.25 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 1

mM $CaCl_2$, 15 mM piperazine-N, N-bis(2-ethanesulfonic acid) (PIPES), pH 6.8, 0.8% v/v Triton X-100, 1% v/v protease inhibitor cocktail [Sigma]). The mixture was centrifuged at 10,000 g for 20 min at 4℃. The supernatant was discarded and the pellet containing nuclei was resuspended in 6 ml of 0.4 M H_2SO_4 for acid extraction of highly basic proteins such as histones. The extract was centrifuged at 22,000 g for 5 min at 4℃. The supernatant was transferred to a new tube and 12 volume of acetone was added, mixed well and kept at -20℃ overnight to precipitate proteins. The proteins were pelleted by centrifugation at 8,000 rpm for 15 min at 4℃. The pellet was dissolved in 600 µl of freshly prepared 4M urea and stored at -80℃

Western blot analysis

i. SDS-PAGE and Coomassie blue staining.

Gels were prepared in the Bio-Rad mini-protean 3 gel electrophoresis system using 10% acrylamide as the separating/resolving phase (10% w/v of acryl/bisacrylamide mixture [37.5:1], 0.375 M Tris, pH 8.8, 0.1% w/v SDS, 0.1% w/v ammonium persulfate [APS], 0.01% v/v N,N,N',N'-Tetramethylethylenediamine [TEMED]). Once the gel was set, the stacking phase (6% w/v of acryl/bisacrylamide miture, 0.125 M Tris, pH 6.8, 0.1% w/v SDS, 0.1% w/v ammonium persulfate [APS, 0.01% v/v TEMED) was poured on top of the resolving phase. A 15% w/v resolving phase was prepared when analysing total histone proteins, due to their small size.

Protein extracts were mixed with an equal volume of 4X sample buffer (120 mM Tris, pH 6.8, 4% w/v SDS, 10% v/v glycerol, 0.025% w/v bromophenol blue, 200 mM dithiothreitol [DTT]) and heated at 95℃ for 10 min to denature proteins. For the protein extract prepared by the crude extraction method, the sample buffer (Laemmli buffer) had been added while extraction. The protein samples (roughly 20 µg of total protein) were loaded, and separated by electrophoresis at a constant voltage (80 V) at room temperature until the bromophenol blue dye migrated out of the bottom of the gel. The electophoresis was run in the running buffer comprising 25 mM Tris, pH 8.3, 192 mM glycine, 0.1% w/v SDS).

Two identical gels were run simultaneously. The first was stained with Coomassie blue to assess whether equal amounts of protein had been loaded for the different samples and also to check the integrity and resolution of the proteins. The gel was soaked in staining solution (45% v/v methanol [MeOH], 10% v/v acetic acid, 0.25% w/v Coomassie Brilliant Blue R-250) overnight at room temperature with gentle shaking (40 rpm). After staining, the gel was de-stained by incubating in several changes of a 30% v/v MeOH, 10% v/v acetic acid solution until bands corresponding to proteins were clearly visible. The second gel was used for Western blotting as described below.

ii. Protein transfer to membrane (Western blot)

The glass plates housing the polyacrylamide gel were prised apart, and the gel was arranged in a “sandwich” for transferring the resolved proteins onto membrane. The

sandwich was arranged as follows: black side of the Biorad MiniProtean western transfer cassette, sponge, 3M filter paper, nitrocellulose membrane (pore size 0.22 μm), the gel, 3MM filter paper, sponge, white side of the cassette. The sandwich was set in the electrophoresis tank with the black side of the cassette toward the anode and immersed in freshly prepared Western transfer buffer (0.1 M Tris, 192 mM glycine, 0.1% w/v SDS, 20% v/v MeOH). Proteins were transferred to the filter at 22 Volts overnight at 4^{o}C. To prevent heating and possible protein denaturation during transfer, the buffer was cooled by immersion of an ice pack (Biorad) previously frozen at -80^{o}C and circulated using a stir bar and magnetic stirrer. In addition, the transfer was performed in a cold room. All gels included prestained protein markers loaded in the first lane. After the Western transfer, the membrane was examined to confirm that the prestained protein markers had transferred to the membrane.

iii. Blocking, antibody incubation and detection

Once the transfer was completed, the membrane was soaked in 10% w/v skimmed milk and shaken at 40 rpm at room temperature in order to "block" the membrane, that is to prevent non-specific binding of antibody to membrane at later stages by first allowing milk proteins to bind to it. The membrane was briefly washed with 0.1% PBST (1X phosphate buffered saline [PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4 \bullet 2\ H_2O$, 2 mM KH_2PO_4, pH=7.4], 0.1% v/v Tween 20) and then the primary antibody was applied. The primary antibody was diluted in blocking buffer (1% skimmed milk in 0.1% PBST) roughly 1:1000~1:5000 depending on the antibody used.

The binding of the primary antibody was done at 4℃ overnight with 40 rpm shaking. After binding, the un-bound antibody was washed out with 0.1% PBST by shaking at 40 rpm for 5 min. This wash step was repeated four times. Then the secondary antibody (typically a mouse monoclonal antibody against rabbit IgG, conjugated to horseradish peroxidase, HRP) diluted (usually 1:5000) in blocking buffer was added to the membrane and it was incubated at room temperature for 1 hr. The membrane was washed with 0.1% PBST for 5 min a total of three times, then washed once in PBS. The bound antibodies were detected by chemiluminescence using a kit (Pierce® Western Blotting substrate, #32109) as follows: The supplied peroxidase and illuminon solutions were mixed 1:1 and applied directly to the membrane. 1 ml of each solution was sufficient for for a typical 5 x 8 cm membrane. Once the ECL solutions were applied, the membrane was briefly dried and wrapped in cling film, and then exposed to films. The films were developed using an automatic developing machine (Konica SR101A).

Chromatin immunoprecipitation (ChIP)

The ChIP procedure was based on the method described by Gendel et al. (Gendrel et al., 2002) with minor modification. In brief, the plant tissue was first fixed with formaldehyde to cross-link proteins through the amino groups on lysine residues to those on neighbouring proteins or DNA molecules. The cross-linking holds protein-DNA or protein-protein complexes together during chromatin extraction procedures. Next, the nuclei were isolated and lysed to obtain chromatin. Chromatin

was sonicated afterward to break long DNA into small fragments. Chromatin was immunoprecipitated by desired antibodies pre-bound on magnetic beads. After several washes, the immunoprecipitated protein-DNA complexes were eluted. The cross-linking of eluted complexes was reversed by heating in the presence of high concentrations of NaCl and the protein parts were digested by proteinase K. Finally, the amount of certain DNA fragments was estimated by real time qPCR amplification.

i. Tissue fixation

In order to fix chromatin proteins such as histones to DNA, cross linking was performed using formaldehyde. Plants were grown on MS plate under LD condition, and 1 g of 2-week-old whole seedlings was harvested and immersed in 1X PBS containing 1% w/v formaldehyde on ice and a vacuum was applied for 30 min to aid penetration of the fixative into plant tissues. Fixation was stopped by quenching with glycine at a final concentration of 0.125M and applying a vacuum for 5 min. Tissue was rinsed with dH_2O and all liquid was removed with tissue paper. Dried tissue was frozen and ground in liquid nitrogen and stored at -80^oC.

ii. Chromatin isolation and sonication

30 ml of ChIP extraction buffer 1 (0.4 M sucrose, 10 mM Tris, pH 8.0], 10 mM $MgCl_2$, 1 mM EDTA, pH 8.0, 5 mM β–mercaptoethanol, 0.2 mM pefabloc, 0.5% v/v protease inhibitor cocktail [Sigma]) was added to the ground tissue, vortexed and the crude extract was filtered twice through Miracloth to remove large debris. The flow-through

was centrifuged at 3,220 g for 10 min at 4^{o}C to separate the dense nuclei from membranes and other organelles. Then the supernatant was carefully discarded and the pellet was thoroughly re-suspended in 1 ml ChIP extraction buffer 2 (0.25 M sucrose, 10 mM Tris, pH 8.0, 10 mM $MgCl_2$, 1 mM EDTA, pH 8.0, 1% v/v Triton X-100, 5 mM β–mercaptoethanol, 0.2 mM pefabloc, 0.5% v/v protease inhibitor cocktail [Sigma]) and transferred to a new 1.5 ml eppendorf tube, followed by centrifugation at 16,000 g at 4^{o}C for 20 min. Again, the supernatant which is enriched for lysed chloroplasts and other organelles due to the high detergent in buffer 2, was removed and the pellet containing nuclei was resuspended by pipeting in 300 µl of ChIP extraction buffer 3 (1.7 M sucrose, 10 mM Tris, pH 8.0, 2 mM $MgCl_2$, 1 mM EDTA, pH 8.0, 0.15% v/v Triton X-100, 5 mM β–mercaptoethanol, 0.2 mM pefabloc, 0.5% v/v protease inhibitor cocktail [Sigma]). The resuspended solution was applied on top of 300 µl clean ChIP extraction buffer 3 in a new eppendorf. The two-layer solution was centrifuged at 16,000 g at 4^{o}C for 60 min as a crude sucrose gradient centrifugation aimed at purifying nuclei from other organelles on the basis of density. The white-ish pellet enriched for nuclei was extracted in 300 µl of nuclear lysis buffer (50 mM Tris, pH 8.0, 10mM EDTA, pH 8.0, 1% w/v SDS, 0.1 mM pefabloc, 0.5% v/v protease inhibitor cocktail [Sigma]) in which the high SDS detergent lyses nuclei and solubilises chromatin. To fragment the chromatin into small pieces (DNA of length 300 – 600 bp) and also to further solubulise the chromatin, it was sonicated using a Diagenode Bioruptor for 20 min (total time) in 20 cycles of the 30 sec ON and 30 sec OFF program. After sonication, the nuclear debris was removed by centrifuging at 16,000 g for 10 min at 4^{o}C. The supernatant is enriched for solubilised fragmented

chromatin containing DNA crosslinked to histones and other proteins and RNA. 20 μl of the chromatin-enriched extract was taken as input control. The remaining chromatin-enriched extract was diluted 10X using ChIP dilution buffer (16.7 mM Tris, pH 8.0, 1.2 mM EDTA, pH 8.0, 1.1% v/v Triton X-100, 167 mM NaCl, 0.1 mM pefabloc, 0.5% v/v protease inhibitor cocktail [Sigma]) to reduce the concentration of SDS to 0.1%, as most antibodies will not bind their targets in 1% SDS. The diluted chromatin was used for immunoprecipitation.

iii. Immuno-precipitation (IP) and chromatin DNA elution

Protein A-conjugated magnetic beads (Dyna beads, Invitrogen) were pre-bound with the desired antibody at 4^{o}C for at least four hrs before IP. For each sample, 20 μl of beads and 2 μg of antibody were applied (α-H3K27me3, Millipore 07-449; α-H3K4me3, Active Motif 39159; α-H3K36me3, Abcam 9050; α-H3ac, Millipore 06-599; α-H4ac, Millipore 06-866). After binding, the antibodies which do not bind to the protein A beads were washed off with ChIP dilution buffer at 4^{o}C, this washing was repeated 3 times The ChIP dilution buffer was removed before 450~650 μl of the diluted chromatin was added to the beads. The IP was performed at 4^{o}C with gentle shaking for 2 hrs up to overnight. Subsequently, the beads were washed at 4^{o}C for 5 min for two washes each in a series of wash buffers of increasing stringency (salt concentration) as follows: low salt buffer (150 mM NaCl, 0.1% w/v SDS, 1% v/v Triton X-100, 2 mM EDTA, pH 8.0, 20 mM Tris, pH 8.0]); high salt buffer (500 mM NaCl, 0.1% w/v SDS, 1% v/v Triton X-100, 2 mM EDTA, pH 8.0, 20 mM Tris, pH

8.0]); finally, LiCl wash buffer (0.25 M LiCl, 1% v/v NP-40, 1% sodium deoxycholate, 1 mM EDTA, pH 8.0, 10 mM Tris, pH 8.0). The beads were washed twice in 1ml of each wash buffer. Finally, the beads were washed with 1 ml of TE buffer once. The bound chromatin was eluted by heating up the beads to 65^{o}C for 30 min with 500 µl elution buffer (1% w/v SDS, 0.1 M $NaHCO_3$). 20 µl 5 M NaCl was added to the eluted chromatin solution and incubated at 65^{o}C for more than 5 hrs to reverse the crosslinking.

iv. Proteinase K treatment and chromatin DNA purification

After the eluted chromatin solution had cooled, 10 µl of 1 M Tris (pH=6.5), 20 µl of 0.5 M EDTA and 1 µl of 10 mg/ml proteinase K was added and mixed well. The mixture was left at 45^{o}C for 1 hr to digest proteins. While the digestion was completed, 500 µl of phenol/chloroform (1:1) was applied to denature proteins and DNA was dissolved in aqua phase. The mixture was vortexed vigorously and centrifuged at 16,000 g for 10 min. The DNA-containing, upper layer was transferred to a new tube and DNA was precipitated by adding 1 ml of 95% EtOH. 2 µl of 10 mg/ml glycogen was also added as a carrier to aid precipitation and visualizing of the minute amounts of DNA involved (less than 5 ng). The DNA was precipitated at -20^{o}C overnight. The precipitated DNA was collected by centrifuging at 16,000 g at 4^{o}C for 10 min and the pellet was washed with 1 ml of 70% EtOH. Subsequently, the EtOH was removed and the pellet was dried at room temperature. The dried pellet was dissolved in 1 ml of T0.1E buffer (10 mM Tris, pH 8.0, 0.1 mM EDTA, pH 8.0). 7 µl of the purified DNA

was used for quantification in realtime PCR reaction (see "Realtime PCR for quantification of gene expression" in RNA analysis section). Percentage of input (%input) was calculated as the relative abundance of the amplified loci in immunoprecipitated sample to input control. Fold enrichment was the the ratio of %input of target divided by % input of the loading controls, depending on the histone marks observed.

Yeast two hybrid assay

i. Construction

The *SOP12* coding sequence (CDS) was cloned into the pGBKT7 vector by performing the LR reaction between the entry clone (*SOP12* cDNA in pENTR™/D-TOPO® vector) and destination vector (modified pGBKT7 vector which contains the Gateway cassette [attL sites for recombination, ccdB gene which only allows the vector exists only in DB3.1, and chloramphenicol resistant gene], a gift from Dr. Daniel Schubert in University of Dusseldorf, Germany). This produced the "bait" construct (pGBKT7-SOP12) which expresses SOP12 protein fused with GAL4 DNA binding domain (BD). By the same method, *SOP12* cDNA was also constructed into modified Gateway compatible pGADT7 vector from Dr. Daniel Schubert, which would give the GAL4 transcription activation domain (AD)-SOP12 fusion protein.

In addition to *SOP12*, CDS of *EFS*, *ULT1* and *ULT2* were also constructed into Gateway compatible pGBKT7 and pGADT7 vectors, respectively.

ii. Yeast transformation by lithium acetate (LiOAc) and polyethylene glycol (PEG) method

The yeast strain used for yeast two hybrid assay was Y2H Gold® (Clontech) *Mata,trp1–901,leu2–3,112,ura3–52,his3–200,gal4Δ,gal80Δ,Lys2::Gal1$_{UAS}$-Gal1$_{TATA}$ ::His3, Gal2$_{UAS}$-GAL2$_{TATA}$::Ade2,Ura3::Mel1$_{UAS}$-Mel1$_{TATA}$::Aur1-C, Mel1*, Clontech), whereas in yeast two hybrid library screening involving mating Y187 (*Matα,ura3-52,his3-200,ade2-101,trp1-901,leu2-3,112,gal4Δ,met–,gal80Δ,Mel1,Ura 3::Gal1$_{UAS}$-Gal1$_{TATA}$-lacZ*) was used since the library was in strain AH109 (*Mata,trp1-901,leu2-3,112,ura3-52,his3-200,gal4Δ,gal80Δ,Lys2::Gal1$_{UAS}$-Gal1$_{TATA}$: :His3,Gal2$_{UAS}$-3Gal2$_{TATA}$::Ade2, Ura3::Mel1$_{UAS}$-Mel1$_{TATA}$-lacZ::Mel1)* with opposite mating type. Those strains carry multiple auxotrophic mutants so the ability to synthesise the nutrients tryptophan, leucine, histidine and adenine can be used to select transformed yeast cells or for interaction in the two hybrid assay as described later.

Yeast competent cells were made on the day of use as follows: yeast cells were cultured in 1X yeast peptone dextrose adenine (YPDA, 1% w/v yeast extract, 2% w/v difco peptone, and 0.003% w/v adenine hemisulfate) overnight at 30^{o}C. The overnight culture was diluted 1:50 and incubated at 30^{o}C until the OD_{600} reached 0.5. The culture was split into 50 ml aliquots and spun down at 450 g for 3 min and the supernatant was removed. The pellet was resuspended in 1 ml of LiOAc mix (100 mM LiOAc, 10 mM Tris, pH 7.5, 1 mM EDTA) by pipetting and transferred to a 1.5 ml tube. LiOAc provides Li+ which may help increase transformation efficiency by its chaotropic

effect (S. Kawai et al., 2010). Cells were spun down at 3800 g for 1 min, resuspended in 1 ml of LiOAc mix, spun down again and resuspended in 500 μl LiOAc mix. The yeast competent cells could be left at room temperature for several hours.

100 μl of yeast competent cells in LiOAc mix were first mixed with 10 μl of 10 mg/ml salmon sperm DNA and 1~2 μg of plasmid DNA. Then 700 μl of PEG mix (40% w/v PEG-3000 [Sigma #P3640], 100 mM LiOAc, 10 mM Tris, pH 7.5, 1 mM EDTA) were added and the solution was mixed thoroughly by gently pipetting. The mixture was heat-shocked at 42 ^{o}C for 15 min in a water bath, and then left in 30 ^{o}C incubator for 30 min to recover. After cells were spun down by centrifugation at 3800 g for 1 min, the pellet was resuspended in 100 μl of 1X YPDA and plated on double drop-out media (DDO media, which is composed of yeast minimum media [YMM, 0.67% w/v yeast nitrogen base without amino acids, 2% agar] and essential amino acids except for leucine [Leu, L] and tryptophan [Trp, W]). The transformants harbouring the bait or the prey construct would obtain the ability to synthesize Trp and Leu, respectively. In other words, the yeast cells growing on media lacking Leu and Trp are transformed by both bait and prey constructs. After 3~5 days of incubation at 30^{o}C, the colonies formed on DDO were replicated to triple drop-out media (TDO media, -L –W, –histidine [His, H]) and quadruple drop-out media (QDO media, -L –W –H, -adenine [Ade, A]) to test interaction between the bait and the prey, because the interaction would bring GAL4 DNA binding domain and transcription activation domain together and drive the expression of reporter genes (*His3* for histidine synthesis, *Ade2* for adenine synthesis, and α-galactosidase [*Mel1*] which metabolises

X-α-gal [5-Bromo-4-chloro-3-indolyl-α-D-galactopyranoside] to produce a blue product).

iii. Library screening by yeast mating

The procedures used were based on the protocols provided in the Matchmaker™ Library Construction & Screening Kits User Manual (Clontech). Instead of yeast transformation, in this method the bait and prey are introduced into opposite yeast haploid mating types (α or a), and then the two type of cells are mated to generate diploid double transformants. The resulting diploid cells are selected on DDO media and then the two hybrid interaction is tested by growing the diploid strain on TDO and QDO media.

The bait, pGBKT7-SOP12 was introduced into Y187 by transformation and was used to screen a library prepared from whole plant, including leaves, flowers, and siliques, etc. The cDNA library was made by Dr. H. Sommer, Max Planck Institute, Germany) and provided as a gift by Dr. B Causier and B Davies (University of Leeds, UK). The prey, cDNA library, was made from random primed cDNA and expressed as fusions to GAL4-AD in the pGADT7-Rec vector introduced into yeast strain AH109. About $5x10^7$ cells of the prey were thawed at room temperature and mixed with 5 ml overnight culture of the bait and 45 ml of 1X YPDA in a 1 liter flask. The two types of cells (Y187 transformed with pGBKT7-SOP12 and AH109 harbouring pGADT7-cDNA library) were mated in the flask at 30^oC for 24 hrs with slow shaking (20 rpm). After mating, the mixture was transferred to a new tube and cells were

collected by centrifugation at 1,000 g for 10 min. The supernatant was removed before cells were resuspended in 10 ml of 1X YPDA and plated on DDO media to select diploid cells containing both the bait and prey. For a 10-cm diameter plate, 90 μl of the suspended cell were plated. The plates were grown at 30^{o}C for 3~5 days, when colonies were formed. The clones grown on TDQ media were replicated to quadruple QDO containing X-α-gal.

iv. Recovery and analysis of prey constructs interacting with SOP12 bait

Colonies which expressed all three reporters for reconstituted GAL4 activtiy indicative of bait prey interaction were identified as those that survived on QDO media and further tested positive for *Mel1* reporter gene activity (blue colonies on plates containing X-α-Gal). The colonies were picked and swirled in 100 μl sterilized water. 1 μl of the resulting yeast cells suspension was used in a 20 μl PCR reaction to amplify the small amounts of plasmid DNA liberated when the cells were heated to 98^{o}C in the PCR reaction. The primers for the PCR reaction were 5'AD (5'-CTATACGATGATGAAGATACCCCACCAAACCC-3') and 3'AD (5'-CTGT-GCATCGTGCACCATCT-3') and are specific for sequences flanking the cDNA insertion site in the pGADT-Rec vector. After PCR, 10 μl of the product were analysed by agarose gel electrophoresis. Once a DNA band was observed, 1 μl of the PCR product was directly sequenced using the 5'AD primer. The resulting sequences were analysed using Blast to query the Arabidopsis genome (http://blast.ncbi.nlm.nih.gov/Blast.cgi.) and so identify the origin of the interacting

prey protein. In addition, the sequence was translated using the Geneious software suite to confirm whether the insert was present as an in frame fusion to the GAL4 AD, as might be expected for a functional protein fusion.

Immunoprecipitation (IP) for Mass spectrometry

i. Nuclear protein extraction

Nuclei were isolated as described in subsection ii in section "chromatin immunoprecipitation". About 2 g of whole plants (tissue above ground) was used. The pellet which formed after centrifugation in extraction buffer 3 is nuclei-enriched. Instead of resuspending in nuclei lysis buffer, the pellet was resuspended in 300 µl benzonase buffer (20 mM HEPES, pH 7.5, 1 mM $MgCl_2$, 10 mM NaCl, 0.1% Triton X-100, 1 mM DTT, 0.5% protease inhibitor cocktail [Sigma]). 1 µl of benzonase (about 250 U) was added to digest DNA into small fragments. The digestion was processed at room temperature for 15 min and then on ice for 45 min. After digestion, the concentration of NaCl was increased to 150 mM by adding 8.66 µl of 5 M NaCl. The solution was left on ice for 20 min to denature benzonase which is sensitive to high salt concentration. Next, the insoluble fraction (nuclear membrane debris) was pelleted by centrifugation at 16,000 g for 20 min at 4^oC. The supernatant was diluted 10X in IP buffer (20 mM HEPES, pH 7.5, 1 mM $MgCl_2$, 150 mM NaCl, 0.1% Triton X-100, 1 mM DTT, 0.5% protease inhibitor cocktail [Sigma]) and was ready for IP.

ii. IP

Before target proteins (GFP-tagged proteins) were immunoprecipitated. 15~20 μl of GFP-trap® beads (ChromoTek) was added into diluted nuclear protein extract. The solution was agitated at 50 rpm at 4°C for two hours. Subsequently, the beads were harvested by spinning and washed three times with 1 ml of IP buffer. Last, the beads were boiled for 10 min in 30 μl of 2X Laemmli buffer to elute the bound proteins/complexes. The eluted protein samples were sent to Prof. J. Rappsilber (University of Edinburgh, UK) to perform Mass spectrometry and the subsequent analysis.

Chapter 3. Molecular cloning of *SOP12*

3.1 Cloning sequences flanking the right border of the T-DNA insertion in *sop12-1*

The *sop12-1* allele was identified as a mutation suppressing the clf phenotype in the M2 population of a T-DNA mutagenesis conducted in a conditional *clf-50 pCLF::CLF-GR* background. The preliminary genetic analysis and molecular cloning were conducted by Dr. J. Goodrich and his results are briefly summarised here. When the suppressor mutant (*sop12-1 clf-50 pCLF::CLF-GR*) was crossed to *clf-50* mutants, the resulting 28 F1 plants (*sop12-1/+ clf-50 pCLF::CLF-GR*) all had the clf phenotype in the absence of DEX, indicating the suppressor mutation was recessive. In addition, when the F1 plants were selfed, the resulting F2 population segregated about 3:1 for plants with *clf* and suppressed clf phenotypes, respectively. This indicates that the suppression was likely caused by a single, recessive mutation, designated *sop12-1*. The T-DNA construct used for the mutagenesis was a modified version of the activation tagging construct, pSKI074 (Figure 3-1) (Weigel et al., 2000), into which a selectable marker for seed fluorescence (*pAT2S3::GFP*) (Kroj et al., 2003) had been inserted. When the F3 seeds from 13 F2 plants with the suppressed phenotypes (i.e. *sop12-1 clf-50* homozygotes) were examined, all were fluorescent, suggesting that the *sop12-1* mutation co-segregated with and was caused by a T-DNA insertion.

To identify the *SOP12* gene, the DNA flanking the T-DNA insertion was cloned using the genome walker technique (Siebert et al., 1995). The basic principle is that genomic DNA is cut with restriction enzymes, and ligated to adaptors. DNA flanking the insertion is then amplified by PCR using primers specific to the left border (LB) of the T-DNA and the adaptor sequence. This revealed that *sop12-1* plants harboured a T-DNA insertion with the left border inserted within the third exon of *AT3G63270*. Furthermore, PCR amplification using primers flanking the T-DNA insertion at *AT3G63270* showed that *sop12-1* mutants were homozygous for the T-DNA insertion. The *AT3G63270* gene was therefore a candidate for the *SOP12* locus.

Because T-DNA insertions can cause rearrangements such as inversions or large deletions that disrupt more than one gene, I first attempted to recover the DNA flanking the RB of the T-DNA insertion at *AT3G63270* in order to confirm that neighbouring genes were not affected. To do this, the plasmid rescue technique was used. Genomic DNA from *sop12-1* plants was cut with *Eco*RI or *Hin*dIII and ligated under dilute conditions favouring circularisation, introduced into *E. coli*. by heat shock transformation and transformants carrying T-DNA sequences were selected on the basis of ampicillin resistance. This was predicted to recover the T-DNA backbone, containing pUC19 plasmid sequences harbouring the ampicillin resistance gene and *E. coli*. origin of replication, together with sequences flanking the plasmid RB.

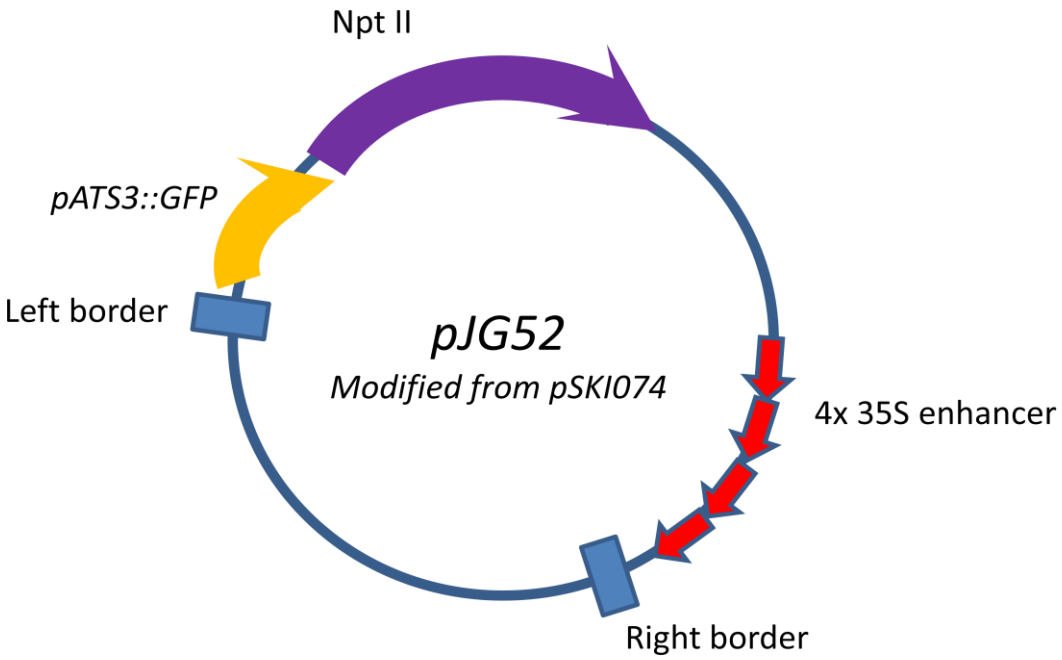


Figure 3-1. Map of pJG52

pJG52 is a derivative of pSKI074 (Weigel et al., 2000). pJG52 carries an additional selectable marker (*AT2S3::GFP*, which gives fluorescent seeds) near the left border. The T-DNA region of pSKI074 is shown in black, the plasmid backbone in grey. The features of the T-DNA region illustrated are the left border and right border, the *nptII* gene encoding kanamycin resistance as the selectable marker for plant transformation, pUC19 sequences containing a plasmid replication origin and ampicillin resistance gene allowing for propagation and selection when plant DNA is circularised and introduced into *E. coli* during plasmid rescue, and four CaMV 35S enhancers near the RB end. To make pJG52, pSKI074 was digested with *Hin*dIII and made blunt ended, followed by ligating with a blunt ended *pAT2S3::GFP* fragment. The construction and further transformation of *Agrobacteria* and *Arabidopsis* were conducted by Dr. J. Goodrich.

However, sequence analysis of the rescued plasmids showed that in all cases the T-DNA RB was flanked not by plant DNA but rather by the RB sequence of a second T-DNA in opposite orientation. This suggested that the T-DNA insertion at *AT3G63270* was complex and contained at least two T-DNAs in inverted head to tail orientation (Figure 3-2). Further attempts at genome walker PCR using either LB or RB specific primers did not help recover plant DNA flanking the other end of the T-DNA. As an alternative strategy, primers to amplify *AT3G63270* sequences predicted to be downstream of the insertion were designed and used for genome walking back towards the T-DNA insertion (Figure 3-2). Sequence analysis of the resulting products indicated that the plant DNA was adjacent to a second LB T-DNA end. The simplest interpretation of these results was that the *sop12-1* mutation resulted from a tandem T-DNA insertion in the third exon of *AT3G63270*, with the two T-DNAs in inverted orientation. Importantly, the sequence analysis indicated that there was no major rearrangement of plant DNA involved, beyond the deletion of 62 bases at the T-DNA insertion site (see Figure 3-2)

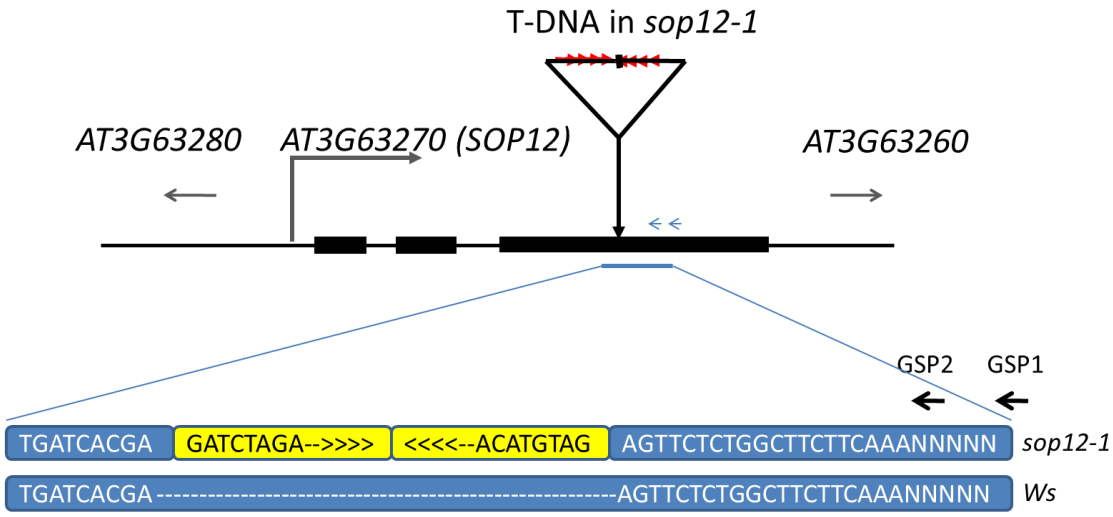


Figure 3-2. Sequences of the T-DNA flanking region in *sop12-1*

Gene structure of *SOP12* and the location of T-DNA insertion in *sop12-1*. Black boxes represent exons of *AT3G63270*, and the solid line indicates non-coding regions. The gray arrows indicate orientation of genes, the big open triangle represents the T-DNA inserted at *AT3G63270* and the red arrows point out the location of CaMV 35S enhancers at the RB of T-DNA. Part of the sequence, indicating by a short blue line, is shown underneath the gene structure diagram, which is the alignment of *sop12-1* and the *Ws* progenitor sequences in the region of the T-DNA insertion. Sequences in the third exon of *AT3G63270* are shown in white text with blue highlight, whereas the ends of the inserted T-DNA are shown in black text with yellow highlight. The T-DNA insertion resulted in deletion of 62 bases, which are indicated in dashed lines in *Ws* the sequence. The four CaMV 35S enhancers near the T-DNA RB end are indicated by > and < signs. The rough positions of the two gene-specific primers (GSP1 and GSP2) used for the genome walker procedure are marked. In this diagram the length of each fragment is not proportional to the length of actual sequences.

3.2 The T-DNA insertion in *sop12-1* does not affect expression of genes neighbouring *AT3G63270*.

The pJG52 construct used for the mutagenesis harbours four copies of the 35S enhancer sequence near the T-DNA RB and might therefore cause over expression of genes neighbouring the insertion ("activation tagging"). Although the molecular analysis described in section 3.1 indicated that the T-DNA inserts at *AT3G63270* did not delete or disrupt neighbouring genes, it was still possible that their expression might be affected. To test this, cDNAs from seedlings of *Ws* (wild type), *clf-50*, *sop12-1* and *sop12-1 clf-50* plants were prepared and the expression of *AT3G63260* and *AT3G63280*, the genes either side of *AT3G63270*, was analysed by RT-PCR. The expression of these two genes does not correlate with *sop12* mutation (Figure 3-3), so it is likely that the T-DNA insert at *sop12-1* affects only *AT3G63270*. A similar analysis of an independent allele harbouring a transposon insertion at *AT3G63270* (*sop12-3*, see next section) also suggested that the expression of the neighbouring genes was unaffected (Figure 3-3). Furthermore, these results are consistent with the genetic analysis of *sop12-1* which showed that its suppression of the clf phenotype is a recessive trait rather than the dominant or semi-dominant behaviour that would be expected to be caused by hyperactivity of *SOP12*.

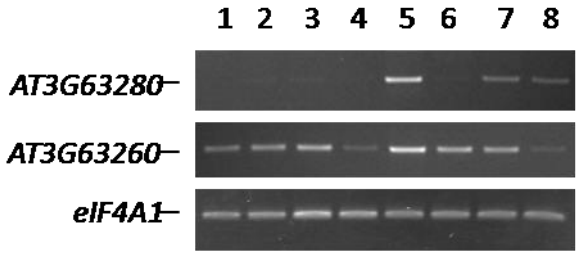


Figure 3-3. Expression of *AT3G63260* and *AT3G63280* is not affected by the *sop12-1* mutation

RT-PCR analysis of cDNA from seedlings. Lane 1 to 8 comprise 1, *Ws*; 2, *clf-50*; 3, *sop12-1*; 4, *sop12-1 clf-50*; 5, L*er*; 6, *clf-2*; 7, *sop12-3*; 8, *sop12-3 clf-2*. *eIF4A1* encodes a translation initiation factor and is used as a reference gene to control for equal cDNA amounts between samples. *AT3G63280* is expressed at a relatively low level in lanes 1-4, *AT3G63260* at a slightly higher level, however in neither case is expression in *sop12-1* (lanes 3 and 4) obviously elevated.

3.3 An independent allele disrupting *AT3G63270* also confers the *sop12* phenotype

The above analysis suggested that *AT3G63270* might correspond to *SOP12*, but did not exclude the possibility that a linked mutation in a nearby gene was actually responsible. To test further the role of *AT3G63270,* databases were searched for additional lines harbouring T-DNA or other inserts. This revealed two available insertions. The first line, named *sop12-2*, harbours a SALK T-DNA insert upstream of *AT3G63270* in the promoter region (SALK_051438c). Although genetic analysis showed that this allele did not suppress the clf phenotype, RT-PCR analysis indicated that it did not obviously reduce *SOP12* expression (data not shown) and therefore is

unlikely to cause a loss of function. The second line (CSHL_ET1398, hereafter designated *sop12-3*) contains a *Ds* transposon insertion in the third exon of *AT3G63270*, and was therefore more likely to cause a loss of function. The *sop12-3* allele was generated in L*er* background. Accordingly, double mutants were constructed between this allele and *clf-2*, a null allele of *clf* also in a L*er* background. Consistent with *AT3G63270* corresponding to *SOP12*, *sop12-3* gave a similar suppression of the clf phenotype as did *sop12-1* (see Figure 3-4)

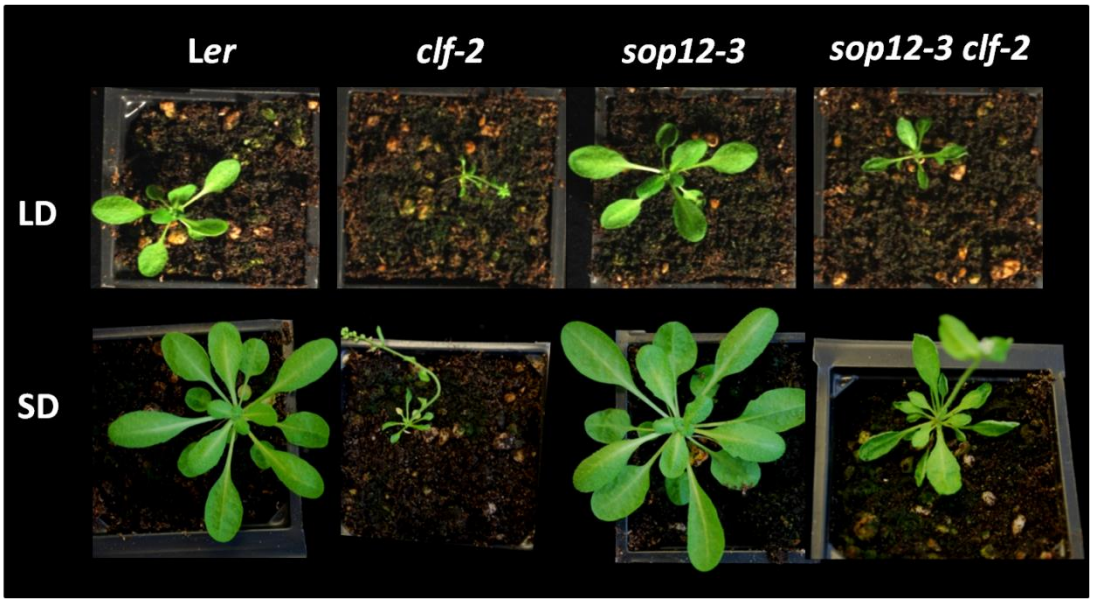


Figure 3-4. The independent *sop12-3* allele in L*er* backgound also suppresses the clf phenotype

sop12-3 is an independent mutation allele of *AT3G63270*. It suppresses the phenotype of a null *clf* mutant in L*er* background (*clf-2*). In *sop12-3 clf-2*, leaves are less curled and the morphology is more similar to wild type (L*er*). Photos were taken when plants were grown under long day (LD) for 26 days or short day (SD) for 51 days. The inflorescence stems were removed from L*er*, *sop12-3* and *sop12-3 clf-2* under LD.

3.4 *AT3G63270* complements *sop12* mutation

To confirm that *AT3G63270* corresponds to *SOP12*, a transgene complementation experiment was performed. Because *sop12* single mutants have barely discernible phenotypes, the strategy was to introduce a genomic clone of *AT3G63270* into the *sop12-1 clf-50* mutant background and to look for complementation of the suppression, in other words restoration of the severe clf phenotype. To do this, primers were designed to amplify the *AT3G63270* genomic locus including -440 bps of upstream sequences and 931 bps downstream of the translational stop codon. Genomic DNA was amplified using a high fidelity polymerase (Phusion® High-Fidelity DNA Polymerase, NEB). However, amplification from genomic DNA of Columbia (Col) ecotype plants was not successful, even though several alternative primers were tried. By contrast, there was no difficulty in amplifying the coding sequences and downstream sequences (data not shown). One possibility is that the reference *Arabidopsis* genome sequence, from which the primers were designed, was rearranged, different from that in the lab Col accession used, or the sequence assembly was incorrect. Since the *sop12-1* mutation was isolated in *Ws* ecotype background, genomic DNA from *Ws* plants was used as the PCR template, and in this case a 2.9 kb fragment was amplified that spanned the *AT3G63270* locus. The fragment was cloned into pENTR™/D-TOPO® vector, to give clone pENTR-SOP12g, introducing attL sequences for Gateway recombination on either side of the insert. Sequence comparison of the insert with the Col-0 reference sequence revealed 3 single nucleotide polymorphisms (-565 CCG to CCC, -547 CCT to CCC, and -153 GTC to

GTG where the ATG start codon is postion 1) and one 186-bp-deletion in *SOP12* promoter in *Ws* (-475 to -288) (Figure 3-5).

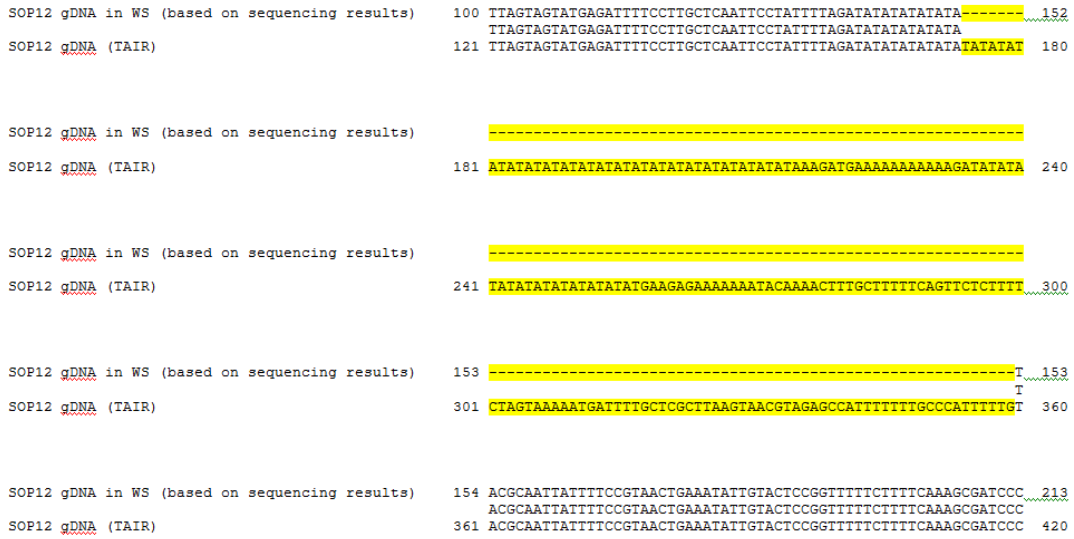


Figure 3-5. Comparison of the *SOP12* promoter sequence between *Ws* and Col ecotypes
The cloned *SOP12* promoter from *Ws* was sequenced and aligned with the Col reference sequence from TAIR. A 186-bp deletion is indicated in grey highlights. The start codon (ATG) is at position 516 and 723 in *Ws* and TAIR reference sequence (Col), respectively.

In addition, there was only one polymorphism found in the gene body (521 GAC to GAT) and this is predicted to be silent with respect to the amino acid composition. To verify that the sequence polymorphisms were not PCR induced mutations, direct sequencing of independently amplified products from *Ws* genomic DNA was conducted and the identical results obtained suggests that the differences represent genuine polymorphisms between the ecotypes. The sequence affected does not obviously contain terminal inverted repeats or target site duplications that are

characteristic of transposon insertions. However, it is very rich in TA repeats, suggesting that it may consist in part of a microsatellite repeat sequence. Microsatellite sequences are highly variable in length, it is thought due to misalignment (slippage) occurring during DNA replication.

To perform the complementation, the cloned *AT3G63270* locus in pENTR-SOP12g was transferred into the binary vector pGWB1 (Nakagawa et al., 2007) by Gateway recombination using the LR reaction. The resulting clone, here designated *pSOP12::SOP12* was transferred into *Agrobacterium tumefaciens* strain GV3101 (pMP90) and the transformed *Agrobacteria* were used for floral dip transformation of *sop12-1 clf-50* plants. The primary transformants were selected on the basis of hygromycin resistance conferred by the pGWB1 binary vector. Because the selection in tissue culture gave small, early flowering plants it was difficult to evaluate the phenotypes of the T1 plants, so this was done in the T2 families after direct sowing of the seed on soil. Among three independent transgenic lines, full complementation of the *sop12-1 clf-50* phenotype was observed (Figure 3-6). The T2 families segregated for plants showing suppressed and non-suppressed clf phenotype. When plants were genotyped using PCR to amplify the transgene, the non-suppressed (complemented) plants all harboured the transgene, whereas those with the suppressed (sop12) phenotype did not. Taken together, the facts that independent alleles that disrupted *AT3G63270* gave a sop12 phenotype and that introduction of wild type *AT3G63270* into the *sop12-1* mutant background complemented the mutant phenotype it was concluded that *AT3G63270* corresponded to *SOP12*.

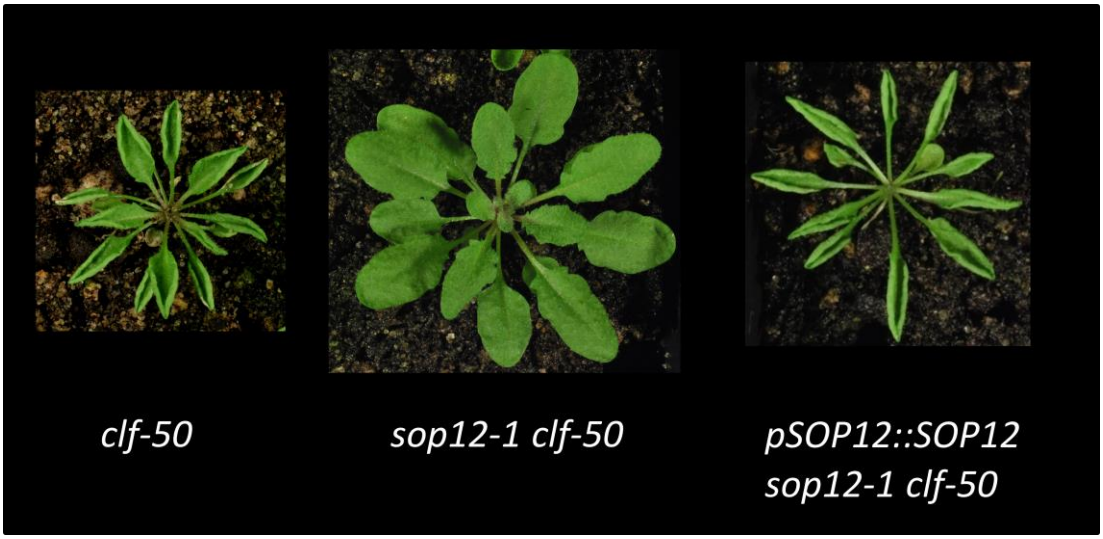


Figure 3-6. *AT3G63270* complements *sop12-1*.

The genomic sequence (including up-stream promoter) of *AT3G63270* was cloned and introduced into *sop12-clf-50*. The transgene complemented *sop12-1* mutation and restores the *clf* leaf curling phenotype in transgenic plants. Plants were grown in SD conditions, in which the suppressed phenotype is most apparent.

3.5 SOP12 encodes a novel protein with similarity to *Harbinger* transposase

Referring to TAIR, a 1,581 bp-long cDNA was predicted to be transcribed from *AT3G63270*. The cDNA consists of 75 bp 5'-untranslated region (UTR), 315 bp 3'-UTR, and 1,191 bp protein coding sequence (CDS). In comparison with genomic DNA sequence, the *SOP12* locus contains three exons. The CDS of *SOP12* was cloned into pENTR/D-TOPO® via PCR using cDNA from Col as the template followed by TOPO cloning. The cloned CDS was sequenced and the result confirmed that the sequence on TAIR is correct. When the CDS was expressed under the CaMV 35S promoter, it complemented the sop12 phenotype (Figure 3-7), which indicates that the

amino acid sequence encoded by this predicted CDS is likely to be correct and full-length.

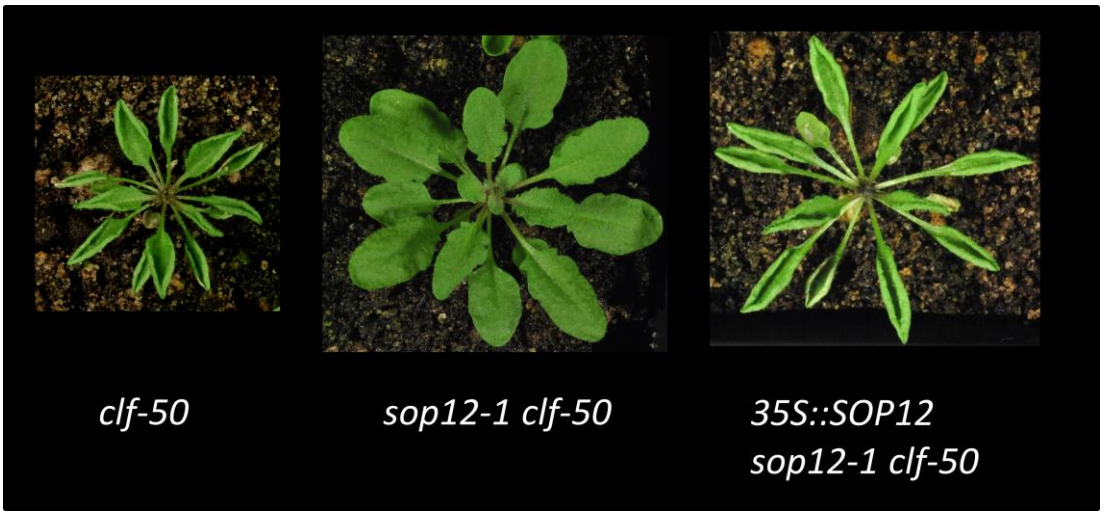


Figure 3-7. Over-expression of CDS of *AT3G63270* complements *sop12-1*.

The CDS of *AT3G63270* was cloned into Gateway® binary vector pGWB2 and introduced into *sop12-clf-50* through floral dipping. The transgene complemented *sop12-1* mutation and gives the clf leaf curling phenotype in transgenic plants as *clf-50*. Plants were grown in SD conditions, in which the sop12 clf phenotype is most apparent.

The mRNA is predicted to encode a protein of 396 amino acids (Figure 3-8). The estimated protein mass is 45.0 kD, and protein isoelectric point (pI) is around 8.0 (Isoelectric Point Calculator, http://isoelectric.ovh.org). Results from PredictProtein server (Rost et al., 2004) suggested that no nuclear localisation signal (NLS), transmembrane domain or DNA binding motif is found in SOP12, however the NLS prediction program NLStradamus (Nguyen Ba et al., 2009) identified a putative NLS from amino acid 5 to 29 (5-KQKKKNKKKPLDKAKKLAKNKEKKR–29), a region

rich in basic amino acids with similarity to bipartite NLS. When the SOP12 protein sequence was used to search the Conserved Domain Database [CDD] on NCBI, position 183 to 348 was predicted to contain a DDE domain of a DDE endonuclease superfamily (Figure 3-9).

```
1          10         20         30         40         50         60
MAPVKQKKKNKKKPLDKAKKLAKNKEKKRVNAVPLDPEAIDCDWWDTFWLRNSSPSVPSD
          70         80         90         100        110        120
EDYAFKHFFRASKTTFSYICSLVREDLISRPPSGLINIEGRLLSVEKQVAIALRRLASGD
          130        140        150        160        170        180
SQVSVGAAFGVGQSTVSQVTWRFIEALEERAKHHLRWPDSDRIEEIKSKFEEMYGLPNCC
          190        200        210        220        230        240
GAIDTTHIIMTLPAVQASDDWCDQEKNYSMFLQGVFDHEMRFLNMVTGWPGGMTVSKLLK
          250        260        270        280        290        300
FSGFFKLCENAQILDGNPKTLSQGAQIREYVVGGISYPLLPWLITPHDSDHPSDSMVAFN
          310        320        330        340        350        360
ERHEKVRSVAATAFQQLKGSWRILSKVMWRPDRRKLPSIILVCCLLHNIIIDCGDYLQED
          370        380        390    396
VPLSGHHDSGYADRYCKQTEPLGSELRGCLTEHLLR
```

Figure 3-8. Predicted amino acid sequence of SOP12

The amino acid sequence translated from predicted CDS of *SOP12*.

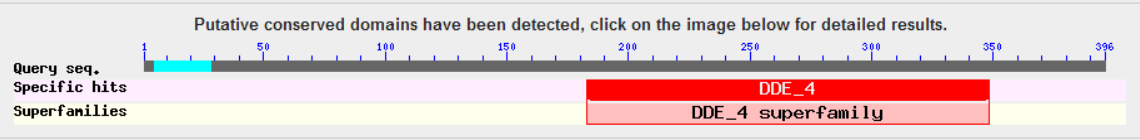


Figure 3-9. DDE containing domain is detected in the C-terminus of SOP12.

While BLAST the SOP12 amino acid sequence to discover homologs within NCBI database, the conserved domain prediction returned that SOP12 harbours a DDE domain, which is the common feature of DDE_4 superfamily.

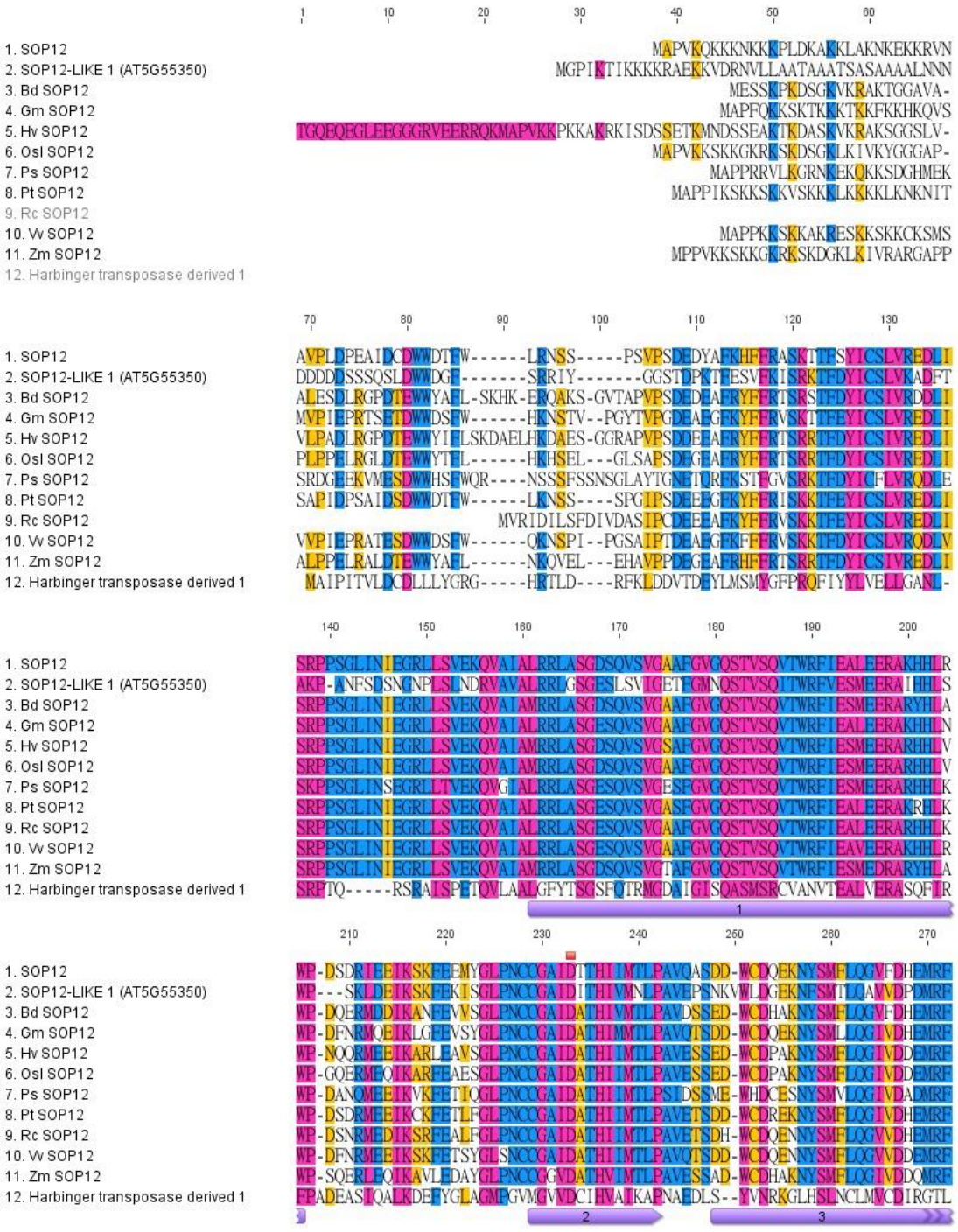
The DDE domain is a common catalytic feature of nucleases encoded by transposases. In particular, SOP12 shares similarity with Harbinger transposase-derived nucleases. Harbinger transposon is a conserved transposon superfamily (Kapitonov and Jurka,

2004). In most cases, transposases contain both a nuclease domain and a DNA binding domain, whereas in the families of Harbinger transposases, the two functional domains are encoded by two separate genes. The nuclease contains a DDE triad which is highly conserved amongst transposases and other nucleases, with four conserved aspartic acid (D) or glutamic acid (E) residues that are thought to coordinate divalent cations needed for nuclease activity (Yuan and Wessler, 2011). In the case of the Harbinger family, a fourth, highly conserved aspartic acid has been identified so the term "DDE triad" has been used to refer to a DDDE motif (Kapitonov and Jurka, 2004). Nevertheless, alignment of SOP12 with Harbinger transposase derived 1 from *Homo sapiens* (human) indicates the 4 catalytic residues in DDE domain are not conserved except for D184 in SOP12 (Figure 3-10), so it is highly unlikely that SOP12 has the nuclease activity.

When BLAST the amino acid sequence of *SOP12* to the non-redundant protein database on NCBI, the best "hits" [(60-70%) amino acid identity with the E value less than $8e^{-175}$] were with proteins, often described as Harbinger derived nucleases, from phylogenetically diverse plant species including *Glycine max* (soybean), *Vitis vinifera* (grape), *Brachypodium distachyon* (a grass), and *Picea satchenis* (a gymnosperm). The high percent identity (>60%) implies those proteins are possibly the orthologues of SOP12 in different plant species, expanding from gymnosperm to monocotyledon and dicotyledon. Importantly, when these putative orthologous protein sequences were BLAST back against the *Arabidopsis* proteome, SOP12 was the top hit – in other words, SOP12 is more similar to the SOP12 orthologues in other plants than it is to any

other harbinger related protein in *Arabidopsis*. Putative HARBI1-like proteins from animals, for example *Strongylocentrotus purpuratus* (purple sea urchin), were also identified but showed much lower similarity (30% identity).

The alignment of various plant SOP12 orthologues with human Harbinger transposase derived 1 (Figure 3-10) showed they, except for *Picea sitchensis*, do not retain all the conserved DDE triad residues. In addition to the D in domain1, monocots (barley, rice and corn, etc.) have the D within the consensus YVNGDXXYP in domain 4, but eudicot species do not (Figure 3-10). This is a piece of evidence that SOP12 has lost nuclease activity on the way of evolution.


1. SOP12
2. SOP12-LIKE 1 (AT5G55350)
3. Bd SOP12
4. Gm SOP12
5. Hv SOP12
6. Osl SOP12
7. Ps SOP12
8. Pt SOP12
9. Rc SOP12
10. Vv SOP12
11. Zm SOP12
12. Harbinger transposase derived 1
1
2
3

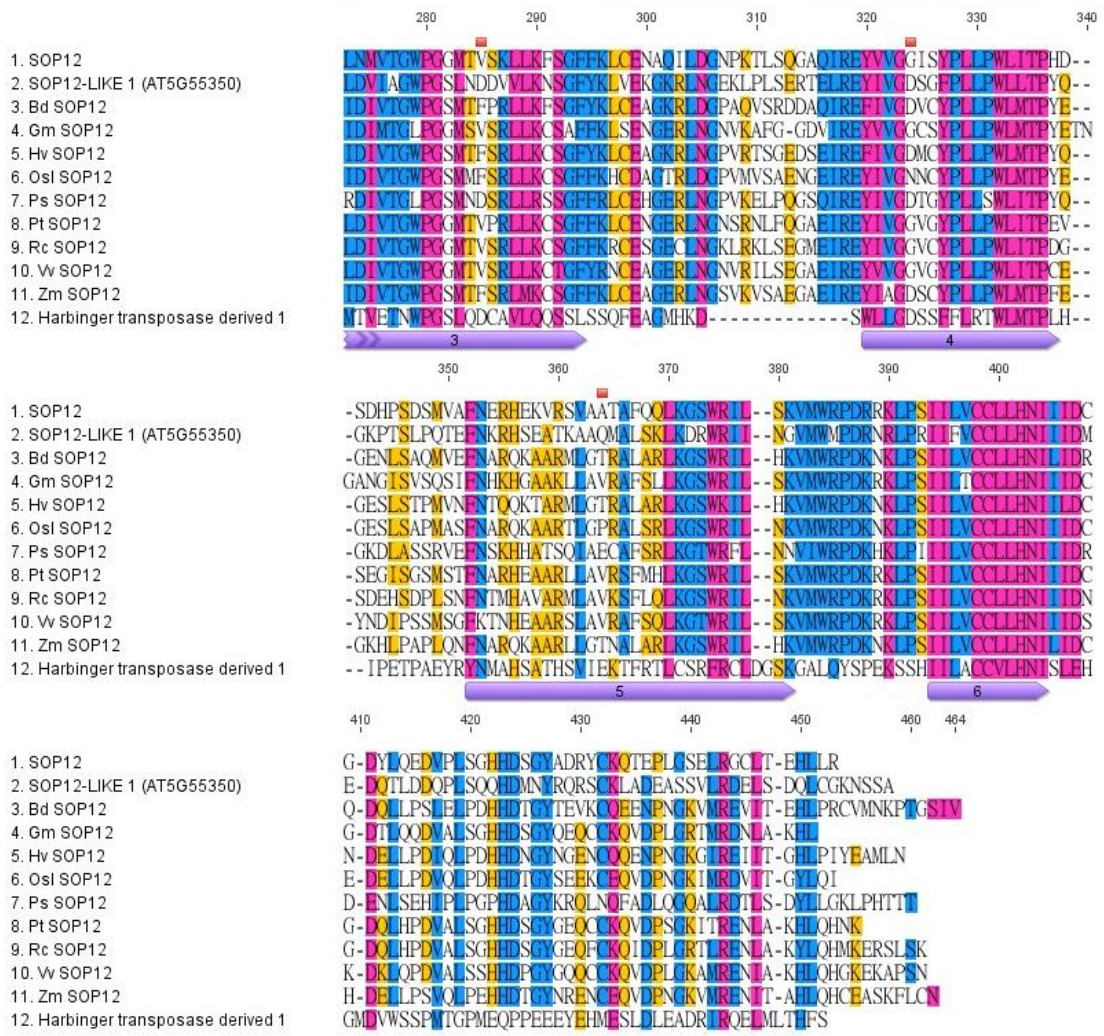


Figure 3-10. Alignment of plant SOP12 orthologues and human HARBI1. (continue)

The amino acid sequences of genes similar to SOP12 in a variety of plant species (3, *Brachypodium distachyon*; 4, *Glycine max*; 5, *Hordeum vulgare*; 6, *Oryza sativa* [rice, Indica group]; 7, *Picea sitchensis*; 8, *Populus trichocarpa*; 9, *Ricinus communis*; 10, *Vitis vinifera*; 11, Zea mays) were aligned with SOP12 (1), SOP12-LIKE1 (AT3G55350, 2), and Harbinger transposase derived 1 from *Homo sapiens* (human, 12) as a reference. Six conserved domain and four catalytic residues of DDE motif are pointed by purple bars and red rectangles, respectively, according to Kapitonov and Jurka (2004). Only one of the catalytic residues is conserved in SOP12 (D184 in SOP12, which is D241 in the alignment) but others are not, which indicates it is likely that SOP12 has lost the nuclease activity. The alignment

was done by Clustal W method using default parameters in Geneious® program. Amino acids shaded in pink are 100% conserved between species, whereas the conservation degrees between 80%~99%, 60%~80% and less than 60% are shaded by blue, orange, and white, respectively.

There are four genes annotated as putative Harbinger transposase derived nucleases in *Arabidopsis* according to TAIR database, including *SOP12*, *AT2G13770*, *AT5G35695*, and *AT5G41980*. Additionally, *AT3G55350* (designated as *SOP12-LIKE1*) and *AT3G19120* are putative P instability factor (PIF)/Ping-Pong family of plant transposases. PIF/Ping-Pong and Harbinger transposases belong to the same PIF/Harbinger superfamily. These 6 genes share low amino acid identity with each other, ranging from 10% to 25%, except for the case of SOP12 with SOP12-LIKE1 (45.9%) (Table 3-1). Even within the DDE domain, those six genes are more diverse compared to the alignment between SOP12 and its homologues in other plant species (data not shown).

Table 3-1. The amino acid sequence identity between putative harbinger transposase-derived nuclease in *Arabidopsis*.

% identity	SOP12	SOP12-LIKE1	AT2G13770	AT3G19120	AT5G41980	AT5G35695
SOP12		45.9	9.8	19.8	15.9	12.7
SOP12-LIKE1	45.9		10.7	19.9	17	13.2
AT2G13770	9.8	10.7		10.4	11.3	11.3
AT3G19120	19.8	19.9	10.4		12.7	13.5
AT5G41980	15.9	17	11.3	12.7		23.4
AT5G35695	12.7	13.2	11.3	13.5	23.4	

3.6 *SOP12* does not show functional redundancy with its close relative *SOP12-LIKE1*

The fact that *sop12* only partially suppressed *clf*, and also that *sop12* single mutant had near wild type phenotype (Figure 3-4), raised the question whether *SOP12* has redundancy with related genes in *Arabidopsis*. Since *SOP12-LIKE1* seems to be the closest gene of SOP12, I tested whether the double mutant *sop12 sop12-like1* shows more severe phenotypes and if *sop12-like1* suppresses clf phenotype. Neither did *sop12-1 sop12-like1* double mutant give a dramatic phenotype (Figure 3-10), nor did *sop12-like1* suppress clf phenotype (data not shown). The abnormal, oval leaf shape is more likely due to mix background of the double mutant because it did not segregate with *sop12-like1* mutation.

Figure 3-11. The plant *sop12-1 sop12-like1*.

This double mutant was generated via crossing of *sop12-1* (*Ws* background) and *sop12-like1*-2 (SALK_122829, Col). The plant showed wild type-like morphology with oval-shaped leaves which is possibly due to the mix-background effect rather than mutations. This was a 40 day-old plant grown in LD condition and the inflorescence stem was removed. The scale bar shown represents 1 cm.

3.7 Summary and discussion

In this chapter, it is shown that *AT3G63270* corresponds to *SOP12*. First of all, the T-DNA insertion in *sop12-1* disrupts *AT3G63270,* but does not cause deletion or rearrangement of adjacent loci. Second, the suppression of the clf phenotype co-segregated with the T-DNA insertion at *sop12-1* as a recessive trait, and expression analysis shows that the insertion at *sop12-1* does not elevate the expression of *AT3G63260* and *AT3G63280* (the genes flanking *AT3G63270*), showing that loss of *AT3G63270* function is the cause of the sop12 phenotype. Third, the independent allele *sop12-3* harbouring a *Ds* insertion in the third exon of *AT3G63270* also suppressed *clf-2*, which is consistent with the conclusion from *sop12-1*. Additionally, this result indicated that the suppression is independent of ecotypic background. Last, the re-introduction of *AT3G63270* locus back into *sop12-1 clf-50* restored the clf phenotype, i.e., *AT3G63270* complements the *sop12-1* mutation. Taken together, these results show that *SOP12* corresponds to *AT3G63270*.

SOP12 may be a nuclear protein. Although PredictProtein Server (Rost et al., 2004) did not identify a NLS in SOP12, another annotation server NLStradamus (Nguyen Ba

et al., 2009) discovered a putative NLS in the N terminus (position 5-29) of SOP12. The high lysine content in SOP12 indeed resembles the consensus sequence of a typical bipartite NLS which is KR[PAATKKAGQA]KKKK (Dingwall et al., 1988). During the latter part of this study, the SOP12 protein was independently identified from characterisation of a mutant suppressing of the Pc-G mutant *like hetrochromatin 1* (*lhp1*) (Hartwig et al, 2012). This study also predicted a helix-loop-helix (HLH) domain in the SOP12 N-terminus (position 110-141), which might contribute to DNA binding activity, based on searches of the CDD database (Hartwig et al., 2012). However, the similarity of SOP12 N-terminus to HLH domain is relatively weak with a high E value ($5.05e^{-3}$), and SOP12 was not identified as a DNA binding proteins in BLAST searches, so it is unclear whether that SOP12 is likely to bind to DNA. It needs to be confirmed whether SOP12 is a DNA-binding protein by experimental DNA binding assay, like electrophoretic mobility shift assay.

To understand more about SOP12 in the first instance, a bioinformatic strategy was applied. The amino acid sequence of SOP12 was used in Blast searches to identify related proteins in other species and potentially to shed light on SOP12 function. In the outset of this project, SOP12 was an unknown protein with no functional domain, so the function of the possible orthologs would provide clues for prospective functions of SOP12. The BLAST results indicated SOP12 is derived from PIF/Harbinger family transposon nuclease. Transposons, also called transposable elements (TE), are "jumping genes" in genome. When a TE inserts at a new site in the genome through transposition, it may interrupt a gene's structure and cause mutations, so normally the

transposition of TE is inhibited by the host, for example by silencing of transposase genes by siRNA mediated transcriptional or post transcriptional silencing pathways (Levin and Moran, 2011). The transposition of TE requires transposases, which have two functions: DNA binding and nuclease activities. The DNA binding activity is to recognise terminal inverted repeat (TIR) sequences found at the termini of cut and paste type transposons, and then the nuclease cuts DNA to release TE from genome, and ligates it to target site (Levin and Moran, 2011). In general, two types activities are allocated to different domains in one protein (transposase), except for transposases in PIF/Harbinger transposase superfamily. PIF/Harbinger transposase superfamily is an ancient group of transposases found in diatom, insects, vertebrates and plants (Kapitonov and Jurka, 2004). In this superfamily, the DNA binding and nuclease activities are encoded by two distinct proteins. In mammals, Harbinger transposases have domesticated to transposition-inactive nuclease HARBI1 (Kapitonov and Jurka, 2004). HARBI1 has retained characteristics of Harbinger nucleases, for instance, the conserved catalytic DDE triad, probably the nuclease activity and the interaction with DNA binding protein (Sinzelle et al., 2008). Among the four D/E catalytic residues in the DDE domain, only one out of four is still conserved in SOP12 (Figure 3-10). Therefore this suggests that SOP12 may be a domesticated transposase that is derived from a PIF/Harbinger nuclease but has lost nuclease activity and acquired a novel role in gene regulation, even though many residues other than the catalytic ones within the DDE domain are conserved in SOP12. This implies that the suppression of *clf* by *sop12* is possibly not associated with the mechanism of TE transposition. The

similarity of SOP12 to Harbinger nuclease gives a hint that SOP12 might not bind to DNA directly but interact with a DNA binding protein.

What is the correlation between transposase and Pc-G or trx-G regulation? An interesting transposase called Metnase, which contains a SET domain delivering DNA binding activity (Lee et al., 2005), caught my attention. Many Pc-G and trx-G members are SET domain containing proteins, and SOP12, a nuclease in transposase, may somehow participate in Pc-G or trx-G regulation pathway. Taken together, it raised the question whether SOP12 interacts with a SET domain protein to form a structure like Metnase but the complex has a distinct function, such as gene activation rather than gene transposition. Unfortunately, SOP12 and the transposase domain of Metnase are very diverse, although they both have DDE domains. It looks unlikely SOP12 participate in a Metnase-like complex, but this does not rule out the possibility that SOP12 interacts with SET domain containing proteins. How a Harbinger nuclease-like protein is involved in Pc-G and trx-G regulation remains mysterious.

SOP12 protein sequence was also BLAST against the *Arabidopsis* protein database (TAIR) to find SOP12 homologues, or gene family. The BLAST result revealed a closely related gene, *SOP12-LIKE1*, which is also a putative transposase in PIF/Harbinger superfamily. This result is concomitant with the result of BLAST on NCBI that SOP12 is a Harbinger nuclease-like protein. However, *SOP12* and *SOP12-LIKE1* did not show genetic interaction or redundancy and *sop12-like1* does not suppress clf phenotypes, which illustrates the two genes have quite different roles

in plants. In addition to *SOP12-LIKE1*, other *Arabidopsis* genes on the BLAST list shared low identity with SOP12, so are less likely to form a gene family with SOP12. Those observations support that SOP12 is a single copy gene, as also suggested by detailed phylogenetic comparisons in an independent study (Hartwig et al., 2012).

In *Arabidopsis*, the Harbinger nuclease activity might be contributed by five putative PIF/Harbinger transposase derived nucleases (*AT3G55350*/*SOP12-LIKE1*, *AT3G19120*, *AT2G13770*, *AT5G35695*, and *AT5G41980*). Their amino acid identity to, for example, human HARBI1 is lower than SOP12 to human HARBI1----the identity of SOP12-LIKE1 to human HARBI1 is 20.3%, the other four to human HARBI1 is between 11.5%-16.8%, whereas SOP12 to human HARBI1 is 21.3%. Even though the disparity in the percentage is minor, those five preserve more catalytic residues in the DDE motif. All four residues of the (D)DDE triad are present in AT2G13770; the first three D/E are conserved in SOP12-LIKE1, AT3G19120, and AT5G41980; the last three D/E could be found in AT5G35695. In spite of the bioinformatic prediction, which suggests that one or more of these other proteins are more likely to provide Harbinger nuclease activity, whether there is Harbinger transposase derived nuclease activity in *Arabidopsis* yet needs to be verified by experiments.

Chapter 4. Phenotypic and genotypic characterisation of *SOP12*

The genetic identification of *SOP12* as a suppressor of a Pc-G mutant raised the possibility that it was a trx-G member. It may be an activator of Pc-G target genes via acting on chromatin. However, in our lab the previous analysis of mutants suppressing the clf phenotype indicated that many suppressors were extremely late flowering and represented transcription factors or RNA binding factors required to activate targets such as *FPA* in leaves (Lopez-Vernaza et al., 2012). To further characterise the role of *SOP12*, I therefore characterised the *sop12* mutant phenotype in the wild type background, and tested its effects on expression of *CLF* target genes. Furthermore, it is curious if the suppression is *clf*-specific, and if *sop12* is able to enhance trx-G mutants. Hence its genetic interactions with other Pc-G and trx-G members were characterised

4.1 *sop12* mutant is slightly late flowering in short day (SD) condition

The *sop12-1* mutation was originally isolated in a *clf-50* mutant background that also carried the conditional, steroid dependent *CLF:GR* transgene. The mutant was backcrossed to the wild type progenitor (*Ws* ecotype) and the resulting F2 and F3 generations were genotyped to identify *sop12* plants that had lost the transgene in the presence of wild type allele CLF. Meanwhile, *sop12-1 clf-50* double mutants that lacked *CLF:GR* were isolated, to ensure that the suppression did not require the transgene. The *sop12-3* mutation was obtained from stock centres in a wild-type (L*er*) background and was crossed to *clf-2*, the *clf* allele in L*er* background. The work

mentioned above was done by Dr. J. Goodrich before I started the project. The *sop12* mutants in *clf* and wild type backgrounds were then grown alongside with their progenitors in LD and SD to compare morphology and flowering time. Results showed that *sop12* single mutants did not have major effects on gross plant morphology, nor were they extremely late flowering like the *fpa*, and *ft* mutants which were previously found to suppress *clf (Lopez-Vernaza et al., 2012)*. However, *sop12* single mutants showed a mild late flowering phenotype in SD. (Table 4-1 and Figure 4-1, upper panel). *sop12-1* plants flowered at 73.8±0.7 days-after-growth (DAG), whereas *Ws* plants flower 64.3±1.0 DAG. The same pattern was obtained from the experiment in L*er* background: in *sop12-3*, inflorescence stems emerged at 80.3± 1.3 days, while in L*er* it was 71.8±0.9 days. The flowering time differences between *sop12* and wild type were very minor in LD (Table 4-1 and Figure. 4-1, lower panel). For example, the distribution of the flowering time in wild type and *sop12* was overlapped, although the statistical tests (one way ANOVA) indicate there were differences. It is difficult to identify *sop12* from wild type plants according to flowering time, unlike suppressors such as *fpa* or *ft*, which are very late flowering in LD.

Comparison of the effects of *sop12* in *clf* mutant backgrounds indicated that the mutants were intermediate between wild type and *clf*. In other words, *sop12* gave a partial suppression of the clf early flowering phenotype. Thus *clf-50* flowered at 39.7±0.4 days, much earlier than *Ws*, consistent with previous studies (Goodrich et al., 1997) whereas *sop12-1 clf-50*, flowered at 54±0.8 days, later than *clf-50* but still

earlier than *Ws*. A partial suppression in *sop12-3 clf-2* was also observed. *sop12-3 clf-2* flowered 10 days later than *clf-2* (48.9±0.5 days in comparison with 38.1±0.5 days) but significantly earlier than L*er*.

Table 4-1. Flowering time of *sop12* mutants in LD and SD conditions.

Condition/ Genotype	LD		SD	
	DAG	SE	DAG	SE
Ws	23.6	0.4	64.3	1.0
clf-50	21.7	0.2	39.7	0.4
sop12-1	25.7	0.2	73.8	0.7
sop12-1 clf-50	22.1	0.2	54.0	0.8
Ler	27.1	0.2	71.8	0.9
clf-2	24.4	0.1	38.1	0.5
sop12-3	29.2	0.5	80.3	1.3
sop12-3 clf-2	25.7	0.2	48.9	0.5

Flowering time was defined as the time interval (days after growth, DAG) from sowing to when the inflorescence stem of each plant was 1 cm-high. The mean of at least 12 plants and the standard error of the mean (SE) were shown.

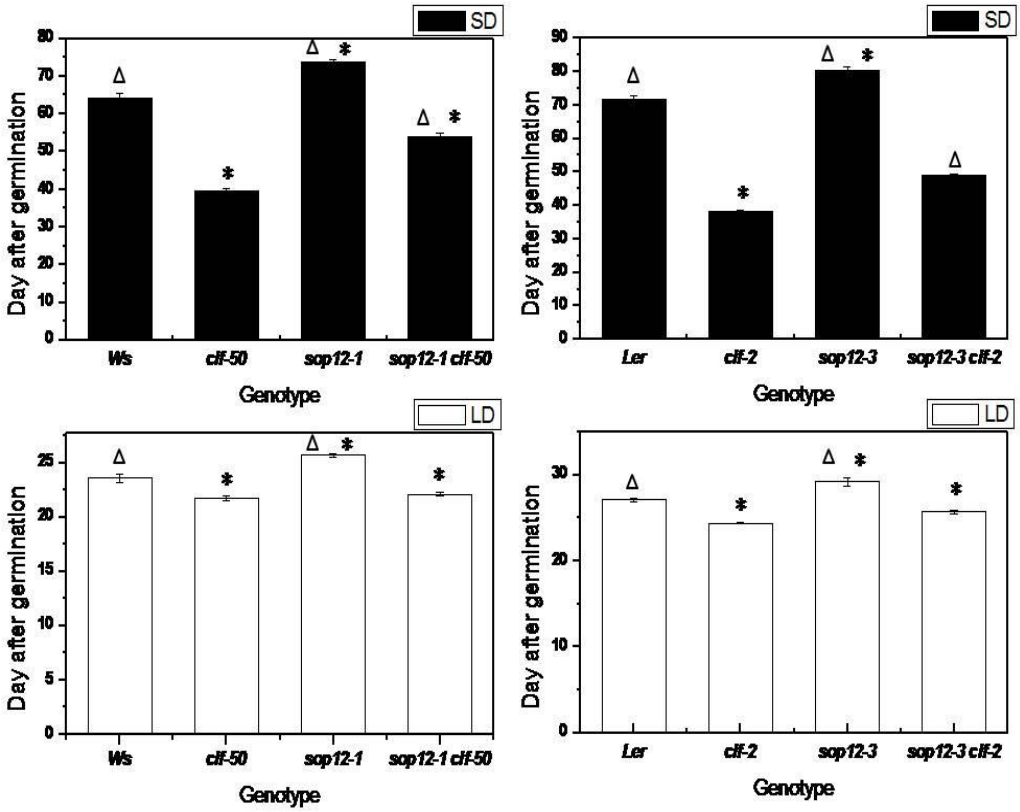


Figure 4-1. Flowering time of *sop12* mutations in LD and SD conditions.

Bar diagram representing the data in Table 2. The black columns in upper panel are data collected in SD condition from all genotypes, and the open columns in lower one show LD data. On average, *sop12* mutations delayed the flowering time by about 9 days in both backgrounds in SD but only contributed a 1-day difference in LD. Error bars are shown as ±1 SE. The stars (*) address the significant differences in the data were supported by ANOVA while comparing to wild type of corresponding background under the same condition with the p value less than 0.001. The empty triangles indicate the significant difference while comparing to *clf* ($p<0.001$).

4.2 *SOP12* over-expression is somehow prohibited

In addition to observing the flowering time phenotype of loss of function mutants, gain of function transgenic lines (*35S::SOP12* in *Ws* background) were also examined under SD. Surprisingly, four independent transgenic lines showed the late-flowering phenotype relative to wild type (data not shown). When *SOP12* expression level was checked via real time RT-PCR, it was actually decreased rather than increased (Figure 4-2). I did not manage to obtain constantly overexpression (*35S::SOP12/Ws*) transgenic lines in the end. It is less likely that the construct is not properly expressed, since the identical construct complements *sop12* (Figure 3-7), even though it does not contain regulatory sequences (5'- and 3'-UTR). It is suspicious maybe there is strict post-transcriptional regulation to control *SOP12* level in *planta*, because overexpression *SOP12* abates the viability of plants, for example, to make transgenic plants more sensitive to hygromycin, the antibiotics used for select plants harbouring T-DNA, so transgenic lines which highly express SOP12 are rare.

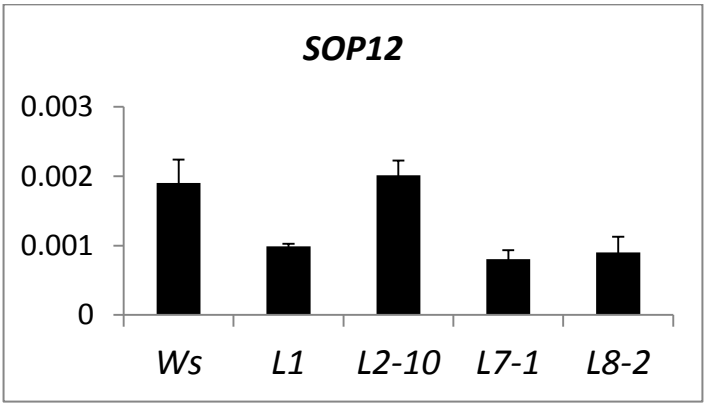


Figure 4-2 *SOP12* expression in putative *SOP12*-overexpressed transgenic lines

The figure shows the real time RT-PCR analysis of SOP12 mRNA levels relative to *eIF4A*, the internal loading standard, in four independent *35S::SOP12* transgenic lines. RNA was extracted from 14-day-old seedlings. Three biological replicates were included. The error bars represent 1 SE. These four transgenic lines do not accumulate more SOP12 mRNA compared to wild type plants (*Ws*).

4.3 *sop12* behaves similar to wild type-like under heat stress

In order to discover phenotypes of *sop12* single mutants, the response of *sop12* at different temperatures was tested. The rationale was that when *sop12-1* was grown in different conditions by Dr. J. Goodrich at CSIRO in Australia, *sop12-1* showed a mild phenotype and mutants could be distinguished from wild type as the rosette was slightly smaller and leaves had a weak down-ward curling along their margins. To test whether this was an effect of temperature, *Ws*, *sop12-1*, L*er* and *sop12-3* were grown at 21^{o}C, 25^{o}C and 30^{o}C under LD conditions. Unfortunately, the same result was not retrieved in our growth room----the leaf morphology of *sop12* was indistinguishable from the one of wild type plants (Figure 4-3).

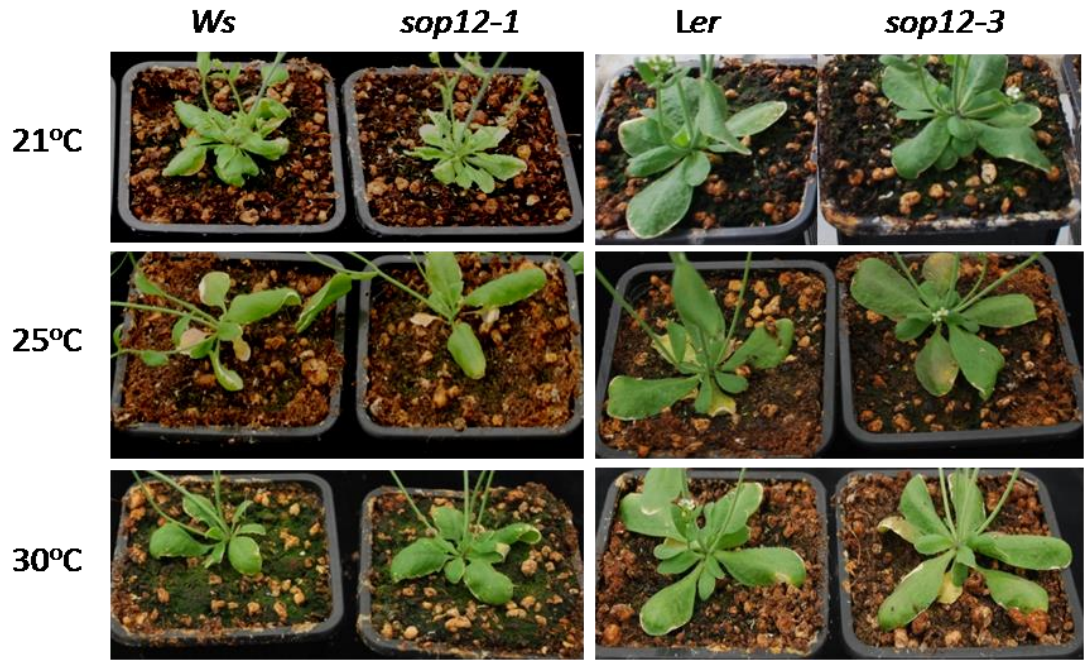


Figure 4-3. The leaf morphology of wild type and *sop12* plants under heat stress.

Plants were grown at different temperature (21°C, 25°C and 30°C) in the LD condition. 21°C is the normal growth temperature, 25°C is intermediate and 30°C is the condition of heat stress. All the seeds were germinated at 21°C before transferred to other growth temperature. 7 plants of each genotype were grown and their responses were concomitant with what was shown in this figure.

4.4 SOP12 is required for the mis-expression of Pc-G target genes in *clf*.

Mutations affecting key target genes of *CLF*, for example *ag* or *sep3* mutants, are known to suppress clf phenotype (Goodrich et al., 1997; Lopez-Vernaza et al., 2012). This raised the question whether *SOP12* was required for the activity of *CLF* target genes, or rather was itself a target gene, or acted downstream of targets like *AG*. To address this, the expression level of known Pc-G targets was examined by real time RT-PCR analysis. If *SOP12* is a Pc-G target or acts downstream, the expression level of other Pc-G targets in *sop12-1 clf-50* is expected to be mis-expressed at similar levels as in *clf*. On the other hand, if *SOP12* acts upstream, then expression of target genes should be reduced. The examination was first done in *Ws*, *clf-50*, and *sop12-1 clf-50* with one sample from each genotype by Dr. J. Goodrich. I repeated the same experiment with three biological replicates, and included *sop12-1* and all 4 genotypes in the L*er* background. Consistent with previous studies, target genes such as *FLC*, *FT*, *SEP3*, and *AG* were mis-expressed in *clf* (Goodrich et al., 1997; Jiang et al., 2008) (Figure 4-4). This mis-expression was down-regulated in *sop12-1 clf-50*. The level of down-regulation was not consistent in all genes---the down-regulation was most obvious in *SEP3*, which decreased 93.3% from *clf-50* to *sop12-1 clf-50*. By comparison, *AG* showed less decrease, about 30%. By comparison, between *Ws* and *sop12-1* single mutants the expression of those four genes did not change much and the pattern is not consistent. In *sop12-1*, *FLC*, *SEP3*, and *AG* decreased and *F*T increased, all in small scales (Figure 4-4). The higher amount of *FT* may reflect the decrease in *FLC*, because *FLC* represses *FT* expression (Chris et al., 2006; Searle et al., 2006).

In the comparison of *sop12-3* and *sop12-3 clf-2* in the L*er* background, the reduction of mis-expression was accordant with the results of *sop12-1* and *sop12-1 clf-50* as all four genes showed reduced mis-expression (Figure 4-5), and the pattern of relative expression when comparing L*er* and *sop12-3* also agreed with the results from *Ws* ecotype. *FT* was expressed lower in L*er*, and *FLC*, *SEP3* and *AG* were higher in *sop12-3* (Figure 4-5)

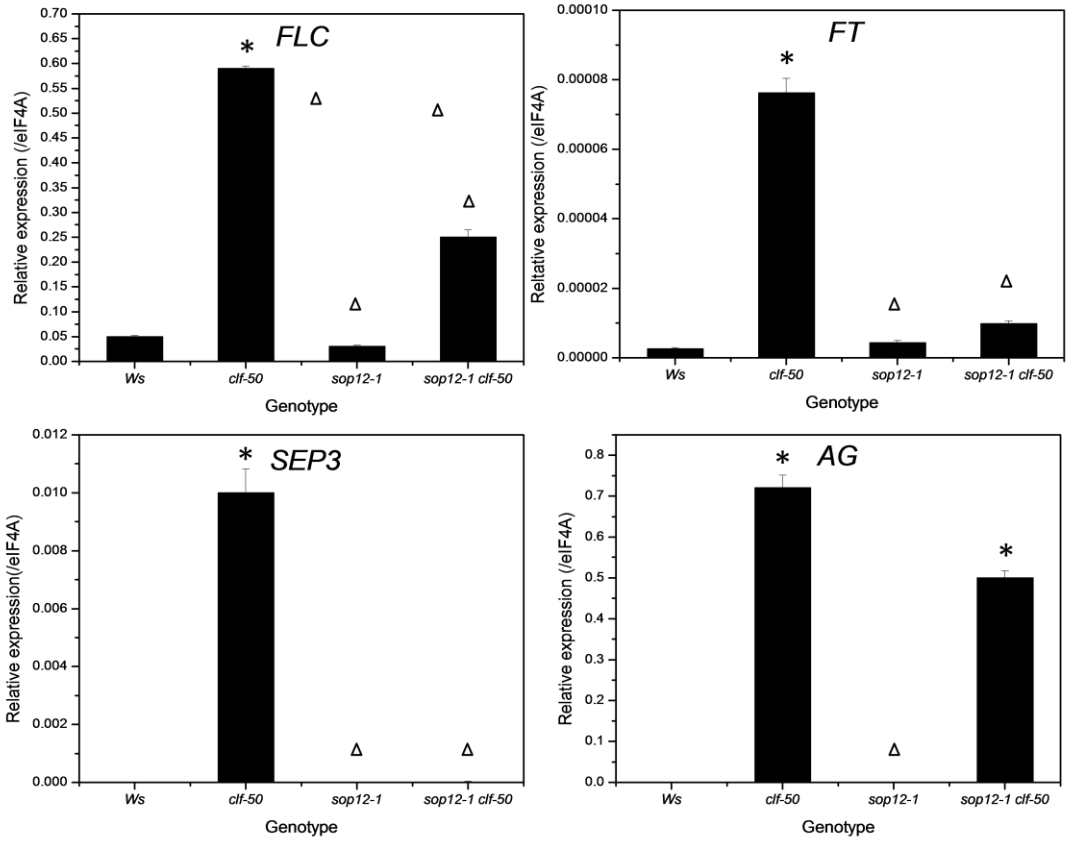


Figure 4-4. The expression of Pc-G targets in *sop12-1 clf-50*.

This is the real time RT-PCR analysis of four Pc-G targets from 14-day-old seedlings of four genotypes (the wild type *Ws*, *clf-50*, *sop12-1*, and *sop12-1 clf-50*) grown in SD. These four Pc-G targets were decreased in the expression level in *sop12-1 clf-50* compared to *clf-50*. Plant samples were harvested about six to eight hr after dawn. Expression level is shown relative to *eIF4A*, a standard which is expressed at high levels and constitutively. The error bars represent 1 SE. Stars and triangles represent significantly difference compared with *Ws* and *clf-50*, respectively ($p<0.005$).

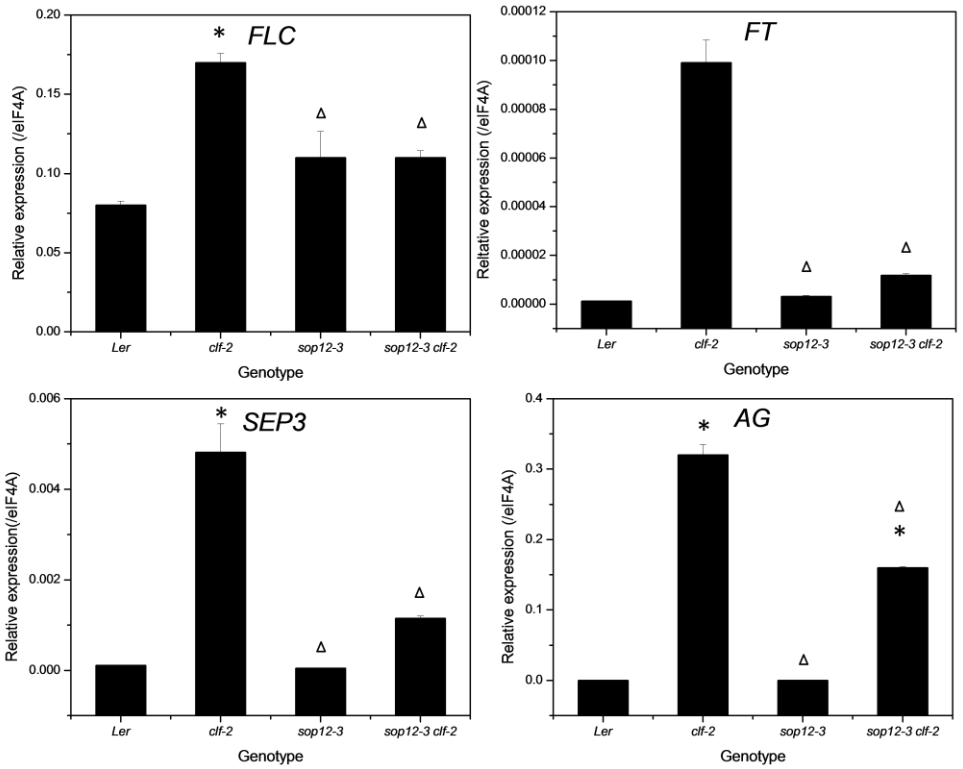


Figure 4-5. The expression of Pc-G targets in *sop12-3* mutant backgrounds.

The real time RT-PCR analysis of four Pc-G targets from 14-day-old seedlings of four genotypes (the wild type L*er*, *clf-2*, *sop12-3*, and *sop12-3 clf-2*) in SD. Plant samples were harvested about six to eight hr after dawn. Relative expression level with respect to *eIF4A* reference gene is shown. The error bars represent 1 SE. Four Pc-G targets were significantly

decreased in the expression level in *sop12-3 clf-2* compared to *clf-2*. Stars and triangles represent significantly difference compared with *Ws* and *clf-50*, respectively ($p<0.005$).

4.5 The genetic interaction of *sop12* with Pc-G mutants

Because *sop12* suppresses *clf*, it is interesting to know whether *sop12* suppresses other Pc-G mutants. In addition, the genetic interaction would give us a hint about the role of SOP12 in the Pc-G working context. The mutants tested included *emf2*, *vrn2*, *mea* (belonging to EMF-PRC2, VRN-PRC2 and FIS-PRC2, respectively), *emf1*, and *lhp1* (belonging to PRC1). Those mutants were crossed with *sop12* and the double mutants were revealed in F2 or F3 population. In general, *sop12* is hypostatic to most of the mutants except for *lhp1*.

Both the strong and weak alleles of *emf2* are epistatic to *sop12*

EMF2 is one of the three *Arabidopsis* homologs of *Drosophila* SU(Z)12, a core component of the PRC2 (Yoshida et al., 2001). It participate in EMF2-PRC2 (Jiang et al., 2008) involved in delaying flowering and repression of Pc-G targets like *AG* and *AP3* (Chanvivattana et al., 2004). In *emf2*, *AG* and *AP3* are de-repressed and H3K27me3 is reduced (Calonje et al., 2008). There are two alleles of *emf2* mutants in our lab: a strong allele, *emf2-3* in Col ecotype, which gives plants that skip the vegetative phase after germination (Yang et al., 1995) (Figure 4-6), and a weaker allele, *emf2-10* in *Ws* background, which displays *clf*-like phenotypes (Chanvivattana et al., 2004). *emf2-3* is caused by a 35-bp deletion in the twentieth exon. The phenotypes are

so severe that *emf2-3* plants could only survive on sterile tissue culture plates and are absolutely sterile. The weak allele, *emf2-10* is able to survive on soil to produce a small, partially fertile plant. The morphology of *emf2-10* resembles *clf-50* with up curled leaves (Figure 4-6), and early flowering but is more extreme so plants are much smaller and less fertile (Chanvivattana et al., 2004).

The phenotypes of double mutants of *sop12* with either strong allele or weak allele were similar to the *emf2* single (Figure 4-6). *sop12-1 emf2-3* was late germinating and formed "embryonic flowers" after germination on MS plates (Figure 4-6). No *emf2-3 sop12-1* plants were found when seeds of *sop12-1 emf2-3/+* were sown directly on soil, in other words *sop12-1* did not rescue the lethality of *emf2-3* on soil. *sop12-1 emf2-10* doubles did survive on soil, like *emf2-10* single mutants, however no morphological differences were observed between the two (Figure 4-6), either in SD or LD conditions.

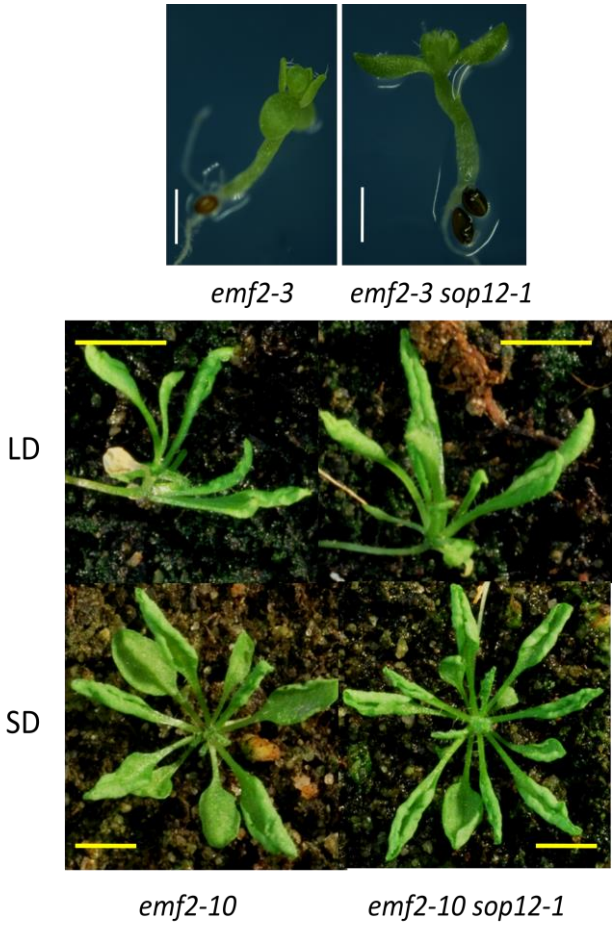


Figure 4-6 *emf2* is epistatic to *sop12*

emf2-3 and *emf2-3 sop12-1* (upper panel) were both late-germinating and developed flowers directly after opening cotyledons. *emf2-10* and *emf2-10 sop12-1* (middle and lower panels) exhibit clf-like phenotypes. *sop12* does not show suppression effects on *emf2*. The double mutants are identical to the strong and weak *emf2* single mutants. The white scale bars represent 1 mm and yellow ones are 1 cm. *emf2-3* and *emf2-3 sop12-1* (upper panel) were grown on MS plates under LD condition, whereas *emf2-10* and *emf2-10 sop12-1* (middle and

lower panels) were grown on SD or LD for 40 days on soil. The inflorescence stems were removed before photos were taken.

sop12 does not interact with *mea*

According to microarray database in eFP browser, the highest level of *SOP12* expression is in seeds (Winter et al., 2007), so it is worth to test if *sop12* suppresses *mea*. MEA, like CLF, shares homology with E(z) (Grossniklaus et al., 1998), and acts as the catalytic component in seed-specific FIS-PRC2 (Köhler et al., 2003b). Mutations in *MEA* cause over-expression of type I MADS-box gene *PHERES1* (*PHE1*) after fertilisation (Köhler et al., 2003a). *MEA* is imprinted, such that only the maternally derived allele is expressed in the seed. Consequently, all seed derived from a *mea* mother are mutants, regardless of the paternal genotype. After fertilization, the embryos inheriting a maternal *mea* allele develop until heart stage and then arrest. The immature seeds turn white and then brown, and become aborted, shrivelled seeds at maturity (Grossniklaus et al., 1998).

The double mutant *sop12-1 mea* was made by crossing *sop12* as female with *mea*/+ as male and isolating *sop12-1* homozygotes that were *mea*/+ in the F2. Self-pollinated progeny from *mea*/+ bear about 50% shrivelled seeds and 50% healthy seeds (42% : 46%, aborted : normal seeds) (Grossniklaus et al., 1998). The ratio of shrivelled seeds to healthy seeds from self-pollinated *sop12-1 mea*/+ was scored. 1778 (53%) among the 3300 progeny of *sop12-1 mea*/+ were shrivelled. This infers that *sop12-1* does not suppress the seed abortion phenotype of *mea* mutants. It cannot be excluded that *sop12*

may give a weak suppression in other aspects. For example, some weak modifiers can allow an increased transmission of mutant *mea* alleles maternally, due to an increased viability of the shrivelled seed. Due to time constraints I was not able to test this.

The *emf1* mutant is largely epistatic to *sop12*

EMF1 is a novel, plant specific Pc-G gene. No homolog has been identified in animal models although it has recently been suggested that EMF1 is functionally analagous to *Drosophila* Psc with a proposed role in chromatin compaction (Beh et al., 2012). EMF1 was assigned to have PRC1-like roles because it is epistatic to the PRC2 mutant *emf2*, and works downstream of it. For instance, H3K27me3 levels at selected targets were less affected in *emf1* than in *emf2* mutants (Calonje et al., 2008). However, recent global epigenomic profiling studies have complicated this interpretation as many Pc-G targets which are mis-expressed in *emf1* show dramatically reduced H3K27me3, more consistent with a PRC2-like role (Kim et al., 2012). Phenotypes of *emf1* are close to severe *emf2*, for instance, no vegetative stages, development of embryonic flower after germination, completely sterile, over-expression of Pc-G targets and reduction of H3K27m3 (Calonje et al., 2008). EMF1 is a DNA binding protein, albeit without sequence specificity, and it binds to Pc-G targets in ChIP assays (Calonje et al., 2008). Its role in the mechanism of Pc-G repression remains to be fully discovered.

The allele *emf1-1* (Col) harbours a non-sense mutation, the point mutation causing a pre-mature stop codon in the first exon. The double mutant *emf1-1 sop12-1* was

slightly different to *emf1-1*----more leaf-like, or sepal-like structures were formed (Figure 4-7). However, it is not clear whether this is due to effects of the mixed genetic backgrounds (Col and *Ws*) of the parents. The *emf2-3 sop12-1* double was also isolated in a Ws/Col mixed background, but no extra leaf-like structure was observed (Figure 4-6). It is possible therefore that *sop12-1* causes a very weak suppression of *emf1-1*.

Figure 4-7. Leaf-like structure in *emf1-1 sop12-1*

After germination, *emf1-1* opened cotyledons and then produced flower-like structure from shoot apical meristem. In *emf1-1 sop12-1*, similar petiole-less cotyledons were produced and in addition to embryonic flower, more leaf-like tissues were emerged. Both mutants were grown on MS plates. The white scale bars represent 1 mm.

sop12-1 suppresses *like heterochromatin 1*(*lhp1*)

LHP1 is the homologue of HP1 in animal models (Gaudin et al., 2001), but the sequence similarity is rather weak and the function of LHP1 is not associated with heterochromatin. The major role of LHP1 in the Pc-G regulation pathway is to

recognise the repressive mark H3K27me3 by its chromodomain (Turck et al., 2007; Zhang et al., 2007; Exner et al., 2009), which mimics the role of Pc in PRC1 of *Drosophila* (Lin Xu, 2008; Bratzel et al., 2010). The recognition of H3K27me3 is essential for repression (Exner et al., 2009)----in *lhp1*, H3K27me3 levels do not change significantly (Turck et al., 2007) but Pc-G targets are de-repressed (Kotake et al., 2003). *lhp1* resembles *clf* in terms of phenotypes with a few differences: *lhp1* mutants are small, early flowering plants with needle-like leaves and inflorescences that bear terminal flowers, but the rosette leaves are curling down-wards (Gaudin et al., 2001).

In the double mutants of *sop12* with *lhp1*, lhp1 phenotypes were partially suppressed: the leaves were flatter (Figure 4-8), flowering time was delayed and more flowers were produced before the inflorescent meristem was completely consumed (Hartwig et al., 2012). This suppression was also identified in Col background by Dr. B. Hartwig in Dr. F. Turck's lab in Max Plunck Institute, Germany (Hartwig et al., 2012).

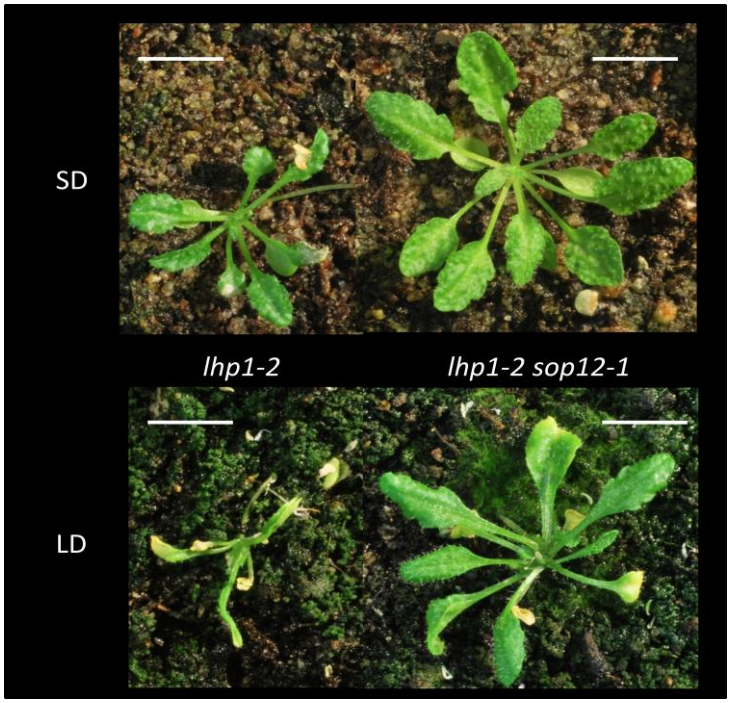


Figure 4-8. *sop12* suppresses *lhp1*

lhp1 phenotypes, for example, the down-ward leaf curling, are less severe in SD than in LD. The *sop12-1 lhp1-2* plant is larger than *lhp1-2*, and the differences seems to be more apparent in LD. Inflorescence stems were removed before photographing. The plants of *lhp1-2* and *lhp1-2 sop12-1* were grown in SD and LD for 40 days. The scale bars, 1 cm.

4.6 The genetic interaction of *SOP12* with trx-G mutants

In contrast to Pc-G genes, fewer trx-G genes have been identified in plants. Several putative homologuess were found through sequence searches but a limited number of them are functionally characterized as trx-G genes (Schuettengruber et al., 2011; Thorstensen et al., 2011). A few years ago, *ULT1*, was identified as a plant-specific trx-G gene through genetic screens (Carles and Fletcher, 2009). According to literature,

characteristics of trx-G genes include, 1) trx-G genes are required for Pc-G target gene expression, 2) the mutant suppresses Pc-G mutants, and 3) the mutant enhances trx-G mutants. As mentioned in previous sections, *SOP12* fits with the first and second ones. In this section, the genetic interaction between *SOP12* and other trx-G genes was examined to show whether SOP12 acts like a trx-G gene genetically. *sop12* was crossed to *atx1*, *atx2*, *efs*, *ult1*, and *ult2*, respectively. The interaction between *sop12* and *atx1* or *atx2* was subtle, whereas *sop12* enhances the phenotypes of *efs*, *ult1* and *ult2*.

sop12 is additive with *atx1* and *atx2* mutants

As mentioned in section 1.2, *ATX1* is a homologue of *TRX* from *Drosophila*. ATX1 contains a SET domain, the functional domain of histone lysine methyltransferases and it targets H3K4 (Alvarez-Venegas et al., 2003). Mutation in *atx1* results in small plants with defect in flower architecture (Alvarez-Venegas et al., 2003). In *atx1* mutant, the expression of several Pc-G targets, including *FLC* and *AG* are decreased (Alvarez-Venegas et al., 2003), accompanied by a reduction of H3K4me3 at these loci (Saleh et al., 2007; Pien et al., 2008). In our growth condition, *atx1-1* was slightly late flowering, which agreed with the result published by R. Alvarez-Venegas (2003). The double mutant *sop12-1 atx1-1* did not show obvious enhancement of single mutant phenotypes (Figure 4-9). 26-day-old *sop12-1 atx1-1* was similar to *atx1-1* (Figure 4-9, upper panel), whereas 40-day-old double mutant was intermediate, but closer to *atx1-1* (Figure 4-9, lower panel).

ATX2 is a paralog of *ATX1* (Saleh et al., 2008). It also possesses SET domain as ATX1 does but it is a putative H3K4 di-methyltransferase rather than tri-methyltransferase. ATX2 shares target genes with ATX1 but the function of ATX2 is sometimes opposite to ATX1 (Saleh et al., 2008). The *ATX2* mutant displayed minor phenotypes, such as slightly delayed abscission of flowers (Saleh et al., 2008), and smaller than wild type (Figure 4-10). The double mutant *atx2-2 sop12-1* was close to *atx2-2* at 26 DAG (Figure 4-10, upper panel), and displayed intermediate phenotypes between *sop12-1* and *atx2-2* at 40 DAG (Figure 4-10, lower panel).

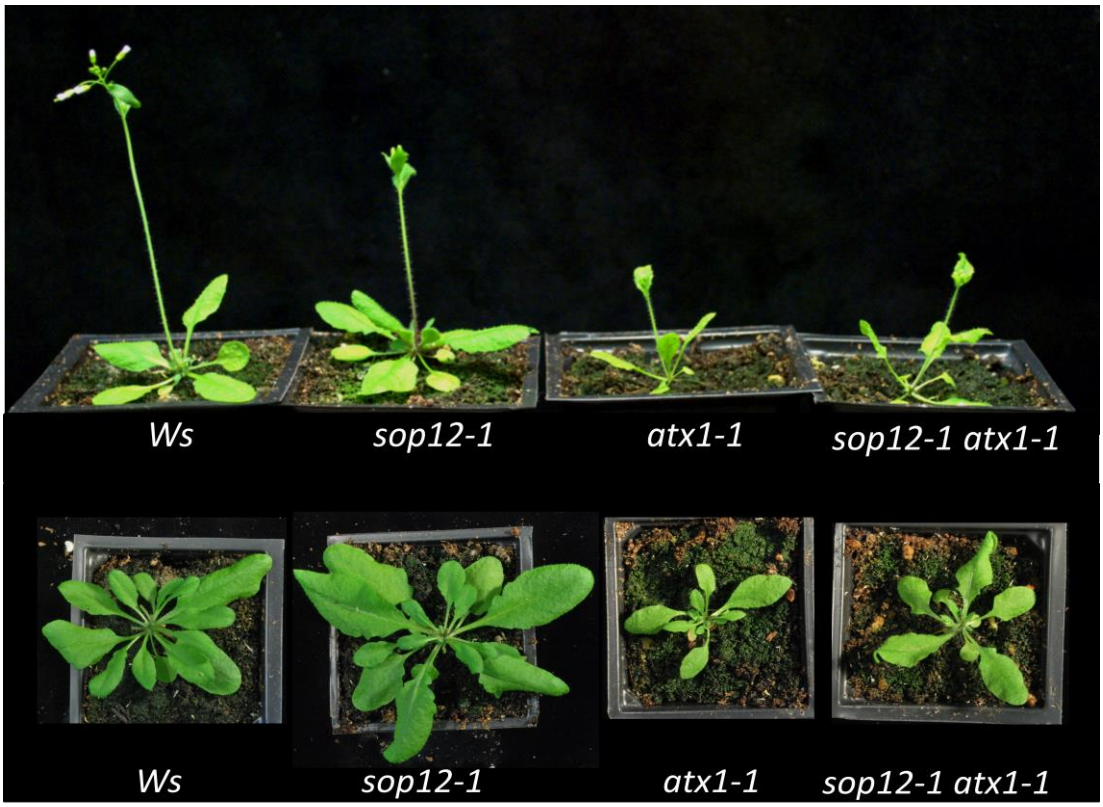


Figure 4-9. The genetic interaction of *sop12-1* and *atx1-1*

The phenotype of *sop12-1 atx1-1* was closer to *atx1-1*, which showed *sop12* does not significantly interact with *ATX1*. Plants were grown under LD conditions for 26 days (upper panel) and 40 days (lower panel). The inflorescence stems were removed in the lower panel. The edge of the pots is 5 cm.

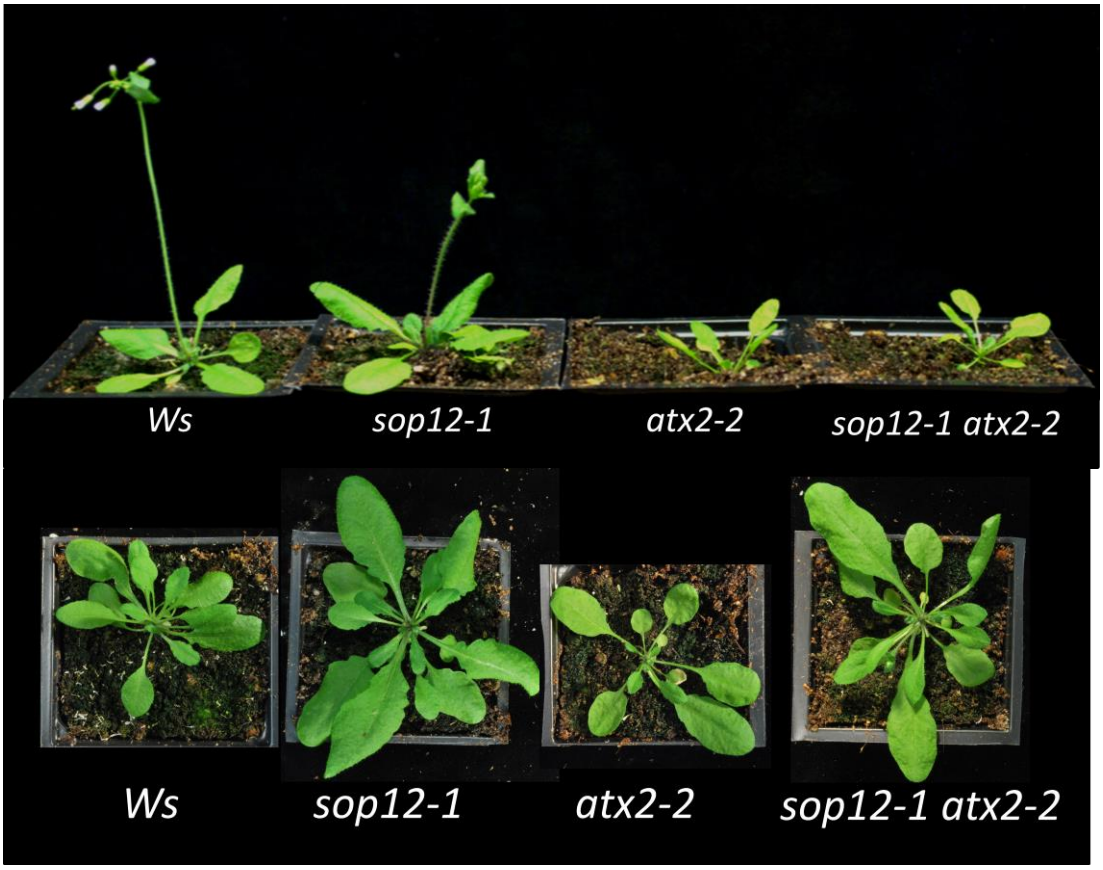


Figure 4-10. Intermediate phenotype of *sop12-1 atx2-2*.

The phenotype of *sop12-1 atx2-2* was closer to *atx2-2* at 26 DAG (upper panel), but shift to *sop12-1*-like at 40 DAG (lower panel). *sop12-1 atx2-2* is mix-backgound (*Ws* and Col, respectively). Plants were grown under LD conditions for 26 days and 40 days. The inflorescence stems were removed in the lower panel. The edge of the pots is 5 cm.

sop12-3 enhances *efs*

EFS contains a SET-domain and shares homolog with *Ash1* in *Drosophila* (Baumbusch et al., 2001), so *EFS* is also named as *Ash1 HOMOLOG 2* (*ASHH2*). EFS

was shown to be involved in H3K4me3 at *FLC* (Kim et al., 2005) and H3K36 di- and tri-methylation (Xu et al., 2008), which are associated with active genes (Santos-Rosa et al., 2002; Krogan et al., 2003; Xiao et al., 2003a). *efs* mutants exhibit pleiotropic phenotypes, including dwarfing, multiple branches, and partial sterility (Soppe et al., 1999) (Figure 4-11). The *efs sop12* double mutant was created using the *sop12-3* mutant, which was isolated in the same genetic background (L*er*) as *efs*. The double mutant *sop12-3 efs* is more dwarfed and almost completely sterile than *efs* single mutants (Figure 4-11). Since *sop12-3* single mutants are wild type with respect to branching, fertility and stature, this is a non additive interaction but *sop12* strongly enhances *efs*. This result was confirmed by phenotyping followed by genotyping in the progeny of an F2 *efs sop12-3/+* individual. The F3 plants segregated less and more severe phenotype within the family. 17 plants with more severe phenotypes and another 39 less severe ones were picked to extract DNA for genotyping *sop12-3*. 94.1% (16/17) of the severe-phenotyped plants were *sop12-3* homozygotes, while 100% of the *efs*-like plants were SOP12 homozygotes or *sop12-3/+* heterozygotes. The results strongly suggested that the enhanced efs phenotype was conferred by the *efs sop12-3* double mutant.

Figure 4-11. The *sop12-3* mutation enhances the *efs* mutant phenotype

Morphologically *sop12-3 efs* is more severe than *efs* single mutant. For instance, the double mutant is more dwarfed than *efs* single mutant, and *efs* is partially sterile, wherease *sop12-3 efs* seldom produce any seeds. Plants were grown in LD for 40 days.

sop12-3 exacerbates the extra-floral organ phenotype of *ultrapetalla 1* (*ult1*) and *ultrapetalla 2* (*ult2*)

ULT1 is thought to represent a plant specific trx-G gene, based on the suppression of *clf* mutants by *ult1* mutants and also the activation of *AG* when *ULT1* is over-expressed (Carles and Fletcher, 2009). *ult1* has a larger meristem than wild type

plants, which results in extra floral organs (Fletcher, 2001). Interestingly, *sop12 ult1* has even more organs compared to *ult1*, despite the fact that sop12 single mutants have normal floral organ numbers (Figure 4-12). For example, in *ult1-1*, the flowers may contain two or three carpels [Figure 4-12 (A)] but never exceed three in my observation, whereas in *sop12-3 ult1-1*, four carpel flowers were found occasionally, especially in first flowers. The average numbers of the four whorls of organs, in the order of sepals, petals, stamens and carpel, in *ult1-1* were 6.1, 6.4, 6.1, and 2.4, while in *sop12-3 ult1-1*, those were 7.0, 6.9, 6.4, and 3.1 (Figure 4-12, lower panel).

(A)

(B)

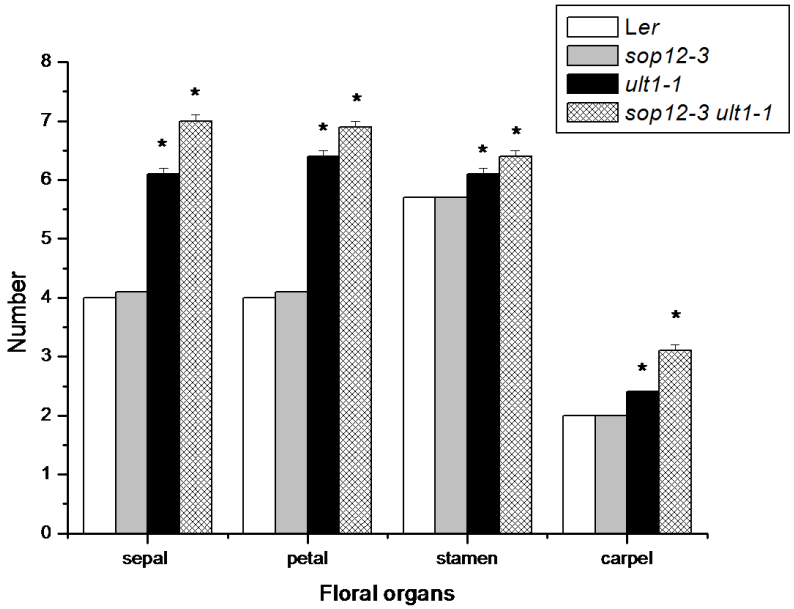
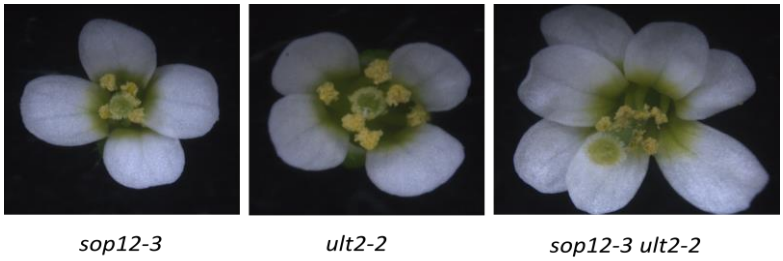


Figure 4-12. The extra floral organs in *ult1-1* and *sop12-3 ult1-1*.

(A) The flower and silique of *sop12-3* and *ult1-1* mutants. In *ult1-1* and *sop12-3 ult1-1*, extra petals were easily noticed. The silique of *sop12-1 ult1-1* was composed of four carpels, while of *ult1-1*, it was usually three. Scale bars, upper panel, 1 mm; lower panel, 0.5 mm.
(B) The statistics analysis of floral organ numbers in *sop12-3* and *ult1-1* mutants. The floral organs of the initial 10 flowers on primary inflorescence stems were counted and the average numbers of each floral organ are shown with 1 SE as error bars. Data were collected from 11-19 individual plants. The stars mark the data that are significantly different with data of wild type plants ($p<0.001$). Note that there was also significant difference between *sop12-3 ult1-1* and *ult1-1* ($p<0.001$).

ULT2 shares 81% amino acid identity with ULT1 and is a paralogue of ULT1 (Carles et al., 2005). The single gene mutation *ult2* does not give extra floral organs, but when *ULT2* was over-expressed, it complemented the ult1 phenotype (Carles et al., 2005). In *ult2*, the numbers of floral organs are the same as the ones in wild type, but in *ult2-2 sop12-3*, the numbers of sepals (s) and petals (p) are significantly increased from 4.0 (s) and 4.0 (p) in *ult2-2* to 4.8 (s) and 4.9 (p) (Figure 4-13). The average stamen and carpel numbers remained unchanged as in wild type plants.

(A)



(B)

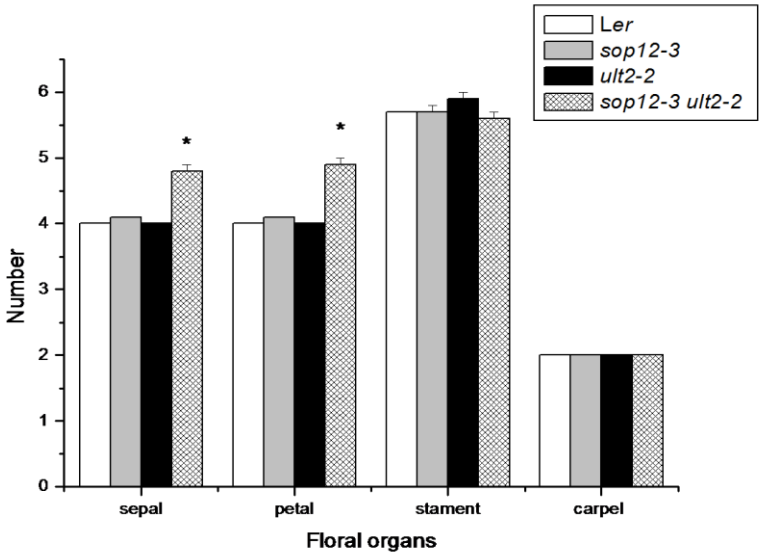


Figure 4-13. The extra floral organs in *ult2-2* and *sop12-3 ult2-2*.

(A) The flower of *sop12-3* and *ult2-2* mutants. In *ult2-2* and *sop12-3*, the numbers of floral organs are normal, whereas the double mutant *sop12-3 ult2-2* displayed extra petals. Photographs were taken under the same scale.

(B) Statistical analysis of floral organ numbers in *sop12-3* and *ult2-2* mutants. The floral organs of initial 10 flowers on primary inflorescence stems were counted and the average numbers of each floral organ are shown with 1 SE as error bars. Data were collected from 11-19 individual plants. The stars mark the data that are significantly different compared with data of wild type plants ($p<0.001$).

4.7 Suppression of *clf* is not caused by altered leaf polarity in *sop12* mutants

One possibility for the suppression of the clf leaf curling phenotype is that *SOP12* is associated with controlling leaf polarity. The growth of the leaf blade is thought to result from the juxtaposition of abaxial (dorsal) and adaxial (ventral) domains in the leaf primordium. Mutations which have weak effects on genes defining adaxial or abaxial identities frequently result in leaf curling, for example adaxialised leaves show upwards curling whereas abaxialised leaves show downwards curling. Although the upwards curling seen in *clf* mutants is not obviously due to alterations in leaf polarity genes, the fact that *sop12* mutants had weakly downward curling leaves in certain growth conditions raised the question whether *sop12* mutants caused weak abaxialisation of leaves. To test this possibility, *sop12-3* was crossed with the *asymmetric leaves 1* (*as1*) mutant. AS1 and ASYMMETRIC LEAVES 2 (AS2) form a complex responsible for controlling/determining leaf adaxial fate (Xu et al., 2003). In the scenario that the function of *SOP12* promotes adaxial fate, an enhanced abaxialised phenotype would be expected in the double mutant *sop12-3 as1*. As a matter of fact, *sop12-3 as1* double mutant is indiscernible from *as1* (Figure 4-14). This

provides the evidence that it is less likely *sop12* suppress leaf-curling phenotype of *clf* by change the leaf polarity.

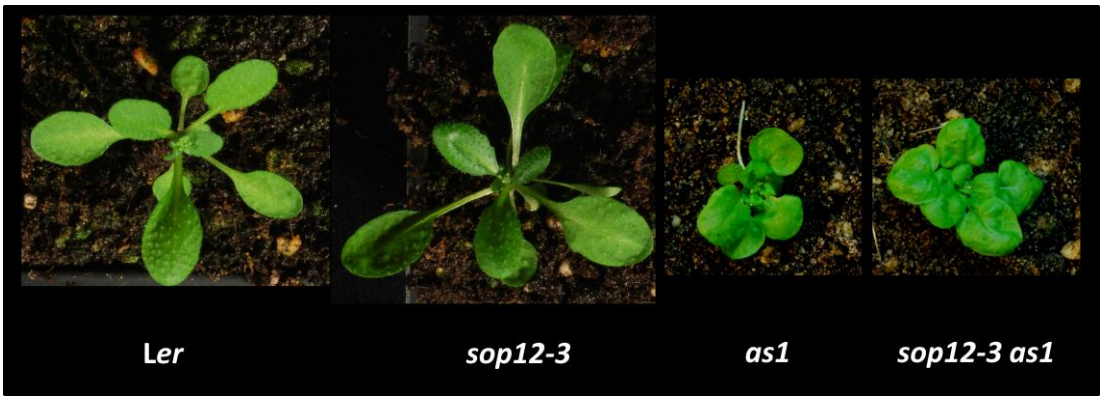


Figure 4-14. *as1* is epistatic *to sop12-3* in double mutant combinations

as1 mutant gives wrinkled, ear-shaped leaves, and *sop12-3 as1* resembles as1 phenotypes. Plants were 28-day-old plants in LD. Photos were taken under the same scale.

4.8 Summary and discussion

Mutation in *SOP12* showed only minor phenotypes, such as slightly late flowering. In all the conditions observed, including LD, SD, tissue culture and heat stress, morphology of *sop12* mutant was almost indistinguishable from wild type, even though, mutation of *SOP12* in *clf* background drastically altered the clf phenotypes. These suggest the function of SOP12 is distinct from late flowering suppressors and may have more interesting and specific roles on Pc-G regulation. The leaf-curling in *clf* is caused by mis-expression of *AG* and *SEP3* in leaves (Goodrich et al., 1997; Lopez-Vernaza et al., 2012). In *sop12-1 clf-50* seedlings, *AG* and *SEP3* were no longer

highly mis-expressed. The down regulation of mis-expressed Pc-G targets agreed with the suppressed phenotypes in *sop12 clf*, and also suggested that *SOP12* is an activator of CLF target genes, like many trx-G genes, rather than acting downstream of the targets or being itself a target. The genetic evidence showed SOP12 is not involved in defining leaf polarity, implying *sop12* suppresses *clf* via other pathway instead of simply altering the leaf morphology. This is associated with the fact that *sop12* suppresses *clf* in multiple aspects including flowering time in addition to leaf curling.

The huge reduction in *SEP3* activity in *sop12 clf* is sufficient to suppress clf phenotype

The loss of repression at Pc-G targets like *AG* and *SEP3* in leaves causes the curled leaf phenotype of *clf*. When *AG* or *SEP3* is over expressed in leaves, the morphology resembles *clf* mutation (Mizukami and Ma, 1992; Pelaz et al., 2001). Moreover, when *AG* or *SEP3* expression is eliminated by mutation in *clf* background, the leaf curling is also abolished (Goodrich et al., 1997; Lopez-Vernaza et al., 2012). AG and SEP3 proteins may form a complex which specifies carpel development (Favaro et al., 2003). Hence, withdrawal of either one of them is able to disrupt the proper function of the complex. Consistent with this, in *clf sep3* mutants, leaf curling is lost despite the fact that *AG* mRNA and protein remain strongly mis-expressed in leaves. In other words *AG* protein activity requires *SEP3* activity to cause the curled phenotype in leaves (Lopez-Vernaza et al., 2012). In *sop12 clf*, notwithstanding the decrease of *AG*

is less dramatic, the expression of *SEP3* reduces merely to wild type level in SD conditions, giving wild type-like leaves.

The weaker suppression of clf phenotype in LD may be caused by higher amount of *FT* accumulation. Light perhaps stabilises *CONSTANS* (*CO*) (Yanovsky and Kay, 2002), a positive regulator of FT (Guo et al., 1998; Kardailsky et al., 1999; Samach et al., 2000). *FT* is required for *SEP3* expression in leaves according to the fact that in *clf ft*, *SEP3* mRNA level is decreased compared to *clf* (Lopez-Vernaza et al., 2012). Thus, higher *FT* accumulate under LD may result in higher amount of *SEP3*, explaining the less suppressed phenotype in LD in *sop12 clf*.

sop12 suppresses *clf* and *lhp1*

The suppression of *lhp1* by *sop12* indicates that the interaction of *sop12* with Pc-G is not confined to *clf*. This suppression was also observed in another independent genetic screen for suppressors of *lhp1* in a different ecotypic background (Col), which identified the *ANTAGONIST OF LHP1* (*ALP1*) gene, which corresponds to SOP12 (Hartwig et al., 2012). LHP1 is the only protein discovered so far to be responsible for H3K27me3 recognition in plants. However, there must be other proteins to bind to H3K27me3 since *lhp1* produces 100% viable seeds and Pc-G activities are not completely depleted in *lhp1*. Thus mutants which lack H3K27me3, such as *clf swn* double mutants, are much more severe than null *lhp1* mutants.

Since in *lhp1* mutants profiling suggests that H3K27me3 levels remains more or less unchanged (Turck et al., 2007), the suppression of *lhp1* by *sop12* implies the mechanism may not simply involve restoring H3K27me3 levels.

sop12 does not strongly interact with many Pc-G mutants tested.

sop12 is a weak suppressor of *clf*. In LD, leaf curling is partially suppressed in *sop12 clf*, while in SD, *sop12 clf* is still early flowering compared to wild type (Figure 4-1). Except for *clf* and *lhp1*, *sop12* did not show strong interaction with other Pc-G mutants tested (*mea*, *emf1*, strong and weak *emf2*). The possible reason is that those mutants might be too severe to be rescued by this weak suppressor. However, it is extremely puzzling that *sop12* does not suppress *emf2-10*. The targets of EMF2-PRC2 are *AG* and *FT* (Chanvivattana et al., 2004; Jiang et al., 2008). According to Figure 4-4, *SOP12* is required for full *AG* and *FT* mis-expression in *clf*, although the effect on *AG* is relatively subtle. It was speculated *sop12* would suppress *emf2*-10 as well. In fact, *emf2-10* is similar to *clf* in morphology but more severe than *clf* in terms of organ size and flowering time, etc. (Chanvivattana et al., 2004). Removing of *AG* in *clf* successfully restitutes the morphology back to wild type but only partially restored phenotypes of *emf2-10* (Chanvivattana et al., 2004). Therefore the minor effect on *AG* caused by *sop12* may be too subtle to alter *emf2-10* phenotypes.

The preliminary examination of the interaction between *sop12* and *mea* was to count the ratio of aborted seeds of *sop12 mea*/+ progeny. Although the result concluded that

sop12 does not suppress seed abortion phenotype of *mea*, it still possible that SOP12 has subtle effects on *mea* phenotypes, such as shrivelled *sop12 mea* seeds might be more viable, or the embryos in those shrivelled seeds develop to further stages than *mea* seeds. To assay whether *sop12* has minor interactions with *mea*, more detailed characterizations of endosperm or of maternal transmission are required.

Extra floral organs and *AG* regulation

ULT1 regulates *AG* expression via participating in H3K4me3 at *AG* locus. *ULT1* is required for mis-expression of *AG* in *clf*. In *ult1* and *ult1 clf* double mutants, H3K4me3 at *AG* is decreased compared to wild type and *clf*, respectively (Carles and Fletcher, 2009). The extra floral organ phenotype of *ult1* is correlated with the down regulation of *AG*. Strong *ag* mutants display indeterminate flowers (Goodrich et al., 1997), indicating the stem cells in floral meristem (FM) are maintained instead of being consumed. It was proposed that *AG* expression terminates the expression of *WUS*, the gene required for determining the stem cell identity (Mayer et al., 1998), leading to losing meristem cells in FM (Liu et al., 2011). In *ult1*, the reduction of *AG* expression reduces or delays the repression of *WUS*, resulting in expansion of the meristem niche and extra floral organs (Carles et al., 2004; Liu et al., 2011). The enhancement of *ult1* by *sop12* is probably due to the requirement of *SOP12* for activating *AG*. It is speculated that, in *sop12-3 ult1-1*, *AG* expression is decreased to a lower amount than in *ult1-1*, giving larger meristem and more organs in the double.

ULT2 is homologous to *ULT1*. The protein sequences share high identity in amino acid sequences (81%) and over-expression of *ULT2* rescues *ult1* mutation (Carles et al., 2005). No significant phenotypes have been characterised in *ULT2* mutant so far, probably because the redundancy of *ULT1* and *ULT2* is so strong that in *ult2*, *ULT1* masks the possible phenotypes of *ult2*. The fact that *sop12 ult2* gives flowers with extra floral organs implies *ULT2* functions largely mimic *ULT1*.

SOP12 acts like a trx-G gene

From the genetic analysis in *Drosophila*, trx-G mutants enhance other trx-G mutants (Calgaro et al., 2002) and suppress Pc-G mutants (Ingham, 1998). In plant model *Arabidopsis thaliana*, those charateristics are powerful tools for identifying more unknown components, because somehow in plants, PRC1 and trx-G are not as conserved as PRC2 is to animal models in the sequence homology. For instance, PRC1-like components EMF1 and trx-G gene-like ULT1 are plant-specific (Calonje et al., 2008; Carles and Fletcher, 2009). Nevertheless, the mechanism is still similar to animal models (Hennig and Derkacheva, 2009; Bemer and Grossniklaus, 2012).

Genetic and molecular evidences support that SOP12 has trx-G like roles. Genetically, *sop12* suppresses Pc-G mutants (*clf* and *lhp1*) and enhances trx-G mutants *ult1* and *efs*. In molecular aspects, SOP12 is a general activator of Pc-G targets. In addition, since the four Pc-G targets regulated by two different PRC2----*FLC* by VRN2-PRC2, *FT*, *SEP3* and *AG* by EMF2-PRC2 (Gendall et al., 2001; Chanvivattana et al., 2004) and

are all repressed again in *sop12-1 clf-50*, the effect of *sop12* is not limited to one certain Pc-G pathway.

The enhancement of weak alleles indicates working on the same pathway

When the enhanced phenotypes occurred in double mutant, in the case these two genes are not redundant homologues, they may participate in the same protein complex, especially if the mutants are weak alleles, or work synergistically to regulate the same biological process (Martienssen and Irish, 1999; Herman and Yochem, 2005). In a weak allele, contrast to completely loss the gene function in a null allele, the mutated gene generates partially functional proteins. Moreover, because gene duplication produces many functionally redundant genes, mutation of a gene member in a family results in partially loss of the overall activity. These mutants could be considered as weak allele as well. The three mutants enhanced by *sop12* are possibly all weak alleles: *ULT1* and *ULT2* are highly redundant, so *ult1* and *ult2* single mutants must retain some ULT activity. EFS is one of the major, but not the only H3K36 methyltransferase. SDG4, SDG7, and SDG24 are putative H3K36 methyltransferases that potentially redundant to EFS (Xu et al., 2008). The enhancement by *sop12* of *ult1* and *efs* implies SOP12 may involve in regulating *AG* or H3K36 methylation, either by forming a complex with one of them or working synergistically.

Chapter 5. Functional analysis of SOP12

SOP12 was revealed from the screening for suppressors of *clf* mutants. It was shown in chapter 4 that it is required for mis-expression of Pc-G targets in *clf*. According to the bioinformatic database research, the protein sequences of SOP12 shares homology with Harbinger transposase-derived nuclease like protein but has likely lost the nuclease activity and obtained novel biological roles. However, it is still puzzling what the mechanism by which *sop12* mutation suppresses *clf* is. In this chapter, I tried biochemical and molecular approaches to investigate the role of *SOP12* in the Pc-G regulation pathway.

5.1 Localisation of SOP12

SOP12 is expressed in multiple tissues

The expression pattern of a gene might provide clues about its function. For example, *AG* is expressed in whorls three and four of the floral meristem where it specifies stamen and carpel development (Yanofsky et al., 1990). In order to have a general idea about *SOP12* expression, a variety of tissues, including seedlings, roots, leaves, flowers and siliques, were collected from wild type plants (*Ws* background) and the mRNA levels were examined by RT-PCR (Figure 5-1). The present of SOP12 protein was also determined by detecting GFP expression in a *pSOP12::SOP12-GFP*/*sop12-1* transgenic line on Western blot (Figure 5-2). Both the results show *SOP12* is widely expressed.

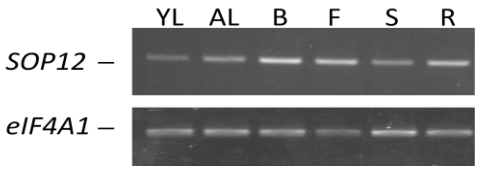


Figure 5-1.

SOP12 expression in different tissues. YL, young leaves; AL, adult leaves; B, flower buds; F, flowers; S, seedlings; R, roots. *eIF4A1* is the internal control indicating cDNA amount loaded in each experiment.

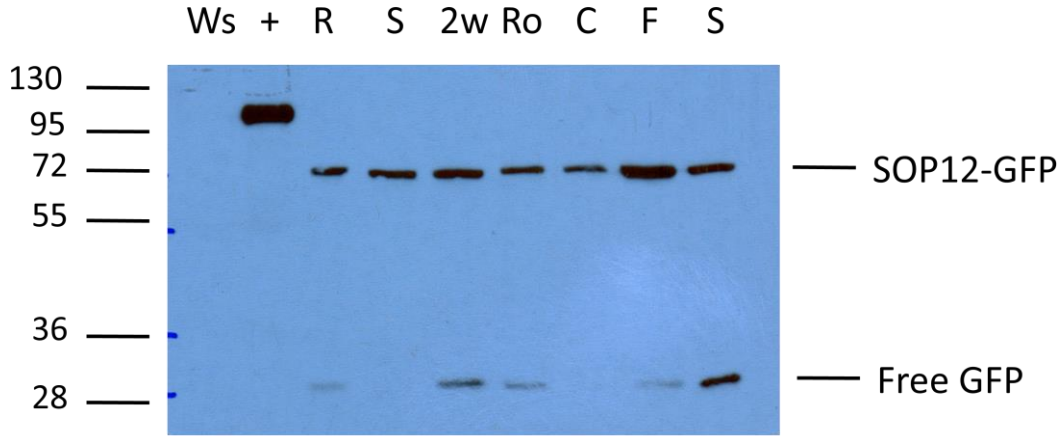


Figure 5-2. SOP12-GFP is widely expressed

Western blot analysis of the presence of SOP12-GFP in tissues. Total crude proteins were extracted from a variety of tissues including roots (R), inflorescence stems (S), 2-week-old seedlings (2w), rosette leaves (Ro), cauline leaves (C), flower bud and inflorescence (F) and siliques (S) of transgenic plant *pSOP12::SOP12-GFP/sop12-1*, and then analysed by Western blotting using a mouse monoclonal antibody against GFP. Protein extracts from *Ws* and *pLHP1::LHP1-GFP* (+) were also included as negative and positive control, respectively.

To inspect the spatial expression pattern, the localisation of β-glucoronidase (GUS) activity in the reporter line *pSOP12::SOP12-GUS/sop12-1* was revealed by histochemical staining using a chromogenic substrate (see chapter 2). Consistent with the observations of the presence of *SOP12* mRNA and SOP12-GFP protein, SOP12-GUS was detected in all the tissues tested (Figure 5-3), and preferentially accumulated in young tissues, like new leaves (Figure 5-3A), and axillary buds. In older leaves, SOP12-GUS appeared in vasculature (Figure 5-3C), trichomes and guard cells (Figure 5-3E)

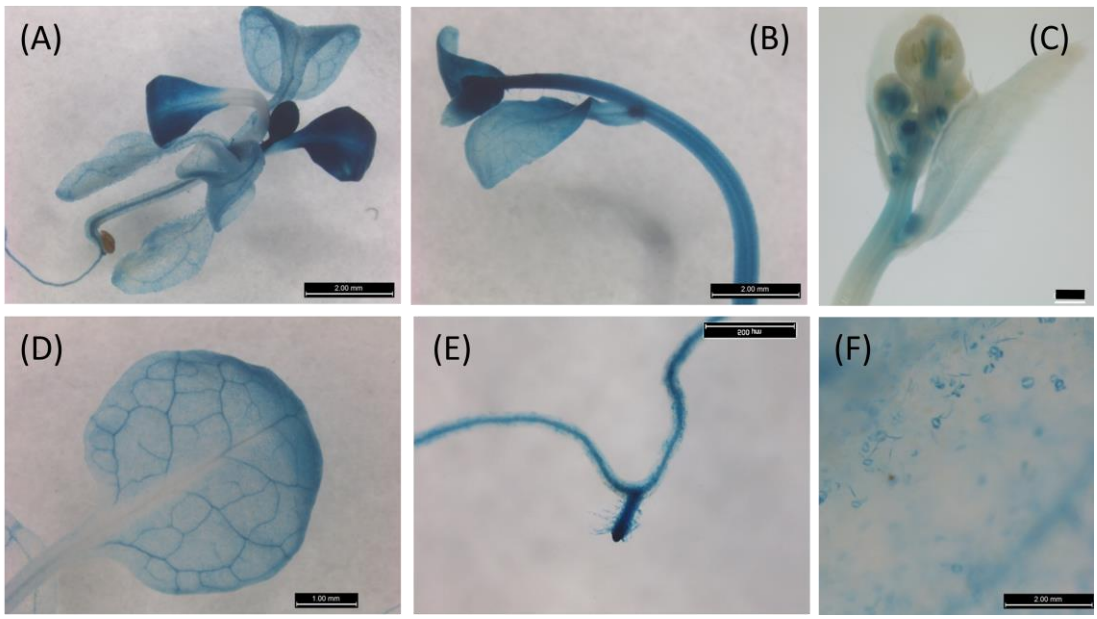


Figure 5-3. Spatial localisation of SOP12-GUS

The *GUS* reporter gene activity in the transgenic plant *pSOP12::SOP12-GUS/sop12-1* was analysed by histochemical staining, to produce a blue product. The pictures show (A) 3-week-old plant grown on MS tissue culture plates, (B) inflorescence, (C) closer look at inflorescence in a weaker expressing line, (D) closer look of a rosette leaf (E) root, and (F) closer look of

SOP12-GUS in the leaf. The signal is strong in guard cells and trichomes. Scale bars in A and B, 2 mm; in C, 0.27 mm; in D 1 mm; in E, 500 µm; in F, 200 µm.

SOP12 localizes in the nucleus

NLStradamus (Nguyen Ba et al., 2009) predicted SOP12 bears putative NLS at amino acid 5 to 29, yet PredictProtein suggested there is no NLS in SOP12. In order to clarify whether SOP12 is a nuclear protein, the roots of the GFP-tagged SOP12 transgenic lines in *sop12-1* mutant background (*pSOP12::SOP12-GFP*/*sop12-1*) were observed using a confocal microscope. Before the observation, the roots were pre-stained with propidium iodide (PI) to reveal the position of cell walls. The green fluorescence indicates the location of SOP12 protein is in nuclei (Figure 5-4). Roots of *Ws* and a *35S::GFP* control line were also examined under the same conditions, which served as no-fluorescence control and indicator of free-GFP signal, respectively. No GFP signal was detected in *Ws*, whereas in *35S::GFP* the GFP spread within the cell and showed nuclear and cytoplasmic localisation.

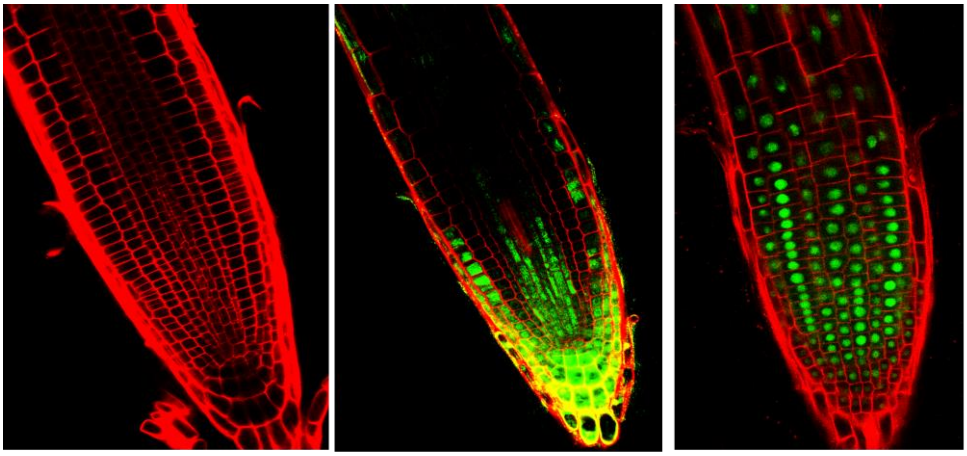


Figure 5-4. Cellular localisation of GFP and SOP12-GFP

The localisation of GFP-tagged SOP12 was inspected by confocal microscopy. The roots of 10 to 14 –day-old *Ws*, 3*5S::GFP*, and *pSOP12::SOP12-GFP/sop12-1* plants on MS plates were stained with PI to mark cell walls and then observed. SOP12-GFP localises in the nuclei and the control free GFP spread in the whole cells. *Ws* is the negative control.

It was noticed that the cellular localisations of SOP12-GUS in Figure 5-3F and SOP12-GFP in Figure 5-4C are different----SOP12-GUS signal was not limited to the nuclei like SOP12-GFP was. Both GUS and GFP were tagged at the C-terminus of SOP12, so the NLS at N terminus should remain unaffected and accessible for importin recognition. Hence, the possible explanation is, GUS is much bigger than GFP in size (90 kD and 27 kD, respectively), somehow resulting in prohibition of transportation of SOP12-GUS into nuclei. Conceivably, this prohibition may further block the normal function of SOP12. A piece of supporting evidence is that the construct *pSOP12::SOP12-GUS* only partially complemented the sop12 phenotype whereas

pSOP12::SOP12-GFP gave full complementation (Figure 5-5), implying SOP12-GUS is impaired perhaps because the fusion protein does not enter nuclei to execute its function.

Figure 5-5. Comparison of the complementation conferred by *pSOP12::SOP12-GFP* and *pSOP12::SOP12-GUS* transgenes.

The genomic DNA of *AT3G63270* including up-stream promoter was cloned into Gateway® binary vector pGWB4 (GFP-tagged) or pGWB3 (GUS-tagged) and introduced into *sop12-clf-50* through floral dipping. *pSOP12::SOP12-GFP* transgene fully complemented *sop12-1* mutation and gives a similar clf leaf curling phenotype in s*op12-1 clf-50* background as *clf-50*, but *pSOP12::SOP12-GUS* does not. Plants were grown in SD conditions.

5.2 Identifying proteins interacting with SOP12

SOP12 may physically interact with EFS

Since the genetic interaction of *SOP12* with *EFS*, *ULT1* and *ULT2* suggest their protein products they may interact in a complex, the yeast two hybrid assay was used to test for physical interaction. The coding sequences of *SOP12*, *EFS*, *ULT1* and *ULT2* were introduced as in-frame fusions into Gateway-compatible pGADT7 and pGBKT7 vectors.

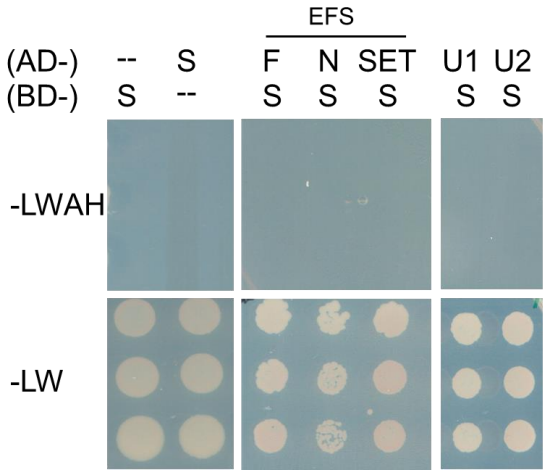
In pGADT7, the constructs encode a Gal4 transcriptional activation domain (AD) fused to the N-terminus of inserted genes, whereas in pGBKT7 there is a Gal4 DNA binding domain (BD) fusion. First of all, the autoactivation of pGBKT7-SOP12 was tested by co-transforming the yeast strain Y2HGold® (Clontech), which carries several GAL4-dependent reporter genes (see chapter 2), with pGBKT7-SOP12 and the pGADT7 empty (lacking an insert) vector. Autoactivation can happen if the proteins which are fused to BD contain transcriptional activation domains so that the fusion proteins activate reporter genes in the absence of any interaction with AD fusions. The combination of BD-SOP12 and AD alone failed to activate reporter genes (Figure 5-6 A). This suggests that SOP12 lacks a transcriptional activation domain, and also that BD-SOP12 does not interact with AD alone. Meanwhile, Y2HGold® (Clontech) was also co-transformed with pGADT7-SOP12 and the pGBKT7 empty vector. If SOP12 binds to DNA, pGADT7-SOP12 might activate reporters on its own. In fact, this combination did not activate reporters, either. This hints either that SOP12 does not generally bind to DNA or that a specific consensus sequence is necessary for its binding.

Once it was clear that BD-SOP12 or AD-SOP12 alone do not activate the reporters, SOP12 and EFS constructs were tested for interaction. In addition to full length EFS (EFS-FL), truncated EFS proteins were also cloned into pGBKT7 and pGADT7. The truncated EFS fragments included the N-terminus (amino acid 1-849, EFS-N), and SET domain (amino acid 843-1204, EFS-SET). When the interaction between AD-SOP12 and BD-EFS constructs were tested, interaction was detected between SOP12 and EFS-FL or EFS-N but not EFS-SET (Figure 5-6 B). However, no interaction was discovered if the

fusion tags were swapped i.e. in the combinations of BD-SOP12 with all AD-EFS constructs, the reporters were not activated (Figure 5-6 A). The interaction was only detected when EFS-FL and EFS-N were fused to BD, which may yet indicate a true interaction; certainly there are several other examples of valid protein-protein interactions which can be detected in one orientation of BD or AD fusions but not in the reciprocal fusions, for example that between the PRC2 members CLF and FIE (Spillane et al., 2000). However, the potential interaction between SOP12 and EFS-FL or EFS-N needs further validation for example to test that BD-EFS-FL and BD-EFS-N do not autoactivate in yeast. Actually those two controls were included in the experiments but unfortunately the yeast transformation failed. Moreover, it would be nice if the interaction could be tested *in vivo*., so I have made crossings between transgenic lines harbouring GFP-tagged SOP12 (*pSOP12::SOP12-GFP/sop12-1*) and FLAG-tagged EFS (*pEFS::EFS-FLAG/efs*) for co-IP testing. However, due to the time limitation I did not manage to repeat or perform those experiments.

The physical interactions between SOP12 and ULT1 or ULT2, and of SOP12 with itself were examined by Y2H. No interaction was detected with ULT1 or ULT2 either as fusions to AD or BD (Figure 5-6). On the other hand, AD-SOP12 interacted with BD-SOP12 (Figure 5-6 B), which suggests SOP12 may form a homo-dimer in plants.

(A)



(B)

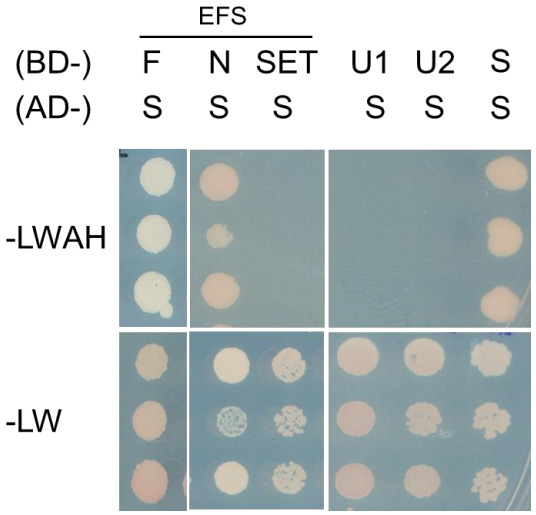


Figure 5-6. Yeast two-hybrid assays for interactions between SOP12 and EFS, ULT1 or ULT2. Full length or truncated cDNA fragments of SOP12, EFS, ULT1 and ULT2 were fused with AD or BD (dashed lines, empty vectors; S, SOP12;, F, EFS-FL; N, EFS-N; SET, EFS-SET; U1, ULT1; U2, ULT2). Media lacking leucine (L) and tryptophan (W) (-LW) are selective for clones

harbouring both AD and BD constructs, and media lacking adenine (A), histidine (H), L and W allow yeast cells to grow when the AD fusion protein and BD fusion interact and activate reporter genes. (A) AD-SOP12 or BD-SOP do not activate reporters when yeast cells were co-transformed with BD and AD, respectively. Neither does BD-SOP12 and AD-EFS-FL, AD-EFS-N, AD-EFS-SET, AD-ULT1 or AD-ULT2. (B) AD-SOP12 interacts with BD-EFS-FL, BD-EFS-N, and BD-SOP12.

Yeast two-hybrid screens identify potential SOP12 interactors with a wide variety of gene functions

In order to discover putative interacting partners of SOP12, further yeast two hybrid (Y2H) screening was performed. The yeast strain Y187 was transformed with pGBKT7-SOP12 and was used as the bait to screen a prey library prepared from *Arabidopsis* cDNA made from a mixture of plant tissues (gift of Prof Brendan Davies, University of Leeds, see chapter 2 for more details). More than 1,000 positive colonies were selected based on adenine and histidine reporters. The number of positive clones was narrowed down by using a third reporter, *MEL1*, which encodes α-galactosidase and produces a blue product when colonies are grown on media containing chromogenic X-α-gal substrate. The positive clones growing on quadruple drop out media (QDO media) which lack four essential amino acids leucine, tryptophan, histidine and adenine (see yeast two hybrid assay in section 2.5), were replicated to new QDO media containing X-α-gal. After two to three days, replicated colonies all grew whereas only a fraction of the colonies turned blue, indicating the expression of *MEL1* reporter. The blue colonies were picked and the cDNA sequences fused with AD were amplified by

PCR and sequenced. The sequences from 312 positive clones were BLAST against the *Arabidopsis* transcriptome database on NCBI to retrieve SOP12 interacting candidates. The reading frame of each candidate was checked to make sure the fusion was in frame with GAL4 AD. Among 238 sequences of which corresponding genes could be found, 22 genes were represented more than three times (Table 5-1). In general, the 22 candidates are not well correlated with Pc-G or trx-G regulation pathways. Furthermore, they can be roughly sorted into three categories including genes in JA signalling and biosynthesis pathway (MYC2, TAT3 and LOX3), stress response proteins (HSC70-1, BIP1, and LOS2, etc.), and chloroplast metabolism (FTSZ2-2, ALDH11A3, and DEGP2, etc.) according to information from TAIR. Among these highly represented clones, only MYC2, PCNA1, LOS2 and ATJ3 are known to be nuclear-localised, so were considered as possible SOP12 interactors based on the cellular localisation of SOP12, MYC2 has been indentified in several other screens with the same library and may be a false positive clone (Dr. B. Causier, University of Leeds, personal communication). To verify whether *SOP12*, like *MYC2*, is involved in the JA signalling pathway, the root length of *sop12* mutants growing in media containing 50 µM methyl-JA (Me-JA) was measured. Because the *myc2* mutant is JA-insensitive (Berger et al., 1996), if SOP12 participates in the same complex as MYC2 does, *sop12* might be JA-insensitive as well. However, the root length on MS plates containing 50 µM Me-JA in *sop12* was similar to *Ws* (data not shown), illustrating that *sop12* is JA-sensitive and SOP12 is unlikely to be required for JA signalling.

The other 128 clones, corresponding to 112 genes, are discrete in a variety of gene categories according to gene ontology analysis of VirtualPlant (Katari et al., 2010). 15.3% of the candidates are the signals response to stimulus (abiotic or biotic stress). 20.1% are involved in metabolic process, 8.5% in biological process (replication, transcription, and phosphorylation, etc.) and 7.9% in developmental process. It was noticed that chloroplast proteins were highly represented (14.2%), and the percentage is higher than nucleus localised protein (6.7%). However, SOP12-GFP did not appear in chloroplasts when the cellular localisation of SOP12-GFP in leaves of *pSOP12::SOP12-GFP/sop12-1* were examined (data not shown). After proteins in different cellular compartment other than nuclear were excluded, the remaining candidates expanded from DNA replication to defence responses (Table5-2). It was noticed one of the interesting candidates was a histone deacetylase, HD2B. The possible interaction between SOP12 and HD2B hints SOP12 may bind and inhibit HD2B activity.

i. Suppression of *clf* by *sop12* is independent on histone deacetylase activity

Histone acetylation at histone H3 (H3ac) is considered as a histone mark associated with transcriptional activity. For example, it was shown that H3ac levels are reduced at *FLC* after vernalisation, accompanying the down-regulation of *FLC* expression (Finnegan et al., 2005). The activity of histone deacetylases is to remove histone acetylation, and reduction of this mark may cause down-regulation of gene expression. Given that *SOP12* promotes transcription of *CLF* targets, the interaction of SOP12 with HD2B, an inhibitor, is difficult to reconcile unless it inhibits HD2B activity. Perhaps when SOP12 is mutated,

HD2B is released from the inhibition, and removes histone acetylation in Pc-G targets, which causes the decrease in Pc-G target mis-expression. In order to verify this hypothesis, *sop12 clf-50* was treated with histone deacetylase inhibitor trichostatin A (TSA) to see whether the suppression is lost upon the treatment. When *clf* or *sop12 clf* plants are grown on MS tissue culture plates, the leaf curling phenotype is more severe than in soil grown plants, most likely because plants are early flowering and have high *FT* levels. Both strong and weak leaf curling were observed in *sop12-1 clf-50* on MS plates (Figure 5-7, 0 μM TSA). TSA treatment further enhanced the early flowering phenotype in both genetic backgrounds, and also induced abaxialisation (Ueno et al., 2007), resulting in down-ward curled leaves. In terms of the leaf curling phenotype, slightly more up-curled leaves were produced in *sop12-1 clf-50* when treated with TSA compared to non-treated *sop12-1 clf-50*. However, *sop12-1 clf-50* still exhibited suppression of clf phenotype under TSA treatment compared to non-treated *clf-50* plants (Figure 5-7). The persistence of suppression in *sop12-1 clf-50* under TSA treatment indicates it is unlikely that clf phenotypes were suppressed by increasing the histone deacetylase activity upon *sop12* mutation.

Table 5-1. Genes that represented more than 3 times amongst SOP12 interacting candidates from Y2H screening.

The ones in bold are proteins annotated by TAIR as having nuclear localisation.

ATG number	**Gene Name**	**Category**	**No. of hits**
AT5G02500	Heat shock 70kDa protein 1/8 (HSC70-1)	Defense to bacterium and fungi. Response to cold, and heat	21
AT1G32640	**MYC2/ Jasmonate insensitive 1 (JIN1)**	**JA signaling**	9
AT5G28540	Luminal-binding protein 1 (BIP1)	Heat response	7
AT1G07370	**Arabidopsis thaliana proliferating cellular nuclear antigen 1 (PCNA1)**	DNA replication	6
AT1G17420	Lipoxygenase 3 (LOX3)	possibly JA biosynthesis	6
AT2G24850	Tyrosine aminotransferase 3 (TAT3)	JA biosynthesis	6
AT3G44110	**Chaperone protein dnaJ 3 (ATJ3)**	Flowering time; salt stress	6
AT3G52750	Tubulin/FtsZ family protein (FTSZ2-2)	Chloroplast organisation	5
AT2G36530	**Low expression of osmotically responsive genes 2 (LOS2)**	**Response to cold, salt stress**.	4
AT3G02260	ATTENUATED SHADE AVOIDANCE 1 (ASA1)	Auxin transport; rood development	4
AT1G58080	ATP phosphoribosyl transferase 1 (ATP-PRT1)	Cysteine, histidine biosynthetic process, root hair elongation	3
AT1G70600	60S ribosomal protein L27a-3	Translation	3
AT1G76080	Thioredoxin-like protein CDSP32 (CDSP32)	Response to oxidative and drought stress	3
AT2G03890	Phosphoinositide 4-kinase gamma 7 (PI4K GAMMA 7)	Phosphatidylinositol phosphorylation	3
AT3G55010	Phosphoribosylformylglycinamidine cyclo-ligase (PUR5)	Nucleotide biosynthesis, etc. ATP binding	3
AT4G24690	Ubiquitin-associated (UBA) zinc-finger and PB1 domain-containing protein	Protein polymerization; ubiquitin binding	3
AT5G42020	Luminal-binding protein 2 (BIP2)	Response to ER stress and heat stress	3
AT5G48570	**Peptidylprolyl isomerase (ROF2**)	**Response to ER stress, heat, and high light intensity.**	3
AT5G55070	Dihydrolipoamide succinyltransferas	Response to oxidative stress; zinc ion binding	3

Table 5-2. SOP12 interacting candidates which are nuclear localised.

Several nuclear localised candidates are list in Table 5-1. Only candidates that represented less than 3 times among the clones sequenced are list here.

ATG number	Gene Name	Category
AT3G01090	SNF1-related protein kinase catalytic subunit alpha KIN10 (KIN10)	ABA mediated signaling pathway
AT5G50670	Squamosa promoter-binding-like protein (SPL13)	Anther development
AT5G43270	Squamosa promoter-binding-like protein 2 (SPL2)	Anther development, vegetative to reproductive phase transition
AT4G31550	Putative WRKY transcription factor 11 (WRKY11)	Defense response to bacterium
AT1G44900	Minichromosome maintenance protein 2 (MCM2)	DNA replication
AT4G02060	PROLIFERA (PRL)	DNA replication, H3K9 methylation
AT3G10270	DNA gyrase subunit B (GYRB1)	DNA topological change
AT2G18510	Splicing factor 3B subunit 4 (emb2444)	Embryo development; ending in seed dormancy
AT1G52380	Nucleoporin 50 protein	Intracellular transport
AT5G22650	**Histone deacetylase HDT2 (HD2B)**	Polarity specification of adaxial/abaxial axis
AT2G21440	RNA recognition motif-containing protein	RNA binding
AT5G41770	Putative crooked neck protein / cell cycle protein	RNA processing
AT1G20960	EMBRYO DEFECTIVE 1507 (emb1507)	Seed dormancy; meiosis
AT5G24500	**Uncharacterized protein (SUP1)**	**Unknown**
AT3G03140	**PWWP domain-containing protein**	**Unknown**

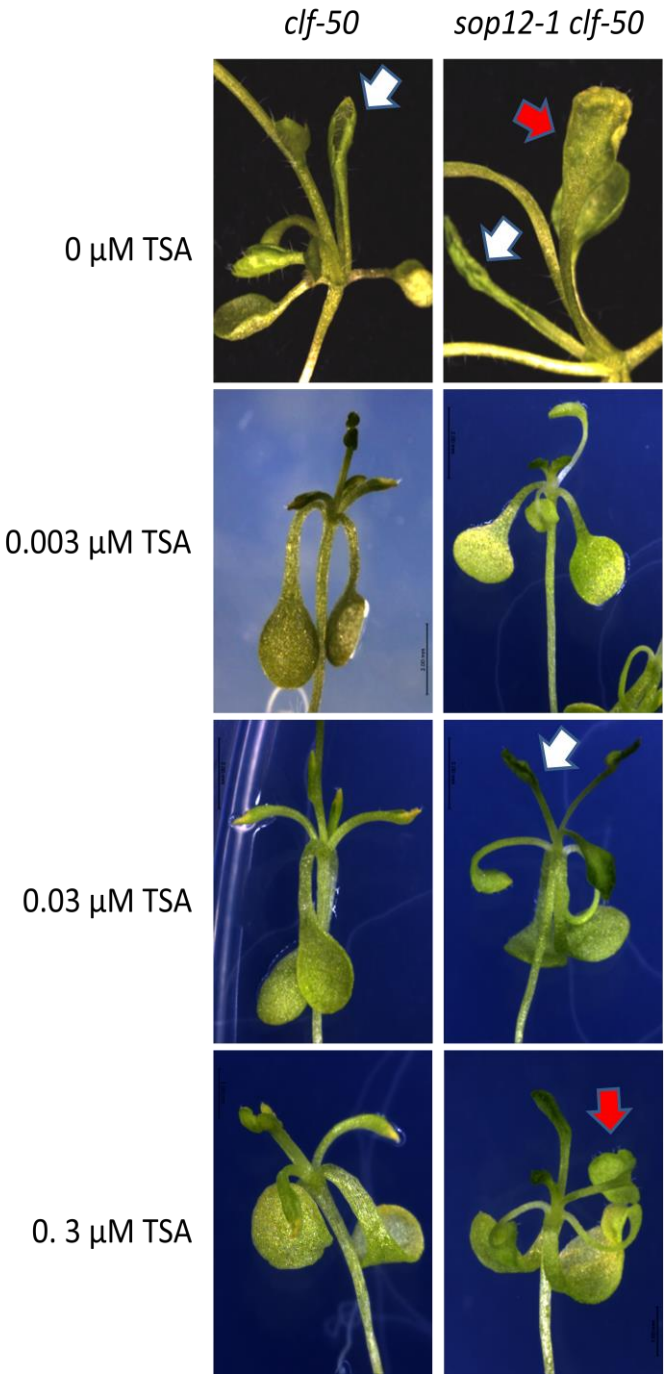


Figure 5-7. TSA treatment enhances leaf curling in *sop12-1 clf-50*.

The effect of histone deacetylase inhibitor TSA on the suppressed phenotype in *sop12-1 clf-50*. TSA treatment affects leaf polarity (Ueno et al., 2007), but did not apparently affect

sop12 phenotypes. White arrows indicate curled leaves, whereas the red ones point out leaves which are not up-curled. *clf-50* and *sop12-1 clf-50* plants were grown on MS plates with or without TSA for 14 days (TSA treatment) or 21 days (0 μM TSA) .

ii. AT5G24500 (SUP1) may be involved in a SOP12-containing complex which is important for Pc-G target expression

AT5G24500 is one of the unknown proteins in Y2H screening of SOP12 interactors. Interestingly, it turned out to be a candidate gene for another *lhp1* suppressor mutation which was identified in Dr. F. Turck's lab (Max Planck Institute, Cologne, Germany). The suppressor, which is not allelelic with *sop12*, has been designated *suppressor1* (*sup1*). Using next generation sequencing to identify the gene disrupted (Hartwig et al., 2012), two candidate genes were identified which were linked to *sup1* and carried DNA sequence alterations, one of which (*AT5G24500*) appears to correspond to the *SUP1* locus (F. Turck, personal communication). According to annotations on TAIR, SUP1 is so far a protein with function. It locates in nucleus at the sub-cellular level and expresses extensively in plant tissues. When the predicted amino acid sequence (Figure 5-8) was BLAST to NCBI database, no conserved domain was predicted by CDD database. SUP1 is conserved in limited species, such as *Arabidopsis lyrata*. The closest proteins in *Vitis vinifera* and *Populus trichocarpa* share less than 40% identity to SUP1. Alignment of SUP1-like orthologues showed potential conserved domains in C-terminus of SUP1 (Figure 5-9).

There were two independent clones that correspond to SUP1 in Y2H screening. Since the cDNA library used was reverse transcribed by random primers, each clone contains various parts of the gene. The fragment included in both clones, amino acid 185-322, may indicate the putative interacting domain (Figure 5-8). The predicted interacting domain in SUP1 overlaps with the potential conserved domains among SUP1-like proteins in other species (Figure 5-9). Further characterisation of SUP1 may shed light on understanding the mechanism how *sop12* suppresses *lhp1* or *clf*.

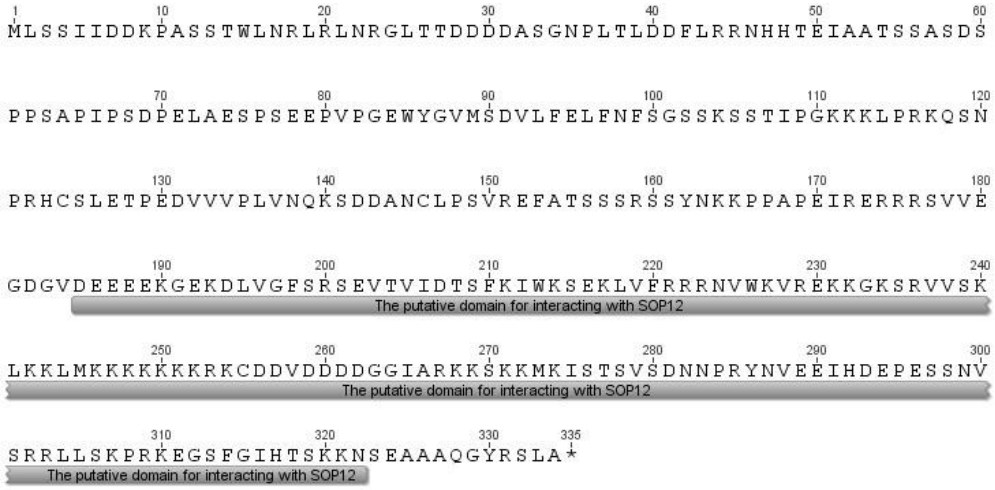


Figure 5-8. The predicted amino acid sequence of AT5G24500 (SUP1)

The amino acid sequence was translated from CDS published on *Arabidopsis* database (TAIR). The putative SOP12-interacting domain is underlined by a grey bar.

Figure 5-9. The alignment of SUP1 orthologues

Multiple alignment of SUP1 orthologues by NCBI. Proteins that share more than 35% identity with coverage more than 50% were chosen. Red texts indicate conserved amino acid residues, and the dash lines are amino acids that not present in proteins. Numbers in the square brackets show how many amino acids are omitted to align with other sequences. NP_197838.2, uncharacterized protein [*Arabidopsis thaliana*]; XP_002874202.1, hypothetical protein [*Arabidopsis lyrata*]; XP_003634173.1, uncharacterized protein [*Vitis vinifera*]; XP_002318455.1, predicted protein [*Populus trichocarpa*]; CAN80175.1, hypothetical protein VITISV_018394 [*Vitis vinifera*]

Surprisingly, another unknown protein identified in a yeast two hybrid screen as a potential CLF-interacting protein in Dr. D. Schubert's lab in University of Dusseldorf, Germany (unpublished data), was also a SOP12 interacting candidate. The interaction with Pc-G related proteins is unexpected because their roles are in the opposite directions, but it is possible that the binding inhibits the function of the partners. This result inspired the idea to test the interaction between SOP12 and Pc-G genes by Y2H. The constructs, including SWN lacking SET domain, LHP1without chromodomain, MSI1, VRN2, EMF2 and CLF in both Gateway compatible pGADT7 and pGBKT7, were kindly provided by Dr. D. Schubert. No interaction was detected in either of the reciprocal combinations of AD-SOP12 with BD-PcG or AD-PcG with BD-SOP12 (data not shown).

Identification of components of native complex by IP-Mass spectrometry (MS)

In Y2H screening experiment, SOP12 pulled out large number of potential interactors, which makes it difficult to identify which proteins are biologically relevant. Thus, to refine the interactors, I tried to purify the native SOP12-containing complex and identify the components by Mass spectrometry (MS) in the cooperation with the expertise in mass spectrometry of Prof. J. Rappsilber's (University of Edinburgh) laboratory. SOP12 complex was purified by using GFP-trap beads (Chromotek). These beads are conjugated with the variable region of camelid antibodies that recognise GFP. GFP-tagged SOP12 in inflorescence stems of *pSOP12::SOP12-GFP/sop12-1* transgenic plants was immunoprecipitated by the

beads for MS analysis. Inflorescence stems were used because there was less cleaved free GFP (Figure 5-2). A *35S::GFP* transgenic line was included as a negative control to exclude proteins interacting with GFP alone in this analysis.

Two independent experiments were performed. In the first experiment, proteins were extracted from non-cross-linked tissues. A small scale IP followed by Western blot analysis was tested first to ensure the SOP12-GFP was successfully pulled down (Figure 5-10). The result showed not only SOP12-GFP but also free GFP presented in the immunoprecipitated proteins.

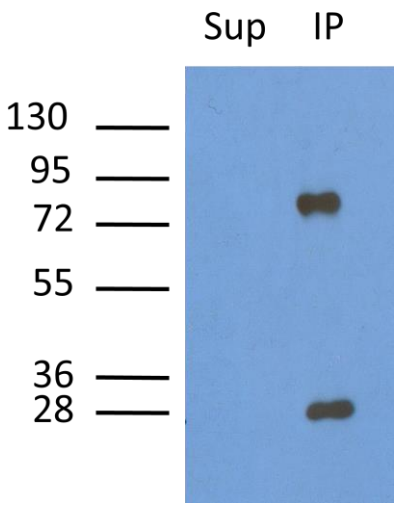


Figure 5-10. Small scale IP test of native protein extract

Western blot analysis of protein samples immunoprecipitated from non-cross-linked, above-ground tissues. The protein extracts of *pSOP12::SOP12-GFP/sop12-1* were immunoprecipitated by GFP-trap agarose beads, and the beads and supernatant after IP were analysed by Western blotting using a mouse monoclonal antibody against GFP. Sup,

proteins extracts after IP. IP, the proteins eluted from GFP-trap agarose beads. A noticeable fraction of free GFP was detected in the IP.

Although free GFP was included in the pulled down samples, the noise could be subtracted by comparing to *35S::GFP* negative controls. The large scale IP was done with, *35S::GFP-CLF/clf* and *pLHP1::LHP1-GFP/lhp1* as positive controls in addition to *35S::GFP*, *pSOP::SOP12-GFP/sop12-1*. The immunoprecipitated proteins were sent to Prof. J. Rappsilber's lab to be analysed by MS. Overall, 762 and 1991 peptides were identified from samples of LHP1-GFP and GFP-CLF, respectively. Among the peptides identified, several Pc-G proteins that are known to interact with LHP1 or CLF (Table 5-3) were included, indicating the immunoprecipitation and MS were successful.

Table 5-3. Several known interactors of LHP1 and CLF were identified by non-crosslinked IP-MS.

The numbers represent the number of peptides detected.

Accession	Description	LHP1-GFP	CLF-GFP
P93831	Histone-lysine N-methyltransferase CLF		64
O64827	Histone-lysine N-methyltransferase SUVR5		1
Q946J8	Chromo domain-containing protein LHP1	21	2
Q8L6Y4	Polycomb group protein EMF2	3	11
Q9LT47	Polycomb group protein FIE		27
Q9LYD9	Protein EMF1		1
O22467	WD-40 repeat-containing protein MSI1	8	13
O22607	WD-40 repeat-containing protein MSI4	3	1
Q9SU78	WD-40 repeat-containing protein MSI5	1	

However, only 372 peptides were identified from SOP12-GFP, and 330 of these peptides overlapped with *35S::GFP* control. To improve, the experimental procedure was then modified in two ways in addition to increasing the amount of material used. First, the samples were cross-linked by 1% formaldehyde to stabilise interactions between proteins. Secondly, nuclei were purified prior to protein extraction. Since the cleaved GFP in cytosol was washed away, less free GFP was detected in the small scale IP followed by Western blot analysis (Figure5-11).

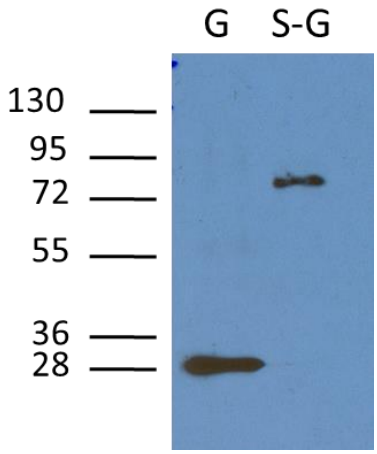


Figure 5-11. Small scale IP test of formaldehyde-cross-linked nuclei protein extract

Western blot analysis of protein samples immunoprecipitated from cross-linked inflorescence tissues. Inflorescence stems of *35S::GFP* and *pSOP12::SOP12-GFP/sop12-1* were cross-linked by 1% formaldehyde before nuclei were isolated for protein extraction. The nuclear extracts were immunoprecipitated by GFP-trap agarose beads and the immunoprecipitated proteins were analysed by Western blot using a mouse monoclonal antibody to GFP. G, proteins immunoprecipitated *35S::GFP*; S-G, proteins immunoprecipitated from *pSOP12::SOP12-GFP/sop12-1*.

In the second experiment, IP was done in *35S::GFP* and two different SOP-GFP transgenic plants, p*SOP12::SOP12/sop12-1* and *35S::SOP12-GFP/sop12-1*. 3104 peptides were identified from samples of *pSOP12::SOP12-GFP/sop12-1*, and 316 are SOP12-specific after subtracting peptides in the GFP negative control. By comparison, 2849 peptides were identified in *35S::SOP12-GFP/sop12-1*. 203 are SOP12-specific interactors, and 103 are overlap with peptides from *pSOP12::SOP12-GFP/sop12-1* (Table 5-4). In total, 122 proteins were identified, but SOP12 itself was missing, which is worrying. Surprisingly, chloroplastic proteins were again over-represented (13.7%) compared with nuclear proteins (5.8%), even though nuclear proteins should be enriched in the input proteins. Gene ontology analysis showed the top three categories in biological process which those SOP12 interactors fall into are metabolic process (23.3%), response to stimulus (14.8%) and cellular component organisation or biogenesis process (8.5%). 4.4% of the candidates are categorised as participating in developmental processes. It is unclear in which developmental process SOP12 is involved. There are several histone related proteins (H2A and H2B variants, HDA15 and CHR11) on the list, which implies SOP12 may participate in histone modification.

Table 5-4. SOP12 interacting proteins identified by Mass Spectrometry.

Protein accession number, the description, and the numbers of peptides identified from immunoprecipitated samples of transgenic plants *pSOP12::SOP12-GFP/sop12-1* (*pSOP12*) and *35S::SOP12-GFP/sop12-1* (*35S*) are shown. Chloroplast and mitochondrion proteins are indicated in the description.

Accession Number	**Description**	***pSOP12***	***35S***
P42732	30S ribosomal protein S13, chloroplastic, RPS13		3
Q94AR8	3-isopropylmalate dehydratase, IIL1	1	
Q94BT2	Auxin-induced in root cultures protein 12, AIR12	4	
O81742	Beta-adaptin-like protein C, BETAC-AD	3	2
O04151	Calreticulin-1, CRT1		5
Q38858	Calreticulin-2, CRT2		1
O22784	Cysteine-rich repeat secretory protein 11, CRRSP11	2	
Q08298	Dehydration-responsive protein RD22, RD22	4	
Q84JU6	**E3 ubiquitin-protein ligase HOS1, HOS1**		1
Q9LTT8	Enhancer of mRNA-decapping protein 4, VCS		1
Q9FK25	Flavone 3~-O-methyltransferase 1, OMT1	3	1
Q9FHX5	Glucan endo-1,3-beta-glucosidase 10, AT5G42100	2	
P55228	Glucose-1-phosphate adenylyltransferase small subunit, chloroplastic, APS1		2
Q42472	Glutamate decarboxylase 2, GAD2	1	
Q42472	Glutamate decarboxylase 2, GAD2	1	
Q9SR40	Laccase-7, LAC7		1
O22898	Long chain acyl-CoA synthetase 1, LACS1	3	
Q9SE60	Methylenetetrahydrofolate reductase 1, MTHFR1	2	2
Q9SIU0	NAD-dependent malic enzyme 1, mitochondrial, NAD-ME1		1
A4GSN8	Nuclear-pore anchor, NUA		3

Q9FR44	Phosphoethanolamine N-methyltransferase 1, NMT1	2	
P50318	Phosphoglycerate kinase 2, chloroplastic, AT1G56190	1	
Q38882	Phospholipase D alpha 1, PLDALPHA1		6
O82660	Photosystem II stability/assembly factor HCF136, chloroplastic, HCF136		1
O04848	Probable histone H2AXa, AT1G08880	4	
Q9M8T0	Probable inactive receptor kinase AT3G02880, AT3G02880	3	2
Q9LVI6	Probable inactive receptor kinase RLK902, RLK902	3	
Q93Z70	Probable N-acetyl-gamma-glutamyl-phosphate reductase, chloroplastic, AT2G19940	3	3
Q42384	Protein pleiotropic regulatory locus 1, PRL1		3
Q9LTR2	Protein TRIGALACTOSYLDIACYLGLYCEROL 2, chloroplastic, TGD2	1	
Q8RWY3	**Putative chromatin-remodeling complex ATPase chain, CHROMATIN-REMODELING PROTEIN 11, CHR11**		9
Q9FK68	Ras-related protein RABA1c, RABA1C	3	
Q9ZPX5	Succinate dehydrogenase [ubiquinone] flavoprotein subunit 2, mitochondrial, SDH1-2	1	1
Q9SJT7	Vacuolar proton ATPase a2, VHA-a2	1	
Q8W4S4	Vacuolar proton ATPase a3, VHA-a3	5	2

It is confusing if the high contents of proteins localised in chloroplast or involved in metabolic process mean that SOP12 has functions in these aspects or it is an artefact happened by chance. When taking candidates in the interactome of GFP alone to be analysed by gene ontology, the top categories were similar (Table 5-5), making it less likely that SOP12 is a metabolic protein. It was noticed that higher percentages of nuclear proteins were discovered in SOP12-GFP MS and Y2H compared to GFP alone. Moreover, the percentage of proteins involving in developmental processes

was higher in candidates of Y2H screening than in IP-MS analysis of SOP12-GFP or GFP alone.

Table 5-5. The comparison of gene ontology analysis results from GFP-MS, SOP12-GFP-MS and Y2H screening.

Cellular compartment	GFP MS	SOP12-GFP MS	Y2H
Chloroplast	11.2% (202)	11.6% (41)	14.2% (36)
Nuclear	3.1% (56)	5.1% (18)	6.7% (17)
Biological process			
Cellular process	31.2% (270)	30.5% (76)	27.0% (51)
Metabolic process	25.3% (219)	23.3% (60)	20.1% (38)
Response to stimuli	14.1% (122)	14.8% (33)	15.0% (29)
Development	4.7% (41)	4.4% (11)	7.9% (15)

5.3 The effect of sop12 upon chromatin modifications

The functions of Pc-G and trx-G are highly associated with histone modification. For instance, the Pc-G complex PRC2, the trx-G members ATX1, and EFS, are all histone methyltransferases (Alvarez-Venegas et al., 2003; Xu et al., 2008; Hennig and Derkacheva, 2009). Mutations in Pc-G histone modifiers like *CLF* result in reduction in the repressive mark H3K27me3 at target genes, and accumulation of active marks such as H3K4me3 together with loss of the transcriptional repression of Pc-G targets.

SOP12 functions as a trx-G member, and is essential for the mis-expression of Pc-G targets in *clf*. One of the possible mechanisms through which *SOP12* regulates Pc-G targets is that *SOP12* also plays a role in histone modification, as many trx-G genes do. In order to test this possibility, H3K27me3, H3K4me3, H3K36me3 and H3 acetylation (H3ac) at Pc-G targets were examined using ChIP.

H3K27me3 is not restored in *sop12 clf*

H3K27me3 is a key mark in the Pc-G repression context. The PRC2 complex is responsible for generating H3K27me3 at targets via the catalytic subunit CLF or the related SWN and MEA. Subsequently, this mark is bound by LHP1 to recruit PRC1 complex to stabilise the repression (Bemer and Grossniklaus, 2012). Both H3K27me3 and repression of target genes is lost in *clf*, whereas the repression is restored in *sop12 clf* (Figure 4-4 and 4-5). I wondered if SOP12 might be involved in removing H3K27me3, for example, by activating H3K27 demethylase activity, or inhibiting CLF/SWN/MEA histone methyltransferases such that loss of SOP12 activity would leads to restored H3K27me3, resulting in repression of Pc-G targets in *sop12 clf*. The H3K27me3 level in two Pc-G targets, *SEP3* and *AG* was examined by ChIP. The result showed this modification was not restored in *sop12-1 clf-50* (Figure 5-12), which is in agreement with the implication revealed from the fact that *sop12* suppresses *lhp1*. In other words, in *lhp1* mutants, Pc-G targets are mis-expressed (Kotake et al., 2003) yet the H3K27me3 level remains unchanged compared to wild type (Turck et al., 2007) , so the suppression of *lhp1* mutants by *sop12* is unlikely to

act by increasing H3K27me3 levels. If *sop12* suppresses *clf* by elevating the H3K27me3 level, it would not suppress *lhp1*. The relative amount of H3K27me3 in *sop12* single mutants was reduced compared to *Ws*. The reduction of H3K27me3 was unexpected because the expression levels of the observed loci *AG* and *SEP3* were not increased (Figure 4-4 and 4-5).

It is also possible SOP12 may activate a demethylase to regulate genes which act upstream of Pc-G target, so that no change of H3K27me3 was observed in Pc-G targets themselves. In this case, H3K27me3 would accumulate in targets of the demethylase in *sop12* background. To test this hypothesis, H3K27me3 level in *TOUCH 4* (*TCH4*) and *AT3G04210*, two targets of the recently identified H3K27me3 histone demethylase RELATIVE OF EARLY FLOWERING 6 (REF6) (Lu et al., 2011), was checked. H3K27me3 in *TCH4* remained unchanged in the comparison of *sop12-1 clf-50* to *clf-50* and *sop12-1* to *Ws* (data not shown). In another target *AT3G04210*, it was decreased rather than increased in *sop12-1* and *sop12-1 clf-50* (data not shown). Together, these results suggest that SOP12 does not act by modulating the H3K27me3 demethylation activity of REF6. Nonetheless, it is not ruled out that SOP12 might have a role with other H3K27demethylases.

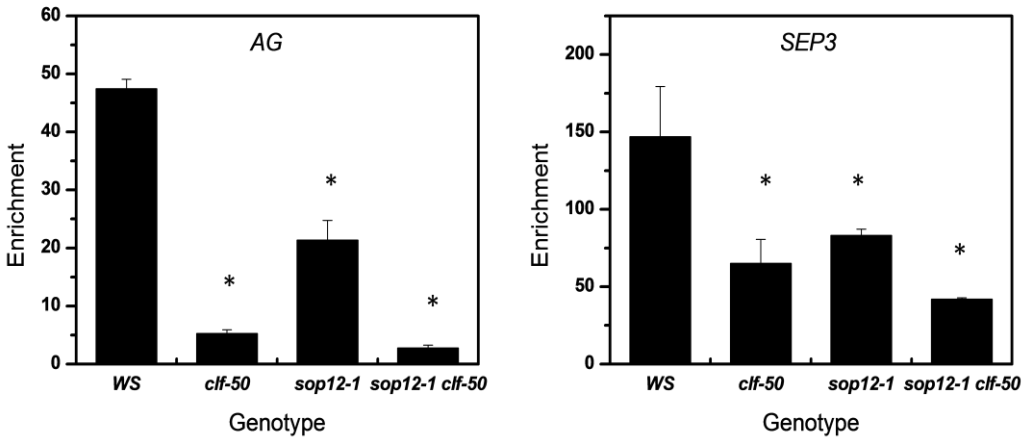


Figure 5-12. H3K27me3 in *AG* and *SEP3*

Fold enrichment of H3K27me3 in two Pc-G targets, *AG* and *SEP3*, relative to that at the non-enriched 5S ribosomal control genes, were demonstrated by ChIP method. The enrichment of H3K27me3 in *clf-50* is strongly reduced in comparison to *Ws*. In *sop12-1 clf-50*, the levels remain low at both loci. 14-day-old seedlings grown in LD were used and three biological replicates were included. Error bars represent 1 SE. Stars represent significantly difference compared with *Ws* ($p<0.005$).

H3K4me3 is reduced at *SEP3* but not at *AG* in *sop12-1 clf-50*

In *sop12 clf*, the reduction of Pc-G target expression may be due to decrease in active marks on chromatin. Because ULT1 was shown to participate in deposition of H3K4me3 at *AG* and *sop12* enhanced *ult1*, H3K4me3 at *AG* and *SEP3* was examined in four genotypes (Figure 5-13). H3K4me3 level was raised in *clf-50* in both loci relative to wild-type (*Ws*), which is consistent with previous observations and the increase in target gene expression (Saleh et al., 2007). The high level of H3K4me3

was decreased in *sop12-1 clf-50* at *SEP3*, but did not change prominently at *AG* considering the error. In comparisons between Ws and *sop12-1* single mutants, the H3K4me3 level remained unchanged at *AG*, and was decreased in a small amount at *SEP3*.

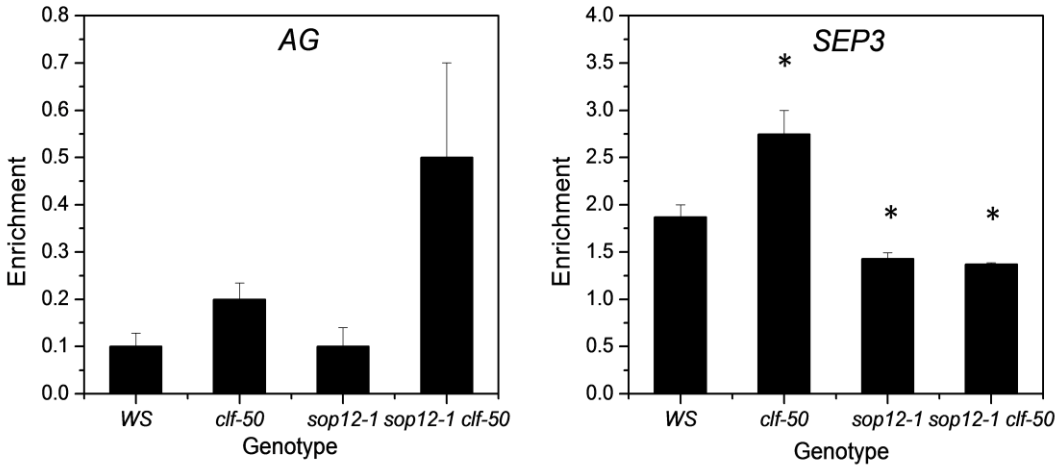


Figure 5-13. ChIP analysis of H3K4m3 levels at *AG* and *SEP3*.

Fold enrichment of H3K4me3 at *AG* and *SEP3*, relative to that at *ACTIN* control, were examined by ChIP. Because *ACTIN* is an H3K4me3-enriched gene but *AG* and *SEP3* are not, the values of fold enrichment are low. The enrichment of H3K4me3 in *clf-50* is increased in the comparison to *Ws* in both loci, whereas in *sop12-1* it shows no significant difference. In *sop12-1 clf-50*, the accumulation is not significantly changed in *AG* but decreased in *SEP3* relative to *clf-50*. The decrease at *SEP3* is correlated with reduced *SEP3* expression in *sop12-1 clf-50* (Figure 4-4). 14-day-old seedlings grown in LD were used and three biological replicates were included. Error bars represent 1 SE. Stars represent significantly difference compared with *Ws* ($p<0.005$).

The accumulation of H3K36me3 is diminished in *sop12 clf*

H3K36me3 is a histone mark located in the gene body of actively transcribed genes. It is predominantly deposited by EFS, since Western blot analysis of total histone extracts indicates that H3K36 methylation is globally reduced in *efs* (Xu et al., 2008). Because *sop12* enhanced *efs*, SOP12 and EFS may participate in the same pathway or work synergistically. Therefore the effect of *sop12* mutation on H3K36me3 was examined. H3K36me2 was also tested but the experiments were not successful probably due to the antibody used was not suitable. Global H3K36me3 was decreased in *efs*, as shown in previous study (Figure 5-14, (Xu et al., 2008), but was not further reduced in *efs sop12* compared to the *efs* single mutant. Neither did it decrease in *sop12* mutants compared to corresponding wild types.

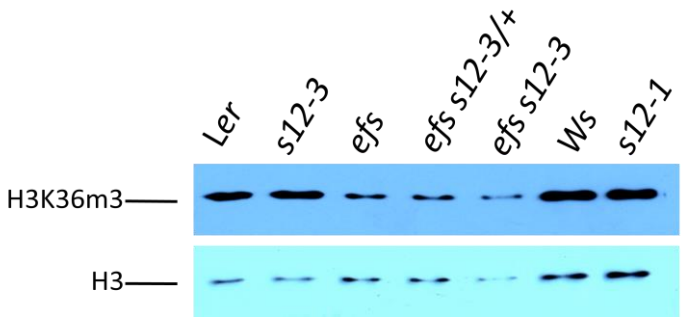


Figure 5-14. H3K36me3 in *efs* and *sop12* mutants.

Western blot analysis of global H3K36me3 changes in *efs* and *sop12* mutants. H3K36me3 is decreased in *efs* and *efs sop12-3*/+ but not clearly reduced in *efs sop12-3* double mutant. *sop12* mutants do not show reduction in H3K36me3 compared to corresponding wild type plants. After detecting H3K36me3, the membrane was stripped and blotted with H3 antibody

to show the amount of loading in each lane. *s12-3*, *sop12-3*; *s12-1*, *sop12-1*. *sop12-3* and *efs* mutants are in L*er* background whereas *sop12-1* is in *Ws* background.

Even though *sop12* mutation did not affect global H3K36me3, it is still possible that SOP12 participates in H3K36 methylation at a sub set of specific gene loci. Thus H3K36me3 at *AG* and *SEP3* was scrutinized. H3K36me3 level was higher in *clf-50* relative to *Ws*, which indicates H3K36me3 responds to the reduction of H3K27me3 in a similar trend as H3K4me3. There was no difference between *sop12* and *Ws*. Interestingly, H3K36me3 was decreased in *SEP3* but not in *AG* in *sop12 clf* compared to *clf* (Figure 5-15), which may reflect that the degree of transcriptional down-regulation was greater for *SEP3* than *AG* (Figure 4-4 and 4-5).

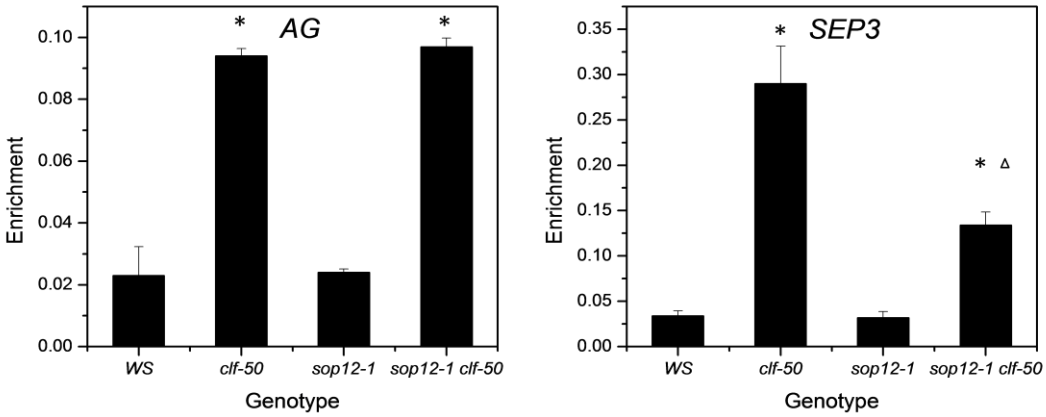


Figure 5-15. ChIP analysis of H3K36me3 at *AG* and *SEP3*

Fold enrichment of H3K36me3 at *AG* and *SEP3* relative to that at *ACTIN*, were examined by ChIP. The enrichment of H3K36me3 in *clf-50* is increased in the comparison to *Ws*. In

sop12-1 clf-50, the accumulation is decreased in *SEP3* but not in *AG*, which is correlated with their gene expression level. 14-day-old seedlings grown in LD were used and three biological replicates were included. Error bars represent 1 SE. Stars and triangles represent significantly difference compared with *Ws* and *clf-50*, respectively ($p<0.005$).

Histone acetylation is decreased in *sop12-1 clf-50*

Because histone deacetylases were implicated as SOP12 interacting candidates both in Y2H and MS, I wondered if *sop12* mutation affects histone acetylation in Pc-G targets. SOP12 is a positive regulator in terms of mis-expression of Pc-G in *clf* mutant, so the possible role of SOP12 is to inhibit histone deacetylases. In this hypothesis, it was expected H3ac would decrease at Pc-G targets in *sop12* compared to wild type plants, or *sop12 clf* to *clf*. *AG*, *SEP3*, *FT*, and *FLC* chromatin was analysed by ChIP. It seemed H3ac were scarce in *FLC* and *FT* since the percentage of DNA precipitated from the input DNA (%input) was less than 0.01% (data not shown). *AG* and *SEP3* are probably not targets, or are very weak targets of H3ac. The percentage of input in *AG* after chromatin had been immunoprecipitated was low (around 0.05%). The percentage of input in *SEP3* reached 0.1% in *Ws*. The enrichment of H3ac in *AG* did not change much in four genetic backgrounds. On the other hand, the relative amount in *clf-50* to *sop12-1 clf-50* at *SEP3* was reduced, which is similar to what observed in H3K4me3 and H3K36me3 (Figure 5-13 and 5-15). However, H3ac was not decreased in *sop12-1* compared to *Ws* (Figure 5-16). These results fail to support the expectation that H3ac level might be lower if SOP12 is an inhibitor of histone deacetylases.

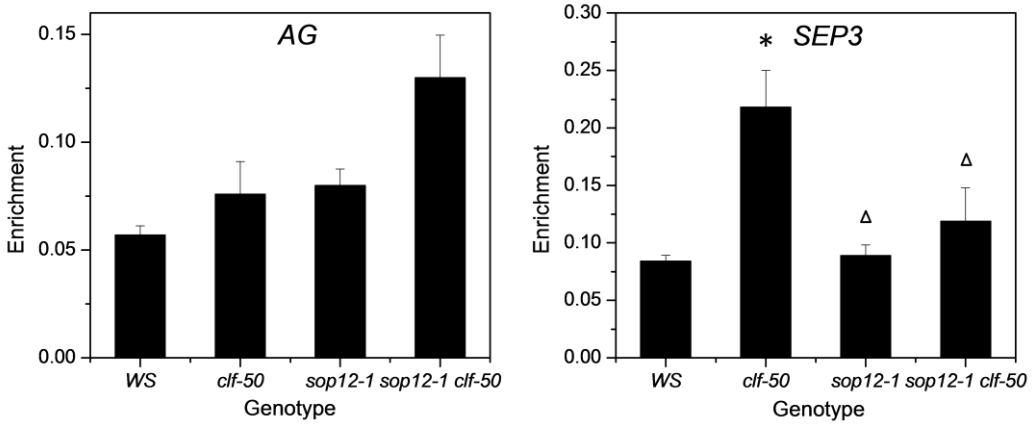


Figure 5-16. ChIP analysis of H3ac in *AG* and *SEP3*

Fold enrichment of H3ac at *AG* and *SEP3*, relative to loading control *SAM2*, were examined by ChIP. *SAM2* was chosen according to previous study (Bond et al., 2009). The enrichment of H3K36me3 in *clf-50* is increased in the comparison to *Ws*. In *sop12-1 clf-50*, the accumulation is decreased in *SEP3* but not in *AG*, which is correated to their gene expression level. 14-day-old seedlings grown in LD were used and three biological replicates were included. Error bars represent 1 SE. Stars and triangles represent significantly difference compared with *Ws* and *clf-50*, respectively ($p<0.005$).

5.4 Global transcriptional profiling of SOP12 targets

The expression levels of Pc-G targets shown by real time RT-PCR suggest SOP12 acts upstream of CLF targets, similar to what trx-G genes do (Figure 4-4 and 4-5). However, I wondered whether SOP12 simply regulates genes affecting flowering time and thus suppresses *clf* through effects on *SEP3*. Global transcriptional

profiling allows us to address the question whether SOP12 is required generally for CLF targets' activity or just for genes associated with regulation of flowering time such as *SEP3* and *FT* etc. Meanwhile, transcriptome analysis may reveal the function of *SOP12* in wild type background. In order to clarify these questions, transcriptomes of *Ws*, *clf-50*, *sop12-1* and *sop12-1 clf-50* double mutant were analysed by RNA-seq. Total RNA was extracted from 10-day-old seedlings growing on MS tissue culture plates in LD, and the RNA samples were processed by the next generation sequencing facility at the Max Planck Institute, Cologne, Germany as part of a collaboration with Dr F. Turck. Biological triplicate samples were included in all four genetic backgrounds. The preliminary analysis of the data including normalisation and statistical analysis comparison of the samples was performed by in house bioinformatics service at the Max Planck Institute. The relative abundance of RNA reads, which was represented as [$\log_2$ (fold change)] ($\log_2$FC), was calculated between the comparison of *sop12-1* to *Ws* (s/W), *clf-50* to *W*s (c/W), *sop12-1 clf-50* double mutant to *Ws* (d/W), the double mutant to *clf-50* (d/c), and the double mutant to *sop12-1* (d/s). The false discovery rate (FDR) is a statistic parameter for multiple hypothesis testing used in multiple comparisons. This cut-off was set as < 0.05.

To validate the quality of the analysis, the genes up-regulated in c/W was used as an indicator to see if known CLF targets show up in the list. In the comparison c/W, 441 genes were up-regulated and 260 were down-regulated. More genes were up-regulated in *clf-50* is in agreement with the role of CLF as a gene expression repressor. The top 20 up-regulated genes are listed in Table 5-6. *AG* and *MEA* are identified as CLF

targets. In addition, several well-known targets, such as *SEP3*, *AP3* and *PISTILLATA* (*PI*), etc. are included in the up-regulated group (data not shown).

Table 5-6. Top 20 up-regulated genes in the transcriptome comparison *clf-50/Ws*

Two well-know CLF targets *AG* and *MEA* are high lighted.

ATG number	**Description**	**Log_2FC**
AT3G46660	UGT76E12, UDP-glucosyl transferase 76E12	9.499
AT3G46658	Function unknown	8.390
AT3G49270	Function unknown	8.113
AT5G35935	Transposable element gene	7.931
AT1G27045	Homeobox-leucine zipper protein family	7.810
AT1G67270	Zinc-finger domain of monoamine-oxidase A repressor R1 protein	7.405
AT3G20210	DELTA-VPE, DELTAVPE, delta vacuolar processing enzyme	7.311
AT1G67455	F-box and associated interaction domains-containing protein	7.305
AT2G46960	CYP709B1, cytochrome P450, family 709, subfamily B, polypeptide 1	7.272
AT5G24910	CYP714A1, cytochrome P450, family 714, subfamily A, polypeptide 1	7.096
AT1G12010	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	7.020
AT2G42830	AGL5, SHP2, K-box region and MADS-box transcription factor	6.596
AT3G20520	SVL3, SHV3-like 3	6.465
AT5G12270	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	6.402
AT2G27390	proline-rich family protein	6.206
AT1G67160	F-box family protein with a domain of unknown function (DUF295)	6.060
AT1G67390	F-box family protein	5.933
AT4G18960	**AG, K-box region and MADS-box transcription factor family protein**	5.878
AT5G56970	ATCKX3, CKX3, cytokinin oxidase 3	5.783
AT1G02580	**EMB173, FIS1, MEA, SDG5, SET domain-containing protein**	5.782

According to the trend of expression change in real time PCR analysis, SOP12 is likely to be a positive gene regulator**.** In s/W, 358 genes are mis-regulated, which is

about half of the number of mis-regulated genes in clf (701 genes). This is consistent with the mild phenotype of *sop12* mutants relative to *clf*. Among the 358 genes, 50.3% (180) of the genes were down-regulated and 49.7% (178) are up-regulated, which indicates SOP12 is perhaps not a gene repressor like CLF.

When comparing up-regulated genes in c/W and d/W, there are 119 genes in common, which means 73% (322) mis-expressed genes in *clf-50* were down-regulated to wild type level in *sop12-1 clf-50* (Figure 5-17). Furthermore, among 119 genes that were up-regulated in c/W and stayed up-regulated in d/W, 20 of them were actually down-regulated in d/c. These 20 genes are less mis-expressed due to *sop12* mutation. To sum up, 77.5% (342) CLF targets were down-regulated by *sop12* mutation in *clf* background (Figure 5-17). Therefore SOP12 is likely to be a gene activator, especially in the absence of CLF. Moreover, the 342 CLF targets that were down-regulated by *sop12* mutation contain not only flowering genes but also genes involved in embryo development such as *MEA*, *SEEDSTICK* (*STK*), and *ABA INSENSITIVE 5* (*ABI5*). This illustrates that SOP12 is a general activator of CLF targets instead of a specific regulator of flowering time.

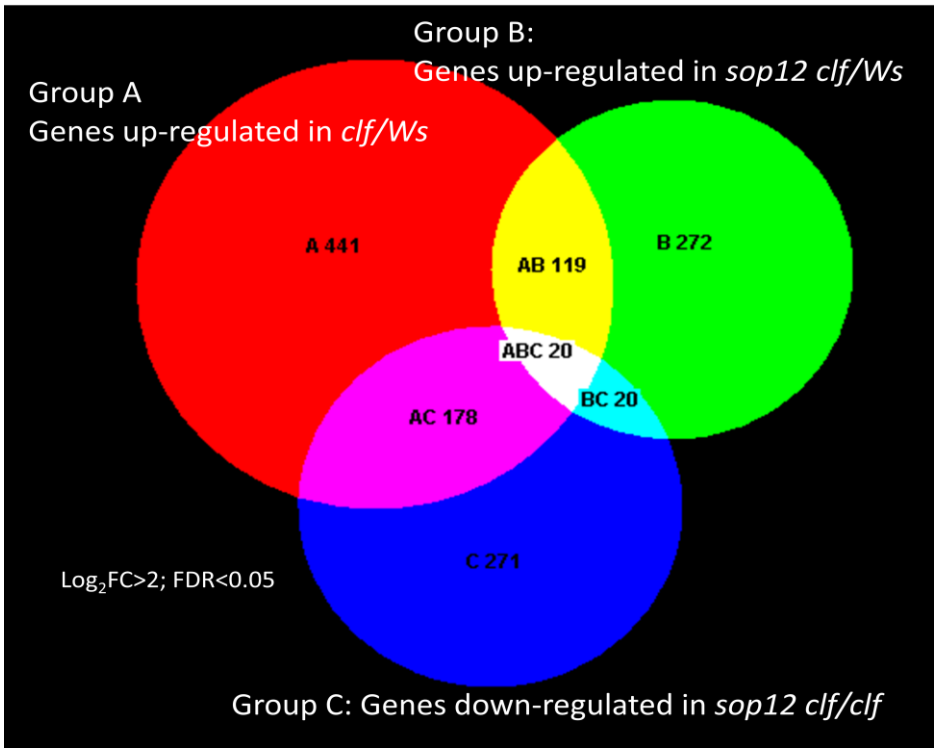


Figure 5-17. Venn diagram of numbers of mis-regulated genes in *clf*, *sop12* and *sop12 clf* transcriptomes.

Group A: genes that up-regulated in the comparison of *clf-50/Ws* (c/W); Group B: genes that are up-regulated in the comparison of s*op12-1 clf-50*/*Ws* (d/W); Group C: genes that down-regulated in the comparison of *sop12-1 clf-50/clf-50* (d/c). Mis-regulated genes were filtered by the cut off $\log_2$FC>2 and FDR<0.05. A large number of genes which up-regulated in *clf* are not up when *sop12* is also mutated (red and pink area). 20 of them are still up-regulated but the amplitude is decreased (white area). Overall, 77.5% of the CLF targets are down-regulated by *sop12* mutation.

Since SOP12 is likely to be an activator, I tried to find genes regulated by SOP12 by comparing the transcriptomes in *sop12* mutants. The targets of SOP12 should decrease in the absence of SOP12. In other words, the expression of those targets is

reduced both in s/W and in d/c. There are 180 and 271 down-regulated genes in s/W and d/c, respectively and only 15 genes are overlapped. 5 of them are outliers, which have 0 read in the raw RNA reads, resulting in massive log_2FC. The other 10 genes are listed in Table 5-7. Noticed that *AP3*, which is a MADS box transcription factor essential in flower development, and also a target of Pc-G (Calonje et al., 2008), is a strong target of SOP12.

Table 5-7. SOP12 target candidates upon the assumption that SOP12 is a gene activator.

ATG number	**Annotation**	**s/W log_2FC**	**d/c log_2FC**
AT1G09932	Phosphoglycerate mutase family protein	-2.01	-2.95
AT1G56060	unknown protein	-2.02	-3.85
AT2G04040	ATDTX1, TX1, MatE efflux family protein	-2.87	-2.12
AT2G20142	Toll-Interleukin-Resistance (TIR) domain family protein	-3.46	-2.78
AT3G22910	ATPase E1-E2 type family protein	-2.60	-3.17
AT3G54340	**AP3, K-box region and MADS-box transcription factor**	-4.14	-3.78
AT3G60920	unknown protein	-2.96	-3.51
AT4G12490	Seed storage 2S albumin superfamily protein	-4.15	-4.36
AT4G12500	Seed storage 2S albumin superfamily protein	-3.08	-4.17
AT4G15620	Uncharacterised protein family (UPF0497)	-2.72	-6.45

It was quite surprising that only few genes were overlapped in the up-regulated genes in s/W and d/c, which made me wonder whether SOP12 regulates genes just when CLF is absent, which is similar to a feature of trx-G proteins that they compete with Pc-G proteins to regulate Pc-G gene expression. Therefore, based on the assumption that SOP12 is a trx-G component mechanistically, SOP12 targets would behave

oppositely in the comparisons of d/c and d/s. d/c represents the effects of *sop12* mutation when *CLF* (a Pc-G gene) is absent, whereas d/s represents the effect of *clf* mutation when *SOP12* (a trx-G gene) is absent. The target candidates which fit the assumption are list in Table 5-8. Two interesting candidates, *AP3* and *SHATTERPROOF2* (*SHP2*), are both involved in floral organ development.

Table 5-8. SOP12 target candidates based on the assumption that *SOP12* acts like a trx-G gene

ATG number	**Annotation**
AT1G08860	BON3, Calcium-dependent phospholipid-binding Copine family protein
AT1G17710	AtPEPC1, PEPC,. Pyridoxal phosphate phosphatase-related protein
AT1G51270	Structural molecules;transmembrane receptors;structural molecules
AT2G23310	ATRER1C1, Rer1 family protein
AT2G23370	Protein of unknown function
AT2G23390	Protein of unknown function
AT2G42830	**AGL5, SHP2, K-box region and MADS-box transcription factor family protein**
AT3G20210	DELTA-VPE, DELTAVPE, delta vacuolar processing enzyme
AT3G46660	UGT76E12, UDP-glucosyl transferase 76E12
AT3G54340	**AP3, ATAP3, K-box region and MADS-box transcription factor family protein**
AT4G24000	ATCSLG2, CSLG2, cellulose synthase like G2
AT4G36740	ATHB40, HB-5_HB40, homeobox protein 40
AT4G37780	ATMYB87, MYB87, myb domain protein 87
AT5G12270	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT5G24910	CYP714A1, ELA1, cytochrome P450, family 714, subfamily A, polypeptide 1
AT5G60140	AP2/B3-like transcriptional factor family protein

By the reason that SOP12 has a role in gene activation, it is intriguing if SOP12 binds directly to the genes which it activates. A ChIP experiment was conducted using the

anti-GFP antibody against GFP tagged SOP12 in a *35S::SOP12-GFP/sop12-1* transgenic line. Samples from *35S::GFP* and *pLHP1::LHP1-GFP/Col FRI*, a gift from Dr. J. Finnegan, were used as the negative and positive control, respectively. A quick ChIP test using seedlings of those lines was performed. The target regions expanded from 5' upstream to 3' down-stream of the examined loci, including *FLC*, *FT*, *AG* and *SEP3*. After the test, the target regions were narrowed down to one region of the 5' region and the one of the gene body based on the preliminary result. However, no consistent enrichment in any of the loci examined was found (one example is shown in Figure 5-18). By contrast, the binding of LHP1-GFP was observed, which suggests the ChIP technique was successful.

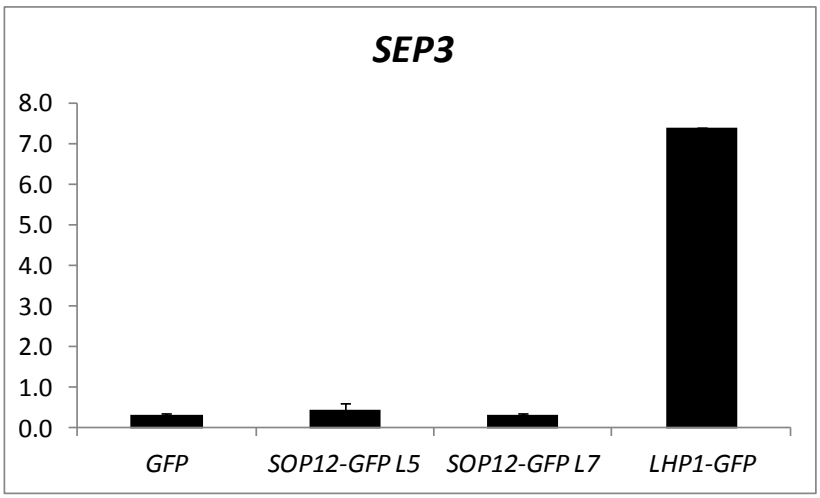


Figure 5-18. The binding of free GFP and GFP-tagged proteins to *SEP3* in seedlings

Enrichment (% input of target divided by % input of 5S ribosomal DNA, the loading control) of GFP and GFP-tagged SOP12 (SOP12-GFP) and LHP1 (LHP1-GFP) at *SEP3*, were demonstrated by ChIP method. Two independent transgenic lines harbouring 35S::SOP12-GFP in sop12-1 background were used (L5 and L7). The enrichment of

LHP1-GFP is significantly increased compared to the free GFP control, but no enrichment is shown of SOP12-GFP. 14-day-old seedlings grown in LD were used and three biological replicates were included except for LHP1-GFP. Error bars represent 1 SE.

After discussing with Dr. J. Goodrich, I tried to look at the binding of SOP12 to Pc-G targets in inflorescence meristem-enriched tissues, because one of the target candidates of SOP12 revealed from RNA-seq is *AP3*, and it expresses in inflorescence. SOP12 may bind to and activate *AP3* expression in inflorescence meristem instead of seedlings. Unfortunately, no significant enrichment was observed, either (Figure 5-19). In inflorescence tissues, the binding of LHP1-GFP was weaker, which is consistent with the scenario that *AP3* is de-repressed in floral tissues. Possibly SOP12 does not bind to these loci, or its gene activation role is not delivered through binding to DNA. Also, SOP12 might bind to its targets only in the condition *CLF* is absent. This assumption will be tested by looking at the binding of SOP12-GFP in *clf* background.

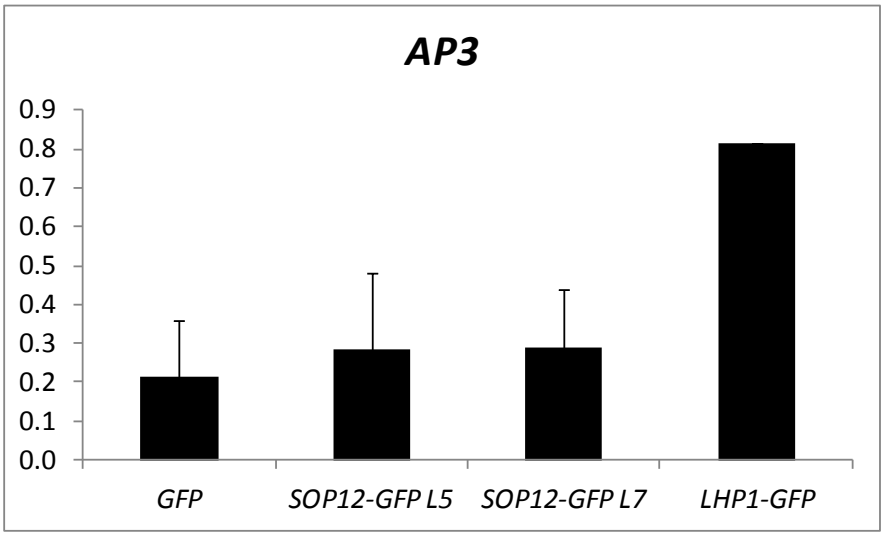


Figure 5-19. The binding of GFP-tagged proteins to *AP3* in inflorescence tips

Enrichment (% input of target divided by % input of 5S ribosomal DNA, the loading control) of GFP and GFP-tagged SOP12 (SOP12-GFP) and LHP1 (LHP1-GFP) at *AP3*, were demonstrated by ChIP method. Two independent transgenic lines (L5 and L7) harbouring *35S::SOP12-GFP* in *sop12-1* background were used. The enrichment of LHP1-GFP is slightly abundant compared to the free GFP control, but no enrichment is shown of SOP12-GFP. 14-day-old seedlings grown in LD were used and three biological replicates were included except for LHP1-GFP. Error bars represent 1 SE.

5.5 Summary and discussion

SOP12 is a nuclear protein and expressed extensively through the developmental stages of plants, especially in young leaves, young flowers, and axillary buds. Y2H test indicated it may interact with EFS, which is consistent with the genetic interaction between *SOP12* and *EFS*. In addition, SOP12 is likely to form homo-dimers or multi-mers within cells, but does not, or at least not directly, interact with ULT1 and

ULT2. A Y2H screening pulled out an unknown protein, SUP1, which was independently identified in a genetic screen as a *lhp1* suppressor. To validate the *in vitro.* interactome suggested by Y2H, GFP tagged native SOP12 complex was purified by IP and analysed by MS. MS data did not fully support the Y2H results: there were no exact matches between the list of SOP12 interactors identified by MS and Y2H, but a few functionally close-related proteins, like HDA15, another histone deacetylase, were on the peptide list in MS.

All three active marks were decreased in *sop12-1 clf-5*0 compared to *clf-50* (Figure 5-13, 5-15 and 5-16), but this did not apply for the level of H3K27me3, which agrees with the hint from *lhp1* suppression. In terms of phenotypes, gene expression and histone marks, the effects of mutation at *SOP12* are more apparent in the *clf* mutant background. The *sop12* single mutant has a very weak phenotype, whereas *sop12 clf* is clearly distinguishable from *clf* morphologically (Figure 1-3 and 3-4). Consistent with the phenotypes, the transcript levels of Pc-G targets do not reduce significantly when mutating *SOP12* in wild type background, but decrease greatly when in *clf* background (Figure 4-4 and 4-5). The activity of SOP12 is highly sensitive to the presence of CLF, which implies SOP12 acts in a similar way to trx-G components.

SOP12 is a global activator of CLF targets

From transcriptome analysis, it was noticed that SOP12 is required for mis-expression of CLF targets when *clf* is mutated. About 80% of mis-expressed

genes in *clf* are down-regulated when *sop12* is mutated simultaneously, and about 70% are down to nearly the same levels as in wild type. Moreover, the role of SOP12 is not limited to activating flowering genes. Among the CLF targets, genes associated with embryo/seed development are also targets of SOP12. This also suggests *sop12* suppresses *clf* by a role similar to trx-G components instead of simply affecting flowering time like *fpa* and *ft* do.

It is surprising that well known Pc-G target genes like *FLC*, *FT*, *AG* and *SEP3* etc. are not in the group of genes that require SOP12 for mis-expression in *clf*, as earlier qPCR data indicated. Those genes did appear on the list of up-regulated genes in c/W, but not on the list of down genes in c/d because the values of $\log_2$FC fell between -2 to 2, or the FDR is higher than 0.05 (*FLC* and *AG*). Indeed, when $\log_2$FC was calculated from the qPCR data shown in Figure 4-4, the values are all lower than 2 except for *SEP3* (4.41) in the comparison of d/c. A possible reason for why the effects on *SEP3* expression are lower in RNA seq experiment than in the realtime-PCR is that the samples for RNA-seq were grown in LD condition in tissue culture whereas those for realtime-PCR were grown in SD on soil. The phenotype of *clf* mutants is more extreme in LD and on tissue culture plates and plants flower earlier. The *SEP3* gene is normally expressed in young floral meristems in wild type plants whereas in *clf* mutants it is also expressed in leaves. Because the plants used for the RNA seq experiments may have initiated flowering, there will have been higher expression of *SEP3* in the shoot apex and less effect of *sop12* on the overall levels of *SEP3* expression in s*op12 clf*.

Identifying the SOP12 complex

To identify the SOP12 interactome, I tried yeast two hybrid screening and MS. The yeast two hybrid assay is a more artificial test in a heterologous host, and might identify interactions which are not biologically relevant, for example, between two proteins which are expressed in completely different tissues or located in distinct cellular compartments. Also, if two proteins do not directly interact but participate in a complex, yeast two hybrid assay cannot show the indirect interaction unless the intermediate proteins are simultaneously expressed in yeast. Potential advantages are that the results are generally more reproducible, especially for identifying transient or weak interactions, and are less dependent on antibodies and protein database. MS method overcomes this by analysing the whole native complex when the complex is successfully purified, but sometimes it is tricky to achieve because it is not known under what condition the intact complex could be purified, and in many cases the results are not reproducible between replicates. Overall, these two techniques are good applications for screening potential interacting partners. The results need to be confirmed by, for instance, co-IP.

If a candidate is identified in both Y2H and IP-MS that provides strong support that the interaction is real. Unfortunately, there were no exact matches in my Y2H and IP-MS screens. Some proteins in the same families were found, like glutathione S-transferase (GST) family proteins, β- galactosidase, and dihydrolipoyllysine-residue acetyltransferase component, etc., yet they provide very

limit information about the role or mechanism of SOP12 in development and Pc-G context. One of the possible reasons that made the differences is, in Y2H screening, Gal4 BD was fused to N terminus of SOP12, whereas in MS, GFP was fused to C terminus. I also made transgenic lines harbouring GFP tagged SOP12 in N terminus (*35S::GFP-SOP12/sop12-1 clf*) to avoid that the potential functional domains at C-terminus of SOP12 are blocked by GFP. However, it turned out the N terminus of SOP12 is more important for protein localisation, as suggested by NL prediction by NLStradamus. GFP-SOP12 showed nuclear localised GFP but relatively rarely among transgenic lines (1 among 5 independent lines). 4/5 show quite strong cytosol signal so that not suitable for further analysis.

Another problem in IP-MS analysis is that the peptides pulled down by SOP12-GFP and over-expressed SOP12-GFP are not well overlapped. Over-expressed SOP12 was included just in case the expression of SOP12-GFP driven by native promoter is too low to pull down decent amount of proteins for MS analysis. The interactome of ectopically over-expressed SOP12-GFP might pick up the proteins that do not interact in naturally, but should be able to include and validate interactome of natively expressed SOP12-GFP. It is puzzling why the two interactomes largely exclude each other and may indicate that one or both datasets are not very reliable.

SUP1 and SOP12 may participate in the same complex to regulate Pc-G target expression

AT5G24500 was identified as a SOP12 interacting candidate. It is intriguing because *at5g24500* is also a candidate that suppresses *lhp1* like *sop12* does. Genetic tests (Dr. F. Turck, personal communication) suggested that *SUP1* and *SOP12* may act in a common pathway as they showed similar single mutant phenotypes and the *sup1 sop12 lhp1* triple mutant was similar to the *sup1 lhp1* and *sop12 lhp1* double mutants. Mapping by sequencing identified two candidate mutations closely linked to the *sup1* mutation, one of which was in gene *AT5G24500* and the other in *AT5G25280*. The *AT5G24500* candidate, but not the other gene, was identified twice as a potential SOP12 interactor in the yeast two hybrid screening. If *AT5G24500* does correspond to the *SUP1* locus, the interaction of SUP1 and SOP12 proteins in a complex would potentially account for their shared mutant phenotypes and genetic epistasis. Unfortunately, the SUP1 protein sequence is also novel and therefore does not give clues as to the mechanism of SOP12 action. It would be interesting to know whether *sup1* suppresses *clf* as well, and it is important to test if the interactome of SUP1 overlaps with the one of SOP12. To further characteristic the genetic, molecular and biochemical features of SUP1 will provide decent information about SOP12 biological functions.

.The roles of SOP12 in development

While analysing the interactome and transcriptome of SOP12, I tried to focus on the proteins or genes that are involved in development in order to discover the possible developmental traits of *sop12*. For instance, if SOP12 interacts with a trans-regulatory factor, and the targets of the regulatory factor are shown in SOP12 target list, it would explain the phenotypes observed or lead the direction of the study in the future.

i. Chromatin regulation

Several histone related proteins, like putative histone H2A.8, probable histone H2AXa, H2B and its variants H2B.1 and H2B.11, histone deacetylase HD2B and HDA15, and putative chromatin-remodeling complex ATPase chain, CHR11. However, there were also H2A, H2B variants, and histone deacetylase pulled down in the GFP negative control (H2A.5, H2A.7, and H2B.10, and HDA14). This makes it unclear whether SOP12 does interact with histone proteins. Excitingly, CHR11, appeared to be an interesting candidate, although it interacted only in ectopically expressed SOP12-GFP. CHR11 bears a chromodomain, like LHP1 does. The function of chromodomain is to alter the chromatin accessibility, such as to induce the formation of heterochromatin or to release the winding of chromatins. CHR11 is highly redundant with CHR17 (Huanca-Mamani et al., 2005). The double mutant exhibits the clf-like, up-curled leaf phenotype due to mis-express of *SEP3*, not *AG*, implying a more-specific role compared to Pc-G genes, such as *CLF* and *LHP1* (Li et al., 2012). It was recently shown to be required for repress *SEP3* and *FT* in vegetative leaves (Li et

al., 2012), indicating CHR11 is a negative regulator of *SEP3* and *FT*. *chr11 chr17* has stronger effects on *SEP3* and *FT*, but less in *AG*, which is similar to the effects of mutation at *SOP12* in *clf* background, but in the opposite direction (Figure 4-4 and 4-5). Since the effects of *sop12* and *chr11* on gene expression are opposite, the interaction between the two if real is likely to be inhibitory. It is conceivable that the mutation of *sop12* releases the inhibition of CHR11, maybe also CHR17, leading to strong repression of *SEP3* and *FT* in *sop12 clf*.

ii. Flowering time

sop12 displayed a very weak late flowering phenotype in SD conditions (Figure 5-1). There are a few interactor candidates which are correlated with flowering time regulation, such as EFS (possibly), ATJ3, ARP4, HOS1 (ESD6) and NUA (Table 5-1 and 5-4). *ATJ3* is a highly-represented gene in the result of Y2H screening. In terms of expression pattern, ATJ3 is a nuclear protein, and it expresses ubiquitously as SOP12 is. ATJ3 was shown to interact with SVP, a floral repressor, to regulate *FLC* expression (Shen et al., 2011). Nonetheless, *ATJ3* mutant exhibits late flowering phenotype in LD conditions, where *sop12* does not. Since it is possible that SOP12 may have functional redundancy with other genes which are not discovered yet so that the phenotype of *sop12* is subtle, it cannot be ruled out that SOP12 and its redundant partners, ATJ3 and SVP may form a complex to regulate *FLC* and affect flowering time.

iii. Flower development

Several genes involving in floral development showed up as candidates in interactome or SOP12 targets in transcriptome, especially in the two inner whorls. In Y2H screening, there are LOX3, SPL2 and SPL3, plus RPT5a identified by MS. These four are all involved, or potentially involved in male reproduction. LOX3, a lipoxygenase, involves in anther development via JA biosynthesis pathway (Caldelari et al., 2011). SPL2 and SPL13 belong to SQUAMOSA-promoter binding protein-like family, which is a plant-specific DNA binding protein family with putative transcription factor activities (Chen et al., 2010). SPL2 and SPL3 are both predicted by TAIR to have potential roles in anther development as well. RPT5a, a component of 26S proteasome regulatory particle, affects both male and female gametophytes when it is mutated (Gallois et al., 2009). Consistently, in transcriptome, genes participating in development of inner whorls of flower were also included. For instance, the potential SOP12 target *AP3* is a B-class floral organ identity gene. B class genes are responsible for petal and stamen development (Alvarez-Buylla Elena et al., 2010). In addition, *SHP1*, *SHP2* and *STK*, the genes mis-expressed in *clf* under the requirement of SOP12, were shown to play roles in carpel development through an AG-independent pathway (Favaro et al., 2003). By contrast to the number of genes found to be involved in inner whorls development, less genes in outer whorls development were found, which suggests SOP12 may specifically associate with anther and carpel development.

iv. Embryo development

According to microarray information on eFB browser, SOP12 accumulates through the maturation of seeds, and reaches the peak in dry seeds (Winter et al., 2007). In a rough test *sop12* did not display deficiency in seed dormancy. Neither do seeds of *sop12* have problems on germination, nor apparent abnormal seed phenotypes. It was concluded that SOP12 has subtle or no effects on embryo and seeds development. Thus, the increase of SOP12 expression along the maturation of seeds is possibly for activating 2S albumin seed storage proteins as RNA-seq suggested. 2S albumin seed storage proteins, together with 12S globulin, are most abundant seed storage proteins in *Arabidopsis* seeds. Seed storage proteins are the nitrogen sources for supporting seeds upon germination. However, it was wondered whether over-expression of SOP12 causes embryo lethality or hypersensitive to antibiotics because of the fact that somehow I did not manage to get one *SOP12* over-expression line. It led to the notice the genes and proteins involved in seed development or germination in the transcriptome and interactome. Among the interacting candidates, EMB2444, EMB1507, AKIN10, MCM2, PDIL2-1 are correlated with seed development (Wang et al., 2008; Carles and Fletcher, 2009; Tsai and Gazzarrini, 2012). AKIN10 is a kinase targeting *FUSCA3* (*FUS3*), and this phosphorylation may protect FUS3 from degradation by 26S proteasome (Tsai and Gazzarrini, 2012). *FUS3* is an essential gene in seed development. *fus3* mutation causes a delay in embryo development and low seed storage protein accumulation (Keith et al., 1994). *FUS3* is a Pc-G target and is repressed by MEA in seeds (Makarevich et al., 2006). Intriguingly, in RNA-seq data,

MEA was up-regulated in *clf* and SOP12 is required for the up-regulation. If the de-misexpression of *MEA* is regulated by CHR11 as mechanism proposed in the previous section, it is speculated that seeds of *sop12 clf*, but not seeds of *clf*, may exhibit stronger seed dormancy. In embryo of *clf*, even if *MEA* is de-repressed, the seeds remain healthy because high level of MEA is necessary for embryo development. However, in embryo of *sop12 clf*, *MEA* may be down-regulated by CHR11 which is free from SOP12 inhibition, resulting in FUS3 accumulation to promote seed maturation and dormancy (Keith et al., 1994).

The role of H3K36me3 in gene regulation

H3K36me3 accumulates in the gene body of actively transcribed genes. Because it is localised in gene body, it is reckoned to be more likely the consequences of transcription rather than the reason (Shilatifard, 2006; Oliver and Denu, 2011). In other words, the reduction of H3K36me3 in gene loci, like in *SEP3* in *sop12 clf*, represents the down-regulation of these genes instead of leading to decrease of the expression level. H3K36me3 is generated by Ash1 in *Drosophila* (Tanaka et al., 2007). It was suggested as an anti-repressor that antagonises Pc-G repression but not a co-activator of gene expression (Klymenko and Müller, 2004). Several studies about the role of H3K36me3 suggested H3K36me3 is pivotal in the cross-talk with H3K27me3. H3K36me3 was shown to prevent histone H3 from H2K27me3 (Yuan and Wessler, 2011). Globally, H3K36me3 and H3K27me3 do not co-exist, and when histone H3 is methylated at K36 followed by treating with PRC2, H3K27me3 is not

added. The repelling of these two marks is probably coordinated by Pcl. Brien et. al showed mammalian Pcl1 PHF19 binds to H3K27me3 and H3K36me3 and recruit H3K36me3 demethylase (Brien et al., 2012). Thus, if H3K36me3 is lost or reduced at a locus, the gene may obtain more H3K27me3 and subsequently be repressed by Pc-G pathway. H3K36me3 may not participate in gene expression initiation, yet very likely to play a role in antagonising H3K27me3 and maintenance of gene expression, which fits the function of Ash1, or trx-G components.

It is not clear if this mechanism is conserved in plant kingdom. At least in *sop12 clf*, what observed was the level of H3K36me3 was reduced at *SEP3* (Figure 5-15), but H3K27me3 was not elevated (Figure 5-12). Maybe this is just an exception by chance, or removing of H3K36me3 results in adding another repressive mark instead of H3K27me3 in *Arabidopsis*. It is also possible that plants have evolved different mechanisms, as they have quite different PRC1 and trx-G genes.

Chapter 6. Discussion

The aims of this thesis were first to clone the *SOP12* gene, second to test the hypothesis that it corresponds to a trx-G member by molecular and genetic characterization, and lastly to find the mechanism by which it suppresses Pc-G mutants. The evidence that *SOP12* corresponds to *AT3G63270* is straightforward, as summarised in the discussion to chapter 3, so will not be discussed further. Here I will review the data for aims two and three and discuss possible future experiments to resolve outstanding issues.

6.1 *SOP12* as a trx-G gene

SOP12 was discovered in a genetic screen for suppressors of *clf* and independently identified as a suppressor of another Pc-G mutant, *lhp1* (Hartwig et al., 2012). The suppression phenotypes in *sop12 clf* suggested various possible roles of *SOP12* including: (1) a target of *CLF*, since mutations in *CLF* targets like *AG* and *SEP3* show suppression phenotypes (Goodrich et al., 1997; Lopez-Vernaza et al., 2012). (2) a factor that promotes flowering, because late flowering mutants such as *ft* and *fpa* suppress *clf* (Lopez-Vernaza et al., 2012). (3) a gene that functions like trx-G genes

which are required for Pc-G target expression and show antagonistic effect on Pc-G gene expression. Because the *sop12* single mutant has extremely weak effects upon flowering time, unlike those of suppressors such as *fpa* or *ft*, it is unlikely that *sop12* suppresses *clf* via altering flowering time. Quantification of CLF target gene expression by real time PCR suggested that *SOP12* acted upstream of CLF targets and was therefore unlikely to be itself a target of CLF. This conclusion was strengthened by global transcriptional profiling, using RNA-seq, which revealed that *SOP12* plays a role similar to trx-G genes because it is required for the mis-expression of about 80% of CLF targets. In addition, the data confirms that *SOP12* is not a target of CLF as there is very little change in its expression between wild type and *clf-50* mutants. A second important feature to emerge from the profiling is that the effects of *sop12* mutation are not simply confined to genes involved in flowering time or flower development. For example, the *MEA* gene involved in endosperm development in seed is mis-regulated in *clf* seedlings but not in *clf sop12*.

If *SOP12* is a trx-G gene, it might be expected to suppress other Pc-G mutants. Importantly, *sop12* also suppresses *lhp1* mutants. The *LHP1* and *CLF* genes have

similar mutant phenotypes but have distinct functions in binding and catalysing histone methylation, respectively. Therefore the function of *SOP12* may not be restricted to antagonising EMF-PRC2. Curiously, *sop12* does not suppress *emf2* despite the fact that CLF and EMF2 proteins interact in yeast and *in vitro* (Chanvivattana et al., 2004) and likely participate in the same complex *in vivo*. – indeed, EMF2 was identified in the IP-MS of CLF-GFP as presented in chapter 5. The mutation phenotypes of *clf* and *lhp1* are subtle compared to the ones of *emf* mutants, presumably due to structural or functional redundancy for *CLF* or *LHP1* leading to partially rescue of the mutants. For instance, *CLF* is largely redundant with the structurally related *SWN* (Chanvivattana et al., 2004). Mutations in *EMF2* may impair the activity of the EMF-PRC2 more severely because there is less redundancy with the related *VRN2* gene, so that causes more or stronger Pc-G target mis-regulation. Indeed, in *emf1* and *emf2* mutants, not only floral identity genes like *AG* and *SEP3* are mis-expressed, but also floral meristem identity genes such as *LEAFY* (*LFY*) and *APETALA1* (*AP1*) (Chou et al., 2001). On the other hand, according to my RNA-seq data, *LFY* and *AP1* were not mis-expressed in *clf* compared to wild type samples. *LFY* and *AP1* are strong activator

of floral identity genes (Alvarez-Buylla Elena et al., 2010), so in *emf* mutants, *AG* and *SEP3* may receive further induction signal from the up-regulated *LFY* and *AP1* in addition to the relief from Pc-G repression. Given that the suppression of *clf* and *lhp1* by *sop12* mutation is partial, the emf phenotypes may be too severe for subtle effects of s*op12* to be visible.

The *sop12* mutation also does not obviously suppress *mea* mutations although *SOP12* is expressed during seed development. However, it is not clear if plant trx-G, such as *SYD* or *ULT1*, can suppress *mea* mutations either – so far most mutants that modify the mea phenotype seem to be involved in DNA methylation and affect the imprinted expression of *MEA* or its targets (Vielle-Calzada et al., 1999; Xiao et al., 2003b). Testing whether *sop12* suppresses *vrn2* mutations is more time consuming, as the effects of *vrn2* are only seen in late flowering, vernalisation requiring backgrounds such as *fca* mutants and lengthy vernalisation treatments are also needed. Comparison of the flowering time of *fca vrn2 sop12* with *fca vrn2* mutants is ongoing. Preliminary results support that *sop12 fca vrn2* is much earlier flowering than *fca vrn2* after

vernalisation treatment, so it is promising that *sop12* suppresses the late flowering phenotype of *vrn2* as well as the early flowering of *lhp1*.

One aspect of trx-G genes is that their mutations can enhance one another. There was little genetic interaction between *sop12* and *atx1* or *atx2* mutants, however *sop12* did enhance *ult1*, *ult2* and *efs* mutants. The enhancement is striking, because the *sop12* single mutant does not itself reveal any aspect of the *ult* or *efs* mutant phenotypes, and could mean that the genes act on common targets, or potentially that their gene products are in a common complex. The enhancement in *sop12 ult1* or *sop12 ult2* was discussed in section 4.7 that it is likely associated with the function of *AG* in floral meristem termination. The enhancement phenotype observed between *efs* and *sop12* could be explained by their positive effect on *AP3* (Grini et al., 2009). *efs* plants are partially sterile due to defect in male and female gametophytes, especially in male structures (stamen and anther), reflecting the down-regulation of floral identity genes such as *AP3* and *PI*. The flowers of *sop12 efs* are very small and nearly entirely sterile. From the preliminary observation, the flowers were comprised of sepals and unhealthy carpels, whereas petals and stamens were almost disappeared, representing the deficiency

phenotype of B class genes. In addition to flower defect, other pleiotropic phenotypes, for example, the dwarf phenotypes, of *efs* are also enhanced by *sop12* as well, illustrating the two genes share targets which regulate developmental processes other than flower development. In *efs,* H3K36me3 at EFS targets is decreased. It is not clear if it is possible SOP12 somehow participate in H3K36 tri-methylation with EFS. In western analysis, *sop12* mutants did not affect the global H3K36me3 level like *efs* does. To look at the H3K36me3 at the common target *AP3* may provide the explanation for this question. It is possible that *SOP12* interacts with *EFS* to form a complex, based on yeast two hybrid assay results, but it needs to be clarified that the EFS bait does not autoactivate, and also *in vivo.* evidence is required to verify the interaction.

Another aspect of the trx-G is that they are required to maintain activation of Pc-G target genes, particularly homeotic genes. The *sop12* single mutant phenotype does not reveal obvious floral homeotic or other other developmental defects, although as mentioned above, the interaction with *efs* and *ult* mutants does argue for a role in activation. It is unlikely that the weak single mutant phenotype is due to redundancy with genes similar to *SOP12,* as my genetic analysis did not show obvious

enhancement in double mutant combinations. More promisingly, the transcriptional profiling of wild type, *clf*, *sop12* and *clf sop12* seedlings did identify some genes which are upregulated in *clf* mutant backgrounds due to *SOP12* activity and conversely down regulated in *sop12* mutant backgrounds due to *CLF* activity. One of these targets was the floral homeotic gene *AP3*. However, *sop12* mutants do not show *ap3* mutant-like phenotypes, and I was unable to find a decrease in *AP3* expression in comparison of *sop12* and wild type floral tissue by realtime PCR (not shown) so the relevance is not clear. It is notable that the plant trx-G member *SYD*, which is involved in *AP3* activation, has very mild single mutant phenotype but strongly enhances weak mutants in genes such as *LFY* and *UFO* which are involved in transcriptional activation of floral homeotic genes (Wagner and Meyerowitz, 2002a). It will therefore be important to test whether *sop12* mutants enhance weak *lfy*, *ufo* or *ap3* mutants. Preliminary experiments involving double mutant combinations of *sop12-3* with a weak allele of *ag*, *ag-4*, did not reveal any enhancement (not shown), but the RNA-seq results suggest that *AP3* may be a better candidate.

In general, my genetic and molecular characterization of *sop12* mutants is consistent with *SOP12* acting as a trx-G member, although the case would be stronger if a clear role in gene activation in a wild type background could be shown. The other important aspect of trx-G genes is a mechanistic link to chromatin modification.

6.2 Mechanism of SOP12 action as an activator of Pc-G targets

One expectation is that SOP12 protein should be nuclear localised. This was met, as SOP12-GFP fusions complemented *sop12* mutations and showed nuclear localisation. The link to chromatin is more difficult, in part because the SOP12 protein sequence is novel and with no obvious similarity to known chromatin factors. A second expectation is that SOP12 protein should localise to chromatin in ChIP experiments. However, no direct evidence has been shown that SOP12 directly binds to CLF targets and activates their expression. SOP12-GFP was not located at the *SEP3* locus showed by ChIP using seedling and inflorescence tips. Since *SOP12* activity is sensitive to the presence of *CLF*, and may compete with *CLF* for target gene binding as trx-G proteins compete with Pc-G proteins, it is intriguing to see whether SOP12 binds to CLF targets in *clf*

mutants. These ChIP experiments are clearly important but are still ongoing due to time required to generate sufficient seed of the *SOP12-GFP* line in *clf* mutant background. It is possible that the difficulties with localising SOP12 in ChIP are technical, but other possibilities are that it is not present on chromatin, for example acting by inhibiting Pc-G proteins or activating trx-G components. Further insight may come from identifying whether or not SOP12 is in a complex with informative proteins.

A general problem with the yeast two hybrid and IP-MS screens is that many possible candidates were found and there is little overlap between the different screens. The best candidate, found in the yeast two hybrid screen, is *AT5G24500* for which there is independent genetic evidence that it may correspond to another *lhp1* suppressor, *SUP1*. Since it was identified as an independent *lhp1* suppressor, and the *lhp1 sup1 sop12* triple mutant resembles the s*op12 lhp1* double mutant (personal communication with Dr. F. Turck), it is highly possible that SOP12 and SUP1 are in a complex that regulates CLF target expression. However, further experimental evidences are necessary to support this hypothesis. First of all, a complementation test is required to ensure *AT5G24500* is corresponding to *SUP1*. The preliminary result from Dr. F.

Turck (personal communication) agreed with this but needs to be verified by generating more transgenic lines. Moreover, whether *SOP12* appears on the list of *SUP1* interactome may confirm the existence of the SOP12-SUP1 complex. In addition, testing whether *sup1* suppresses *clf* as well as *lhp1* will provide a good hint if *sop12* suppresses *clf* and *lhp1* via similar mechanism. Unfortunately, the SUP1 protein sequence is also novel and does not provide any obvious hints as to its function. Nonetheless, if the SUP1 results are confirmed for example by co-IP, then use of SUP1 in MS-IP may be informative and help to validate results with SOP12. In other word, if candidates are found that interact with both SUP1 and SOP12 then that would support that the interaction is real.

It was puzzling to me that the effects of SOP12 are highly dependent on *clf* mutation. Excitingly CHR11 was pulled down as a potential component in the SOP12 complex, which has brought in a tempting prospect. Even though CHR11 was identified by MS-IP as a candidate to interact with the over-expressed SOP12-GFP, which makes it controversial whether SOP12 and CHR11 interact *in vivo.,* the possibility is intriguing as CHR11 is a chromatin remodelling factor and genetic analysis shows that it acts to

repress similar targets as CLF. CHR11 shares sequence homology with chromatin remodelling proteins in *Iswi* family (Knizewski et al., 2008), and was shown to regulate CLF targets *SEP3* and *FT* in redundancy with CHR17. The *chr11 chr17* double mutant showed clf-like phenotype (Li et al., 2012). In this scenario, SOP12 would have to act as an inhibitor of CHR11 in order to account for their different effects on CLF targets. I proposed that CHR11 represses *SEP3* and *FT* in a CLF or LHP1-independent manner. SOP12 inhibits the repression of CHR11, or a protein which is functionally and sequence homologous to CHR11, by physical interaction. Thus, in *clf* background, Pc-G targets, including *SEP3* and *FT*, are released from CLF repression, whereas in *sop12 clf*, a repressor like CHR11 is relieved from inhibition of SOP12, leading to repression of *SEP3* and *FT*, which leads to suppressed phenotypes in *sop12 clf*. By contrast, in the presence of CLF when Pc-G genes are repressed, even if CHR11 is free of inhibition and is able to repress *SEP3* and *FT* in *sop12* mutant, their expression does not change much, so that *sop12* mutation only shows little effect. To verify this hypothesis, it is pivotal that the interaction between SOP12 and CHR11 is shown clearly by Co-IP. One other important test would be to see if the suppression of

clf and *lhp1* by *sop12* is dependent on *CHR11* activity, for instance by making *clf sop12 chr11* or *sop12 lhp1 chr11* triple mutants.

In the aspect of histone modifications, *sop12* mutation causes little effects, but counteracts the accumulation of active marks in *clf*. As with the morphological phenotype, *SOP12* is sensitive to the presence of *CLF*. In the presence of *CLF*, the influence of *SOP12* on histone modifications appears limited. In *clf sop12* double mutants H3K27me3 levels remain low, which suggests that *SOP12* does not regulate H3K27me3, for example by inhibiting the HMTase SWN or activating a demethylase. *sop12* mutants do appear to partially decrease H3K4me3 and H3K36me3 and H3Kac levels at CLF targets. Although these results are consistent with SOP12 in some way promoting activating histone modifications, the difficulty is that the changes may simply be a consequence of altered transcription of CLF targets rather than a cause (Shilatifard, 2006). It has been a long debate whether histone modification is the cause or effect of transcription. An early study in yeast showed that when K123 in H3 is mutated into arginine (R), which abolishes the potential of ubiquitylated, the transcription is affected (Xiao et al., 2005), suggesting histone modifications are not

simply the consequence of transcription. Although a recent study indicated the histone modification happens later than transcription of *FT* and is not prerequisite (Adrian et al., 2010), it does not reject that histone modifications affect transcription later on instead of initiation. In a ABI3-like factor (ALF)-*phaseolin* (*phas*) activation system in which histone modification and gene transcription could be un-coupled, altering histone modifications are necessary for stable and high level of expression, but not initial activation (Ng et al., 2006). In addition, histone modifications were shown to influence the transcription by enhancing the recruiting of RNA pol II or facilitating elongation and splicing, etc. (Krogan et al., 2003; Hartzog and Tamkun, 2007; Luco et al., 2010). Thus it is likely the reduction of active marks in *sop12 clf* results in down-regulation of CLF targets.

6.3 Future experiments

As discussed above, it would strengthen the case that *sop12* is a general suppressor of Pc-G if it can be shown to suppress *vrn2* mutations, particularly since the effects of *vrn2* on flowering time are opposite to those of *clf*. It would also help if genetic tests

could support the role of *SOP12* in activating *AP3* or other Pc-G targets. The main outstanding question is how *SOP12* regulates transcription and whether this is through chromatin modification. Given that the IP-MS results with LHP1-GFP and CLF-GFP were promising, in terms of identifying several known interactors, it would be worthwhile to optimise the procedure for SOP12-GFP in unfixed tissues. The IP-MS with SOP12-GFP was less successful, as SOP12 itself was not identified, although it was detected in Western analysis of similar IP experiments. Given that some cleavage of the GFP tag was observed, it is possible that alternative tags would work better, or that an antibody to native SOP12 would help. Although an antibody was raised against SOP12 protein (not shown), so far I was not able to show that it recognised native SOP12 in Western blot experiments.

6.4 SOP12 represents a domesticated transposase

SOP12 encodes a Harbinger transposase derived nuclease like protein but has lost nuclease catalytic residues and likely the activity as well. A question remaining is the link between the biological roles of SOP12 and its Harbinger nuclease-like

characteristics. Since SOP12 has likely lost nuclease functions, it is possibly incorporated in a protein complex functionally distinct from transposase and has completely novel roles. In other words, SOP12 is very likely to be part of a domesticated transposon with new functions. It is quite common that domesticated transposons obtain new functions (reviewed by (Sinzelle et al., 2009). Many of them are participated in transcription regulation and DNA repair. For instance, FAR-RED ELONGATED HYPOCOTYLS 3 (FHY3), which shares homology with MURA, the transposase of maize *Mutator* element (Hudson et al., 2003), was found to associate with transcription activation of light signalling genes *FHY1* and *FAR-RED-ELONGATED HYPOCOTYL1-LIKE* (*FHL*) (Lin et al., 2007). Another example is SETMAR, which is a fusion protein of transposase derived from *Mariner* transposons and SET domain. SETMAR does not mobile transposons, but functions in H3 methylation and nonhomologous end-joining DNA repair mechanism (Lee et al., 2005). On the other hand, the sequence homology sometimes does not absolutely match the actual conformation of the gene product. Thus, it is possible that the folding of SOP12 protein is actually diverged from Harbinger nucleases. Those speculations

cannot be proved without crystal structures of SOP12. Thinking outside of the box, under the presupposition that SOP12 is a trx-G gene based on its genetic response and antagonise role of Pc-G gene, it is a great novelty that the transposase nuclease-like protein is involved in trx-G complexes.

Chapter 7. References

Ach, R.A., Taranto, P., and Gruissem, W. (1997). A Conserved Family of WD-40 Proteins Binds to the Retinoblastoma Protein in Both Plants and Animals. Plant Cell **9,** 1595-1606.

Adrian, J., Farrona, S., Reimer, J.J., Albani, M.C., Coupland, G., and Turck, F. (2010). Cis-Regulatory Elements and Chromatin State Coordinately Control Temporal and Spatial Expression of *FLOWERING LOCUS T* in *Arabidopsis*. Plant Cell **22,** 1425-1440.

Ahmad, K., and Henikoff, S. (2002). The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. Mol. Cell **9,** 1191-1200.

Aichinger, E., Villar, C.B.R., Farrona, S., Reyes, J.C., Hennig, L., and Köhler, C. (2009). CHD3 Proteins and Polycomb Group Proteins Antagonistically Determine Cell Identity in *Arabidopsis.* PLoS Genet **5,** e1000605.

Alvarez-Buylla Elena, R., Benítez, M., Corvera--Poiré, A., Chaos Cador, l., de Folter, S., Gamboa de Buen, A., Garay-Arroyo, A., García-Ponce, B., Jaimes-Miranda, F., Pérez-Ruiz Rigoberto, V., Piñeyro-Nelson, A., and Sánchez-Corrales Yara, E. (2010). Flower Development. In The Arabidopsis Book (The American Society of Plant Biologists), pp. 1-57.

Alvarez-Venegas, R., and Avramova, Z. (2001). Two *Arabidopsis* Homologs of the Animal Trithorax Genes: a New Structural Domain is a Signature Feature of the Trithorax Gene Family. Gene **271,** 215-221.

Alvarez-Venegas, R., Pien, S., Sadder, M., Witmer, X., Grossniklaus, U., and Avramova, Z. (2003). *ATX1*, an *Arabidopsis* Homolog of *Trithorax*, Activates Flower Homeotic Genes. Curr. Biol. **13,** 627-637.

Ardehali, M.B., Mei, A., Zobeck, K.L., Caron, M., Lis, J.T., and Kusch, T. (2011). Drosophila Set1 is the Major Histone H3 lysine 4 Trimethyltransferase with Role in Transcription. EMBO J **30,** 2817-2828.

Armstrong, J.A., Papoulas, O., Daubresse, G., Sperling, A.S., Lis, J.T., Scott, M.P., and Tamkun, J.W. (2002). The *Drosophila* BRM Complex Facilitates Global Transcription by RNA Polymerase II. EMBO J **21,** 5245-5254.

Aubert, D., Chen, L., Moon, Y.-H., Martin, D., Castle, L.A., Yang, C.-H., and Sung, Z.R. (2001). EMF1, a Novel Protein Involved in the Control of Shoot Architecture and Flowering in *Arabidopsis*. Plant Cell **13,** 1865-1875.

Bantignies, F., and Cavalli, G. (2011). Polycomb Group Proteins: Repression in 3D. Trends Genet. **27,** 454-464.

Bantignies, F., Goodman, R.H., and Smolik, S.M. (2000). Functional Interaction between the CoactivatorDrosophila CREB-Binding Protein and ASH1, a Member of the Trithorax Group of Chromatin Modifiers. Mol. Cell. Biol. **20,** 9317-9330.

Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., and Dean, C. (2004). Vernalization Requires Epigenetic Silencing of *FLC* by Histone Methylation. Nature **427,** 164-167.

Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I., Reuter, G., and Aalen, R.B. (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.

Beh, L.Y., Colwell, L.J., and Francis, N.J. (2012). A Core Subunit of Polycomb Repressive Complex 1 is Broadly Conserved in Function but not Primary Sequence. Proc. Nat. Acad. Sci. **109,** e1063-e1071.

Beisel, C., Imhof, A., Greene, J., Kremmer, E., and Sauer, F. (2002). Histone Methylation by the *Drosophila* Epigenetic Transcriptional Regulator ASH1. Nature **419,** 857-862.

Bemer, M., and Grossniklaus, U. (2012). Dynamic Regulation of Polycomb Group Activity During Plant Development. Curr. Opin. Plant Biol. **15,** 523-529.

Berger, S., Bell, E., and Mullet, J.E. (1996). Two Methyl Jasmonate-Insensitive Mutants Show Altered Expression of AtVsp in Response to Methyl Jasmonate and Wounding. Plant Physiol. **111,** 525-531.

Biggin, M.D., and Tjian, R. (1988). Transcription Factors that Activate the *Ultrabithorax* Promoter in Developmentally Staged Extracts. Cell **53,** 699-711.

Biggin, M.D., Bickel, S., Benson, M., Pirrotta, V., and Tjian, R. (1988). Zeste Encodes a Sequence-specific Transcription Factor that Activates the *Ultrabithorax* Promoter *in vitro.* Cell **53,** 713-722.

Bond, D.M., Dennis, E.S., Pogson, B.J., and Finnegan, E.J. (2009). Histone Acetylation, VERNALIZATION INSENSITIVE 3, FLOWERING LOCUS C, and the Vernalization Response. Mol. Plant **2,** 724-737.

Bouyer, D., Roudier, F., Heese, M., Andersen, E.D., Gey, D., Nowack, M.K., Goodrich, J., Renou, J.-P., Grini, P.E., Colot, V., and Schnittger, A. (2011). Polycomb Repressive Complex 2 Controls the Embryo-to-Seedling Phase Transition. PLoS Genet **7,** e1002014.

Bratzel, F., López-Torrejón, G., Koch, M., Del Pozo, J.C., and Calonje, M. (2010). Keeping Cell Identity in *Arabidopsis* Requires PRC1 RING-Finger Homologs that Catalyze H2A Monoubiquitination. Curr. Biol. **20,** 1853-1859.

Breen, T.R., and Harte, P.J. (1991). Molecular Characterization of the Trithorax Gene, a Positive Regulator of Homeotic Gene Expression in *Drosophila*. Mech. Develop. **35,** 113-127.

Brien, G.L., Gambero, G., O'Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L., Lohan, A.J., Ferguson, N., Shi, X., Sinha, K.M., Loftus, B.J., Cagney, G., and Bracken, A.P. (2012). Polycomb PHF19 Binds H3K36me3 and Recruits PRC2 and

Demethylase NO66 to Embryonic Stem Cell Genes During Differentiation. Nat Struct Mol Biol **19,** 1273-1281.

Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.-L., and Kassis, J.A. (1998). The *Drosophila* Polycomb Group Gene Pleiohomeotic Encodes a DNA Binding Protein with Homology to the Transcription Factor YY1. Mol. Cell **1,** 1057-1064.

Caldelari, D., Wang, G., Farmer, E., and Dong, X. (2011). *Arabidopsis lox3 lox4* Double Mutants are Male Sterile and Defective in Global Proliferative Arrest. Plant Mol. Biol. **75,** 25-33.

Calgaro, S., Boube, M., Cribbs, D.L., and Bourbon, H.-M. (2002). The *Drosophila* Gene taranis Encodes a Novel Trithorax Group Member Potentially Linked to the Cell Cycle Regulatory Apparatus. Genetics **160,** 547-560.

Calonje, M., Sanchez, R., Chen, L., and Sung, Z.R. (2008). *EMBRYONIC FLOWER1* Participates in Polycomb Group-Mediated *AG* Gene Silencing in *Arabidopsis*. Plant Cell **20,** 277-291.

Cao, R., Tsukada, Y.-i., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A Ubiquitylation and Hox Gene Silencing. Mol. Cell **20,** 845-854.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. Science **298,** 1039-1043.

Carles, C.C., and Fletcher, J.C. (2009). The SAND Domain Protein ULTRAPETALA1 Acts as a Trithorax Group Factor to Regulate Cell Fate in Plants. Gene. Dev. **23,** 2723-2728.

Carles, C.C., Lertpiriyapong, K., Reville, K., and Fletcher, J.C. (2004). The *ULTRAPETALA1* Gene Functions Early in Arabidopsis Development to Restrict Shoot Apical Meristem Activity and Acts Through WUSCHEL to Regulate Floral Meristem Determinacy. Genetics **167,** 1893-1903.

Carles, C.C., Choffnes-Inada, D., Reville, K., Lertpiriyapong, K., and Fletcher, J.C. (2005). *ULTRAPETALA1* Encodes a SAND Domain Putative Transcriptional Regulator that Controls Shoot and Floral Meristem Activity in *Arabidopsis*. Development **132,** 897-911.

Chan, C.S., Rastelli, L., and Pirrotia, V. (1994). A Polycomb Response Element in the *Ubx* Gene that Determines an Epigenetically Inherited State of Repression. EMBO J. **13,** 2553-2564.

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.-H., Sung, Z.R., and Goodrich, J. (2004). Interaction of Polycomb-group Proteins Controlling Flowering in *Arabidopsis*. Development **131,** 5263-5276.

Chen, I. (2012). The Silence of the Plants. Nat. Struct. Mol. Biol. **19,** 663-663.

Chen, X., Zhang, Z., Liu, D., Zhang, K., Li, A., and Mao, L. (2010). SQUAMOSA Promoter-Binding Protein-Like Transcription Factors: Star Players for Plant Growth and Development. J. Integrative Plant Biol. **52,** 946-951.

Chou, M.-L., Haung, M.-D., and Yang, C.-H. (2001). EMF Genes Interact with Late-flowering Genes in Regulating Floral Initiation Genes during Shoot Development in *Arabidopsis thaliana*. Plant Cell Physiol. **42,** 499-507.

Chris, A.H., Craig, C.W., Masumi, R., Peacock, W.J., and Elizabeth, S.D. (2006). The Arabidopsis *FLC* Protein Interacts Directly *in vivo* with *SOC1* and *FT* Chromatin and is Part of a High-molecular-weight Protein Complex. Plant J. **46,** 183-192.

Clough, S.J., and Bent, A.F. (1998). Floral Dip: a Simplified Method for *Agrobacterium*-mediated Transformation of *Arabidopsis thaliana*. Plant J. **16,** 735-743.

Corona, D.F.V., Siriaco, G., Armstrong, J.A., Snarskaya, N., McClymont, S.A., Scott, M.P., and Tamkun, J.W. (2007). ISWI Regulates Higher-Order Chromatin Structure and Histone H1 Assembly In Vivo. PLoS Biol **5,** e232.

Crosby, M.A., Miller, C., Alon, T., Watson, K.L., Verrijzer, C.P., Goldman-Levi, R., and Zak, N.B. (1999). The Trithorax Group Genemoira Encodes a Brahma-Associated Putative Chromatin-Remodeling Factor in *Drosophila melanogaster*. Mol. Cell Biol. **19,** 1159-1170.

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.

Czermin, B., Schotta, G., Hulsmann, B.B., Brehm, A., Becker, P.B., Reuter, G., and Imhof, A. (2001). Physical and functional association of SU(VAR)3-9 and HDAC1 in Drosophila. EMBO Rep **2,** 915-919.

Daubresse, G., Deuring, R., Moore, L., Papoulas, O., Zakrajsek, I., Waldrip, W.R., Scott, M.P., Kennison, J.A., and Tamkun, J.W. (1999). The *Drosophila Kismet* Gene is Related to Chromatin-remodeling Factors and is Required for Both Segmentation and Segment Identity. Development **126,** 1175-1187.

Davis, A., Hall, A., Millar, A., Darrah, C., and Davis, S. (2009). Protocol: Streamlined Sub-protocols for Floral-dip Transformation and Selection of Transformants in *Arabidopsis thaliana*. Plant Methods **5,** 3.

de Ayala Alonso, A.s.G.n., Gutie´rrez, L., Fritsch, C., Papp, B., Beuchle, D., and Müller, J. (2007). A Genetic Screen Identifies Novel Polycomb Group Genes in *Drosophila*. Genetics **176,** 2099-2108.

De Lucia, F., Crevillen, P., Jones, A.M.E., Greb, T., and Dean, C. (2008). A PHD-Polycomb Repressive Complex 2 Triggers the Epigenetic Silencing of *FLC* During Vernalization. Proc. Nat. Acad. Sci. **105,** 16831-16836.

de Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M., Koseki, H., and Brockdorff, N. (2004). Polycomb Group Proteins Ring1A/B Link

Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. Dev. Cell **7,** 663-676.

Desfeux, C., Clough, S.J., and Bent, A.F. (2000). Female Reproductive Tissues Are the Primary Target of *Agrobacterium*-Mediated Transformation by the *Arabidopsis* Floral-Dip Method. Plant Physiol. **123,** 895-904.

Dingwall, A.K., Beek, S.J., McCallum, C.M., Tamkun, J.W., Kalpana, G.V., Goff, S.P., and Scott, M.P. (1995). The *Drosophila* Snr1 and Brm Proteins are Related to Yeast SWI/SNF Proteins and are Components of a Large Protein Complex. Mol. Biol. Cell **6,** 777-791.

Dingwall, C., Robbins, J., Dilworth, S.M., Roberts, B., and Richardson, W.D. (1988). The Nucleoplasmin Nuclear Location Sequence is Larger and More Complex than that of SV-40 large T Antigen. J. Cell Biol. **107,** 841-849.

Edwards, K., Johnstone, C., and Thompson, C. (1991). A Simple and Rapid Method for the Preparation of Plant Genomic DNA for PCR Analysis. Nucleic Acids Res. **19,** 1349.

Elfring, L.K., Deuring, R., McCallum, C.M., Peterson, C.L., and Tamkun, J.W. (1994). Identification and Characterization of *Drosophila* relatives of the Yeast Transcriptional Activator SNF2/SWI2. Mol Cell Biol **14,** 2225-2234.

Enderle, D., Beisel, C., Stadler, M.B., Gerstung, M., Athri, P., and Paro, R. (2010). Polycomb Preferentially Targets Stalled Promoters of Coding and Non-coding Transcripts. Genome Res. **21,** 216-226.

Eshed, Y., Baum, S.F., and Bowman, J.L. (1999). Distinct Mechanisms Promote Polarity Establishment in Carpels of *Arabidopsis*. Cell **99,** 199-209.

Exner, V., Aichinger, E., Shu, H., Wildhaber, T., Alfarano, P., Caflisch, A., Gruissem, W., Ko¨ hler, C., and Hennig, L. (2009). The Chromodomain of LIKE HETEROCHROMATIN PROTEIN 1 Is Essential for H3K27me3 Binding and Function during *Arabidopsis* Development. PLoS ONE **4,** e5335.

Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H., and Karch, F. (1994). The Trithorax-like Gene Encodes the *Drosophila* GAGAFactor. Nature **371,** 806-808.

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M.F., Kater, M.M., and Colombo, L. (2003). MADS-Box Protein Complexes Control Carpel and Ovule Development in *Arabidopsis*. Plant Cell **15,** 2603-2611.

Finnegan, J., Kovac, K.A., Jaligot, E., Sheldon, C.C., James Peacock, W., and Dennis, E.S. (2005). The Downregulation of *FLOWERING LOCUS C* (*FLC*) Expression in Plants with Low Levels of DNA Methylation and by Vernalization Occurs by Distinct Mechanisms. Plant J. **44,** 420-432.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., 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,** 1870-1881.

Fletcher, J.C. (2001). The *ULTRAPETALA* Gene Controls Shoot and Floral Meristem Size in *Arabidopsis*. Development **128,** 1323-1333.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin Compaction by a Polycomb Group Protein Complex. Science **306,** 1574-1577.

Gallois, J.-L., Guyon-Debast, A., Lécureuil, A., Vezon, D., Carpentier, V., Bonhomme, S., and Guerche, P. (2009). The *Arabidopsis* Proteasome RPT5 Subunits are Essential for Gametophyte Development and Show Accession-Dependent Redundancy. Plant Cell **21,** 442-459.

Gaudin, V., Libault, M., Pouteau, S., Juul, T., Zhao, G., Lefebvre, D., and Grandjean, O. (2001). Mutations in *LIKE HETEROCHROMATIN PROTEIN 1* Affect Flowering Time and Plant Architecture in *Arabidopsis*. Development **128,** 4847-4858.

Gendall, A.R., Levy, Y.Y., Wilson, A., and Dean, C. (2001). The *VERNALIZATION 2* Gene Mediates the Epigenetic Regulation of Vernalization in *Arabidopsis*. Cell **107,** 525-535.

Gendrel, A.-V., Lippman, Z., Yordan, C., Colot, V., and Martienssen, R.A. (2002). Dependence of Heterochromatic Histone H3 Methylation Patterns on the *Arabidopsis* Gene *DDM1*. Science **297,** 1871-1873.

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.

Grini, P.E., Thorstensen, T., Alm, V., Vizcay-Barrena, G., Windju, S.S., Jørstad, T.S., Wilson, Z.A., and Aalen, R.B. (2009). The ASH1 HOMOLOG 2 (ASHH2) Histone H3 Methyltransferase Is Required for Ovule and Anther Development in *Arabidopsis*. PLoS One **4,** e7817.

Grossniklaus, U., Vielle-Calzada, J.-P., Hoeppner, M.A., and Gagliano, W.B. (1998). Maternal Control of Embryogenesis by MEDEA, a Polycomb Group Gene in *Arabidopsis*. Science **280,** 446-450.

Guo, H., Yang, H., Mockler, T.C., and Lin, C. (1998). Regulation of Flowering Time by *Arabidopsis* Photoreceptors. Science **279,** 1360-1363.

Gutiérrez, L., Oktaba, K., Scheuermann, J.C., Gambetta, M.C., Ly-Hartig, N., and Müller, J. (2012). The Role of the Histone H2A Ubiquitinase Sce in Polycomb Repression. Development **139,** 117-127.

Hagstrom, K., Muller, M., and Schedl, P. (1997). A Polycomb and GAGA Dependent Silencer Adjoins the Fab-7 Boundary in the *Drosophila* Bithorax Complex. Genetics **146,** 1365-1380.

Hallson, G., Hollebakken, R.E., Li, T., Syrzycka, M., Kim, I., Cotsworth, S., Fitzpatrick, K.A., Sinclair, D.A.R., and Honda, B.M. (2011). dSet1 Is the Main H3K4 Di- and Tri-Methyltransferase Throughout *Drosophila* Development. Genetics **190,** 91-100.

Hartwig, B., James, G.V., Konrad, K., Schneeberger, K., and Turck, F. (2012). Fast Isogenic Mapping-by-Sequencing of Ethyl Methanesulfonate-Induced Mutant Bulks. Plant Physiol. **160,** 591-600.

Hartzog, G.A., and Tamkun, J.W. (2007). A New Role for Histone Tail Modifications in Transcription Elongation. Gene Dev. **21,** 3209-3213.

Hennig, L., and Derkacheva, M. (2009). Diversity of Polycomb Group Complexes in Plants: Same Rules, Different Players? Trends Genet. **25,** 414-423.

Herman, R.K., and Yochem, J. (2005). Genetic enhancers. WormBook.

Ho, L., and Crabtree, G.R. (2010). Chromatin Remodelling During Development. Nature **463,** 474-484.

Horard, B., Tatout, C., Poux, S., and Pirrotta, V. (2000). Structure of a Polycomb Response Element and In vitro Binding of Polycomb Group Complexes Containing GAGA Factor. Mol. Cell. Biol. **20,** 3187-3197.

Huanca-Mamani, W., Garcia-Aguilar, M., León-Martínez, G., Grossniklaus, U., and Vielle-Calzada, J.-P. (2005). CHR11, a Chromatin-remodeling Factor Essential for Nuclear Proliferation During Female Gametogenesis in *Arabidopsis thaliana*. Proc, Nat. Acad. Sci. **102,** 17231-17236.

Hudson, M.E., Lisch, D.R., and Quail, P.H. (2003). The FHY3 and FAR1 Genes Encode Transposase-related Proteins Involved in Regulation of Gene Expression by the Phytochrome A-Signaling Pathway. Plant J. **34,** 453-471.

Hurtado, L., Farrona, S., and Reyes, J. (2006). The Putative SWI/SNF Complex Subunit BRAHMA Activates Flower Homeotic Genes in *Arabidopsis thaliana*. Plant Mol. Biol. **62,** 291-304.

Ingham, P.W. (1981). Trithorax: A New Homeotic Mutation of Drosophila melanogaster II. The Role of *trx+* After Embryogenesis. Wilhelm Roux's Archives **190,** 365-369.

Ingham, P.W. (1983). Differential Expression of Bithorax Complex Genes in the Absence of the *Extra Sex Combs* and *Trithorax* Genes. Nature **306,** 591-593.

Ingham, P.W. (1988). The Molecular Genetics of Embryonic Pattern Formation in *Drosophila*. Nature **335,** 25-34.

Ingham, P.W. (1998). *Trithorax* and the Regulation of Homeotic Gene Expression in *Drosophila*: a Historical Perspective. Int. J. Biol. **42,** 423 - 429

Ingham, P.W., and Whittle, R. (1980). Trithorax: A New Homeotic Mutation of *Drosophila melanogaster* Causing Transformations of Abdominal and Thoracic Imaginal Segments. Mol. Genet. Genomics **179,** 607-614.

Inoue, H., Nojima, H., and Okayama, H. (1990). High Efficiency Transformation of *Escherichia coli* with Plasmids. Gene **96,** 23-28.

Jacob, Y., Feng, S., LeBlanc, C.A., Bernatavichute, Y.V., Stroud, H., Cokus, S., Johnson, L.M., Pellegrini, M., Jacobsen, S.E., and Michaels, S.D. (2009). ATXR5 and ATXR6 are H3K27 Monomethyltransferases Required for Chromatin Structure and Gene Silencing. Nat Struct Mol Biol **16,** 763-768.

Jenuwein, T., Laible, G., Dorn, R., and Reuter, G. (1998). SET Domain Proteins Modulate Chromatin Domains in Eu- and Heterochromatin. Cell. and Mol. Life Sci. **54,** 80-93.

Jiang, D., Wang, Y., Wang, Y., and He, Y. (2008). Repression of *FLOWERING LOCUS C* and *FLOWERING LOCUS T* by the *Arabidopsis* Polycomb Repressive Complex 2 Components. PLoS ONE **3,** e3404.

Jurgens, G. (1985). A Group of Genes Controlling the Spatial Expression of the Bithorax Complex in *Drosophila*. Nature **316,** 153-155.

Köhler, C., and Hennig, L. (2010). Regulation of cell identity by plant Polycomb and trithorax group proteins. Current Opinion in Genetics & Development **20,** 541-547.

Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W., and Grossniklaus, U. (2003a). The Polycomb-group Protein MEDEA Regulates Seed Development by Controlling Expression of the MADS-box Gene *PHERES1*. Gene Dev. **17,** 1540-1553.

Köhler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U., and Gruissem, W. (2003b). *Arabidopsis* MSI1 is a Component of the MEA/FIE Polycomb Group Complex and Required for Seed Development. EMBO J **22,** 4804-4814.

Kal, A.J., Mahmoudi, T., Zak, N.B., and Verrijzer, C.P. (2000). The *Drosophila* Brahma Complex is an Essential Coactivator for the Trithorax Group Protein Zeste. Gene Dev. **14,** 1058-1071.

Kapitonov, V.V., and Jurka, J. (2004). Harbinger Transposons and an Ancient HARBI1 Gene Derived from a Transposase. DNA Cell Biol. **23,** 311-324.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., and Weigel, D. (1999). Activation Tagging of the Floral Inducer FT. Science **286,** 1962-1965.

Kasajima, I., Ide, Y., Ohkama-Ohtsu, N., Hayashi, H., Yoneyama, T., and Fujiwara, T. (2004). A Protocol for Rapid DNA Extraction from *Arabidopsis thaliana* for PCR Analysis. Plant Mol. Biol. Rep. **22,** 49-52.

Katari, M.S., Nowicki, S.D., Aceituno, F.F., Nero, D., Kelfer, J., Thompson, L.P., Cabello, J.M., Davidson, R.S., Goldberg, A.P., Shasha, D.E., Coruzzi, G.M., and Gutiérrez, R.A. (2010). VirtualPlant: A Software Platform to Support Systems Biology Research. Plant Physiol. **152,** 500-515.

Kaufman, T.C., Tasaka, S.E., and Suzuki, D.T. (1973). The Interaction of Two Complex Loci, Zeste and Bithorax in *Drosophila melanogaste*. Genetics **75,** 299-321.

Keith, K., Kraml, M., Dengler, N.G., and McCourt, P. (1994). *fusca3*: A Heterochronic Mutation Affecting Late Embryo Development in *Arabidopsis*. Plant Cell **6,** 589-600.

Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent Modifiers of Polycomb and Antennapedia Mutations in *Drosophila*. Proc. Nat. Acad. of Sci. **85,** 8136-8140.

Kennison, J.A., and Tamkun, J.W. (1992). Trans-regulation of Homeotic Genes in *Drosophila*. New Biol. **4,** 91-96.

Ketel, C.S., Andersen, E.F., Vargas, M.L., Suh, J., Strome, S., and Simon, J.A. (2005). Subunit Contributions to Histone Methyltransferase Activities of Fly and Worm Polycomb Group Complexes. Mol. Cell. Biol. **25,** 6857-6868.

Kim, S.Y., Zhu, T., and Sung, Z.R. (2009). Epigenetic Regulation of Gene Programs by EMF1 and EMF2 in *Arabidopsis*. Plant Physiol. **152,** 516-528.

Kim, S.Y., He, Y., Jacob, Y., Noh, Y.-S., Michaels, S., and Amasino, R. (2005). Establishment of the Vernalization-Responsive, Winter-Annual Habit in Arabidopsis Requires a Putative Histone H3 Methyl Transferase. Plant Cell **17,** 3301-3310.

King, I.F.G., Francis, N.J., and Kingston, R.E. (2002). Native and Recombinant Polycomb Group Complexes Establish a Selective Block to Template Accessibility To Repress Transcription in Vitro. Mol. Cell. Biol. **22,** 7919-7928.

King, I.F.G., Emmons, R.B., Francis, N.J., Wild, B., Müller, J., Kingston, R.E., and Wu, C.-t. (2005). Analysis of a Polycomb Group Protein Defines Regions That Link Repressive Activity on Nucleosomal Templates to In Vivo Function. Mol. Cell. Biol. **25,** 6578-6591.

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Anat, K., Margossian, L., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Control of Fertilization-Independent Endosperm Development by the MEDEA Polycomb gene in *Arabidopsis*. Proc. Nati. Acad. Sci. **96,** 4186-4191.

Klymenko, T., and Müller, J. (2004). The Histone Methyltransferases Trithorax and ASH1 Prevent Transcriptional Silencing by Polycomb Group Proteins. EMBO Rep. **5,** 373–377.

Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb Group Protein Complex with Sequence-specific DNA-binding and Selective Methyl-lysine-binding Activities. Gene Dev. **20,** 1110-1122.

Knizewski, L., Ginalski, K., and Jerzmanowski, A. (2008). Snf2 Proteins in Plants: Gene Silencing and Beyond. Trends plant Sci. **13,** 557-565.

Konev, A.Y., Tribus, M., Park, S.Y., Podhraski, V., Lim, C.Y., Emelyanov, A.V., Vershilova, E., Pirrotta, V., Kadonaga, J.T., Lusser, A., and Fyodorov, D.V. (2007). CHD1 Motor Protein Is Required for Deposition of Histone Variant H3.3 into Chromatin in Vivo. Science **317,** 1087-1090.

Kotake, T., Takada, S., Nakahigashi, K., Ohto, M., and Goto, K. (2003). *Arabidopsis TERMINAL FLOWER 2* Gene Encodes a Heterochromatin Protein 1 Homolog and Represses both *FLOWERING LOCUS T* to Regulate Flowering Time and Several Floral Homeotic Genes. Plant Cell Physiol. **44,** 555-564.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., Shilatifard, A., Buratowski, S., and Greenblatt, J. (2003). Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. Mol. Cell. Biol. **23,** 4207-4218.

Kroj, T., Savino, G., Valon, C., Giraudat, J., and Parcy, F. (2003). Regulation of Storage Protein Gene Expression in *Arabidopsis*. Development **130,** 6065-6073.

Kwong, C., Adryan, B., Bell, I., Meadows, L., Russell, S., Manak, J.R., and White, R. (2008). Stability and Dynamics of Polycomb Target Sites in *Drosophila* Development. PLoS Genet **4,** e1000178.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A.A., and Verrijzer, C.P. (2008). dKDM2 Couples Histone

H2A Ubiquitylation to Histone H3 Demethylation During Polycomb Group Silencing. Gene Dev. **22,** 2799-2810.

LaJeunesse, D., and Shearn, A. (1996). E(z): a Polycomb Group Gene or a Trithorax Group Gene? Development **122,** 2189-2197.

Laney, J.D., and Biggin, M.D. (1992). Zeste, a Nonessential Gene, Potently Activates *Ultrabithorax* Transcription in the *Drosophila* Embryo. Gene. Dev. **6,** 1531-1541.

Laurent, B.C., Treich, I., and Carlson, M. (1993). The Yeast SNF2/SWI2 Protein Has DNA-stimulated ATPase Activity Required for Transcriptional Activation. Gene. Dev. **7,** 583-591.

Lawrence, P.A., and Morata, G. (1994). Homeobox genes: Their function in *Drosophila* segmentation and pattern formation. Cell **78,** 181-189.

Lee, S.-H., Oshige, M., Durant, S.T., Rasila, K.K., Williamson, E.A., Ramsey, H., Kwan, L., Nickoloff, J.A., and Hromas, R. (2005). The SET Domain Protein Metnase Mediates Foreign DNA Integration and Links Integration to Nonhomologous End-joining Repair. Proc. Nat. Acad. Sci. USA **102,** 18075-18080.

Levin, H.L., and Moran, J.V. (2011). Dynamic Interactions Between Transposable Elements and Their Hosts. Nat. Rev. Genet. **12,** 615-627.

Levine, S.S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P., and Kingston, R.E. (2002). The Core of the Polycomb Repressive Complex Is Compositionally and Functionally Conserved in Flies and Humans. Mol. Cell. Biol. **22,** 6070-6078.

Levy, Y.Y., Mesnage, S., Mylne, J.S., Gendall, A.R., and Dean, C. (2002). Multiple Roles of Arabidopsis *VRN1* in Vernalization and Flowering Time Control. Science **297,** 243-246.

Lewis, E.B. (1978). A Gene Complex Controlling Segmentation in *Drosophila*. Nature **276,** 565-570.

Li, G., Zhang, J., Li, J., Yang, Z., Huang, H., and Xu, L. (2012). Imitation Switch Chromatin Remodeling Factors and Their Interacting RINGLET Proteins Act Together in Controlling the Plant Vegetative Phase in *Arabidopsis*. Plant J. **72,** 261-270.

Li, W., Wang, Z., Li, J., Yang, H., Cui, S., Wang, X., and Ma, L. (2011). Overexpression of *AtBMI1C*, a Polycomb Group Protein Gene, Accelerates Flowering in *Arabidopsis*. PLoS ONE **6,** e21364.

Libault, M., Tessadori, F., Germann, S., Snijder, B., Fransz, P., and Gaudin, V. (2005). The *Arabidopsis* LHP1 Protein is a Component of Euchromatin. Planta **222,** 910-925.

Lin, R., Ding, L., Casola, C., Ripoll, D.R., Feschotte, C., and Wang, H. (2007). Transposase-Derived Transcription Factors Regulate Light Signaling in *Arabidopsis*. Science **318,** 1302-1305.

Lin Xu, R.M., Alexandre Berr, Jörg Fuchs, Valérie Cognat, Denise Meyer Wen-Hui Shen. (2008). The E2 ubiquitin-conjugating enzymes, AtUBC1 and AtUBC2, play redundant roles and are involved in activation of FLC expression and repression of flowering in Arabidopsis thaliana. The Plant Journal **57,** 279-288.

Liu, X., Kim, Y.J., Müller, R., Yumul, R.E., Liu, C., Pan, Y., Cao, X., Goodrich, J., and Chen, X. (2011). AGAMOUS Terminates Floral Stem Cell Maintenance in Arabidopsis by Directly Repressing *WUSCHEL* through Recruitment of Polycomb Group Proteins. Plant Cell **23,** 3654-3670.

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. Nat. Aca. Sci. **90,** 10370-10374.

Lopez-Vernaza, M., Yang, S., Müller, R., Thorpe, F., de Leau, E., and Goodrich, J. (2012). Antagonistic Roles of *SEPALLATA3*, *FT* and *FLC* Genes as Targets of the Polycomb Group Gene *CURLY LEAF*. PLoS ONE **7,** e307-e315.

Lu, F., Cui, X., Zhang, S., Jenuwein, T., and Cao, X. (2011). *Arabidopsis* REF6 is a Histone H3 Lysine 27 Demethylase. Nat. Genet. **43,** 715-719.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of Alternative Splicing by Histone Modifications. Science **327,** 996-1000.

Luo, M., Ilodeau, P.B., Koltunow, A., Dennis, E.S., Peacock, W.J., and Chaudhury, A.M. (1999). Genes Controlling Fertilization-Independent Seed Development in *Arabidopsis thaliana*. Proc. Nati. Acad. Sci. **96,** 296-301.

Müller, J., Müller, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone Methyltransferase Activity of a *Drosophila* Polycomb Group Repressor Complex. Cell **111,** 197-208.

Makarevich, G., Leroy, O., Akinci, U., Schubert, D., Clarenz, O., Goodrich, J., Grossniklaus, U., and Köhler, C. (2006). Different Polycomb Group Complexes Regulate Common Target Genes in *Arabidopsis*. EMBO Rep. **7,** 947–952

Marfella, C.G.A., and Imbalzano, A.N. (2007). The Chd family of chromatin remodelers. Mutat. Res.-Fund. Mol. M. **618,** 30-40.

Martienssen, R., and Irish, V. (1999). Copying out our ABCs: the Role of Gene Redundancy in Interpreting Genetic Hierarchies. Trends Genet. **15,** 435-437.

Mayer, K.F.X., Schoof, H., Haecker, A., Lenhard, M., J gens, G., and Laux, T. (1998). Role of WUSCHEL in Regulating Stem Cell Fate in the *Arabidopsis* Shoot Meristem. Cell **95,** 805-815.

Mihaly, J., Mishra, R.K., and Karch, F. (1998). A Conserved Sequence Motif in Polycomb Response Elements. Mol. Cell **1,** 1065-1066.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a

Complex of Proteins Associated with a Trithorax-related SET Domain Protein. Proc. Nat. Acad. Sci. **98,** 12902-12907.

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

Moazed, D. (2009). Small RNAs in Transcriptional Gene Silencing and Genome Defence. Nature **457,** 413-420.

Mohan, M., Herz, H.-M., Smith, E.R., Zhang, Y., Jackson, J., Washburn, M.P., Florens, L., Eissenberg, J.C., and Shilatifard, A. (2011). The COMPASS Family of H3K4 Methylases in *Drosophila*. Mol. Cell. Biol. **31,** 4310-4318.

Mohd-Sarip, A., Venturini, F., Chalkley, G.E., and Verrijzer, C.P. (2002). Pleiohomeotic Can Link Polycomb to DNA and Mediate Transcriptional Repression. Molecular and Cellular Biology **22,** 7473-7483.

Murawska, M., Kunert, N., van Vugt, J., Längst, G., Kremmer, E., Logie, C., and Brehm, A. (2008). dCHD3, a Novel ATP-Dependent Chromatin Remodeler Associated with Sites of Active Transcription. Mol. Cell. Biol. **28,** 2745-2757.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T. (2007). Development of Series of Gateway Binary Vectors, pGWBs, for Realizing Efficient Construction of Fusion Genes for Plant Transformation. J. Biosci. Bioeng. **104,** 34-41.

Nakahigashi, K., Jasencakova, Z., Schubert, I., and Goto, K. (2005). The *Arabidopsis* HETEROCHROMATIN PROTEIN1 Homolog (TERMINAL FLOWER2) Silences Genes Within the Euchromatic Region but not Genes Positioned in Heterochromatin. Plant Cell Physiol. **46,** 1747-1756.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Kocher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Muller, J. (2007). Pcl-PRC2

is Needed to Generate High Levels of H3-K27 Trimethylation at Polycomb Target Genes. EMBO J **26,** 4078-4088.

Ng, D.W.K., Chandrasekharan, M.B., and Hall, T.C. (2006). Ordered Histone Modifications are Associated with Transcriptional Poising and Activation of the phaseolin Promoter. Plant Cell **18,** 119-132.

Ng, J., Hart, C.M., Morgan, K., and Simon, J.A. (2000). A *Drosophila* ESC-E(Z) Protein Complex Is Distinct from other Polycomb Group Complexes and Contains Covalently Modified ESC. Mol. Cell. Biol. **20,** 3069-3078.

Nguyen Ba, A., Pogoutse, A., Provart, N., and Moses, A. (2009). NLStradamus: a Ssimple Hidden Markov Model for Nuclear Localization Signal Prediction. BMC Bioinformatics **10,** 202.

Nislow, C., Ray, E., and Pillus, L. (1997). SET1, A Yeast Member of theTrithorax Family, Functions in Transcriptional Silencing and Diverse Cellular Processes. Mol. Biol. Cell **8,** 2421-2436.

Noh, Y.-S., and Amasino, R.M. (2003). PIE1, an ISWI Family Gene, is Required for FLC Activation and Floral Repression in *Arabidopsis*. Plant Cell **15,** 1671-1682.

Ogas, J., Kaufmann, S., Henderson, J., and Somerville, C. (1999). PICKLE is a CHD3 Chromatin-remodeling Factor that Regulates the Transition from Embryonic to Vegetative Development in *Arabidopsis*. Proc. Nat. Acad. Sci. **96,** 13839-13844.

Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Mutations in FIE, a WD Polycomb Group Gene, Allow Endosperm Development without Fertilization. Plant Cell **11,** 407-416.

Oktaba, K., Gutiérrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E.M., and Müller, J. (2008). Dynamic Regulation by Polycomb Group Protein Complexes Controls Pattern Formation and the Cell Cycle in *Drosophila*. Dev. Cell **15,** 877-889.

Oliver, S.S., and Denu, J.M. (2011). Dynamic Interplay between Histone H3 Modifications and Protein Interpreters: Emerging Evidence for a "Histone Language". Chem. BioChem. **12,** 299-307.

Orlando, V., and Paro, R. (1993). Mapping Polycomb-repressed Domains in the Bithorax Complex Using in vivo Formaldehyde Cross-linked Chromatin. Cell **75,** 1187-1198.

Papoulas, O., Beek, S.J., Moseley, S.L., McCallum, C.M., Sarte, M., Shearn, A., and Tamkun, J.W. (1998). The *Drosophila* Trithorax Group Proteins BRM, ASH1 and ASH2 are Subunits of Distinct Protein Complexes. Development **125,** 3955-3966.

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

Pelegri, F., and Lehmann, R. (1994). A Role of Polycomb Group Genes in the Regulation of Gap Gene Expression in *Drosophila*. Genetics **136,** 1341-1353.

Petruk, S., Sedkov, Y., Smith, S., Tillib, S., Kraevski, V., Nakamura, T., Canaani, E., Croce, C.M., and Mazo, A. (2001). Trithorax and dCBP Acting in a Complex to Maintain Expression of a Homeotic Gene. Science **294,** 1331-1334.

Phillips, M.D., and Shearn, A. (1990). Mutations in Polycombeotic, a *Drosophila* Polycomb-group Gene, Cause a Wide Range of Maternal and Zygotic Phenotypes. Genetics **125,** 91-101.

Pien, S., Fleury, D., Mylne, J.S., Crevillen, P., Inze, D., Avramova, Z., Dean, C., and Grossniklaus, U. (2008). ARABIDOPSIS TRITHORAX1 Dynamically Regulates *FLOWERING LOCUS C* Activation via Histone 3 Lysine 4 Trimethylation. Plant Cell **20,** 580-588.

Poux, S., McCabe, D., and Pirrotta, V. (2001). Recruitment of Components of Polycomb Group Chromatin Complexes in *Drosophila*. Development **128,** 75-85.

Reynolds, N., Salmon-Divon, M., Dvinge, H., Hynes-Allen, A., Balasooriya, G., Leaford, D., Behrens, A., Bertone, P., and Hendrich, B. (2012). NuRD-Mediated Deacetylation of H3K27 Facilitates Recruitment of Polycomb Repressive Complex 2 to Direct Gene Repression. EMBO J **31,** 593-605.

Rost, B., Yachdav, G., and Liu, J. (2004). The PredictProtein Server. Nucleic Acids Res. **32,** W321-W326.

Saleh, A., Al-Abdallat, A., Ndamukong, I., Alvarez-Venegas, R., and Avramova, Z. (2007). The *Arabidopsis* Homologs of *Trithorax* (*ATX1*) and *Enhancer of zeste* (*CLF*) Establish 'Bivalent Chromatin Marks' at the Silent *AGAMOUS* Locus. Nucl. Acids Res. **35,** 6290-6296.

Saleh, A., Alvarez-Venegas, R., Yilmaz, M., Le, O., Hou, G., Sadder, M., Al-Abdallat, A., Xia, Y., Lu, G., Ladunga, I., and Avramova, Z. (2008). The Highly Similar *Arabidopsis* Homologs of Trithorax *ATX1* and *ATX2* Encode Proteins with Divergent Biochemical Functions. Plant Cell **20,** 568-579.

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., and Coupland, G. (2000). Distinct Roles of CONSTANS Target Genes in Reproductive Development of Arabidopsis. Science **288,** 1613-1616.

Sanchez-Pulido, L., Devos, D., Sung, Z., and Calonje, M. (2008). RAWUL: A new Ubiquitin-like Domain in PRC1 Ring Finger Proteins that Unveils Putative Plant and Worm PRC1 Orthologs. BMC Genomics **9,** 308.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature **419,** 407-411.

Sarnowski, T.J., wiez?ewski, S., Pawlikowska, K., Kaczanowski, S., and Jerzmanowski, A. (2002). AtSWI3B, an *Arabidopsis* Homolog of SWI3, a Core Subunit of Yeast Swi/Snf Chromatin Remodeling Complex, Interacts with FCA, a Regulator of Flowering Time. Nucleic Acids Res. **30,** 3412-3421.

Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Muller, J. (2010). Histone H2A Deubiquitinase Activity of the Polycomb Repressive Complex PR-DUB. Nature **465,** 243-247.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., Bunker, R.D., Wirth, U., Bouwmeester, T., Bauer, A., Ly-Hartig, N., Zhao, K., Chan, H., Gu, J., Gut, H., Fischle, W., Müller, J., and Thoma¨, N.H. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Mol. Cell **42,** 330-341.

Schubert, D., Primavesi, L., Bishopp, A., Roberts, G., Doonan, J., Jenuwein, T., and Goodrich, J. (2006). Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. EMBO J **25,** 4638-4649.

Schuettengruber, B., Martinez, A.-M., Iovino, N., and Cavalli, G. (2011). Trithorax Group Proteins: Switching Genes ON and Keeping Them Active. Nat Rev Mol Cell Biol **12,** 799-814.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb Silencing Mechanisms and the Management of Genomic Programmes. Nat Rev Genet **8,** 9-22.

Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Krober, S., Amasino, R.A., and Coupland, G. (2006). The Transcription Factor *FLC* Confers a Flowering Response to Vernalization by Repressing Meristem Competence and Systemic Signaling in *Arabidopsis*. Genes Dev. **20,** 898-912.

Sedkov, Y., Tillib, S., Mizrokhi, L., and Mazo, A. (1994). The Bithorax Complex is Regulated by Trithorax Earlier During *Drosophila* Embryogenesis than is the Antennapedia Complex, Correlating with a Bithorax-like Expression Pattern of Distinct Early Trithorax Transcripts. Development **120,** 1907-1917.

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

Shearn, A. (1989). The *Ash-1*, *Ash-2* and *Trithorax* Genes of *Drosophila melanogaster* are Functionally Related. Genetics **121,** 517-525.

Shen, L., Kang, Y.G.G., Liu, L., and Yu, H. (2011). The J-Domain Protein J3 Mediates the Integration of Flowering Signals in *Arabidopsis*. Plant Cell **23,** 499-514.

Shilatifard, A. (2006). Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. Annu. Rev. Biochem. **75,** 243-269.

Shilatifard, A. (2008). Molecular Implementation and Physiological Roles for Histone H3 Lysine 4 (H3K4) Methylation. Curr. Opin. Cell Biol. **20,** 341-348.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A., and Lukyanov, S.A. (1995). An Improved PCR Method for Walking in Uncloned Genomic DNA. Nucleic Acids Res. **23,** 1087-1088.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the Drosophila Bithorax Complex That Mediate Repression by Polycomb Group Products. Developmental Biology **158,** 131-144.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of Polycomb Gene Silencing: Knowns and Unknowns. Nat Rev Mol Cell Biol **10,** 697-708.

Sinzelle, L., Izsvák, Z., and Ivics, Z. (2009). Molecular Domestication of Transposable Elements: From Detrimental Parasites to Useful Host Genes. Cell.Mol. Life Sci. **66,** 1073-1093.

Sinzelle, L., Kapitonov, V.V., Grzela, D.P., Jursch, T., Jurka, J., Izsvák, Z., and Ivics, Z. (2008). Transposition of a Reconstructed Harbinger Element in Human Cells and Functional Homology with Two Transposon-derived Cellular Genes. Proc. Nat. Acad. Sci. **105,** 4715-4720.

Soppe, W.J., Bentsink, L., and Koornneef, M. (1999). The Early-flowering Mutant *efs* is Involved in the Autonomous Promotion Pathway of *Arabidopsis thaliana*. Development **126,** 4763-4770.

Spillane, C., MacDougall, C., Stock, C., Köhler, C., Vielle-Calzada, J.P., Nunes, S.M., Grossniklaus, U., and Goodrich, J. (2000). Interaction of the *Arabidopsis* Polycomb Group Proteins FIE and MEA Mediates Their Common Phenotypes. Curr. Biol. **10,** 1535-1538.

Srinivasan, S., Dorighi, K.M., and Tamkun, J.W. (2008). *Drosophila* Kismet Regulates Histone H3 Lysine 27 Methylation and Early Elongation by RNA Polymerase II. PLoS Genet **4,** e1000217.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated Ubiquitination of H2A Restrains Poised RNA Polymerase II at Bivalent genes in Mouse ES Cells. Nat Cell Biol **9,** 1428-1435.

Struhl, G. (1981). A Gene Product Required for Correct Initiation of Segmental Determination in *Drosophila*. Nature **293,** 36-41.

Struhl, G., and Akam, M. (1985). Altered Ddistributions of *Ultrabithorax* Ttranscripts in Extra Sex Combs Mutant Embryos of *Drosophila*. EMBO J **4,** 3259-3264.

Sung, S., and Amasino, R.M. (2004). Vernalization in *Arabidopsis thaliana* is Mediated by the PHD Finger Protein VIN3. Nature **427,** 159-164.

Sung, Z.R., Belachew, A., Shunong, B., and Bertrand-Garcia, R. (1992). EMF, an *Arabidopsis* Gene Required for Vegetative Shoot Development. Science **258,** 1645-1647.

Takahashi, Y.-h., Westfield, G.H., Oleskie, A.N., Trievel, R.C., Shilatifard, A., and Skiniotis, G. (2011). Structural Analysis of the Core COMPASS Family of Histone H3K4 Methylases from Yeast to Human. Proc. Nat. Acad. Sci. **108,** 20526-20531.

Tanaka, Y., Katagiri, Z.-i., Kawahashi, K., Kioussis, D., and Kitajima, S. (2007). Trithorax-group Protein ASH1 Methylates Histone H3 Lysine 36. Gene **397,** 161-168.

Thorstensen, T., Grini, P.E., and Aalen, R.B. (2011). SET Domain Proteins in Plant Development. Biochim. Biophys. Acta - Gene Regul. Mech. **1809,** 407-420.

Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E., and Harte, P.J. (2001). The *Drosophila* Polycomb Group Proteins ESC and E(Z) are Present in a Complex Containing the Histone-binding Protein p55 and the Histone Deacetylase RPD3. Development **128,** 275-286.

Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated Acetylation of Histone H3 Lysine 27 Antagonizes *Drosophila* Polycomb Silencing. Development **136,** 3131-3141.

Tillib, S., Petruk, S., Sedkov, Y., Kuzin, A., Fujioka, M., Goto, T., and Mazo, A. (1999). Trithorax- and Polycomb-group Response Elements within an *Ultrabithorax* Transcription Maintenance Unit Consist of Closely Situated but Separable Sequences. Mol. Cell. Biol. **19,** 5189-5202.

Tsai, A.Y.-L., and Gazzarrini, S. (2012). AKIN10 and FUSCA3 Interact to Control Lateral Organ Development and Phase Transitions in *Arabidopsis*. Plant J. **69,** 809-821.

Tsukiyama, T., and Wu, C. (1995). Purification and Properties of an ATP-dependent Nucleosome Remodeling Factor. Cell **83,** 1011-1020.

Tsukiyama, T., Daniel, C., Tamkun, J., and Wu, C. (1995). ISWI, a Member of the SWl2/SNF2 ATPase Family, Encodes the 140 kDa Subunit of the Nucleosome Remodeling Factor. Cell **83,** 1021-1026.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.-L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007). *Arabidopsis* TFL2/LHP1 Specifically Associates with Genes Marked by Trimethylation of Histone H3 Lysine 27. PLoS Genet **3,** e86.

Ueno, Y., Ishikawa, T., Watanabe, K., Terakura, S., Iwakawa, H., Okada, K., Machida, C., and Machida, Y. (2007). Histone Deacetylases and

ASYMMETRIC LEAVES2 Are Involved in the Establishment of Polarity in Leaves of Arabidopsis. The Plant Cell Online **19,** 445-457.

van der Vlag, J., and Otte, A.P. (1999). Transcriptional Repression Mediated by the Human Polycomb-group Protein EED Involves Histone Deacetylation. Nat Genet **23,** 474-478.

Vastenhouw, N.L., and Schier, A.F. (2012). Bivalent Histone Modifications in Early Embryogenesis. Curr. Opin. Cell Biol. **24,** 374-386.

Vielle-Calzada, J.-P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M.A., and Grossniklaus, U. (1999). Maintenance of Genomic Imprinting at the *Arabidopsis medea* Locus Requires Zygotic DDM1 activity. Gene Dev. **13,** 2971-2982.

Wagner, D., and Meyerowitz, E.M. (2002a). SPLAYED, a Novel SWI/SNF ATPase Homolog, Controls Reproductive Development in Arabidopsis. Current Biology **12,** 85-94.

Wagner, D., and Meyerowitz, E.M. (2002b). SPLAYED, a Novel SWI/SNF ATPase Homolog, Controls Reproductive Development in *Arabidopsis*. Curr. Biol. **12,** 85-94.

Wang, H., Boavida, L.C., Ron, M., and McCormick, S. (2008). Truncation of a Protein Disulfide Isomerase, PDIL2-1, Delays Embryo Sac Maturation and Disrupts Pollen Tube Guidance in *Arabidopsis thaliana*. Plant Cell **20,** 3300-3311.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of Histone H2A Ubiquitination in Polycomb Silencing. Nature **431,** 873-878.

Weigel, D., Ahn, J.H., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Malancharuvil, E.J., Neff, M.M., Nguyen, J.T., Sato, S., Wang, Z.-Y., Xia, Y., Dixon, R.A., Harrison, M.J., Lamb, C.J., Yanofsky, M.F., and Chory, J. (2000). Activation Tagging in *Arabidopsis*. Plant Physiol. **122,** 1003-1014.

Wilkins, R.C., and Lis, J.T. (1997). Dynamics of Potentiation and Activation: GAGA Factor and its Role in Heat Shock Gene Regulation. Nucleic Acids Res. **25,** 3963-3968.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. PLoS ONE **2,** e718.

Wood, C.C., Robertson, M., Tanner, G., Peacock, W.J., Dennis, E.S., and Helliwell, C.A. (2006). The *Arabidopsis thaliana* Vernalization Rresponse Rrequires a Polycomb-like Protein Complex that also Includes *VERNALIZATION INSENSITIVE 3*. Proc, Nat. Acad. Sci. **103,** 14631–14636.

Wu, M.-F., Sang, Y., Bezhani, S., Yamaguchi, N., Han, S.-K., Li, Z., Su, Y., Slewinski, T.L., and Wagner, D. (2012). SWI2/SNF2 Chromatin Remodeling ATPases Overcome Polycomb Repression and Control Floral Organ Identity with the LEAFY and SEPALLATA3 Transcription Factors. Proc. Nat. Acad. Sci. **109,** 3576-3581.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., Wu, C., and Allis, C.D. (2006). A PHD finger of NURF Couples Histone H3 lysine 4 Trimethylation with Chromatin Remodelling. Nature **442,** 86-90.

Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. (2003a). Phosphorylation of RNA Polymerase II CTD Regulates H3 Methylation in Yeast. Gene Dev. **17,** 654-663.

Xiao, T., Kao, C.-F., Krogan, N.J., Sun, Z.-W., Greenblatt, J.F., Osley, M.A., and Strahl, B.D. (2005). Histone H2B Ubiquitylation is Associated with Elongating RNA Polymerase II. Mol. Cell. Biol., 637–651.

Xiao, W., Gehring, M., Choi, Y., Margossian, L., Pu, H., Harada, J.J., Goldberg, R.B., Pennell, R.I., and Fischer, R.L. (2003b). Imprinting of the MEA

Polycomb Gene Is Controlled by Antagonism between MET1 Methyltransferase and DME Glycosylase. Dev. Cell **5,** 891-901.

Xu, L., and Shen, W.-H. (2008). Polycomb Silencing of *KNOX* Genes Confines Shoot Stem Cell Niches in *Arabidopsis*. Curr. Biol. **18,** 1966-1971.

Xu, L., Xu, Y., Dong, A., Sun, Y., Pi, L., Xu, Y., and Huang, H. (2003). Novel *as1* and *as2* defects in Leaf Adaxial-abaxial Polarity Reveal the Requirement for ASYMMETRIC LEAVES1 and 2 and ERECTA Functions in Specifying Leaf Adaxial Identity. Development **130,** 4097-4107.

Xu, L., Zhao, Z., Dong, A., Soubigou-Taconnat, L., Renou, J.-P., Steinmetz, A., and Shen, W.-H. (2008). Di- and Tri- but Not Monomethylation on Histone H3 Lysine 36 Marks Active Transcription of Genes Involved in Flowering Time Regulation and Other Processes in *Arabidopsis thaliana.* Mol. Cell. Biol. **28,** 1348-1360.

Yang, C.-H., Chen, L.-J., and Sung, Z.R. (1995). Genetic Regulation of Shoot Development in *Arabidopsis*: Role of the EMF Genes. Dev. Biol. **169,** 421-435.

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., and Meyerowitz, E.M. (1990). The Protein Encoded by the *Arabidopsis* Homeotic Gene *AGAMOUS* Resembles Transcription Factors. Nature **346,** 35-39.

Yanovsky, M.J., and Kay, S.A. (2002). Molecular Basis of Seasonal Time Measurement in *Arabidopsis*. Nature **419,** 308-312.

Yoshida, N., Yanai, Y., Chen, L., Kato, Y., Hiratsuka, J., Miwa, T., Sung, Z.R., and Takahashi, S. (2001). EMBRYONIC FLOWER2, a Novel Polycomb Group Protein Homolog, Mediates Shoot Development and Flowering in *Arabidopsis*. Plant Cell **13,** 2471-2481.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 Methylation Antagonizes PRC2-mediated H3K27 Methylation. J. Biol. Chem. **286,** 7983-7989.

Yuan, Y.-W., and Wessler, S.R. (2011). The Catalytic Domain of all Eukaryotic Cut-and-paste Transposase Superfamilies. Proc. Nat. Acad. Sci. **108,** 7884-7889.

Yun, J.-Y., Tamada, Y., Kang, Y.E., and Amasino, R.M. (2012). *ARABIDOPSIS TRITHORAX-RELATED3/SET DOMAIN GROUP2* is Required for the Winter-Annual Habit of *Arabidopsis thaliana.* Plant Cell Physiol. **53,** 834-846.

Zeng, L., and Zhou, M.-M. (2002). Bromodomain: an Acetyl-lysine Binding Domain. FEBS Letters **513,** 124-128.

Zhang, X., Germann, S., Blus, B.J., Khorasanizadeh, S., Gaudin, V., and Jacobsen, S.E. (2007). The Arabidopsis LHP1 Protein Colocalizes with Histone H3 Lys27 Trimethylation. Nat Struct Mol Biol **14,** 869-871.

Zhao, Z., Yu, Y., Meyer, D., Wu, C., and Shen, W.-H. (2005). Prevention of Early Flowering by Expression of *FLOWERING LOCUS C* Requires Methylation of Histone H3 K36. Nat. Cell. Biol. **7,** 1256-1260.

Zheng, B., and Chen, X. (2011). Dynamics of histone H3 lysine 27 trimethylation in plant development. Current Opinion in Plant Biology **14,** 123-129.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation. Mol. Cell **29,** 69-80.

Appendix. Genes that mis-expressed in *sop12-1* compared to wild type (*Ws*) seedlings

Genes that expressed differently in *sop12-1* with wild type plants were list. logFC stands for log (fold change). Only the genes that showed logFC>2 or logFC<-2 were presented. P-value is the statistic parameter which indicates the significance of the difference between *sop12-1* and wild type samples. FDR stands for false discovery rate.

Gene	logFC	pValue	FDR
AT2G04050	-6.40	7.88E-31	2.57E-27
AT1G32350	-5.78	1.07E-10	2.65E-08
AT4G25200	-5.03	2.12E-07	2.49E-05
AT1G18410	-4.99	9.56E-18	5.98E-15
AT3G28580	-4.97	2.96E-12	9.93E-10
AT1G09080	-4.95	5.04E-08	6.83E-06
AT1G05250	-4.93	6.67E-07	6.89E-05
AT3G48850	-4.79	1.70E-07	2.07E-05
AT1G05240	-4.66	1.20E-05	0.00085
AT5G64810	-4.66	8.67E-11	2.17E-08
AT5G24110	-4.58	2.15E-05	0.001384
AT4G31690	-4.55	6.92E-05	0.003604
AT1G60130	-4.55	1.49E-10	3.33E-08
AT2G32140	-4.49	3.86E-05	0.002252
AT1G18280	-4.31	0.000223	0.009055

AT4G12490	-4.15	5.24E-07	5.57E-05
AT4G25820	-4.14	0.000402	0.01418
AT3G54340	-4.14	0.000722	0.022096
AT1G08670	-4.12	0.000722	0.022096
AT4G23230	-4.03	5.05E-18	3.29E-15
AT4G37990	-4.03	0.001299	0.034477
AT2G32020	-4.02	0.000722	0.022096
AT1G08860	-4.01	0.001299	0.034477
AT1G48930	-3.95	3.79E-15	1.96E-12
AT3G48630	-3.91	0.001299	0.034477
AT5G62480	-3.88	3.92E-15	1.99E-12
AT5G52760	-3.88	1.22E-08	2.00E-06
AT1G24147	-3.81	6.02E-23	8.16E-20
AT1G55600	-3.79	2.36E-05	0.0015
AT1G13310	-3.72	4.07E-05	0.002346
AT3G01830	-3.65	1.78E-07	2.16E-05
AT3G12320	-3.62	1.08E-22	1.41E-19
AT1G57780	-3.62	3.95E-10	8.39E-08
AT3G23250	-3.49	0.000209	0.008659
AT3G48640	-3.49	0.000209	0.008659
AT2G20142	-3.46	8.07E-20	7.30E-17
AT3G46090	-3.42	3.29E-08	4.69E-06
AT1G19250	-3.42	0.00062	0.019797
AT4G31630	-3.39	0.000124	0.005805
AT3G57460	-3.38	2.44E-10	5.29E-08
AT5G24640	-3.33	0.000124	0.005805
AT2G32130	-3.32	0.001066	0.02993
AT4G26200	-3.32	0.00062	0.019797
AT1G35730	-3.25	0.000223	0.009055
AT5G42380	-3.24	0.001828	0.044499
AT1G31580	-3.22	5.02E-15	2.51E-12
AT5G41310	-3.19	2.06E-15	1.10E-12
AT1G31640	-3.16	0.000223	0.009055

AT2G29350	-3.15	1.66E-06	0.000155
AT1G30370	-3.14	0.000402	0.01418
AT5G37490	-3.14	0.001828	0.044499
AT1G23110	-3.14	0.001828	0.044499
AT4G19430	-3.10	1.62E-19	1.29E-16
AT3G07970	-3.08	4.07E-05	0.002346
AT4G12500	-3.08	1.15E-05	0.00082
AT3G50930	-3.07	9.67E-21	1.05E-17
AT2G05635	-3.03	0.001299	0.034477
AT5G48657	-3.02	3.07E-05	0.001867
AT5G24206	-3.02	2.12E-07	2.49E-05
AT2G13810	-3.01	3.54E-07	3.91E-05
AT4G04490	-3.01	2.59E-12	8.88E-10
AT3G60920	-2.96	4.01E-10	8.47E-08
AT3G21040	-2.93	0.000722	0.022096
AT1G48940	-2.93	0.000209	0.008659
AT4G01360	-2.91	0.000795	0.023952
AT1G57670	-2.89	0.000103	0.004935
AT5G24150	-2.89	3.62E-11	1.02E-08
AT2G04040	-2.87	5.53E-16	3.10E-13
AT3G15670	-2.77	0.000172	0.007505
AT4G16590	-2.75	0.00062	0.019797
AT4G15480	-2.74	0.001066	0.02993
AT4G31940	-2.72	1.33E-06	0.000127
AT4G15620	-2.72	1.42E-11	4.36E-09
AT2G18690	-2.69	1.85E-08	2.88E-06
AT5G02780	-2.66	4.66E-07	4.99E-05
AT5G47740	-2.66	0.000172	0.007505
AT1G73810	-2.65	4.89E-06	0.000397
AT4G00390	-2.65	0.000287	0.010951
AT2G18193	-2.64	1.46E-13	5.76E-11
AT5G23830	-2.60	1.46E-05	0.000989
AT3G22910	-2.60	3.86E-08	5.42E-06

AT3G09600	-2.59	1.00E-08	1.66E-06
AT5G37690	-2.58	5.45E-11	1.48E-08
AT3G09520	-2.57	0.000478	0.016142
AT1G18830	-2.57	2.83E-13	1.04E-10
AT5G52690	-2.56	0.001828	0.044499
AT5G20240	-2.43	3.86E-08	5.42E-06
AT1G26380	-2.42	4.75E-06	0.000387
AT4G21440	-2.42	0.001318	0.034734
AT1G61800	-2.42	4.65E-14	1.99E-11
AT4G08040	-2.39	1.70E-08	2.71E-06
AT2G46830	-2.37	1.18E-15	6.38E-13
AT5G52750	-2.36	7.50E-12	2.37E-09
AT1G57800	-2.35	1.77E-09	3.34E-07
AT3G03726	-2.34	0.000919	0.026865
AT5G44420	-2.32	6.54E-08	8.65E-06
AT2G30750	-2.29	6.09E-10	1.25E-07
AT4G14370	-2.28	1.46E-05	0.000989
AT3G09940	-2.25	6.76E-07	6.97E-05
AT5G24205	-2.24	1.92E-09	3.58E-07
AT1G18860	-2.23	0.001481	0.038142
AT1G01580	-2.22	7.47E-10	1.53E-07
AT1G66700	-2.21	4.24E-07	4.59E-05
AT5G18720	-2.20	0.000999	0.028596
AT2G16835	-2.19	2.34E-08	3.55E-06
AT4G12545	-2.19	0.000999	0.028596
AT2G16832	-2.18	0.001054	0.02976
AT2G16840	-2.18	0.001054	0.02976
AT3G25250	-2.17	0.000673	0.021075
AT1G13470	-2.16	4.52E-10	9.43E-08
AT4G27250	-2.16	9.01E-05	0.004421
AT5G41690	-2.16	1.58E-11	4.76E-09
AT1G17960	-2.16	4.25E-08	5.91E-06
AT4G19690	-2.15	0.000206	0.008592

AT1G31485	-2.12	0.000252	0.009903
AT1G51890	-2.11	0.000206	0.008592
AT3G21050	-2.09	8.27E-05	0.004173
AT1G12600	-2.07	0.000464	0.01578
AT5G09570	-2.06	9.01E-05	0.004421
AT4G36610	-2.04	9.97E-05	0.004799
AT1G05680	-2.02	7.87E-13	2.73E-10
AT1G56060	-2.02	3.22E-06	0.000279
AT1G18140	-2.01	9.57E-05	0.004627
AT1G09932	-2.01	6.12E-11	1.63E-08
AT2G38250	-2.01	1.44E-06	0.000137
AT5G26220	-2.01	9.93E-20	8.51E-17
AT2G35730	2.02	0.00031	0.011645
AT5G65070	2.02	3.53E-05	0.002094
AT5G65800	2.04	1.20E-06	0.000117
AT1G23930	2.07	1.62E-08	2.60E-06
AT2G35770	2.07	1.05E-07	1.33E-05
AT1G63060	2.09	2.96E-06	0.000259
AT1G62940	2.09	3.63E-08	5.13E-06
AT3G47040	2.12	0.000143	0.006602
AT3G11870	2.17	1.52E-08	2.45E-06
AT2G35890	2.17	2.92E-07	3.28E-05
AT5G04970	2.18	5.42E-07	5.75E-05
AT2G34890	2.23	0.000795	0.023952
AT1G23910	2.24	0.001481	0.038142
AT1G52690	2.25	7.89E-10	1.57E-07
AT5G43980	2.32	3.22E-10	6.95E-08
AT4G30097	2.32	0.000569	0.018672
AT1G67420	2.33	7.15E-27	1.55E-23
AT3G63270	2.36	6.52E-30	1.93E-26
AT1G03445	2.56	5.69E-07	6.01E-05
AT3G47760	2.59	1.85E-08	2.88E-06
AT3G47660	2.68	1.63E-06	0.000152

AT1G30100	2.68	0.001828	0.044499
AT1G67260	2.83	3.92E-05	0.002272
AT1G28590	2.87	1.03E-14	5.06E-12
AT4G15710	2.88	8.17E-05	0.004141
AT1G55800	2.93	2.99E-09	5.41E-07
AT3G54940	3.00	0.00036	0.012987
AT5G43290	3.03	9.61E-16	5.30E-13
AT3G16390	3.04	8.39E-19	5.81E-16
AT1G63030	3.06	1.44E-06	0.000137
AT1G62978	3.06	0.001318	0.034734
AT1G63005	3.06	0.001318	0.034734
AT4G00580	3.08	8.17E-05	0.004141
AT5G04210	3.12	3.07E-05	0.001867
AT5G43490	3.30	9.16E-18	5.84E-15
AT3G63260	3.37	1.98E-64	3.71E-60
AT3G47790	3.38	5.06E-12	1.65E-09
AT3G07250	3.39	1.31E-05	0.000913
AT1G67365	3.41	2.05E-08	3.15E-06
AT1G66950	3.45	3.65E-07	4.00E-05
AT1G67000	3.48	6.16E-05	0.003248
AT5G05060	3.58	8.31E-09	1.39E-06
AT2G18720	3.60	0.000722	0.022096
AT5G04275	3.65	0.001299	0.034477
AT2G34315	3.65	0.001299	0.034477
AT5G14980	3.77	0.000223	0.009055
AT5G40290	3.82	9.14E-11	2.27E-08
AT2G05240	3.84	0.000402	0.01418
AT2G34360	3.90	4.83E-12	1.60E-09
AT3G54070	3.96	0.000223	0.009055
AT2G34100	4.03	0.000124	0.005805
AT5G43840	4.11	2.22E-16	1.31E-13
AT1G73300	4.15	7.60E-08	9.81E-06
AT1G67220	4.17	7.57E-39	4.11E-35

AT1G67640	4.18	1.39E-11	4.31E-09
AT5G43580	4.26	2.59E-15	1.36E-12
AT2G34030	4.28	1.20E-05	0.00085
AT1G67150	4.49	1.51E-20	1.53E-17
AT3G53065	4.58	1.47E-11	4.48E-09
AT2G34350	4.62	7.92E-06	0.000592
AT2G34210	4.63	4.22E-23	5.97E-20
AT5G43935	4.66	3.76E-07	4.11E-05
AT5G43660	4.70	3.29E-44	2.68E-40
AT1G67160	4.71	1.55E-06	0.000146
AT1G67148	4.93	1.08E-24	1.67E-21
AT5G05150	5.22	4.22E-09	7.34E-07
AT1G67623	5.24	2.67E-17	1.61E-14
AT4G27160	5.49	2.50E-38	1.16E-34
AT1G67270	5.69	4.40E-22	5.51E-19
AT5G43650	5.81	5.45E-24	8.06E-21
AT5G43300	5.93	3.99E-42	2.60E-38
AT1G67390	6.18	1.10E-29	2.98E-26
AT5G43403	6.22	1.50E-18	1.02E-15
AT1G67455	7.48	1.29E-36	5.26E-33