ALP1 and ELP1 – novel genetic modifiers of LIKE HETEROCHROMATIN PROTEIN 1 participate in Polycomb mediated gene repression in Arabidopsis

Inaugural-Dissertation

Zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Benjamin Hartwig

aus Winsen/Luhe

Die vorliegende Arbeit wurde am Max-Planck-Institut für Pflanzenzüchtungsforschung in Köln in der Abteilung für Entwicklungsbiologie der Pflanzen (Direktor: Prof. Dr. G. Coupland) angefertigt.

Prüfungsvorsitzender:	Prof. Dr. Martin Hülskamp
Berichterstatter:	Prof. Dr. George Coupland
	Prof. Dr. Ute Höcker
Prüfungsdatum:	16.04.2012

Abstract

Polycomb Group (PcG) proteins are involved in the inheritance of phenotypic traits by repressing the expression of target genes through binding of chromatin. In *Arabidopsis thaliana* several Polycomb Repressive Complexes (PRCs) are active during different stages of development. Distinct components of PRCs2, such as CURLY LEAF, SWINGER and MEDEA, contain a SET-domain for the trimethylation of lysine 27 at histone 3.

This trimethylationmark is bound by PRC1 which leads to stable repression of the corresponding euchromatic locus. LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) is a plant PRC1 component and an orthologue of HETEROCHROMATIN PROTEIN 1 (HP1) of humans, flies and yeast. Two of the PcG proteins that bind LHP1 catalyze mono-ubiquitylation of histone 2A but do not display homology to PRC1 PcG proteins in other organisms. The aim of this study is the identification of enhancers and suppressors of the *lhp1* mutant phenotype to unravel PRC1 characteristics and functions in plants.

Double mutants were generated with *lhp1* seeds in the Col-0 background through application of EMS and selected by phenotype and quantitative real-time PCR of known PRC1 target genes. Next-generation-sequencing of backcrossed, isogenic, bulked DNA samples of candidates was conducted. This isogenic mapping-by-sequencing approach revealed causal amino acid changes for two suppressors and three enhancers of LHP1. One enhancer candidate encodes a WD 40 domain protein predicted to function as an interaction hub preferentially for nuclear proteins. A suppressor gene product has similarities to *Harbinger* transposases and likely contains a functional DNA-binding domain.

De-regulated genes in those double mutants were analyzed *via* mRNA-seq, to link candidate genes to a possible function in chromatin regulation. In the suppressor the majority of de-regulated genes in *lhp1* were again expressed at wild-type (WT) levels. Some genes in the enhancer were de-regulated to a greater extend, than in *lhp1*. MADS-box transcription factors were up-regulated in *lhp1* compared to WT. The up-regulation of those genes was moderated in the suppressor and higher in the enhancer. Consequently the suppressor and the enhancer were named ANTAGONIST OF LHP1 1 (ALP1) and ENHANCER OF LHP1 1 (ELP1) respectively. ALP1 and ELP1 are nuclear proteins that both interacted with LHP1 in split-YFP assays. Further, ELP1 interacted with INCURVATA 2 (ICU2), a subunit of DNA-polymerase alpha that is binding LHP1 *in vitro*.

ALP1 and ELP1 present a possible link to epigenetic regulation. ALP1 could be a direct repressor of LHP1, while ELP1 might be part of PRC1.

Zusammenfassung

Polycomb Group (PcGs) Proteine nehmen auf die Vererbung phänotypischer Merkmale Einfluss, indem sie Chromatin binden und somit Zielgene reprimieren. In *Arabidopsis thaliana* sind mehrere Polycomb Repressive Komplexe (PRCs) in unterschiedlichen Entwicklungsphasen aktiv. Einzelne Komponenten von PRCs2, wie CURLY LEAF, SWINGER und MEDEA, beinhalten eine SET-Domäne, mittels derer sie Lysin 27 von Histon 3 tri-methylieren.

Diese Histonmarkierung wird von PRC1 gebunden, wodurch entsprechende, euchromatische Loci stabil reprimiert werden. LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) ist Teil von PRC1 in Pflanzen und ein Ortholog von HETEROCHROMATIN PROTEIN 1 (HP1) aus Menschen, Fliegen und Hefe. Zwei LHP1 bindende PcG Proteine katalysieren die Monoubiquitinierung von Histon 2A, zeigen jedoch keine Homologie zu PRC1 PcG Proteinen anderer Organismen. Das Ziel dieser Studie war die Indentifizierung von Verstärkern (Enhancer) und Unterdrückern (Suppressoren) des *lhp1* Mutantenphänotyps, um die Charakteristika und Funtionen von PRC1 in Pflanzen zu entschlüsseln.

Doppelmutanten wurden mit *lhp1* Samen im Col-0 Hintergrund durch die Anwendung von EMS generiert und phänotypisch, sowie durch quantitative Echtzeit-PCR bekannter PRC1 Zielgene selektiert. Isogene Mischproben der DNA von Kandidaten des Screens wurden mit Sequenziern der neuen Generation sequenziert. Diese isogene Katrierung durch Sequenzierungsanwendung enthüllte kausale Aminosäureänderungen für zwei Suppressoren und drei Enhancer von LHP1. Ein Enhancerkandidat kodiert für ein Protein mit einer WD 40 Domäne, welches als ein Protein mit vielen Interaktionen, preferentiell mit nuklearen Proteinen, prognostiziert wird. Ein Genprodukt eines Suppressors zeigt Ähnlichkeiten zu *Harbinger* Transposasen und enthält wohlmöglich eine funktionelle DNA-Bindedomäne.

Deregulierte Gene in diesen beiden Doppelmutanten wurden durch mRNA-Sequenzierung analysiert, um die Zielgene mit einer möglichen Funktion in der Regulation von Chromatin zu verbinden. Im Suppressor waren die Mehrheit der in *lhp1* de-regulierten Gene wieder auf Wildtyp (WT) Niveau exprimiert. Einige Gene des Enhancers waren in höherem Maße de-reguliert als in *lhp1*. MADS-box Transkriptionsfaktoren in *lhp1* waren hochreguliert im Vergleich zum WT. Die Hochregulation dieser Gene war abgeschwächt im Suppressor und erhöht im Enhancer. Aus diesem Grund wurden der Suppressor ANTAGONIST OF LHP1 1 (ALP1) und der Enhancer ENHANCER OF LHP1 1 (ELP1) genannt. ALP1 und ELP1 sind nukleare Proteine, die beide mit LHP1 in gespaltenem YFP Untersuchungen interagierten.

Weiterhin interagierte ELP1 mit INCURVATA 2 (ICU2), einer Untereinheit der DNA Polymerase Alpha, welche LHP1 *in vitro* bindet.

ALP1 und ELP1 zeigen eine mögliche Verbindung zu epigenetischer Regulation. ALP1 könnte als direkter Repressor von LHP1 fungieren, während ELP1 eventuell Teil eines PRC1 ist.

Table of Contents

Abstract **I**

Zusammenfassung **III**

Table of Contents **V**

1 Introduction **1**

1.1 Chromatin and transcription ... 1

1.2 Polycomb repressive complexes in different organisms ... 2

1.2.1 Composition and function of PRC1 ... 3

1.2.2 Composition and function of PRC2 ... 5

1.2.3 Targeting of specific DNA sequences by PRCs ... 6

1.2.4 Influence of histone modifications on transcription ... 7

1.2.5 Propagation of the chromatin state through DNA replication ... 8

1.3 LHP1 isolation, function and binding partners ... 9

1.3.1 Isolated *LHP1* mutant lines and their phenotypes ... 9

1.3.2 Impact on flowering time ... 11

1.3.3 Protein structure and function of LHP1 ... 12

1.4 Mapping-by-sequencing ... 13

2 Aim of the Study **17**

3 Material and Methods **19**

3.1 Plant material ... 19

3.1.1 Plant growth conditions ... 19

3.1.2 Flowering time measurements and phenotyping of plants ... 19

3.2 EMS treatment of seeds ... 20

3.3 Selection of potential mutants (putants) ... 20

3.3.1 Selection of enhancers and suppressors of lhp1 ... 21

3.3.2 Sanger sequencing to select enhancer candidates ... 21

3.4 Classical mapping of EMS induced SNPs ... 23

3.5 Next-generation mapping ... 24

3.6 Plasmid construction ... 25

3.6.1 Transgenic plants ... 27

3.7 Genotyping of T-DNA insertion lines and EMS mutant lines ... 27

3.8 DNA extraction ... 29

3.9 Quantification of transcription ... 30

3.10 mRNA-Seq 32
3.11 Western Blotting 33
3.12 Co-infiltration of tobacco leaves for split-YFP detection 34
3.13 Acceptor Photo Bleaching (APB) 34
3.14 Neighbor Joining analysis 35
3.15 Buffers, Solutions and Medium 35
4 Results **39**
4.1 Classification of EMS mutants 39
4.1.1 Altered conditions improved screening for aberrant phenotypes 40
4.1.2 Division of Class I into Class Ia and Class Ib 41
4.1.3 Sorting putants into subclasses 41
4.1.3.1 Enhancers of lhp1 43
4.1.3.2 Suppressors of *lhp1* 44
4.2 Analysis of enhancers and suppressors of *lhp1* 46
4.2.1 Expression analysis of enhancers and suppressors 46
4.2.2 Description of *sup3* and *enh9* phenotypes 47
4.2.3 Comparative global gene expression analysis of *sup3*, *enh9*, *lhp1* and WT 49
4.2.4 Confirmation of mRNA-seq results 52
4.2.5 H3K27me3 decoration of de-regulated genes 54
4.2.6 Active and repressive marks do not change globally 55
4.3 Isogenic mapping of causal mutations by bulked re-sequencing 56
4.3.1 Validation of the SNP underlying the *alp1* phenotype 59
4.3.2 Phylogenetic analysis of ALP1 61
4.3.3 Validation of the SNP underlying the *elp1* phenotype 63
4.3.4 Single mutant analysis 64
4.3.5 *ALP1* and *ELP1* expression are not influenced by LHP1 65
4.3.6 Analysis of de-regulated MADS domain transcription factors 67
4.4 Nuclear localization and binding studies 68
4.4.1 ELP1 is a nuclear protein 68
4.4.2 ALP1 and ELP1 interact with LHP1 in tobacco leaf cells 69
4.4.3 YFP-ALP1 does not co-localize with CFP-LHP1 in tobacco nuclei 71
5 Discussion **73**
5.1 Selection of Enhancers and Suppressors of *lhp1* 73
5.1.1 Enhancers of *lhp1* fall into four categories 73

5.1.2 Flowering time of selected enhancers and *FT* regulation 75
5.1.3 Effect of ambient temperature changes 76
5.2 Isogenic bulked mapping-by-sequencing 76
5.2.1 Advantages of isogenic mapping-by-sequencing over conventional approaches 76
5.2.2 Enhancers of LHP1 reveal a relation to telomere regulation 78
5.3 *ALP1* and *ELP1* – two sides of *LHP1* 79
5.3.1 ALP1 is a domesticated transposase-like gene 79
5.3.2 Localization and binding properties of ALP1 81
5.3.3 ALP1 does not alter selected histone marks globally 81
5.3.4 A dual function of ALP1 in gene activation 82
5.3.5 *ELP1* is neither regulated by LHP1 nor by its neighboring gene 84
5.3.6 ELP1 is an interaction hub binding LHP1 in the nucleus 85
5.3.7 Interaction with ICU2 and influence on chromatin remodeling 86
6 Conclusions and Perspectives **87**
6 Bibliography **89**
7 Appendix **103**
7.1 Supplemental Figures 103
7.2 Supplemental Table 106
7.3 Abbreviations 109
7.3.1 General abbreviations 109
7.3.2 Gene and protein names 112
Acknowledgements **114**
Erklärung **116**
Lebenslauf **118**

1 Introduction

1.1 Chromatin and transcription

Plants can adapt their developmental programs according to internal and external stimuli they perceive. Different layers of epigenetic memory enable them to mold a stimulus into a developmental or responsive program without changing the DNA code. Those layers are methylation of DNA, exchange of histone variants, modification of histone tails and the positioning and higher-order structure of nucleosomes and DNA (Berger, 2007). Evidence accumulates that a mixture of all layers is responsible for the success of an adapted developmental program that can be inherited through multiple cell cycles (Ahmad et al., 2010; Beisel and Paro, 2011; Roudier et al., 2009).

In eukaryotes DNA is usually densely packed in the nuclei of cells. Packaging of DNA into higher-order is achieved by nucleosomes, which consist of 147 base pairs (bp) of DNA wrapped around histone octamers in a 1,65 super-helical turn (Luger et al., 1997). Histone core complexes are composed of four heterodimers of histone 2A and 2B (H2A and H2B), as well as histone 3 (H3) and histone 4 (H4). Histone 1 (H1) facilitates higher-order packaging by linking nucleosomes and DNA which results in tightly packed nucleosomal structures. Nucleosomes, DNA-binding factors, the basal transcription machinery and its emerging transcripts, DNA-repair and replication machineries as well as proteins interacting with all of those factors are referred to as chromatin (van Steensel, 2011).

Chromatin is viewed as a highly dynamic structure that affects numerous DNA transactions, such as transcription, replication, repair, recombination and transposition, as well as chromosome segregation (Margueron and Reinberg, 2010; Roudier et al., 2009; Simon and Kingston, 2009). The tails of histones are exposed and can be post-translationally modified, even when they are incorporated into nucleosomes. Modifications are particularly abundant on N-terminal tails of H3 and H4 and are highly conserved across kingdoms. Lysine residues can be modified by acetylation, methylation, ubiquitinylation or sumoylation. Arginine residues are only associated with symmetric or asymmetric methylation. Serine and threonine residues are prone to be phosphorylated (Bedford and Clarke, 2009; Kouzarides, 2007). Depending on the modified amino acid residue, methylation of histones can either be associated with activation or repression of transcription (Berger, 2007). Complexes binding histone marks associated with repression can aid in presenting a chromatin barrier for RNA

Polymerase II (Pol II). Complexes binding marks associated with activation facilitate transcriptional elongation and likely contain an anti-silencing function by relieving promoters of repressive marks (King et al., 2002; reviewed in Schuettengruber et al., 2011).

In plants, chromatin dynamics have been most extensively studied in *Arabidopsis thaliana* (Arabidopsis). This model organism has a genome size of 125 mega bases (Mb), containing about 28,000 genes (Kaul et al., 2000). Most of those genes are localized within euchromatin (~100 Mb), which is a gene-rich and less condensed form of chromatin. In contrast, heterochromatin is relatively gene-poor, stays highly condensed throughout most of the cell cycle and contains the highest number of transposable elements and other repetitive regions in Arabidopsis (Beisel and Paro, 2011; Grewal and Elgin, 2007). Epigenetic silencing of heterochromatic loci is generally associated with high levels of DNA methylation that interplays with histone modifications on a genome wide scale (Gendrel et al., 2002; Luo and Lam, 2010; Sridhar et al., 2007).

Euchromatin and heterochromatin also differ in the composition of their methylation marks. For instance, heterochromatic elements are prone to be stably silenced and therefore enriched in H3K9 di-methylation (H3K9me2) while this mark is rather elusive of euchromatic regions (Jenuwein and Allis, 2001; Ringrose and Paro, 2004). Instead, euchromatic genes that are repressed are covered by H3K27me3. Epigenetic repression through H3K27me3 can be at a dynamic equilibrium, as shown for *FLOWERING LOCUS C* (FLC), where the mark is accumulating in response to vernalization (Crevillen and Dean, 2011; Lafos and Schubert, 2009).

1.2 Polycomb repressive complexes in different organisms

The association of Polycomb repressive complexes (PRCs) and H3K27me3 is conserved from animals to plants (Beisel and Paro, 2011). H3K27 tri-methylation marks are set at specific target loci by PRC2 and subsequently bound by PRC1 to stabilize repression. Especially PRC2 components are highly conserved, down to unicellular eukaryotes. Despite a likely common ancestor of the plant and animal kingdom with a conserved role of PRC-mediated repression, this mechanism has probably been lost several times during development. For instance, H3K27me3 does not seem to be present in *Chlamydomonas reinhardtii* and PRC2 seems to be entirely absent in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (Shaver et al., 2010).

Many PRC targets are genes involved in development, which are tissue specifically repressed. It is still an open question, to what extend the methylation of histones has contributed to the evolution of multicellular organization and patterning of organs.

1.2.1 Composition and function of PRC1

Polycomb group (PcGs) genes were first described as genetic modifiers of homeotic (*Hox*) genes in *Drosophila melanogaster* (Drosophila) (reviewed in Kennison, 1995). The founding member of PcG proteins, Polycomb (PC), is a core protein of the PRC1 in Drosophila, which might consist of up to 11 components (Shao et al., 1999). Four Drosophila PRC1 components are best described. PC, Posterior Sex Combs (PSC) and Polyhomeotic (PH) interact with each other (Strutt and Paro, 1997). Additionally, PC and PSC interact with the Ring finger protein Sex Combs Extra (SCE) (Gorfinkiel et al., 2004). The complex binds to H3K27me3 or H3K9me3 through the chromodomain (CD) of PC. PH is aiding PC in binding and repression of target genes. SCE catalyzes mono-ubiquitylation of H2A at lysine 118 (H2A-K118ub) and PSC, another Ring finger protein, is involved in complex formation and the inhibition of transcription, likely through chromatin compaction (Francis et al., 2004; King et al., 2005; Zhou et al., 2008).

A family of PcGs in mammals is called Chromobox protein homologues (CBXs). CBX2, CBX4 and CBX6 to CBX8 are homologues of PC, while CBX5, CBX1 and CBX3 are alternative names for the three HETEROCHROMATIN PROTEIN 1 (HP1) proteins HP1α, HP1β and HP1γ, respectively and hence not part of PRC1 in mammals (Beisel and Paro, 2011; Bernstein et al., 2006). The other PRC1 components in mammals are also organized in small gene families (Margueron and Reinberg, 2011; Figure 1).

Since the description of PRC1 in 1999 almost a decade had past until a PRC1-like complex was described in plants, where no evident homologues of the core components of PRC1 seemed to exist. The exception was LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), which is a functional homologue of PC in that it binds to H3K27me3 on a genome wide scale (Turck et al., 2007; Zhang et al., 2007). Recently, two studies demonstrated the association of five RING proteins with LHP1 (Bratzel et al., 2010; Xu and Shen, 2008). AtRING1A and AtRING1B bind LHP1 *in vitro* and are the closest homologues to SCE in Arabidopsis (Sanchez-Pulido et al., 2008). AtBMI1A and AtBMI1B, homologues of PSC, also interact with LHP1 *in vitro*, while the role for the lowly expressed *AtBMI1C* gene has not been

demonstrated. The striking finding that H2A ubiquitylation is catalyzed by the RING finger proteins AtBMI1A and AtBMI1B, completes the emerging picture that one or several distinct

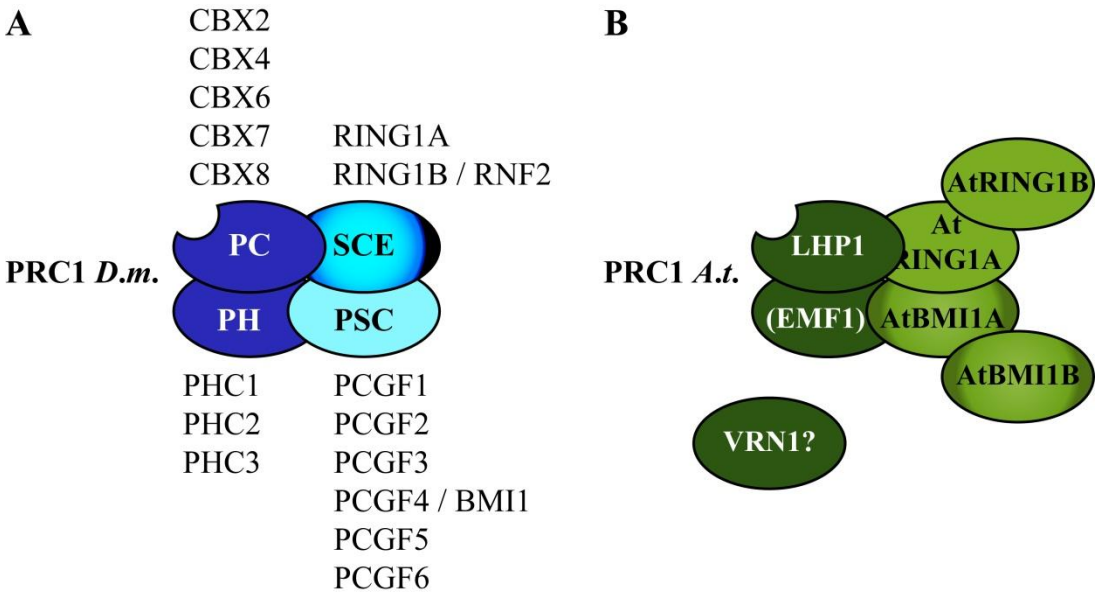


Figure 1: PRC1 core complexes in animals and plants
(**A**, **B**) Only the PRC1 core components are shown. Proteins with main catalytic activity are displayed in different coloring. RING proteins are in brighter shades. (**A**) Drosophila PRC1 contains four core components which have several homologues in humans (listed above or below). The PC protein binds H3K9me3 or H3K27me3 through its CD and SCE catalyzes H2A-K119ub. RING1B and BMI1 catalyze H2A-K118ub in mammals. (**B**) Not all core components of Arabidopsis PRC1 are known and there might be several active PRCs1 in plants. LHP1 binds H3K27me3 exclusively *in vivo* and H2Aub is catalyzed by AtBMI1A and AtBMI1B. EMF1 and VRN1 have not finally been placed into PRC1. Figure is adapted from (Beisel and Paro, 2011).

PRCs1 exist in Arabidopsis (Bratzel et al., 2010 and Figure 1). Notably, in mammals H2A-K119ub instead of H2A-K118ub is part of a stabilized repressive state. H2A-K119ub is mostly catalyzed by RING1B and stimulated through BMI1 (reviewed in Morey and Helin, 2010), which could be a possible scenario in plants as well.

EMBRYONIC FLOWER 1 (*EMF1*) in Arabidopsis might reside within one or several PRCs1. Mutants of the gene display a strong effect on flowering and floral architecture mainly through the up-regulation of the MADS (*MCM1* / *AGAMOUS* / *DEFICIENS* / *SRF*) domain transcription factor and flower homeotic gene *AGAMOUS* (*AG*) (Aubert et al., 2001; Calonje et al., 2008). Interaction of EMF1 with MSI1 (MULTICOPY SUPPRESSOR OF IRA 1, an Arabidopsis component of PRC2) as well as negative influence on transcription, much like PSC, has been demonstrated *in vitro*. In addition, EMF1 has been shown to bind AtRING1A, AtRING1B, AtBMI1A and AtBMI1B *in vitro* (Bratzel et al., 2010; Xu and Shen, 2008). Therefore, EMF1 can be placed in a plant PRC1 complex (Figure 1).

REDUCED VERNALIZATION RESPONSE 1 (VRN1) is a plant specific, B3-domain containing protein that binds double stranded DNA, localizes to euchromatin, stays associated with chromosomes through mitosis and like LHP1 has roles in flowering-time and vernalization response (Mylne et al., 2006). Thus, it might be a protein involved in PRC1 function. To date, no binding of VRN1 to other plant PRC1 components has been demonstrated (Figure 1).

LHP1 has been shown to directly repress MADS domain transcription factors (Germann et al., 2006; Mylne et al., 2006; Sung et al., 2006). This mechanism reminds of the PRC1 repression of *Hox* genes in Drosophila. MADS and *Hox* genes are both tightly interconnected with development but do not share sequence homology, which indicates that the repressive mechanism through H3K27me3 is functional at different target genes in the plant kingdom.

It is possible that several PRCs1 exist in parallel in plants to ensure regulation of expression patterns in different tissues and at different developmental stages. A hint that multiple PRCs1 act in parallel has been demonstrated for the globally mapped distribution of PRC1 in human tumor cell lines *via* ChIP-Seq (personal conference communication with Dr. Zhonghua Gao, Postdoctoral fellow in the lab of Dr. Danny Reinberg, New York). Supporting information for this hypothesis in plants comes from mutant studies of PRC1 components. AtRING1A and AtRING1B do not always influence transcription of the same target genes as LHP1 and phenotypes of triple *lhp1*;*Atring1a*;*Atring1b* mutants are not as severe as double mutants of *Atbmi1a*;*Atbmi1b* (Bratzel et al., 2010; Chen et al., 2010; Xu and Shen, 2008).

1.2.2 Composition and function of PRC2

The best described PRC2 complex in Drosophila consists of four core components. H3K27me3 (or H3K27me2) is established by Enhancer of Zeste (E(Z)), which has two homologues in humans (EZH1 and EZH2) and three in Arabidopsis (CURLY LEAF (CLF), SWINGER (SWN) and MEDEA (MEA)). All of these proteins contain a SET (SUVAR 3-9/E(Z)/Trithorax) domain and function like E(Z), which indicates a high conservation of H3K27me3 mediated gene repression in different organisms.

The other three core components in Drosophila are Extra Sex Combs (ESC), Suppressor of Zeste 12 SU(Z)12 and a subunit of Nucleosome Remodeling Factor (NURF), Nurf55/p55. Homologues for ESC and SU(Z)12 in humans are embryonic ectoderm development (EED) and SUZ12. Arabidopsis contains only one ESC homologue, FERTILIZATION INDEPENDENT ENDOSPERM (FIE), which is present in each described PRC2 complex

and three homologues for SU(Z)12, namely EMBRYONIC FLOWER 2 (EMF2), REDUCED VERNALIZATION RESPONSE 2 (VRN2) and FERTILIZATION INDEPENDENT SEED (FIS). Nurf55 has five described homologues in Arabidopsis (MSI1 - MSI5), of which MSI1 and MSI4 have been shown to directly associate with PRCs2. In humans, homologues of Nurf55 are histone chaperones RbAP48 and RbAP46. All three proteins bind to SU(Z)12 or SUZ12 respectively (Adrian et al., 2009; Beisel and Paro, 2011; Hennig and Derkacheva, 2009; Margueron and Reinberg, 2011; Pazhouhandeh et al., 2011).

Polycomblike (PCL) and JARID2 have been found to interact with the core components of PRC2 in Drosophila and humans. They might reside within the complex to diversify or enhance its function. It is also speculated that at least JARID2 could be involved in the establishment of PcG response elements (PREs). PREs are thought of as a conglomerate of *cis*-elements, non-coding RNAs (ncRNAs), sequence-specific DNA binding proteins and likely other elements that tether PRC2 to its targets (Margueron and Reinberg, 2011; Ringrose and Paro, 2007).

Two additional PRCs2 have been characterized in Drosophila. The DNA-binding PHO-repressive complex (PHORC) is associated with binding of mono- and di-methylated H3K9 and H4K20 marks. Secondly, Polycomb repressive deubiquitinase (PR-DUB) complex contains the protein Calypso which removes mono-ubiquitin specifically from H2A. Together with PRC1 and other ubiquitylating complexes, PR-DUB likely establishes a dynamic balance of ubiquitylation at H2A (Klymenko et al., 2006; Scheuermann et al., 2010). Homologues of these complexes exist in mammals, but their function has not yet been elucidated (Margueron and Reinberg, 2011).

Most PRC2 target genes are subsequently bound by PRC1 at H3K27me3 to stabilize repression. Hence, PCR1 has been placed down-stream of PRC2. Notably, in mammals there are genes methylated by PRC2 that lack H2A-K118ub (Ku et al., 2008) and genes targeted by PRC1 in the absence of PRC2 function (Schoeftner et al., 2006). Thus, both complexes do fulfill independent function at few loci. If this model accounts for plants as well remains to be determined.

1.2.3 Targeting of specific DNA sequences by PRCs

To initiate H3K27me3 at a specific, targeted region through the catalytic core of a PRC2 complex, there needs to be a decision-making process. In Drosophila, it has been suggested that PRC1 and PRC2 bind to PREs. However, those factors, such as the GAGA factor (GAF)

or Zeste are also more generally involved in activation and repression of genes apart from PcG repression (Ringrose and Paro, 2007).

Recent findings suggest that long non-coding RNAs (lncRNAs) and ncRNAs interact with PRC2 to guide them to target genes in *cis* or *trans* and possibly contribute in PRE function in mammals and plants. In Drosophila to date, there is no strong evidence that ncRNAs are involved in PRC target recognition (Heo and Sung, 2011; Kanhere et al., 2010; Rinn et al., 2007; Swiezewski et al., 2009).

More scientific proof accumulates that PREs exist in vertebrates and possibly in plants. Two PREs have been identified in mice that might be bound by the transcription factor yin-yang 1 (YY1), a homologue of the PcG protein Pleiohomeotic (PHO) in Drosophila (Sing et al., 2009; Woo et al., 2010). In Arabidopsis a PRE might be active at *FLC* influencing PcG recruitment and transcription (Buzas et al., 2011). *LEAFY COTYLEDON 2* (*LEC2*) is repressed by the FIE-PRC2 complex in Arabidopsis. Through studies of *LEC2* promoter a regulatory GAGA-element has been found that is bound by BASIC PENTACYSTEINE (BPC) and might be part of a functional PRE (Berger et al., 2011).

A factor that has yet not been considered much in the establishment of repression or activation of genes is the formation of *cis* or *trans* (different chromosomes) DNA loops. Such structures, have been demonstrated to aid PcG function and might contribute to functional PREs (Bantignies and Cavalli, 2011).

All described PRE layers make the process of repression dynamic and flexible. Given that several PRC complexes are active at different times of development add to the complexity of repression and silencing. The demonstration that PcG protein turn-over at target sites is rapid (Ficz et al., 2005), makes it surprising that repression and silencing can be stably established at all. Taken together, the model of a static repression by histone methylation has changed into one that is rather resembled by a “hamster running in a wheel”. Chromatin snapshots so far have only shown the apparently stationary “running wheel” while this state was kept up by constant changes that researchers are only beginning to understand.

1.2.4 Influence of histone modifications on transcription

Histone modifications could be either cause or consequence of target gene repression and silencing. It is not well understood how transcription could be influenced by the binding of PRC2 and if this process is rather a mechanical or an enzymatic one.

The compaction of chromatin through the establishment of H2Aub by PRC1 is thought to directly influence transcription by Pol II, which would strengthen the argument that the effect on transcription is steric (Zhou et al., 2008). On the same line, exchange of histone variants is another factor that influences transcription. Histone variants possibly affect the packaging of chromatin, which allows for a changed accessibility of transcription factors or a different binding affinity of Pol II to a locus (Morey and Helin, 2010). Interestingly, it was shown that incorporation of the histone variant H2A.Z is influenced by ambient temperature in Arabidopsis (Kumar and Wigge, 2010). This particular variant was incorporated into nucleosomes by chromatin remodelers at a higher rate, when temperature was lower. Conclusively, chromatin was packed more densely and affected transcription of genes. However, rate of transcription was not correlated to the amount of H2A.Z at promoters. Thus, it was concluded that H2A.Z incorporation presents a steric hindrance at binding sites of transcriptional activators and repressors (Franklin, 2010; Kumar and Wigge, 2010).

In contrast, JARID2 has been demonstrated to co-localize with PRC2 components and target genes (Pasini et al., 2010). It was demonstrated to be essential for early embryonic development of mammals and is binding DNA. JARID1 family proteins are specific de-methylases of histone marks associated with active transcription. Likely, JARID2 is not acting as a histone de-methylase, due to the lack of two conserved residues in a domain essential for metal-binding (Pasini et al., 2010). A link to transcription was established by the finding that S5-phosphorylated Pol II is missing from promoters when JARID2 is withdrawn (Landeira et al., 2010). Presence of S5-phosphorylated RNA Pol II defines the state of a poised promoter on genes that are not actively transcribed but can be rapidly up-regulated upon signal reception. The relationship between PcGs, histone modifications and transcription might therefore be a mixture of steric hindrance through chromatin compaction and a more direct influence on Pol II.

1.2.5 Propagation of the chromatin state through DNA replication

Once a methylation mark is set at a target site it spreads across a locus by a mechanism that has recently been described in mammals. Mitotic inheritance is a hallmark for *Hox* gene repression and plays an essential role in the establishment of organ development. The process of inheritance of methylation marks is fairly enigmatic. Though, some recent progress has been made in discussing the epigenetic part of Polycomb mediated gene repression.

The key finding that PRC2 is not only setting methylation marks, but can also bind to them was the foundation for a model of propagated methylation marks after DNA replication (Hansen et al., 2008). At least in mammals the PRC2 subunit EED is able to bind H3K27me3 (Margueron et al., 2009). When DNA-replication takes place and nucleosomes are evenly distributed among daughter strands the binding of EED to H3K27me3 might then quickly tether PRC2 to either strand. The methyl-transferase function of EZH1 or EZH2 would then de-novo methylate every other nucleosome in a self-reinforcing loop (Beisel and Paro, 2011). The combination of reading and writing H3K27me3 in one complex likely increases the speed of methylation beyond that of already rapid nucleosome turn-over (Deal et al., 2010), which makes a self-reinforcing model probable.

1.3 LHP1 isolation, function and binding partners

LHP1 is the only Arabidopsis homologue of the small HP1 family from metazoans (Gaudin et al., 2001). It is also the only Arabidopsis protein that contains an amino-terminal CD and a carboxy-terminal chromo-shadow domain (CSD) separated by a 'hinge' region. The CD is a shared domain between HP1 and PC which have a unified role in the ancient unicellular ciliate *Tetrahymena thermophila* (Liu et al., 2007). During evolution the role of PcGs and HP1 seem to have been separated to a euchromatic and a heterochromatic one. In plants, the role of LHP1 is again different from HP1, since it is mainly associated with repression of euchromatic genes (Libault et al. 2005) and is rather a functional homologue of PC (Exner et al., 2009; Turck et al., 2007; Zhang et al., 2007).

1.3.1 Isolated *LHP1* mutant lines and their phenotypes

Mutants of *LHP1* have been isolated several times in ethyl methanesulfonate (EMS) and fast neutron mutagenesis screens (Gaudin et al., 2001; Haughn et al., 1991; Kotake et al., 2003; Larsson et al., 1998). The first *LHP1* mutant identified in an EMS screen for altered glucosinolate levels in leaves was named *tu8* (Haughn et al., 1991). Later, *TU8* was described as a novel allele of *TERMINAL FLOWER 2* (*TFL2*), which now is referred to as *LHP1*, due to its function as a chromatin binding protein (Gaudin et al., 2001; Kim et al., 2004; Table 1).

Plants mutated in *lhp1* are early flowering independent of day length, show enhanced leaf curling, an overall reduced plant size, altered inflorescence meristem development and terminal flower formation (Kotake et al., 2003). In Table 1 the six published alleles of *LHP1* are listed. For all experiments conducted in this study, *lhp1-3*/*tfl2-1* mutants generated in the

Columbia (Col-0) background have been used. The $C_{1363}T$ mutation, at bp 1363 after the transcriptional start site (TSS), in *lhp1-3/tfl2-1* leads to a Q_{280}* (* = stop) amino acid change resulting in a truncated protein lacking the CSD. For simplification, *lhp1-3/tfl2-1* will hereafter be referred to as *lhp1*.

Table 1: Mutant alleles of *LHP1*

LHP1 mutant	Accession	Mutation	Published
lhp1-1 / tfl2-4	Ws-2	T-DNA insertion / Promoter indel	Gaudin et al. (2001)
lhp1-2 / tfl2-5	Ws-2	Deletion of T_{775} / Stop codon [I_{210}*]	Gaudin et al. (2001)
lhp1-3 / tfl2-1	Col-0	EMS / Stop codon $C_{1363}T$ [Q_{280}*]	Larsson et al. (1998)
lhp1-4 / tfl2-2	Col-0	Fast neutron / 1.3 kb promoter deletion	Larsson et al. (1998)
lhp1-5 / tfl2-3	Col-0	T-DNA insertion	Kotake et al. (2003)
lhp1-6 / tfl2-6 / tu8	Col-0	EMS / $C_{1144}T$ [$L_{207}V$]	Haughn et al. (1991)

When *lhp1* plants are compared to wild-type (WT) cellular organization, cell size and cell number of vegetative meristems are not affected (Larsson et al., 1998). There is also no visible effect on the distribution of stomata and leaves of *lhp1*. Mutant plants have about the same number of cells as WT plants. However, leaf cells in *lhp1* mutants are smaller, indicating that the expansion and growth of the cells rather than cell division is affected in mutants (Kotake et al., 2003; Larsson et al., 1998).

Arabidopsis plants normally form an indeterminate number of flowers and keep the apical meristem in an inflorescent state. Floral meristems are produced at its flanks. The mutation of *LHP1* results in the differentiation of the apical inflorescence meristem to a floral meristem. Terminal flower structures without a perianth and with a varying number of carpels and stamens are consequently observed (Penin et al., 2005). The severity of the phenotype was found to vary among Arabidopsis accessions and was comparatively mild in *lhp1* mutants used in this study (personal communication with Dr. Sara Farrona, Uni Düsseldorf).

The *lhp1* phenotype is relatively mild compared to double mutants of the PRC2 components *CLF* and *SWN* or the two PCR1 subunits *AtBMI1A* and *AtBMI1B* that grow as callus like structures (Chanvivattana et al., 2004; Chen et al., 2010; Farrona et al., 2011; Hariganeya et al., 2007). Plants mutated in either *AtRING1A* or *AtRING1B* are indistinguishable from WT. The *Atring1a;Atring1b* double mutant plants display fasciation of the stem and ectopic meristem formation. KNOTTED-like Homeobox (KNOX) genes, which are transcriptional

regulators that act in stem cell maintenance at the shoot apical meristem (SAM) and floral abscission, are de-repressed in double mutants (Scofield et al., 2008; Shi et al., 2011; Xu and Shen, 2008). Interestingly, the *Atring1a*;*Atring1b*;*lhp1* triple mutant still displays a plant phenotype, which is strongly hampered in its growth and development (Xu and Shen, 2008).
These findings are in favor of the hypothesis that multiple PRC1 complexes exist in plants. *LHP1* mutants show a different pleiotropic phenotype from *Atring1a*;*Atring1b* double mutants and *Atbmi1a*;*Atbmi1b* double mutants. Further, *Atring1a*;*Atring1b*;*lhp1* triple mutants grow as plants and not as callus-like structures. Finally, Arabidopsis PRC1 components repress different target genes which results in diverging phenotypes (Figure 2).
More detailed observation of histone marks and gene expression in the callus like structures formed by *clf*;*swn* and *Atbmi1a*;*Atbmi1b* double mutant plants are needed. Evidence for differences between PRC1 and PRC2 targets that have been found in mammals is still scarce in Arabidopsis (Ku et al., 2008).

1.3.2 Impact on flowering time

The change from vegetative to reproductive growth in Arabidopsis is regulated by four main pathways (Boss et al., 2004). The photoperiodic pathway and the vernalization pathway integrate light and temperature signals while the autonomous pathway and the gibberellic acid (GA) pathway integrate various different stimuli to the development of floral organs. Finally all pathways lead to the up-regulation of floral integrators or transcription factors initiating flowering programs at the SAM.
LHP1 is expressed in leaf cells, roots, hypocotyls, SAMs as well as in the vascular tissues of leaves and cotyledons (Kotake et al., 2003; Takada and Goto, 2003). The transcription factor *CONSTANS* (*CO*) and the floral integrator *FLOWERING LOCUS T* (*FT*) also display a vascular expression pattern in leaves. *CO* is expressed at a similar level in all leaf veins, whereas *FT* is rather expressed in the distal part and *LHP1* in the basal part of the leaves. All three genes share overlapping expression in phloem companion cells (Takada and Goto, 2003). While CO protein promotes the expression of *FT* when stabilized in long day (LD) conditions (Turck et al., 2008), LHP1 represses *FT* independent of day length and throughout development (Larsson et al., 1998; Takada and Goto, 2003). H3K27me3 is not sufficient to mediate repression of *FT* alone and needs to be bound by LHP1 (Adrian et al., 2010). Once *FT* is translated, FT protein travels to the SAM where it interacts with *FLOWERING LOCUS D* (*FD*) to trigger the up-regulation of *SUPPRESSOR OF OVEREXPRESSION OF*

CONSTANS 1 (*SOC1*), one of the earliest markers for the initiation of flowering (Fornara et al., 2010).

Apart from *FT*, LHP1 binds and represses the MADS box transcription factor *FLC* directly (Mylne et al., 2006; Sung et al., 2006). This finding is particularly interesting, because FLC together with the MADS box factor SHORT VEGETATIVE PHASE (SVP) in turn is a direct repressor of *FT* (Helliwell et al., 2006; Searle et al., 2006). Thus, LHP1 is involved in the repression of a floral repressor and a floral integrator gene at the same time and consequently both genes are up-regulated in *lhp1* mutant plants with a stronger effect on *FT* (Adrian et al., 2010; this study). *FT* up-regulation and the direct repressive effect of LHP1 on *AG*, *PISTILLATA* (*PI*), *APETALA 3* (*AP3*) and possibly other MADS domain transcription factors explain the early flowering phenotype and the aberrant leaf shape and floral organ structures of the *lhp1* mutant (Germann et al., 2006; Kotake et al., 2003).

1.3.3 Protein structure and function of LHP1

HP1, the homologue of LHP1, is involved in a diversity of cellular processes such as telomere organization, chromosome condensation and segregation, nuclear architecture and maintaining a silent or repressed chromatin structure (Li et al., 2002; Maison and Almouzni, 2004). Even though LHP1 was shown to be able to bind heterochromatic methylation marks, such as H3K9me2 *in vitro*, in Arabidopsis, it is associated with H3K27me3 to keep euchromatic genes in a repressed state (Exner et al., 2009; Turck et al., 2007; Zhang et al., 2007). The CD of LHP1 associates with chromatin whereas the CSD coordinates dimerization and binding to other proteins (Gaudin et al., 2001).

The protein contains five nuclear localization signals (NLS) of which at least three were found to be functional *in planta* (Libault et al., 2005). When truncated constructs of LHP1 either lacking the CD or the CSD were fused to GFP the nuclear localization pattern of the protein was altered. Both domains were necessary to achieve euchromatic localization of LHP1. Truncated constructs containing the hinge region of LHP1 localized to the nucleolus. This compartment was not targeted by the full LHP1 protein (Libault et al., 2005), but this does not hold true for LHP1 homologues of other plant species (Guan et al., 2011).

Genome wide profiling studies of H3K27me3 in different Arabidopsis tissues at different developmental stages are still scarce. First results from tissue specific Chromatin Immunoprecipitation (ChIP) studies followed by genome-wide tiling arrays (ChIP-Chip) confirmed that differential targets exist in Arabidopsis when comparing two tissues (Lafos et

al., 2011). It has been demonstrated that about 20 % of Arabidopsis genes are decorated with H3K27me3 and consequently to a large extend LHP1 targets (Lafos et al., 2011; Lu et al., 2011; Oh et al., 2008; Turck et al., 2007; Zhang et al., 2007). Among the genes directly targeted by LHP1 are almost 50 % of all micro RNAs (miRNAs) present in Arabidopsis and various transcription factors. Hence, de-regulation and mis-expression of genes in *lhp1* mutant plants can be a direct effect or a secondary effect mediated through primary LHP1 targets. Interestingly, Lafos et al. also found the entire Auxin pathway, meaning biosynthesis and transport pathways, to be covered by H3K27me3.

On the same line, a recent study has demonstrated the role of LHP1 in promotion of *YUCCA* (*YUC*) gene expression. *YUC* genes are involved in an early biosynthesis step of the auxin compound indole-3-acedic acid (IAA). Also some early auxin-induced genes, were down-regulated in *lhp1* and auxin levels found in *lhp1* plants were low (Rizzardi et al., 2011). Previously, a function of LHP1 in the positive regulation of genes had not been shown for plants. Positive gene regulation is a common feature which is also demonstrated for HP1γ and HP1a in Drosophila and humans (Piacentini et al., 2003; Vakoc et al., 2005). Thus, LHP1 retained some functionality of HP1.

Further, LHP1 is likely to be involved in RNA processing through the interaction with LHP1-Interacting Factor 2 (LIF2) (Latrasse et al., 2011), replication and DNA repair through interactions with INCURVATA 2 (ICU2) and EARLY IN SHORT DAYS 7 (ESD7) (Barrero et al., 2007; del Olmo et al., 2010), which are subunits of DNA polymerases α and ε respectively, in cell proliferation of roots through binding of SCARECROW (SCR) (Cui and Benfey, 2009) and in various chromatin remodeling processes by binding AtCYP71, which in turn binds to Chromatin Assembly Factor-1 (CAF-1) (Li and Luan, 2011).

1.4 Mapping-by-sequencing

Enhancer/suppressor screens have been successfully used to identify genes that play a role in chromatin-mediated gene repression and activation. For example, many components of the repressive PcG pathway were isolated as genetic enhancers or suppressors of homeotic mutations, whereas components of the Trithorax Group (TrxG) protein pathway were originally identified as suppressors of PcG related mutations (Alonso et al., 2007; Gildea et al., 2000; Landecker et al., 1994).

In Arabidopsis research, EMS mutagenesis is a powerful tool, which has been widely explored to uncover functionality of many genes in a broad spectrum of pathways (Page and

Grossniklaus, 2002). Recent advances in sequencing technology have greatly reduced the time required to pinpoint induced mutations. In a proof of principle experiment, mapping-by-sequencing (SHOREmapping) was first demonstrated on a mutant in the background of the Arabidopsis reference accession Col-0 crossed to the diverged accession Landsberg *erecta* (L*er*). A pool of DNA isolated from bulked segregants was sequenced and used for the simultaneous mapping and mutant identification (Schneeberger et al., 2009b). The simple application was followed by other studies successfully applying mapping-by-sequencing and different analysis pipelines (Austin et al., 2011; Cuperus et al., 2010; Schneeberger et al., 2009a)

Although all described approaches are straightforward and extremely fast, their application is hindered by the requirement for inter-accession crosses that impedes the success rate of screens based on quantitative traits, such as screens for genetic modifiers. The major obstacle is caused by the considerable phenotypic variation in F_2 populations from crosses between accessions impairing recognition of mutants with subtle phenotypic alterations. In addition, if genetic screens involve modifiers of a pre-existing mutant, such as *lhp1*, the mapping depends on the availability of the primary mutant in another suitable accession, the introgression of the mutation in such a background or the laborious additional genotyping for the presence of the first-site mutation.

Avoiding these disadvantages, Ashelford and colleagues have demonstrated that the isolation of a causative EMS-induced change is possible by direct re-sequencing of the complete mutant genome (Ashelford et al., 2011). However, their approach initially resulted in 103 putative causal SNPs that had the potential to change the amino acid sequences of 48 putative proteins. In addition, the SNPs were clustered in two separate regions of the genome, even though the mutant had been backcrossed four times to the parental line.

Uchida et al. argued that chromosomal regions near the causal mutation would not be expected to have crossovers during only a few backcrosses (BCs). Thus, isogenic crosses would not reduce EMS background mutations sufficiently that had been induced close to the causal mutation (Uchida et al., 2011).

First recent results in rice demonstrate how effectively a bulked segregant analysis can counter the obstacle of nearly linked background mutations. Abe et al. have used isogenic backcrossed rice cultivars that had been treated with EMS and used DNA pools of bulked individuals with the selected phenotype for re-sequencing. Assuming an unlinked mutation would appear with a concordance or probability parameter of 0,5 *e.g.* 50 %, Abe et al. showed that they could apply SNP distribution studies with a bulk of as little as 20 plants to find a

cluster of seven linked mutations with a probability parameter of 1 (Abe et al., 2012). It remains to be demonstrated how EMS load, bulk size, number of BCs and the size of the sequenced bulk affect the size of the target region and the number of potentially causal SNPs. It is however likely that with advances of sequencing technology many plant bulks, containing independent mutant alleles, can be sequenced at once to simultaneously find causal mutations of multiple phenotypes.

2 Aim of the Study

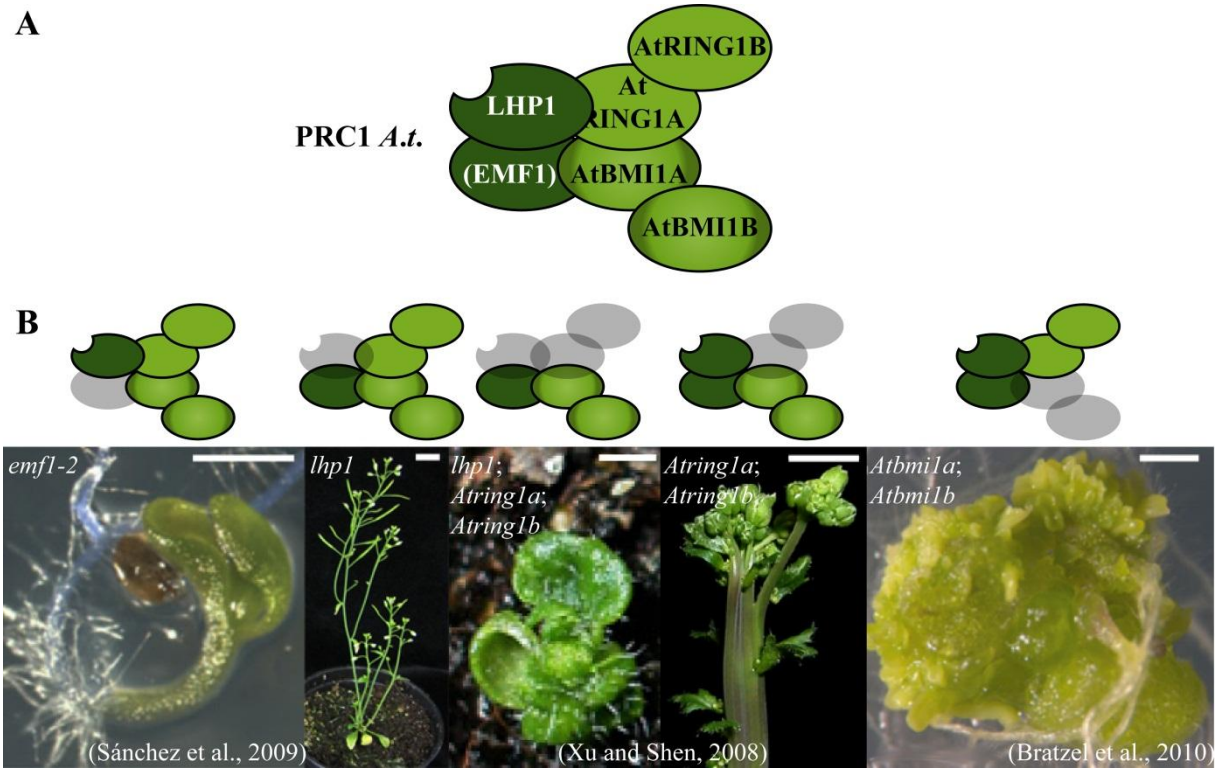


Figure 2: Phenotypes lacking PRC1 components
(**A**) Putative plant PRC1 complex. (**B**) Plants lacking PRC1 components in different combinations (in grey) and the corresponding phenotype below. Except for *lhp1* pictures were adapted from (Bratzel et al., 2010; Sanchez et al., 2009; Xu and Shen, 2008).

In this study, bulked segregant analysis was combined with isogenic back-crossing to facilitate mapping-by-sequencing of genetic modifiers. Mutants identified by this fast isogenic mapping approach were isolated as suppressors or enhancers of developmental aberrations caused by defects in Arabidopsis LHP1.

Several lines of evidence indicate the presence of unknown genetic modifiers participating in PRC1 function. First PRC1 was pulled down as a large complex, consisting of at least eleven components, in Drosophila. Second, multiple PRC1 complexes of changing composition seem to exist in mammals. Third, two of the four core component homologues, found in Arabidopsis, are present in multi-gene families with redundant function. Further, Arabidopsis plants lacking PRC1 core components display distinct phenotypes, such as cotyledons without petioles (*emf1-2*), fasciation of the stem (*Atring1a*;*Atring1b*) or callus formation (*Atbmi1a*;*Atbmi1b*) (Figure 2). Finally, PRC1 core components do not exclusively act on the same target genes in Drosophila (Gutierrez et al., 2012). Thus, it was crucial to search for additional genetic modifiers of PRC1 in order to rank the role of epigenetic repression in plants.

3 Material and Methods

3.1 Plant material

Mutant plants with a Col-0 background used in this study were either generated with EMS or ordered from the Nottingham Arabidopsis Stock Centre (NASC). Enhancer trap line ET1398 in L*er* background was also ordered from NASC (Rojas-Pierce and Springer, 2003). T-DNA insertion lines had been produced and selected by the Salk Institute Genomic Analysis Laboratory (SIGnAL) or the Gabi-Kat consortium (Rosso et al., 2003; Sessions et al., 2002). DGU13 is in Wassilewskija (Ws-2) background and was ordered from the Institute National de la Recherche Agronomique (INRA).
Transgenic plants were exclusively created via floral dip of mutants or WT plants in the Col-0 background. Tobacco plants for transient expression studies were grown for at least two weeks in a greenhouse at long day (LD) conditions.

3.1.1 Plant growth conditions

Plants were either grown on soil or on ½ strength Murashige & Skoog medium (GM) plates after seed sterilization. Short day (SD) conditions were 8 h of light and 16 h of darkness at 20 °C. LD conditions were 16 h of light and 8 h of darkness at 20 °C in the greenhouse and at 22 °C in a growth chamber.
For screening and growth of plants at lower temperature a Percival growth chamber was used. Conditions were set to 12 h of light and 12 h of darkness with 16 °C during the day and 14 °C during the night at 60 % humidity, which resulted in 95 % towards the end of the night. Day temperature was about 3 °C higher than the set 16 °C due to the radiation of heat from light bulbs and fluorescent tubes in the chamber. Plants were subjected to an average of 1.500 Lux of light during the day (measured with KIMO® data logger according to manufacturer's instructions). All seeds were stratified for at least two days at 4 °C prior sowing.

3.1.2 Flowering time measurements and phenotyping of plants

To measure flowering time, number of rosette and cauline leaves on the main shoot of at least nine individuals was counted. Experiments were conducted in LDs as well as in a climate chamber at 12 h of light, 16 °C during the day and 14 °C at night.

Size of rosette leaves, cauline leaves and siliques were measured using a ruler. Length and width were determined for leaves and length for siliques. Initial sorting of M_2 double mutant candidates into classes was done visually by comparing phenotypes to each other and to control plants.

3.2 EMS treatment of seeds

Germination rates of *lhp1* and WT seeds were scored on GM plates after ten days in LD conditions at 22 °C in a Percival plant growth chamber (CLF Plant Climatics). For EMS treatment, 200 mg of seeds were wrapped into miracloth and imbibed on a shaker at 4 °C in 0.1 % KCl solution for 14 h. Seeds were then washed with dH_2O and treated with 100 ml of 30 mM EMS diluted in dH_2O on a magnetic stirrer for 12 h.

Two washing steps with 100 ml of 100 mM sodium thiosulfate for 15 min and three washing steps with 500 ml dH_2O for 30 min followed. After washing, seeds were equally divided into 5 bottles containing 500 ml of 0.1 % Universal agarose (Bio-Budget Technologies GmbH). Seeds were sown in 7.5 ml aliquots onto 9 × 9 cm pots using plastic pipettes.

3.3 Selection of potential mutants (putants)

M_1 plants were bagged in small bulks averaging three plants. For each M_1 plant approximately ten M_2 seeds were sown onto 9 × 9 cm pots in two rows of five plants each. Potential mutants were primarily screened in SD conditions and grown together with *lhp1* (Col-0), *lhp1*;*clf* (Salk 006658 / Col-0), *lhp1* (Col-0);*emf2-10* (Ws-2, Chanvivattana et al., 2004) double mutants and Col-0 WT plants as controls.

Bulks with interesting putants were re-screened in a Percival chamber at 60 % humidity, 12 h light, 16 °C day and 14 °C night temperature. When a putant was confirmed in this secondary screen, seeds and leaf material were collected. At least ten plants of each putant were grown in the M_3 at the same conditions to confirm stability of the previously recorded phenotype.

Seed bulks containing putants that were screened as enhancers of the *lhp1* phenotype but did not survive until their reproductive phase, were sown on plates. After two days of stratification at 4 °C plants were grown in a Percival growth chamber in SD conditions (22 °C). Two weeks after germination surviving putants were transferred to soil and grown at 12 h light, 16 °C day and 14 °C night temperature.

3.3.1 Selection of enhancers and suppressors of lhp1

All selected putants were backcrossed reciprocally (as ♀ and ♂) to *lhp1* (Col-0) and Col-0 WT. Putants with a 3:1 segregation pattern when backcrossed to *lhp1* were kept as candidates. F_2 generations of putant crosses with Col-0 WT were used to screen for individual phenotypes of the candidates. Additionally, quantitative real-time polymerase chain reaction (qRT-PCR) of M_2 plants was carried out to determine mRNA levels of the LHP1 target genes *FT* and *FLC*.

When candidates displayed *FT* and *FLC* mRNA levels similar to *lhp1* (no more than threefold differences were tolerated), had a phenotype close to Col-0 WT for suppressors or *lhp1*;*clf* e.g. *lhp1*;*emf2-10* for enhancers, had no obvious individual phenotype after being backcrossed to Col-0 WT and segregated in a 3:1 ratio after backcross to *lhp1* (Col-0), they were selected for further characterization.

3.3.2 Sanger sequencing to select enhancer candidates

Genomic DNA of enhancer candidates was extracted with a Qiagen DNeasy® kit according to the manufacturer's instructions. A set of known and potential enhancers of *lhp1* was amplified in PCRs using Phusion® High-Fidelity DNA Polymerase (Finnzymes) for Sanger sequencing in selected enhancer putants (Table 2).

PCR products were controlled on 1 % agarose gels, purified with polyethylene glycol (PEG; PEG8000 diluted in 2,5 M NaCl) or cut from the gel and purified with Nucleospin® Extract II kit (Macherey & Nagel). Sequences were determined by the Max Planck Institute for Plant Breeding Research (MPIPZ) DNA core facility on Applied Biosystems (Weiterstadt, Germany) Abi Prism 377, 3100 and 3730 sequencers using BigDye-terminator v3.1 chemistry. Premixed reagents came from Applied Biosystems. Sequencing primers were purchased from Invitrogen and spaced at about 500 bp from each other (Table 2).

Candidates with single nucleotide polymorphisms (SNPs) responsible for non-synonymous amino acid changes in one of the tested PcGs were not analyzed further. The others were subjected to next-generation sequencing for SNP detection.

Table 2: Sequenced genes and sequencing primer

AGI	Gene Name	Primer Number	Sequence 5‘ to 3‘
At2g22540	SVP	139	CAAAATCAATCCCGTTCTCG
		140	GCCTCTTCCATAGGCAGAAA
		141	TTGTGGATTTGGTTTGCATC
		142	AAATCCATGCATGCACACTC

AGI	Gene Name	Primer Number	Sequence 5‘ to 3‘
		143	AAACATTAATCATGACCATAAATGACA
		144	TACCCGAGCCTAAGGGAAGT
At2g22540	SVP	169	ATCCCGTTCTCGAAAGATCC
		170	GGCTTCTTCTTCCTCCTCCT
		172	ATGGCGAGAGAAAAGATTCAGA
		195	TGACATTGACGGCTTAAACA
		196	TGTGGTCGTTTCGAGAATGT
At2g23380	CLF	47	CCACATGGGTTTTTCTGGAC
		48	CCCTTTGTTGTTGTTTCAGGA
		52	TGGCACCATTGAAGGTAAAA
		97	GTTTATAACGACCCGCCAAA
		98	CAATGTGTTCTTCCGTGTGG
		99	GGGATGCAAATCCATCCTAA
		100	AGCTCTGACGCAACCTCATT
		101	CCTTTCCCAATTGGCTTTTT
		118	ATGAGGTTGCGTCAGAGCTT
		128	TTACACGCTTCCCAATTCGT
		186	TCTCCTTCGACCCACTACAGA
		187	AAACAAAAATAGCAACCTTTATTGA
		188	TCATTCTTGCATTTTGGTCTCTT
At3g20740	FIE	122	GTGGTACAAGGGTGGACGTT
		123	CACCAGTTGTTGTTGCATGA
		124	GCATTGCAATCCTATGCTGA
		125	TGTGGGGAATTTTGATGGAT
		126	GAGGGCAACTTTTCAAATCC
		127	GGTTCCTTTTGAAACGGTCA
At4g29730	MSI5	133	TTTCCTCGCAAAAACTGACA
		134	GCCTTTAGAACAGCGTTTGC
		135	AGGCAAGAATCGTCACCAAC
		136	GGCACACTCTCCATTTGTGA
		137	GCTCAGACTTCTTGCCAACC
		138	TTTGGTCAGGCGATTGAGTT
		171	ACTGGTGGTCCTTCTGTTGG
		193	ACATCAGGACTGTCGGTGTG
		194	TGGTCAAGTGTTATCAACGACA
At5g11530	EMF1	102	TGTAGCTCTAGTTGCGGAGAGA
		103	TCATTGATTTCAGACGAACCA
		104	CAACCTCTGAAAATGCTTCCA
		105	TGTGATGGCTGTTTGTTGCT
		106	GCAGGTATGCATCTGCTGAG
		107	AATTTGGCCATCTTCCATGA
		108	TCCAAACTGGTGTGCTTCAG
		165	CAAATCCCTCCCACACACAT
		166	TCTGTTAATCCCTCTGCCTCA
		190	ATAGGTCGACCACGTGCTTT
		191	CTAACCCCAGCAGCATGTTT
		192	AAGAGTGCGCAGGAGTTGAT
At5g51230	EMF2	109	TGCACTGATTCTGGGTATGG
		110	AAAAGATGAGGACAAATTCATGC

AGI	Gene Name	Primer Number	Sequence 5‘ to 3‘
		111	GAAGTTGGGTCAACGGAAAA
		112	GGAGACATTTACCGCCTGAA
At5g51230	EMF2	113	CCCCTGTAGCCCTTGTTTCT
		114	TAGTCAAGCCCAAGCCAACT
		115	CGTAATGCTTTCCAGGTGGT
		116	CCAACTACTGATGCGGTGAA
		117	ACCTTTACATGCGCGACTCT
		167	TGCAGAGGTATTTTGCTCCA
		168	ACCACCTGGAAAGCATTACG
		180	CGACAGGCTTAGGAGAAACG
		181	CATCAACCCACGATCTCCTC
		189	TCAGATGCTCGATGATTTCG
At5g57380	VIN3	145	GTGTGTAATCTCCCGGCTCT
		146	TCATCCATCAGAGGGTTTCC
		147	TCACGGTCATCTTGTCACCT
		148	TCCACGCATTGTCTAACCAG
		149	TGCAGGGTGTTACAAGCAAG
		150	TTGAACCTTTTGTGTGAAATCG
		173	TTACCATCGAAACGCCAGAT
		197	TTCTTCCTCGAGTTTCTACGG
		198	GCCCTTTGACAAACTCAAGC
At5g58230	MSI1	129	ACGGGCTTTTCAAAAATGG
		130	CCATCCTCTCATGCCTCATT
		131	GTCTCGTTCTTTGGGTTCCA
		132	CAACTGGAAAGGTGCGATTT

3.4 Classical mapping of EMS induced SNPs

For classical mapping, candidates were crossed with *lhp1* mutants in the Ws-2 background (DGU13). Plants with a double mutant phenotype (*alp1*;*lhp1*) were selected in the F_2. The DNA of those F_2 individuals was then compared to Col-0 WT at 24 different positions with single-strand length polymorphism (SSLP) markers distributed evenly over the five Arabidopsis chromosomes (Table 3; by courtesy of Dr. Seth Davis, MPIPZ, Cologne).

Comparisons were conducted *via* PCR followed by high resolution melting (HRM) on the LightCycler® 480 II system (Roche) or real-time gel imaging with the QIAxcel system (Qiagen). Data was analyzed and visualized with graphical genotype (GGT; van Berloo, 2008).

Table 3: SSLP markers (m**arkers in grey were analyzed by HRM)**

Chromosome	Marker	Position [Mbp]	Forward primer$_1$ Reverse Primer$_2$ [5‘ to 3‘]
1	JV28/29mod	3,9	1: ACACTTGTTCTGAAACGAAATTGAG 2: CACTCCTTTGTGTGTTTTTTCTTACC
1	F20D23ext	5,8	1: TTATGCCAACTCATGTGGAAAGAAGAC 2: TGTCAAAGCGTCTGGTTCTGTTAAG

Chromosome	Marker	Position [Mbp]	Forward primer$_1$ Reverse Primer$_2$ [5‘ to 3‘]
1	AthSO392mod	10,8	1: GGAGTTAGACACGGATCTGTGAGA 2: TTTCTCAAATTGAAGAGATGATCTGG
1	ciw1	18,3	1: ACATTTTCTCAATCCTTACTC 2: GAGAGCTTCTTTATTTGTGAT
1	NF5I14a	24	1: GTTGAGTCTTGGCATCACAGTTC 2: CTGCCTGAAATTGTCGAAAC
2	msat2_5_mod	0,21	1: GCATCGATCTTCAATGCAGAAAAG 2: CAAAAGGGATACTGACAAAGCTAAA
2	msat2.41	11,09	1: GACTGTTTCATCGGATCCAT 2: ACAAACCATTGTTGGTCGTG
2	msat2.4 mod	13,8	1: GAGGATCACCTAACCAACTCATGGAC 2: TCTTTCCTGTAATCTGGGTTTTTGTG
2	LUGSSLP41	18,3	1: TGCATCAGTTTTGGTTGTGTGATCT 2: GCTGTATTTTCCATAGGGGGCA
3	nga172modB	0,78	1: AGCACATCAAGCTGCTTCCTTATAG 2: CCATTGTTCCTATCATTGTGTTGTC
3	ciw11a	9,8	1: GTTTTTTCTAATCCCCGAGTTGAG 2: GAAGAAATTCCTAAAGCATTC
3	msat3.10ext	17,2	1: CTCCATTGGGCAGAGAGAACTAAAAAAGC 2: TGGCATTGTCCCTATGGGATCCAC
3	F27K19	20,8	1: TGCTTTTGAAGAGATGGTTATTAGG 2: CCCCATTTCACTTATCATTGG
3	T20 O10	23,3	1: GTTGCACGATCATGCGTTTAC 2: CCCCCTTCATTTACGCTGTAG
4	sd4-13	0,36	1: AAAATGTGTGGTCAGGAATTAACAAATA 2: GAGGAGGAATCTGATACGTATTAAAAATG
4	msat4.17modb	7,1	1: CCAAGAACATCATCTTGAAGTAGAA 2: CATAAAACATAGAAGATATTAGGTCTGTTG
4	F26K10	13,9	1: AGAGAGCACGATGCCTGATAG 2: AATGCTTCAGCGATTGAGAAC
4	msat4.12modb	16	1: AGAGAGAGAGGGAGAGGTGCTTTTT 2: TTCTCTTCCTCCTGTATATCCTCCA
4	msat4.28modb	18,5	1: AGATGAAGGAGAAGCCATCACAG 2: ACGGCAGATTCAGAGAGAAAGTG
5	MOJB	2,2	1: GAAGATGAAAGATTTTAGGAGGAC 2: GTTTGTAGGAGAAGGGGACAAG
5	nga139modB	8,4	1: GTCTCTCTCTCTCGGGTCAAAATTAG 2: CTACCAGATCCGATGGTAAGATGAAG
5	SO191ext	15	1: CTCCACCAATCATGCAAATGTTTTTG 2: TGATGTTGATGGAGATGGTCAGAT
5	nga129mod	20,1	1: GTCGACTACAACACTGAAGATGGTCT 2: GTTCTTCAGGAGGAACTAAAGTGAGG
5	MNC17	23,9	1: GTACCGGATCTGTGTTGTGAAG 2: GTGCTCAAGGAAATGGGATAG

3.5 Next-generation mapping

Selected candidates for re-sequencing were backcrossed twice to *lhp1* (several crosses of a single plant) and selected again in the F_2 generation (BC_2F_2). Leaf samples of equal size of

selected plants in the BC_2F_2 were harvested. The leaves were bulked prior to DNA extraction with DNeasy® Plant Maxi Kit (Qiagen).

The DNA was eluted with 500 µl H_2O in four steps. DNA concentration and quality was determined with a nanodrop 1000 (Peqlab) and on a 1 % agarose gel. DNA samples were concentrated to more than 50 ng × μl^{-1} with a speed-vac when necessary.

Samples of more than 3 µg of total high-quality DNA extract (260/280 ratio > 1,8) were sent to the Cologne Center for Genomics (CCG). At the CCG a quality check of the samples was performed with a bioanalyzer (Agilent 2100). Libraries were created and their DNA concentration measured with the DNA 1000 kit as well as the DNA high sensitivity kit for diluted libraries (both Agilent). The samples were sequenced on a Solexa Genome Analyzer II (GAII) in a 96 bp paired end run. For bar-coded sequencing, libraries were prepared and sequenced on one lane of an Illumina Hi-Seq2000 flowcell at the MPIPZ.

Data was analyzed applying the SHORE pipeline (Ossowski et al., 2008) to identify the causal mutation. Single (*lhp1*) and double mutant reads were aligned to the Col-0 reference genome (TAIR 9). By using SHORE import, raw reads were trimmed or discarded based on quality values. GenomeMapper (Schneeberger et al., 2009a) was used for aligning filtered reads. Paired-end correction was done to improve the alignment quality and homozygous SNPs and INDELs were called. Homozygous SNPs found in the *lhp1* reference were used for background correction in the double mutant candidates. Canonical EMS changes with high quality (SHORE score > 24) were selected as the putative causal mutations and annotated using TAIR 10 gene annotation. The mapped regions were compared to intervals determined by classical mapping or used directly for confirmation of SNPs within promoters or exons of potential target genes.

3.6 Plasmid construction

Cloned DNA fragments were amplified using either the Expand High Fidelity PCR system (Roche) or the Phusion® High-Fidelity DNA Polymerase (Finnzymes) with Col-0 DNA or cDNA as templates. For cloning of all fragments forward primers contained Gateway™ (GW) extensions attB1 (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTA-3') and reverse primers had attB2 (5'- GGGGACCACTTTGTACAAGAAAGCTGGGTA-3') tails (Table 4).

All fragments were recombined with GW pDONR201 (Invitrogen™) plasmids in a BP reaction to generate entry clones. Plasmid DNA was recovered using a Nucleospin® Plasmid kit (Machery-Nagel) according to the manufacturer's instructions and sequenced to exclude

Table 4: Cloned DNA fragments amplified with GW primers

AGI	Primer Name / Number		Sequence 5‘ to 3‘ w/o GW extension
At3g42660	att1ENH8	292	ATGAAATCACGATCTCTAAAG
	att2ENH8stop	293	TTAAACAGTTGATTTGAG
	att2ENH8	294	AACAGTTGATTTGAGGAAC
	att1ENH8prom	295	AATCAAGGAACCTCCAAA
	att2ENH8prom	296	CGGAGATGTTGATTGTTTT
At3g63270	att1ALP1	248	ATGGCTCCGGTGAAGCAGAA
	att2ALP1stop	249	TCACCTAAGCAAATGCTCA
	att2ALP1	250	CCTAAGCAAATGCTCAGT
	att1ALP1prom_short	367	TTATTCTGTGCGTTGCGTCT
	att1ALP1prom	255	CTGACAAACACACTTGCTT
	att2ALP1prom	290	CGACGAGTACGACAAGGGTCAAG

PCR or cloning errors. Entry clones were recombined with a set of different destination vectors in LR reactions to create expression clones (Table 5).

Backbones for the 35S:GW::HA, 35S:GW::GFP and GW::RFP vectors were created in the MPIPZ in Cologne. The Cauliflower Mosaic Virus (CaMV) 35S promoter vectors are binary vectors with the GW fusion being under the control of a 2 × 35S promoter. In case of the 35S:GW::HA vector a spliced WRKY33 intron is inserted into the translational leader.

To create the binary destination vector GW::GUS a *rfA* vector conversion fragment followed by a nopaline synthase terminator (NOS) was cloned to the multiple cloning site (MCS) of pGREEN0229 giving rise to the vector GW-MCS-NOS-pGREEN (Corbesier et al., 2007).

Table 5: Vectors used for transformation

Destination Vector	Backbones	Application	Resistances
35S:GW::HA	pAM-Kan	Overexpression with HA-tag	Amp/Kan
35S:GW::GFP	pAM-Kan	Overexpression with GFP-tag	Amp/Kan
GW::GUS	pGREEN0229 (Corbesier et al., 2007), pBT10-GUS (Sprenger-Haussels & Weisshaar, 2000)	Expression pattern of target genes	Kan/BASTA
GW::RFP	pCZN654	Complementation and localization studies	Spec/BASTA
35S:GW::YFP	pAM-PAT	Localization studies	Amp
35S:GW::CFP	pAM-PAT	Localization studies	Amp
Split-Y-N-fusion	pCL112sYFP-N-RfA	Interaction of proteins	Spec/BASTA
Split-Y-C-fusion	pBat-TL-B-sYFP-N	Interaction of proteins	Spec/BASTA
Split-FP-N-fusion	sYFP-C-RfA-pCL113	Interaction of proteins	Spec/BASTA
Split-FP-C-fusion	pBat-TL-B-sYFP-C	Interaction of proteins	Spec/BASTA

The *GUS* coding sequence was amplified from pBT10-GUS and cloned into the multiple cloning site of GW-MCS-NOS-pGREEN (constructed by Dr. Sara Farrona; Adrian et al., 2010; Sprenger-Haussels and Weisshaar, 2000).
Split yellow fluorescent protein (YFP) backbones were kindly provided by Andrea Schrader at the University of Cologne. Transformations were carried out with competent DH5α *Escherichia coli* (*E. coli*) cells *via* heat shock.

3.6.1 Transgenic plants

Plasmids containing destination vectors with a pGREEN0229 backbone were transformed into *Agrobacterium tumefaciens* (Agrobacterium) strain GV3101 (pSOUP). All other plasmids were introduced into strain GV3101 (pRSK90) (based on Koncz and Schell, 1986). Arabidopsis Col-0 WT, *lhp1* single and double mutant plants were transformed by floral dip (Clough and Bent, 1998).
Agrobacteria were grown in 350 ml lysogeny broth (LB) medium until the solution had reached an optical density of at least 0,8 measured at a wavelength of 600 nm. They were centrifuged for 20 min at 5.000 rpm in a JLA 10.500 rotor (Beckman / centrifuge Avanti™ J-25) and taken up in 300 ml of a 5 % sucrose solution. Plant shoots were submerged in the bacterial solution for 1 min before they were covered with a transparent plastic bag. After one day corners of the bags were cut. Another day later they were removed and plants were grown at LD conditions in the greenhouse.
T_1 seeds transformed with a vector based on pAM-Kan (Table 5) were sterilized and screened on plates containing 50 µg × ml^{-1} Kanamycin. All other T_1 transformants were identified on soil on the basis of herbicide BASTA™ resistance. In the next generation T_2 lines were selected based on their segregation ratios under kanamycine or BASTA selection. Lines with a 2:1 to 4:1 segregation ratio for survival were considered to contain single locus insertions of the transgene. Homozygous lines were subsequently identified in the T_3 generation.

3.7 Genotyping of T-DNA insertion lines and EMS mutant lines

All T-DNA insertion mutants were genotyped by using a set of three primers. Two primers were flanking the site of T-DNA insertion and a third primer was on the inserted sequence, close to its left border (Table 6 primer LBb1.3). Primers were designed with a tool on the Salk institute's website (http://signal.salk.edu/tdnaprimers.2.html) or on the Primer3Plus website (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi).

For detection of the *lhp1-1* mutation two primers with a T→A mutation one base before the target SNP were generated using Primer3Plus. The last base of the two primers was either a C for the WT sequence or a T for the mutant sequence. A third primer was used to amplify short fragments of 442 bp with the mutant and the WT primer in two separate PCR reactions:

Initial Denaturation:	94 °C	30 s	1 ×
Denaturation:	94 °C	30 s	
Annealing:	56 °C	30 s	30 × for WT / 32 × for mutant primer
Extension:	72 °C	30 s	
Final Extension:	72 °C	5 min	1 ×
Hold:	12 °C		

When WT DNA was amplified the mutant primer had two mismatches at the 3'-end and *vice versa* resulting in reduced binding of the primer to the template and weaker bands on an agarose gel.

For the point mutation responsible for the *alp1* (At3g63270) phenotype and the SNP in At3g57940 (both found in *lhp1*;*alp1* plants) HRM primers were created with Primer3Plus. Primers amplified a short DNA fragment of 110 bp including the point mutation. Due to the difference in hydrogen bonds, WT DNA fragments had a different melting temperature than mutant fragments which was visualized using the standard two-step real-time program on the LightCycler® 480 II (Roche) with EvaGreen® (Biotium) in 20 µl reactions as described below. After the PCR, samples were heated from 60 °C to 95 °C with an increment of 0,02 °C between each fluorescence measurement. Data was analysed with the LightCycler® software package.

Table 6: Primers for genotyping of T-DNA and point mutant lines

Locus	Mutation	Primer	Forward primer$_1$ Reverse Primer$_2$ [5' to 3']
At5g17690	Point mutation	23	2: ATGAAATCACGATCTCTAAAG
LHP1		39	1: TTAAACAGTTGATTTGAG
		40	1: AACAGTTGATTTGAGGAAC
At3g42660	Point mutation	306	1: GATCTCTAAAGCTCCGGGAAG
ENH8		307	2: CGCGAATCGGAAGTGTAAAT
At3g63270	Point mutation	287	1: CTATCCCAAGGAGCCCAGAT
ALP1		288	2: ACCATGGAATCAGAGGGATG
At3g57940	Point mutation	387	1: CAGTTGGAGAGGGGACAAAT
		388	2: TCAAACGAGGGAGAGCTGAT
LBb1.3	T-DNA insertion Salk lines	Salk	1 AND 2: ATTTTGCCGATTTCGGAAC
At3g57940	T-DNA insertion	222	1: AAAGGCTTTGATGCACTTGAA
	Salk_067163	230	2: TTTGGGACTCATCAACAGGA

Locus	Mutation	Primer	Forward primer$_1$ Reverse Primer$_2$ [5‘ to 3‘]
At3g57940	T-DNA insertion	223	1: CCTTCTCAGCAGCCTCAGTC
	Salk_071904	229	2: GGCTCCTGCTACTGGTCTTG
At3g57940	T-DNA insertion	222	1: AAAGGCTTTGATGCACTTGAA
	Salk_011936	230	2: TTTGGGACTCATCAACAGGA
At3g57940	T-DNA insertion	222	1: AAAGGCTTTGATGCACTTGAA
	Salk_110812	233	2: CAACCTGTCCGAAGAAACTTG
At3g52490	T-DNA insertion	234	1: CCCTACACAGCTCTTCACGAG
	Salk_024706	235	2: TGCCTCTCTCACAAGAAAAGC
At3g60320	T-DNA insertion	319	1: ATTCTTGGCTTTGAAGCTGTG
	Salk_079096	318	2: TGAGAGATCCAATTTCATGCC
At3g55350	T-DNA insertion	320	1: CAACGGATCCAAAGACATTTG
ALP1hom.	Salk_122829	321	2: ACCGGGAGAGAGTAATCAAGC
At3g55350	T-DNA insertion	322	1: AGAGACCTTTGGGATGAACC
ALP1hom.	Salk_134187	323	2: GTTTCTCACGCAAACCAATTC
GABI T-DNA	T-DNA insertion Gabi-Kat lines	366	1 AND 2: ATAATAACGCTGCGGACATCTACATTTT
At3g42260	T-DNA insertion	300	1: GCGCGAGAGAAGTTTAGGTTT
ENH8	GK_382A06	371	2: GTGTGGCCCTTTCTTTCTGA
Ds5-4 T-DNA	T-DNA insertion Enhancer traps	236	1 AND 2: TACGATAACGGTCGGTACGG
At3g63270	T-DNA insertion	237	1: CCCTACACAGCTCTTCACGAG
ALP1	ET_1398	238	2: CAGTCCCCTCAGACGAAGAC

The point mutation causal for the *lhp1*;*elp1* double mutant phenotype was suitable for detection with cleaved amplified polymorphic sequence (CAPS) markers. When the restriction enzyme AvaII was used, only the WT band could be cut, since the recognition sequence of the enzyme was mutated in *lhp1*;*elp1*. DNA fragments of 398 bp including the point mutation site were amplified via PCR using specific primers (Table 6). PCR samples were digested with 25 units of AvaII (New England Biolabs® Inc.) at 37 °C for 1 to 2 h. Reactions were loaded and visualized on a 1 % agarose gel.

3.8 DNA extraction

All plant DNA samples were extracted from young leaf tissue or whole seedlings. Low amounts of DNA were extracted with a BioSprint 96 robotic work station (Qiagen) with the DNA-plant-100 program. Higher amounts of DNA were extracted with a Qiagen DNeasy® kit following the manufacturer’s instructions.

Bacterial plasmids were purified with Nucleospin® Plasmid kits (Macherey & Nagel). DNA fragments for cloning were either precipitated with the same volume of PEG8000 (in 2,5 M NaCl) via centrifugation (1 h at 14.000 rpm / FA 45-30-11 rotor Eppendorf / centrifuge 5417R), washed with 70 % Ethanol and eluted in water or cut from a 1 % agarose gel and purified using a Nucleospin® Extract II kit (Macherey & Nagel).

3.9 Quantification of transcription

Roots of ten day old seedlings grown on soil or on plates were cut and total RNA was extracted from the remaining plant tissue with an RNeasy® Plant Mini kit (Qiagen) according to the manufacturer's instructions. 1 µg RNA of each sample was loaded onto a 1 % agarose gel to control for RNA degradation by visualizing the two distinct rRNA bands.

When quality control was passed, 5 µg of RNA was treated with DNase I using the DNA-free™ kit (Ambion). After DNase treatment cDNA was synthesized with a dT18 primer and a Superscript II reverse transcriptase kit (Invitrogen) according to the manufacturer's instructions. Samples were diluted to 150 µl with dH_2O and 1-3 µl of cDNA was used for qRT-PCR. All qRT-PCRs were either performed in a BioRAD iCycler iQ5™ or a LightCycler® 480 II (Roche) platform with EvaGreen® (Biotium) as a chelating fluorescent dye to quantify the real-time signal. Total reaction volume was either 20 µl (10 µl 2 × EvaGreen® buffer, 0,8 µl Primer mix [10 mM each], 2 µl cDNA and 7,6 µl H_2O) for 96 well plates or 10 µl for 384 well plates. The following PCR programs were used for real-time reactions on the iQ5™ and the LightCycler® respectively:

iQ5™:

Initial Denaturation:	95 °C	3 min	1 ×
Denaturation:	95 °C	10 s	
Annealing:	54-60 °C	20 s	50 ×
Extension:	72 °C	20 s	
Denaturation:	94 °C	3 min	1 ×
Annealing:	50 °C	1 min	1 ×
Melting curve:	50 °C – 95 °C		steps of 0,5 °C

LightCycler®:

Initial Denaturation:	95 °C	4 min	1 ×
Denaturation:	95 °C	10 s	
Annealing:	54 - 60 °C	15 s	45 – 50 ×
Extension:	60 or 72 °C	10-40 s	
Denaturation:	95 °C	1 s	1 ×
Annealing:	40 °C	1 min	1 ×
Melting curve:	60 °C – 95 °C		steps of 0,02 °C

When available a dilution series of a plasmid containing the measured gene was run as a standard for each primer pair. *Actin 2* (*ACT2*) and *Protein Phosphatase 2* (*PP2A*) were used as controls to account for variation between samples respectively. In table 7 primers are listed that were used to quantify gene expression levels.

Table 7: Primers used to quantify the level of transcipt

AGI	Gene Name	Forward primer$_1$ Reverse Primer$_2$ [5‘ to 3‘]
At1g29660	GDSL-like lipase (*F15D2_21*)	1: CAGGGACCAGCTTAATGCAC 2: GCAACACGCAGTGTTCGTAT
At2g26580	*YABBY 5* (*YAB5*)	1: AGCTCTGCTACATCCCTTGC 2: GGAAAGTGTGCCCAATTCTT
At2g46830	*CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*)	1: ATCCTGAGATGGCCAATGAA 2: TAGCACAGGGATATGCATCG
At5g36910	*THIONIN 2.2* (*THI2.2*)	1: TGCTGTCCTACCAAGGATGA 2: ACTTGTGTCAAAGTTTTGGAGAG
At2g47520	*ETHYLENE RESPONSE FACTOR 71* (*ERF71*)	1: GGACTCGGTTCAAAAACGAA 2: TTCACTCACTGAGACGGGTTT
At4g31940	*CYTOCHROME P450, F. 82, S.F. C, P.P. 4* (*CYP82C4*)	1: CTACCTGCCTGGCACTGATT 2: AGGGCCTAAGAGAGGACCAG
At5g52830	*WRKY DNA-BINDING PROTEIN 27* (*WRKY27*)	1: TCTAGAAAAAGAAAGAATCAGCAAA 2: ACGGGCTGAGATTTGTTACG
At5g56080	*NICOTIANAMINE SYNTHASE 2* (*NAS2*)	1: CGCTTCAAACCTCGTTTCTC 2: TAGCATCACCACAGCTCCAG
At3g42660	*ENHANCER OF LHP1 (ELP1)*	1: TGTTGTAGTAGTGATAGTTTTGCAAGG 2: CTTCTCCGTTTGGATTCAGC
At3g42660	*ELP1_2ndprimer*	1: GAACATGAGGGGAACTCACG 2: CAAATCTCATTGTCCACTCACTAT
At3g42670	*CHROMATIN REMODELING 38* (*CHR38*)	1: TTGAGTGAGGAGGAGGAGGA 2: CGCAAAAGGTGCAGAAACAT
At5g10140	*FLOWERING LOCUS C* (*FLC*)	1: ACAAAAGTAGCCGACAAGTCACCT 2: GGAAGATTGTCGGAGATTTGTCCA
At1g13320	*PROTEIN PHOSPHATASE 2A* (*PP2A*)	1: CAGCAACGAATTGTGTTTGG 2: AAATACGCCCAACGAACAAA
At3g63270	*ANTAGONIST OF LHP1* (*ALP1*)	1: GCCTCGATGAATCTCCATGT 2: TCCCTCAGTCCCCTCAGAC
At1g65480	*FLOWERING LOCUS T* (*FT*)	1: CGAGTAACGAACGGTGATGA 2: CGCATCACACACTATATAAGTAAAACA
At3g18780	*ACTIN 2* (*ACT2*)	1: ACTGAGCACAATGTTAC 2: GGTGATGGTGTGTCT
At3g54340	*APETALA 3* (*AP3*)	1: GAGCTCACGGTTTTGTGTGA 2: TCCCAAGAGATTTGAACTTGC
At1g24260	*SEPALLATA 3* (*SEP3*)	1: TATGACGCCTTACAGAGAACC 2: ATACCCATCAGCTAACCTTAGTC
At4g18960	*AGAMOUS* (*AG*)	1: TCCGAGTATAAGTCTAATGCC 2: GCCTATATTACACTAACTGGAGAG

3.10 mRNA-Seq

Total RNA was extracted from ten day old seedlings with a QIAGEN RNeasy® Plant Mini kit according to the manufacturer's instructions. Four biological replicates for each of the genotypes were used. RNA concentration was measured with a nanodrop 1000 (Peqlab) and quality was controlled on a 1 % agarose gel.

Three RNA samples of each genotype were selected for mRNA enrichment with oligo dT beads using the manual of Invitrogen's Dynabeads® mRNA Purification kit. Pure mRNA was used for reverse transcription, adapter ligation and library creation with the New England Biolabs (NEB) Next™ mRNA Sample Prep Master Mix Set 1 following the manufacturer's instructions.

Table 8: Bar-coded primers used for genome wide transcript profiles

Name	Primer [5' to 3']/bar-code in bold/complementary region underlined
PE1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
PE2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCG ATCT
Adapter1a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AATCAC***T
Adapter1b	p**GTGATT**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter2a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ACGCGG***T
Adapter2b	p**CCGCGT**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter3a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AGCAAT***T
Adapter3b	p**ATTGCT**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter4a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**CATTCA***T
Adapter4b	p**TGAATG**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter5a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**CGAGGT***T
Adapter5b	p**ACCTCG**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter6a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**CTGTTA***T
Adapter6b	p**TAACAG**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter7a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GCCAGA***T
Adapter7b	p**TCTGGC**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter8a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GGTAAC***T
Adapter8b	p**GTTACC**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter9a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GTAGCT***T
Adapter9b	p**AGCTAC**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter10a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**TACGTG***T
Adapter10b	p**CACGTA**AGATCGGAAGAGCG GTTCAGCAGGAATGCCGAG
Adapter11a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**TCACCG***T
Adapter11b	p**CGGTGA**AGATCGGAAGAGCG GTTCAGCAGGAATGCCGAG
Adapter12a	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**TTGTTC***T
Adapter12b	p**GAACAA**AGATCGGAAGAGCG GTTCAGCAGGAATGCCGAG

For purification of cDNA, the MinElute® kit from QIAGEN was used. Forward adapters with a binding region for sequencing primers, a complementary region (underlined) to create double stranded adapters, a 6 bp barcode (in bold and also complementary) and a T-overhang with a phosphorothioate bond to protect it from cleavage, were created. Reverse adapters had a phosphorylated base at the 5'-end and no T-overhang (Table 8).

Adapter pairs were diluted to 15 µM in polynucleotid kinase (PNK) buffer and annealed by heating the solution to 95 °C for 5 min and letting it cool down at room temperature. For the enrichment of the cDNA libraries we used the following PCR program:

Initial Denaturation:	98 °C	10 s	1 ×
Denaturation:	98 °C	10 s	
Annealing:	65 °C	30 s	17 ×
Extension:	72 °C	30 s	
Final Extension:	72 °C	5 min	1 ×
Hold:	12 °C		

Enriched libraries were eluted, quality-controlled with a 2100 bioanalyzer (Agilent), loaded on one lane of a GAII flow cells and sequenced in a 36 bp paired end run with primers PE 1 and PE 2 which were able to bind to all adapters (Table 8).

Filtered and trimmed mRNA reads were mapped to the Arabidopsis Col-0 cDNA sequence (TAIR 10) using GenomeMapper (Schneeberger et al., 2009a). For mapping, default values were used except that a maximum of 10 % mismatches and 7 % gaps to the length of trimmed reads were allowed. Raw count reads were scored for each data set (library) and further analyzed with the Bioconductor package edgeR.

For edgeR the default workflow was used, which was based on calculating normalization factors (Robinson and Oshlack, 2010), estimating the dispersion of reads via quantile-adjusted conditional maximum likelihood and doing exact tests on each pair of compared data sets. The p-Values were adjusted by default in edgeR (Benjamini and Hochberg, 1995).

3.11 Western Blotting

Plant material was harvested and frozen in liquid nitrogen. For protein extraction, 100 mg of frozen tissue was ground and 200 µl of 5 × Laemmli buffer heated to 95 °C was added to the samples. Samples were incubated at 95 °C for 10 min, spun down at 10.000 rpm in a FA 45-30-11 rotor (Eppendorf / centrifuge 5417R) for 10 min at room temperature and the supernatant was transferred to a new tube. Centrifugation and transfer of the supernatant were

repeated once. Samples were quantified with amido black and an equal amount of protein was loaded on 16 % polyacrylamide gels.

Gels were run in TGS buffer (supplied by Bio-Rad) at 80 V for 2 h and blotted on PVDF membranes (Millipore) in transfer buffer at 30 V overnight. After blotting, membranes were blocked in TBS-T buffer containing 5 % milk for 1 h. The membranes were incubated with primary (H3K4me3: Cat. No. CS-003-100 / H3K27me3: Cat. No. CS-069-100 both from Diagenode) and secondary antibodies followed by three washing steps, for 20 min each, with TBS-T for the first two and TBS for the last washing step.

Blots were developed with a mixture of SuperSignal® West Femto (350 µl) and SuperSignal® West Dura (150 µl) solutions (Thermo Scientific®). Chemiluminescence pictures were taken with a LAS 4000 imager (Fujifilm) and the blots were quantified with the ImageJ software.

3.12 Co-infiltration of tobacco leaves for split-YFP detection

Entry vectors containing the DNA sequence of interest were created with GW pDONR201 and cloned in DH5α bacteria as previously described. BP cloning was followed by LR reactions with destination vectors carrying either the N- or C-terminal part of YFP, resulting in split-YFP expression vectors (Table 5).

Expression vectors were cloned into GV3103 (pRSK90) Agrobacteria which were grown in 10 ml of LB medium including antibiotics at 28 °C for one day. Bacteria were spun down and pellets were taken up in infiltration buffer, adjusted to an OD_{600} of 3. For each ml of the solution, 1 µl of acetosyringone (150 µM in DMSO) was added. Solutions were kept in darkness at room temperature for a least 3 h prior to infiltration of *Nicotiana tabacum* leaves. For infiltration, equal amounts of two Agrobacterium solutions with YFP constructs and one solution of pK19 Agrobacteria (Voinnet et al., 2003) were mixed. Of each three to six week old tobacco plant the 3rd and 4th leaves were infiltrated at two to four spots. Plants were kept at SD conditions (22 °C) for two days. YFP signals were detected with a confocal laser scanning spectral microscope (Leica TCS SP2 AOBS).

3.13 Acceptor Photo Bleaching (APB)

Tobacco leaves were co-infiltrated with YFP and cyan fluorescent protein (CFP) constructs as described above for split-YFP constructs. Nuclei that gave of YFP and CFP signal were magnified with a 40 × oil objective (HCX PL APO). CFP emission was measured at 450 –

505 nm and YFP emission was measured at 518 - 590 nm with a laser scanning spectral microscope (Leica TCS SP2 AOBS).

Regions of interest were defined manually around nuclei and YFP was bleached to 40 % with a laser at 514 nm. Fluorescence Resonance Energy Transfer (FRET) was determined by comparing fluorescence intensities (D) of CFP before and after bleaching according to the following formula: $FRET_{Eff} = (D_{post} - D_{pre}) / D_{post}$ for all $D_{post} > D_{pre}$

3.14 Neighbor Joining analysis

The evolutionary history was inferred using the Neighbor Joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 10.000 replicates represented the evolutionary history of the proteins analyzed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the number of differences method (Nei and Kumar, 2000) and are in the units of the number of base differences per sequence. The analysis involved 105 amino acid sequences. There were a total of 3082 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011).

3.15 Buffers, Solutions and Medium

½ strength Murashige & Skoog medium

4,4 g × l^{-1}	Murashige & Skoog basal salt mixture
1 ml × l^{-1}	Murashige & Skoog vitamines
3 ml × l^{-1}	2-(*N*-morpholino)ethanesulfonic acid (MES) buffer
10 g × l^{-1}	sucrose
pH 5,7 with KOH	
0,9 %	Agar for solid media

Lysogeny Broth (LB) medium

10 g	tryptone
5 g	yeast extract
5 g	NaCl
ad. to 1 l with dH_2O	
pH 7,2 with NaOH	
8 g	Agar for solid media

MES buffer

16,6 g 2-(*N*-morpholino)ethanesulfonic acid
add H_20 to 100 ml
pH 5,7 with KOH

10 × PNK buffer

500 mM Tris-HCl, pH 8,2
100 mM $MgCl_2$
50 mM DTT
1 mM EDTA
1 mM spermidine

5 × KCM buffer

500 mM KCl
150 mM $CaCl_2$
250 mM $MgCl_2$

5 × Laemmli buffer

Dissolve 2g SDS in 6,25 ml Tris-HCl (1 M; pH 6,8)
9 ml glycerol
12 mg bromphenol blue
5 ml β-mercaptoethanol
100 µl protease inhibitor cocktail

1 × Transfer buffer

3,03 g Tris
14,4 g glycine
200 ml methanol
fill up to 1 l with dH_2O

10 × TBS buffer

500 mM Tris
1,5 M NaCl
fill up to 1 l with dH_2O
pH 7,4 with HCl
for TBS-T add 0,1 % Tween 20

Infiltration buffer

1 mM $MgCl_2$
1 mM MES salt

2 × EvaGreen® buffer (for 100 reactions)

100 µl	20× Reaction buffer (0,8 M KCl, 50 mM $MgCl_2$, 0,2 M Tris-HCI, pH 8,8)
100 µl	20× EvaGreen® dye
40 µl	dNTPs (10mM each)
20 µl	Taq DNA Polymerase (5U/µl)
740 µl	dH_2O

4 Results

4.1 Classification of EMS mutants

The *lhp1* mutant displays a pleiotropic, early flowering phenotype. While searching for new PcG genes that would enhance or suppress the *lhp1* phenotype, three important parameters were identified. Firstly, the expected great abundance of EMS-induced mutants needed to be classified to focus on the mutants that most likely influence LHP1. Secondly, enhancers and suppressors of *lhp1* needed to be defined and selected. This was not a trivial task, since the aim of the study was a search for novelty. Measures had to be taken to exclude known PcG genes and previously described genes in the flowering time pathway, even though they were also expected to enhance or suppress *lhp1*. Thirdly, a pipeline for the fast and reliable recognition of potentially novel PcG genes had to be established. Classical mapping of mutations in a pleiotropic mutant background can lead to false positive or false negative phenotype recognition and be labor intensive as well as time consuming.

Prior to EMS treatment the germination rate of lhp1 seeds was examined on plates in LD conditions. On average 82% *lhp1* seeds had germinated after 10 days at 22 °C. Approximately 10.000 *lhp1* seeds (200 mg) were treated with EMS and about 4.500 seeds germinated in the M_1 generation. In the M_2 generation 10 seeds for each M_1 plant were sown and 21.290 M_2 plants were visually screened in SD conditions.

For a fraction of 555 potential M_2 mutants (not represented in Table 9), hereafter called putants, a set of nine parameters, namely flowering, leaf size, leaf curling, leaf lobing, leaf serration, leaf number, rosette diameter, shoot branching and plant height, were measured or scored visually after growing for 60 days on soil in SD conditions. At the time of scoring 142 of the 555 putants had no rosette leaves left, 42 were just beginning to bolt and 18 had died. Clustering of all gathered data was discarded as a method for classification, since an aberrant phenotype could not be linked to EMS alone, but might have been enhanced by insect feeding or application of insecticide. Alternatively, the M_2 data of the remaining, unharmed putants was used to establish a scoring system for M_3 putants based on three classes (Table 9).

Class III putants either had an unstable phenotype in the M_3, or were no longer distinguishable from *lhp1* in that generation. Class II putants displayed phenotypic redundancy with putants from the same M_1 bulk, which might have the same causal mutation underlying. Other Class II putants showed only subtle changes compared to *lhp1* (scored

visually). Finally, Class I putants displayed strong changes compared to *lhp1* in at least two of the scoring parameters mentioned above. Those changes had to be stable over two generations.

4.1.1 Altered conditions improved screening for aberrant phenotypes

Some enhancers of *lhp1* did not survive or set seeds in SD conditions (Supplemental figure S1). M_1 bulks containing interesting candidates were therefore re-screened in a climate chamber at lower temperature, which was known to alter the *lhp1* phenotype (Personal communication with Dr. Sara Farrona, Uni Düsseldorf).

The longer vegetative growth phase in the chamber increased phenotypic differences between putants and *lhp1*. Some Class I and Class II putants became accessible by using alternative growth conditions and for others it was confirmed that they had been classified correctly using SD conditions (Supplemental Figure 1). A total of 144 M_1 bulks from all classes were re-screened in these conditions (Table 9).

Table 9: Classification of potential EMS mutants

Total Number of Screened M_1 Putants (SDs / Chamber / Plates)
21.290 / 1.444 / 730

Number of M_2 Putants
716

Class Ia	Class Ib	Class II	Class III	Lethal or No Germination
46	37	69	356	208

Subclasses	Name	Number of Candidates
SC 1	**Enhancers of *lhp1***	**7 (20)**
SC 2	**Enhancers of *lhp1* on plates**	**2**
SC 3	**Suppressors of *lhp1***	**11**
SC 4	Later flowering and leaf serration	5
SC 5	Yellowing of leaves	5
SC 6	Twirled leaves	3
SC 7	Smaller and compact architecture	4
SC 8	Slower development	9

4.1.2 Division of Class I into Class Ia and Class Ib

Through the application of three different screening conditions a total of 716 candidates had been found and was divided into classes (Table 9). Class II, Class III and the lethal or not germinated putants were excluded from further experiments, since putants of those classes were considered to be less relevant or accessible for the aim of the study (Table 9 numbers in grey). The focus was laid on the remaining 83 Class I putants. First, plants were tested for stable enhancement or suppression of *lhp1* in two different screening conditions over at least two generations. Second, all Class I putants were confirmed to harbour the homozygous *lhp1* mutation via PCR (data not shown). Third, they were reciprocally backcrossed with *lhp1* and confirmed to have a 3:1 segregation pattern in the BC_1F_2 (second generation of the first backcross), to ensure that the putant phenotype was caused by a single, recessive mutation. Fourth, known PcG genes and described genes of the flowering time pathway were ruled out by direct sequencing of PcG genes and qRT-PCR in the search for altered *FT* and *FLC* levels (Supplemental figure S2). Fifth, Class I putants were backcrossed to Col-0 to search for individual mutant phenotypes in the F_2. Last, Class I was divided into Class Ia and Class Ib based on the following criteria:

Class Ib putants showed high *FLC* (> 5 × Col-0) and / or low *FT* (< ¼ × *lhp1*) levels, while late flowering compared to *lhp1* (Supplemental Figure S2). Alternatively, they had a mutation in previously described, Sanger-sequenced PcG genes, or had an individual mutant phenotype after the backcross with Col-0. Finally, putants with multiple mutations underlying the phenotype were also assigned to Class Ib. Putants in Class Ib are still potentially interesting for the study of PcG-mediated gene repression, but they were discarded at this point to increase the possibility of finding novel enhancers and suppressors of *lhp1*.

In real-time PCR experiments the threshold for exclusion of late flowering putants was set to ¼ times *lhp1* levels for *FT* and to 5 times Col-0 for *FLC*. The *FT* threshold was set slightly above the wild-type (WT) *FT* level and the *FLC* threshold was set lower than up-regulation previously described for autonomous pathway mutants, but higher than WT to allow for variation and possible indirect enhancement (Pazhouhandeh et al., 2011).

4.1.3 Sorting putants into subclasses

The remaining 46 Class Ia putants were sorted into eight subclasses (SC) based on their appearance, flowering time and development in different screening conditions (Table 9). The first three SCs were of greatest interest and will be described in further detail below. SC 4 to

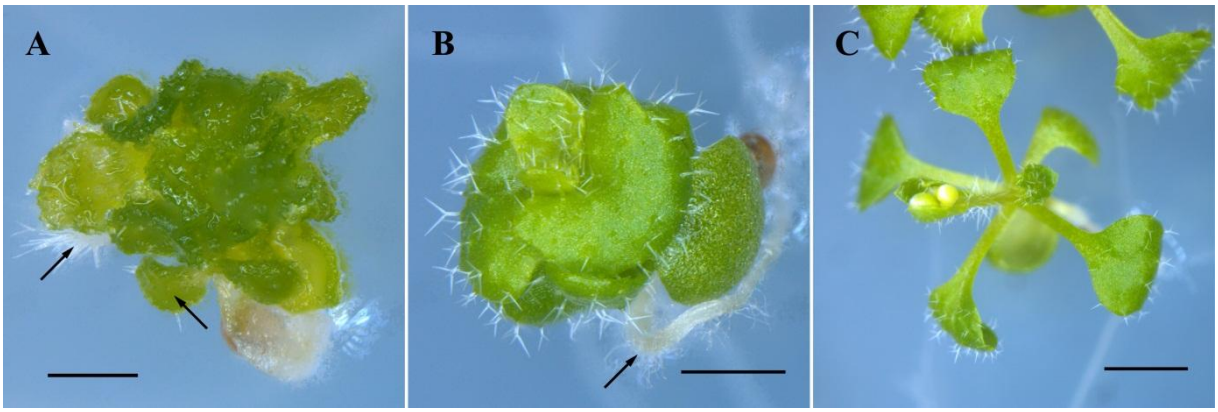


Figure 3: Severe enhancers of *lhp1*
(**A** - **C**) Plants were grown on plates for 14 days in LD conditions. (**A**) *enhs1* displayed a callus like structure with mostly undifferentiated cells. Some parts of the callus still showed organ like patterns such as roots or leaves (arrows). (**B**) *enhs2* was not yet flowering and showed a dense root with root hair clusters (arrow). (**C**) Bolting of the single *lhp1* mutant. (**A**, **B**) Bars = 0,1 cm (**C**) bar = 0,2 cm.

SC 8 were not terminally defined. All putants in those classes are in the process of being tested for their segregation patterns in the F_2 and have already been backcrossed reciprocally to *lhp1* and to Col-0 as the maternal parent.

Especially SC 7 and SC 8, but also SC 1 contained putants that had to be crossed to *lhp1* more than four times until one of those crosses was successful. Some putants could only be used as the pollen donor. Their floral organs were extremely small, fused or producing non fertile pollen. In rare cases, crosses to a different accession, such as Ws-2 *lhp1*, have so far failed to produce seeds. This prohibits mapping of a locus by conventional approaches.

More than 70 M_1 bulks containing putants that did not germinate or survive in both screening conditions were grown on plates (Table 9). Two of the selected double mutants did not produce seeds, when they were transferred to soil. Those enhancers of *lhp1* were named enhancer severe 1 (*enhs1*) and enhancer severe 2 (*enhs2*). After germination *enhs1* plants did start to diverge and evolve into a callus like structure. Leaf-like and root-like organs were still produced (Figure 3 arrows).

The other severe enhancer showed a milder phenotype than *enhs1*. Plants were small, compact and of a dark green colour. They did produce a flower at about the same time as *lhp1* mutants on plates but never grew enough to set seeds. The floral stem did not extend and the overall plant size remained smaller than *lhp1* throughout development (Figure 3B). *FT* and *FLC* expression levels could not be measured for *enhs1* and *enhs2*, because plant material for RNA extraction could not be sampled in sufficient amounts due to the small size of the putants.

4.1.3.1 Enhancers of lhp1

Originally 20 enhancers of *lhp1* (*enh1* - *enh20*) had been identified (Table 9 number in bracets). Most of those were excluded from Class Ia by analyzing the F_2 generations of the backcrosses and by Sanger sequencing of previously described or likely enhancers of *lhp1*, namely *SVP*, *CLF*, *FIE*, *MSI1*, *MSI5*, *EMF1*, *EMF2* and *VIN3*.

Enh1 to *enh3*, *enh5* to *enh7*, *enh9* and *enh18* had one or more individual phenotype/s in the BC_1F_2 generation when crossed with Col-0 as the maternal parent and were therefore assigned to Class Ib (Table 10). In *enh18*, a $G_{359}A$ (guanine at position 359 after ATG start codon mutated to adenine) mutation in *EMF1* was identified that led to the non-synonymous amino acid change $D_{83}N$. However, when *EMF1* was sequenced in 13 F_2 plants with the individual early flowering phenotype of *enh18*, resulting from a backcross with Col-0, the mutation was only present in 11 plants, therefore segregating, and not causal (data not shown).

Phenotypic differences were found twice when *lhp1* was used as a crossing partner. In case of *enh8*, multiple phenotypes were observed in the F_2 of an *enh8* × *lhp1* cross. Due to its phenotypic instability this enhancer was assigned to Class III. All F_1 plants, resulting from a cross between *enh17* as the maternal partner and *lhp1* as the pollen donor, displayed an inter-

Table 10: Classification of enhancers of *lhp1*

Name	Status	Classification
enh1	Individual phenotype after BC with Col-0	Class Ib
enh2	Individual phenotype after BC with Col-0	Class Ib
enh3	Multiple phenotypes after BC with Col-0	Class Ib
enh4	Not tested	Class Ia
enh5	Multiple phenotypes after BC with Col-0	Class Ib
enh6	Individual phenotype after BC with Col-0	Class Ib
enh7	Individual phenotype after BC with Col-0	Class Ib
enh8	Multiple phenotypes after BC with *lhp1*	Class III
enh9	**No individual phenotype after BC**	**Class Ia**
enh10	Multiple phenotypes after BC with Col-0	Class Ib
enh11	**No individual phenotype after BC**	**Class Ia**
enh12	Mutation in PcG	Class Ib
enh13	**No individual phenotype after BC**	**Class Ia**
enh14	Not tested	Class Ia
enh15	**No individual phenotype after BC**	**Class Ia**
enh16	Not tested	Class Ia
enh17	Individual phenotype Col-0 BC / No 3:1 segregation	Class Ib
enh18	Individual phenotype after BC with Col-0	Class Ib
enh19	Mutations in PcG	Class III
enh20	Mutation in TERMINAL FLOWER 1	Class Ib

mediate phenotype between the single and the double mutant. Therefore, *enh17* was sorted into Class Ib.

For *enh12* and *enh19*, mutations in the PRC2 component *CLF* were found by Sanger sequencing. Surprisingly, ten base pair changes were found in the *CLF* gene of *enh19*. Most changes, apart from $G_{3834}A$ and $A_{4164}G$, could be assigned to the *CLF* allele of the Arabidopsis accession Ws-2. Since all other SNPs matched Ws-2 *CLF* it was concluded that *enh19* was either a contaminant of the original mutagenized seed pool (Ws-2 *lhp1*) or resulted from an accidental cross with an *lhp1* mutant in the Ws-2 background. The *lhp1-1* (Ws-2) allele had been used as a control cross to assess natural variation of the *lhp1-3* (Col-0) allele (Supplemental Figure S3). Due to those crosses, a contamination of the screen with *lhp1-1* (Ws-2) might have taken place. *Enh19* was assigned to Class III (Table 10).

In the *CLF* gene of *enh12* the possibly causal mutation for its phenotype was a $G_{4024}A$ change which resulted in an amino acid change of glutamate to lysine at position 790 of the amino acid sequence. The *FT* and *FLC* expression of *enh12* could not be analyzed. Plants were too small to generate enough material for RNA extraction.

The last Class I putant that was not analyzed for its *FT* and *FLC* expression apart from *enhs1*, *enhs2* and *enh12*, was *enh20*. This putant displayed a stunted growth, early flowering and apical dominant phenotype. When the single mutant of *sup20* was generated by backcrossing it to Col-0, it resembled the phenotype of *tfl1* mutant plants. Sequencing of *TFL1* in the double mutant revealed a $G_{729}A$ mutation in the third exon of the gene which led to a $G_{105}S$ amino acid change. Putant *enh20* was hence likely to be a *lhp1*;*tfl1* double mutant plant. It was not further examined in this study and assigned to Class Ib (Table 9).

Putants *enh4*, *enh14* and *enh16* have not been tested in the F_2 generations of their backcrosses and therefore remain in Class Ia (Table 9). The four remaining putants, *enh9*, *enh13*, *enh11* and *enh15* did not display an obvious individual phenotype in the BC_1F_2, when crossed with Col-0 (Table 9). No mutation in the sequenced PcGs could be found and they showed a 3:1 segregation pattern when backcrossed with *lhp1*. Therefore, they remained the most interesting enhancer putants of the screen and are hereafter referred to as candidates.

4.1.3.2 Suppressors of *lhp1*

Like the enhancers, all of the suppressors (*sup1* – *sup11*) have been backcrossed to *lhp1* and to Col-0. Eight suppressors have not yet been tested for their segregation patterns in the

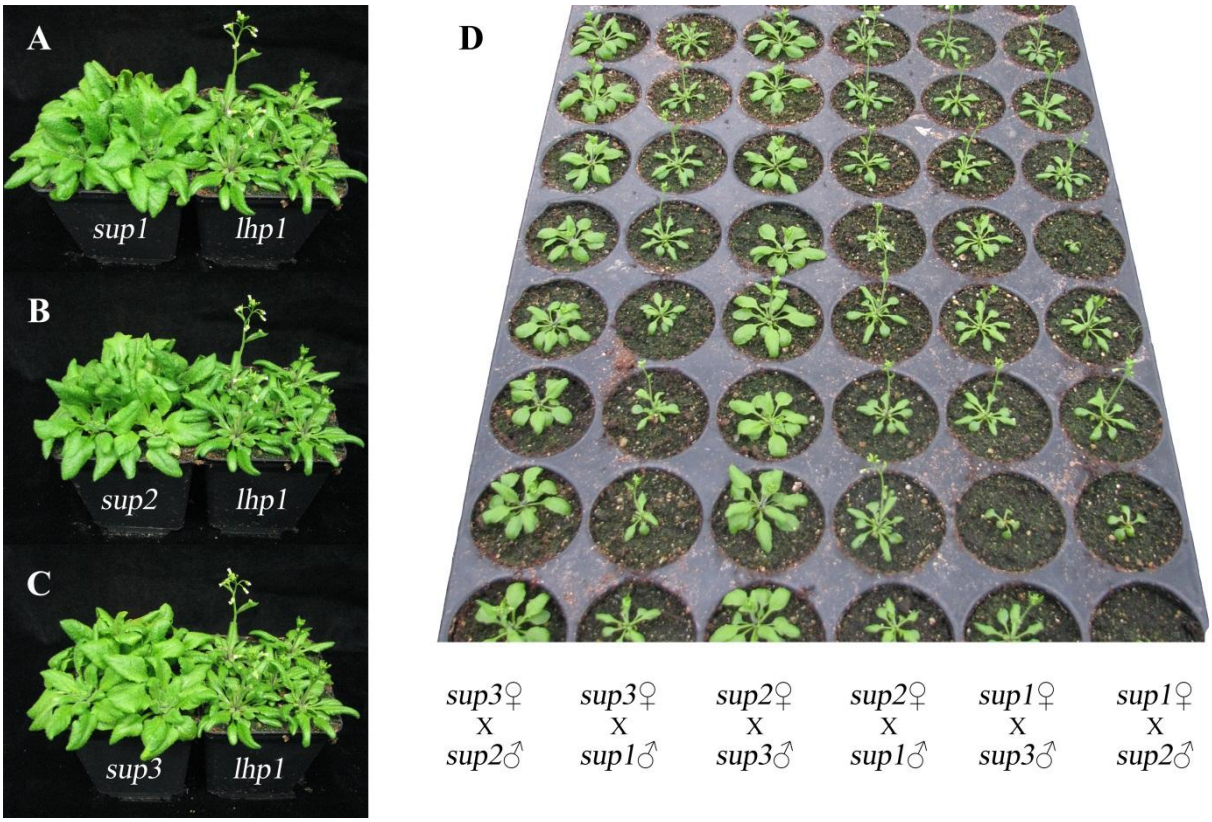


Figure 4: Suppressors 2 and 3 are allelic
(**A** - **C**) Suppressors 1-3 (*sup1*-*sup3*) grown in a climate chamber for 35 days are shown next to *lhp1* mutant plants. Plants were later flowering and had bigger rosette leaves than the single mutant. (**D**) All three suppressors were reciprocally crossed. F_1 plants were grown in LDs and *sup2* × *sup3* crosses shared the double mutant phenotype.

BC_1F_2 with Col-0 used as a maternal crossing partner and are assigned to Class Ia until further examination.

The remaining three suppressors, *sup1* to *sup3*, displayed an intermediate phenotype between WT and *lhp1* plants (Figure 4). When they were reciprocally crossed to each other, *sup2* and *sup3* were found to be most likely allelic, whereas crosses of either putant with *sup1* displayed an *lhp1*-like phenotype in the F_1 generation (Figure 4).

None of the tested suppressors displayed an obvious individual phenotype in the BC_1F_2 when crossed with Col-0 and grown in LD conditions. They also segregated 3:1 when backcrossed with *lhp1*, indicating that a single, recessive mutation was underlying the double mutant phenotype. *Sup1* to *sup3* are hereafter referred to as candidates.

4.2 Analysis of enhancers and suppressors of *lhp1*

4.2.1 Expression analysis of enhancers and suppressors

Three enhancers and three suppressors were selected for a more detailed study of their phenotype and mapping of the causative mutation. Apart from the reasons mentioned above *enh9*, *enh13*, *enh15*, *sup1*, *sup2* and *sup3* were not drastically altered in their *FT* and *FLC* expression when measured in 10 day old seedlings (Figure 5).

FT had a tendency to be expressed higher in *enh9* and *enh13*, matching the early flowering phenotype of *enh9* (Figures 5A, 10B and 10I). *FLC* levels were significantly lower in *enh9* and *enh13* compared to *lhp1*, but significantly higher in *enh15* (2,5 fold / letters of statistical comparison above bars). None of the tested enhancers showed a significant, simultaneous up-regulation of *FT* and *FLC*, as it is found in *lhp1* plants. Since *FLC* represses the expression of *FT* directly (Helliwell et al., 2006), but both are repressed by PRC1, results are challenging to interpret.

For suppressors, *FT* and *FLC* expression fluctuated within a range of about 2,5 fold up- or down-regulation (Figure 5B). *FT* levels had a tendency to be slightly lower and *FLC* levels to

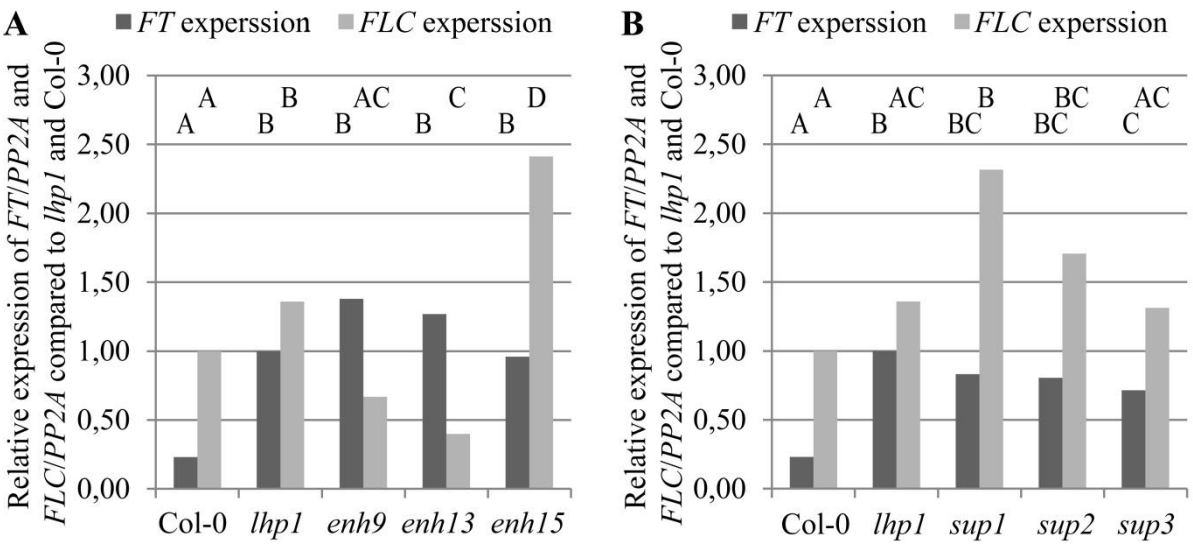


Figure 5: *FT* and *FLC* expression of selected enhancers and suppressors of *lhp1*
(**A**, **B**) 10 day old F_3 seedlings were grown on plates and sampled at ZT16. Expression levels of enhancers and suppressors were normalized with PP2A, displayed on the y-Axis and compared to *FT* levels of *lhp1* and for *FLC* levels of Col-0, which were set to 1. The average of at least two biological replicates per line with three technical replicates each is shown. Single factor ANOVA followed by an honesty significant difference Tukey test confirmed differences between expression levels with a probability of 95 %. Letters above bars indicate similarities and differences between samples. There were 14, 14, 3, 3, 2, 3, 2 and 3 biological replicates for Col-0, *lhp1*, *enh9*, *enh13*, *enh15*, *sup1*, *sup2*, *sup3* respectively. ANOVA was performed separately for enhancers and suppressors.

be slightly higher than in *lhp1*. In *sup1* plants, *FLC* was about 2,5 fold higher expressed than in *lhp1*, which was found to be a significant up-regulation, however still residing below the threshold for possible autonomous pathway mutants (Supplemental figure S2). Hence, the genes underlying *sup1* to *sup3* were likely not part of the well-studied autonomous pathway for flowering.

Taken together, the changes in expression are minor compared to putants that were assigned to Class Ib (Supplemental Figure S2). Thus, flowering phenotypes of all tested candidates are likely independent of *FLC* and *FT* expression. Expression changes of genes downstream of *FT* are probable. Those could either result from additionally altered flowering pathways, or a direct influence on floral integrators, such as *SOC1*, in candidates.

4.2.2 Description of *sup3* and *enh9* phenotypes

Two of the candidates, *sup3* and *enh9*, were chosen to be described in greater detail. Plant height and size were increased in *sup3* compared to *lhp1* but still smaller than WT (Figure 6A). Cauline leaves, siliques and the diameter of the rosette were also of intermediate size compared to Col-0 and *lhp1* (Figures 6C, 6D, 6J and 6K). In contrast to *sup3* plants, leaf size is not increased in *ft;lhp1* double mutants (Kotake et al., 2003), which made it unlikely that the suppression was caused by a mutation in *FT*. Double-mutant plants flowered earlier than Col-0 WT plants but later than *lhp1* mutants (Figure 6I).

Overall, *sup3* mutant plants displayed an intermediate phenotype between Col-0 and *lhp1*. This observation was also statistically supported by an analysis of variance (ANOVA) followed by an honesty significant difference Tukey-test (Figures 6I – 6K). Only the width of the oldest cauline leaf was not significantly different between Col-0 and *sup3* as well as between *lhp1* and *enh9* (Figures 6C, 6E and 6K). The floral organ structures of *sup3* were neither affected in LD nor in SD conditions. However, *sup3* flowers had larger petals and sepals than *lhp1* flowers in SD conditions (Figures 6F and 6G).

A different pattern emerged for *enh9* plants. Plants were earlier flowering and shorter with smaller siliques, cauline and rosette leaves than *lhp1* (Figures 6B, 6D and 6I-K). Mutants in *LHP1* do form terminal flowers with a range of phenotypes, such as carpelloid sepals or lacking petals (Larsson et al., 1998). When the floral architecture was observed in detail some flowers in *enh9* were found that were more severely affected than in *lhp1*. Fused organs or the fusion of entire flowers with two or three fused stamens were observed (Figures 6G and 6H). Quantity of appearance of those flowers was not scored.

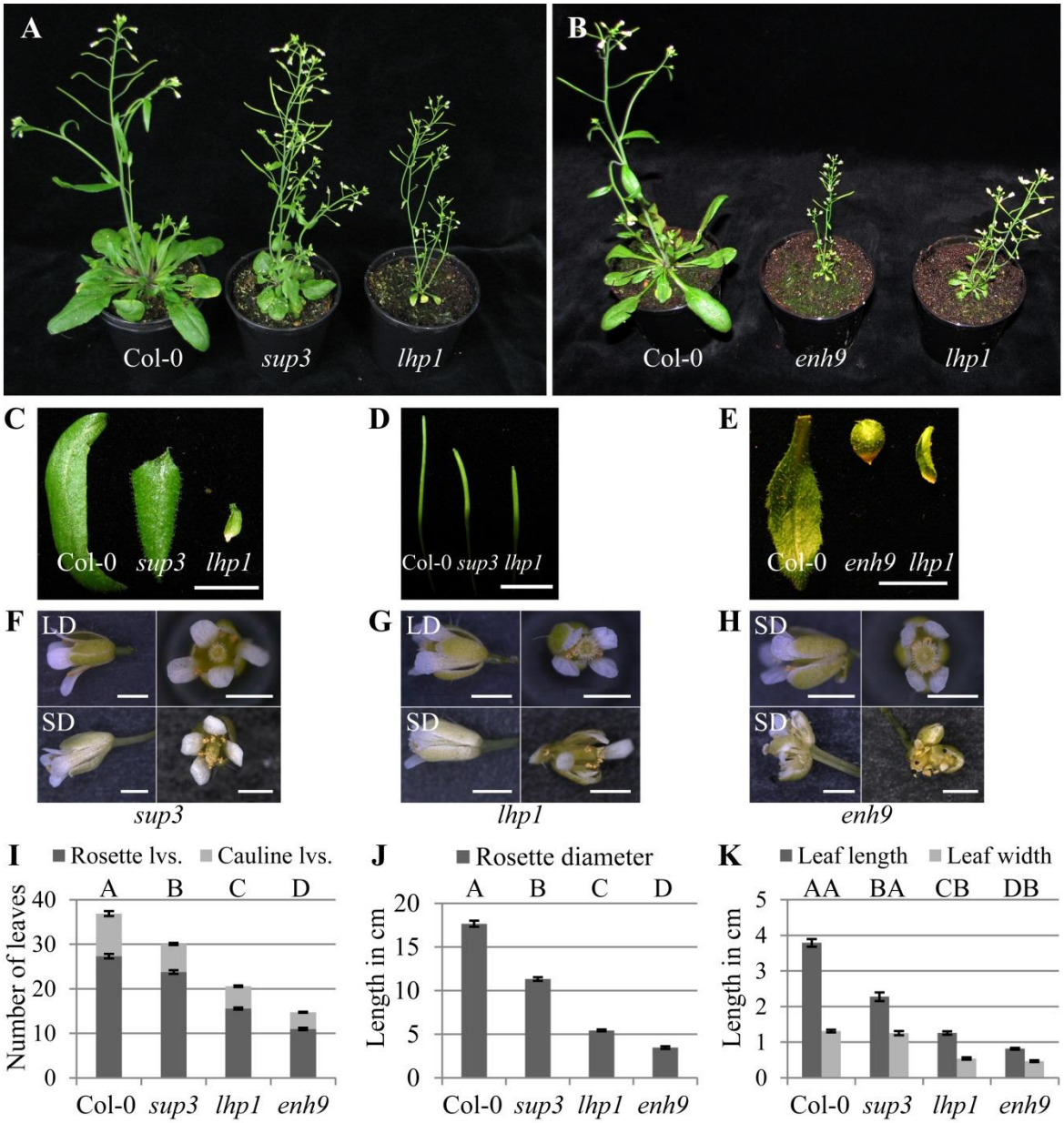


Figure 6: The *sup3* and *enh9* mutants display different phenotypes than Col-0 and *lhp1*
(**A**, **B**) Phenotypic comparisons between Col-0 WT, *sup3* or *enh9* double and *lhp1* single mutant plants are shown 36 days after germination. (**C**, **E**) The 3rd oldest cauline leaf of the main shoot and (**D**) the 7th silique of the main shoot (counted from top) are displayed. (**F - H**) Close-up images of flowers grown in indicated conditions are shown for *sup3*, *lhp1* and *enh9*. (**H**) Some of the *enh9* flowers displayed morphological aberrations, such as fused floral organs. (**I**) Flowering time of 9 Col-0 and *sup3* plants as well as 12 *lhp1* and *enh9* plants was scored after bolting when the main shoot had grown about 1 cm high. The leaf number is indicated on the y-axis. (**J**) For the same number of plants per genotype average rosette diameters, (**K**) average width and length of the oldest cauline leaf of each plant are plotted in cm. (**I - K**) Single factor ANOVA followed by an honesty significant difference Tukey test confirmed differences between genotypes with a probability of 95 %. Error bars represent the standard error of the mean. (**A - K**) Plants were grown in climate chamber conditions (12 h light / 16 °C day / 14 °C night) if not indicated otherwise. (**C**, **D**, **E**) Bars = 1 cm and (**F**, **G**, **H**) bars = 1 mm.

4.2.3 Comparative global gene expression analysis of *sup3*, *enh9*, *lhp1* and WT

For further characterization of *enh9* and *sup3* a high-throughput sequencing approach of whole messenger RNA (mRNA) samples was conducted. To gain significant insight on the molecular footprint of the candidates in comparison to Col-0 and *lhp1*, 12 bar-coded samples, with three biological replicates per genotype were sequenced on a single lane of an Illumina GAII flow cell. This resulted in a total number of ~ 63 M raw reads. After quality control ~ 59 M high quality reads remained and ~ 47 M could be mapped to chloroplast, mitochondrial and nuclear genes.

Reads were evenly distributed with about 7 % of the reads assigned to each bar-code (Figure 7A). Most technical variation was observed within the control groups of Col-0 and *lhp1*. This technical variation was expected, since it had been observed when Actin expression levels of the libraries had been measured in a qRT-PCR experiment prior to the sequencing run (Supplemental Figure S4). Sample *lhp1* c could not be considered in down-stream analysis, since the distribution of raw hits was skewed. A number of genes in that sample had a comparably high number of hits, whereas most of the genes did not have any reads assigned to them (data not shown). Sample Col-0 c was discarded as well, due to the skewed distribution of raw hits, likely resulting from the library preparation, which had to be repeated separately (Figure 7A accentuated samples).

When gene expression was compared to Col-0 WT, most de-regulated genes were discovered in *lhp1* and in the *enh9* double mutant. About 40 to 45 % of the mis-regulated genes were overlapping in both comparisons (*lhp1*/WT and *enh9*/WT / Figure 7B). Not nearly as many genes were de-regulated in the *sup3* double mutant compared to WT. Of the 279 genes about 1/3 was overlapping with one or the two other data sets (Figure 7B). A set of commonly up-regulated genes in all three comparisons were the MADS box transcription factors *SEPALLATA 1* (*SEP1*), *SEP2*, *SEP3*, *AP3*, *PI* and *AG*, which, except for *SEP1* and *SEP2*, have been demonstrated to be direct targets of LHP1 (Adrian et al., 2010; Germann et al., 2006). The diagram was generated considering a threshold of a logarithmic fold change (log_2FC) greater than 2 and a false discovery rate (FDR) smaller than 0,05 for all comparisons. This was found to be a stringent cut-off showing only the most de-regulated genes. In further analysis, to gain insight on the global de-regulation of genes only the FDR was used as threshold (Figure 7D - 7F).

Cross-comparisons of commonly de-regulated genes reinforced the observation that the *sup3* data set resembles that of Col-0 much more than *lhp1* or *enh9*. The largest number of de-regulated genes (2376) was found in the cross-comparison of Col-0 *vs*. *lhp1* and *lhp1 vs*. *sup3*

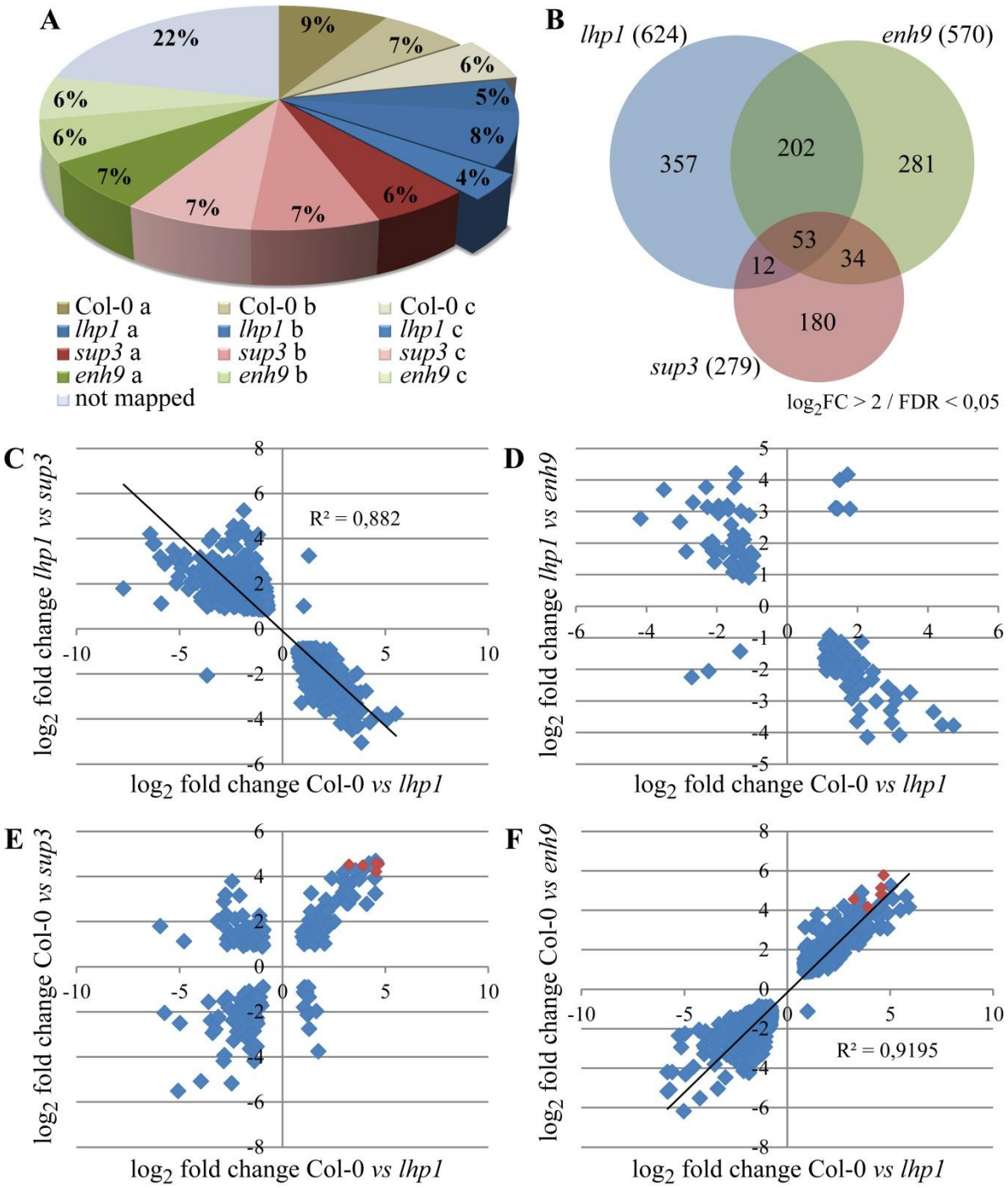


Figure 7: Analysis of mRNA-seq results

(**A**) Distribution of reads in % for the different bar-coded samples that could be mapped to the Arabidopsis genome is shown. Sample Col-0 c and *lhp1* c have not been included in the analysis and are shown as a beveled fraction of the diagram. (**B**) The number of up- and down-regulated genes in *sup3*, *enh9* and *lhp1* was compared to Col-0 and plotted in a Venn diagram. Total numbers of de-regulated genes compared to Col-0 are in brackets. (**C** - **F**) Graphs show commonly de-regulated genes in cross comparisons. Each blue diamond represents a gene in two comparisons *e.g.* Col-0 compared to *lhp1* and *lhp1* compared to *sup3*. X- and y-axis show natural logarithmic values of the fold change to the base of 2. Genes with a FDR < 0,05 were considered (**C**, **F**) Regression lines have been plotted with corresponding values when applicable. (**E**, **F**) Smaller red diamonds represent five commonly up-regulated MADS box transcription factors.

(Figure 7C). When a gene was up-regulated in *lhp1* compared to WT it was plotted on the right side of the y-axis and when the same gene was also up-regulated in *sup3* compared to *lhp1* it was potted on top of the x-axis. Only two genes were found to be up-regulated (upper-right square) and only one gene was down-regulated in both comparisons (Figure 7C). It became evident that most of the genes up-regulated in *lhp1* compared to Col-0 are down-regulated to the same extend in *sup3* compared to *lhp1* (Figure 7C). Therefore, the suppressing gene mutated in *sup3* was termed and from here on referred to as *ANTAGONIST OF LHP1 1* (*ALP1*).

Interestingly, a large number of genes (1581) were commonly de-regulated in the *enh9* and *lhp1* data sets compared to Col-0 (Figure 7F). The direction of de-regulation was the same for all but one gene that was down-regulated in *enh9* when up-regulated in *lhp1*. The correlation of mis-regulated genes compared to Col-0 was high (Figure 7F). Some of the up-regulated genes in *lhp1* were up-regulated to a larger extend in *enh9*. Among those genes were the MADS box transcription factors *SEP1, SEP2, SEP3, AG* and *PI* (red diamonds in Figure 7F). The same genes were also up-regulated in *sup3*, but not generally to a higher extend as in *enh9* (red diamonds in Figure 7E). Due to this finding, the enhancing gene mutated in *enh9* was here after termed *ENHANCER OF LHP1 1* (*ELP1*).

Only 130 and 246 de-regulated genes were found, when *lhp1* was compared to *lhp1*;*elp1* (*enh9* / Figure 7D) or Col-0 to *lhp1*;*alp1* (*sup3* / Figure 7E) respectively. Those two cross-comparisons strengthen the arguments that gene expression in *lhp1*;*alp1* is WT-like, while gene expression in *lhp1*;*elp1* is *lhp1*-like. Taken together, *ALP1* acts as a global repressor of the transcriptomal de-regulation caused by *lhp1*. *ELP1*, on the other hand, enhances expression of a subset of de-regulated genes in *lhp1*.

Notably, splice variants and outliers were removed prior to the analysis (Figures 7C – 7F). Outliers were defined as genes that had a $\log_2$FC > 8 or < -8, which resulted from a statistical bias in the mRNA-seq data. The bias could be explained by lowly expressed genes in all biological replicates of one genotype. When zero raw hits were counted in all replicates of one genotype, the fold change could not be calculated correctly. This could either mean that outliers were not expressed in one genotype at the examined developmental stage, or that some lowly expressed genes could not be studied due to a too low sequencing depth.

There were 57, 9, 27 and 62 outliers defined for each comparison (Figure 7C – 7F) respectively (data not shown). Many of them were overlapping in two or more comparisons, which led to a final list of 93 outliers (Supplemental table 1). Interestingly, *AP3* was an outlier

that was found in three of the comparisons, due to its up-regulation in *lhp1*, *sup3* and *enh9* accompanied by zero hits in all Col-0 samples.

Finally, genes with a FDR < 0,05 that were only de-regulated in one of the two cross-compared datasets were not plotted, but considered in down-stream Gene Ontology (GO) analysis. GO analysis was performed with genes de-regulated compared to Col-0 in *lhp1*, *lhp1*;*alp1* and *lhp1*;*elp1*. In the analysis either only up-regulated or down-regulated genes were considered in groups defined in Figure 7B. Genes were compared to the standard TAIR annotation of the AgriGO visualization platform (Du et al., 2010).

The genes up-regulated in *lhp1*, *lhp1*;*alp1* and *lhp1*;*elp1* were enriched for post-embryonic development, reproductive structure development, reproduction, transcription factor activity and DNA binding, likely due to the large fraction of MADS domain transcription factors in that group. These were the only significantly enriched terms found for up-regulated genes.

Considering the genes solely down-regulated in *lhp1* there was significant enrichment for lipid localization. Genes down-regulated only in *lhp1*;*alp1* were significantly enriched in defence response, response to stimulus, response to bacteria, electron carrier activity and oxygen binding. Last, strongly significant enrichment (p-Value < 10^{-10}) for the response to auxin stimulus was found for the genes down-regulated in *lhp1* and *lhp1*;*elp1*. Among those were 10 SAUR-like auxin response genes, one auxin inducible gene, *AtMYB60,* a guard-cell specific transcription factor and *YUC2*, a gene involved in auxin biosynthesis bound by LHP1 (Rizzardi et al., 2011).

4.2.4 Confirmation of mRNA-seq results

A set of eight genes were chosen to confirm the results obtained by mRNA-sequencing. Four of those had been found to be significantly up-regulated and four had been significantly down-regulated in *lhp1*;*alp1* compared to Col-0, *lhp1* and *lhp1*;*elp1* samples (FDR < 0,05). The up- and down-regulation of six of those genes was confirmed by qRT-PCR in two biological replicates (Figure 8).

Two other genes, At1g29660 and At5g56080, were also tested, but are not plotted here. At1g29660 encodes a lipase, which had been found to be exclusively up-regulated in *lhp1*;*alp1*. In qRT-PCR experiments this gene was only expressed in *lhp1*;*alp1*. At5g56080 encodes *NICOTIANAMINE SYNTHASE 2* (*NAS2*) and was exclusively down-regulated in *lhp1*;*alp1* according to the mRNA-seq results. Real-time data confirmed that finding, with

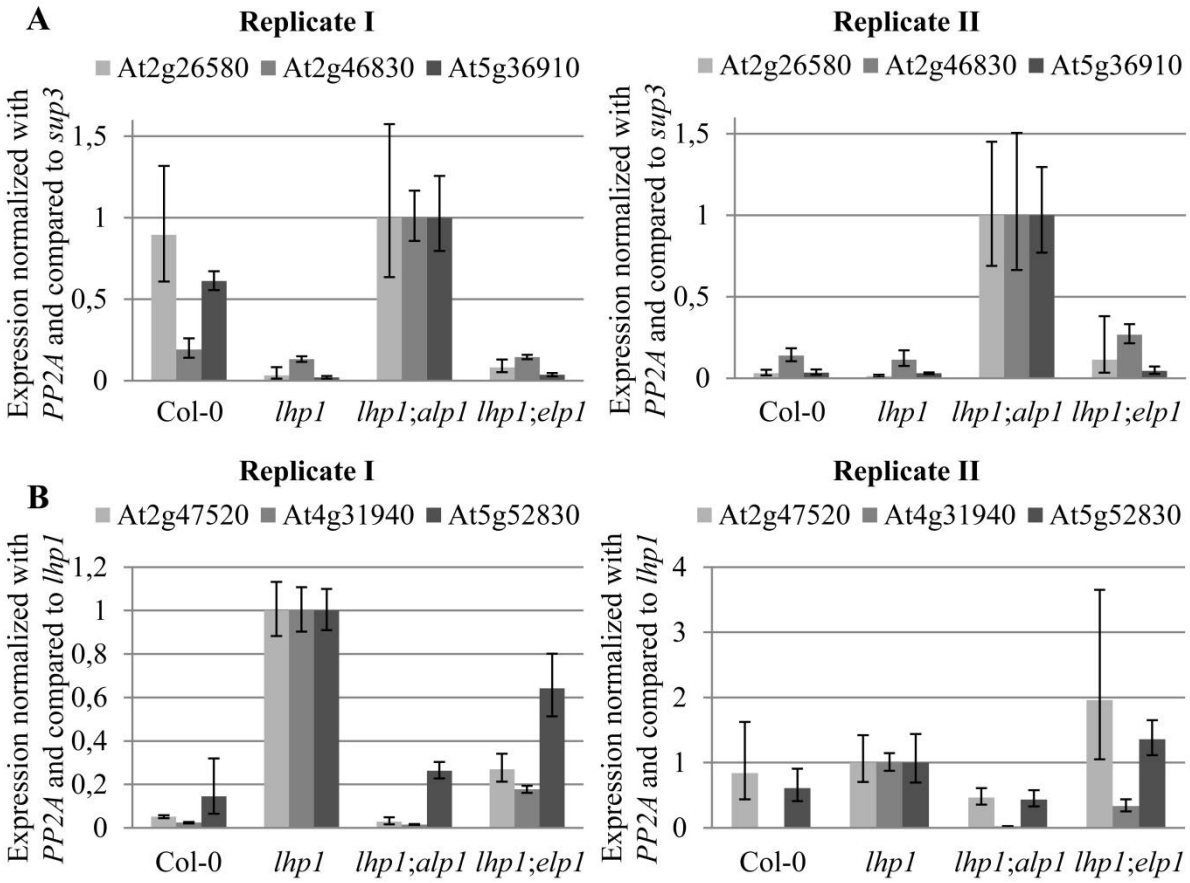


Figure 8: Confirmation of mRNA-seq results
(**A**, **B**) The same RNA samples used for mRNA-seq libraries were used to generate cDNA for qRT-PCR experiments. (**A**) Three genes found to be up-regulated in *lhp1*;*alp1* and (**B**) three genes down-regulated in *lhp1*;*alp1* compared to all other samples are shown in two biological replicates. Error bars represent the standard deviation of three technical replicates compared to the *lhp1*;*alp1* value, which was set to 1.

Col-0, *lhp1* and *lhp1*;*elp1* having had 1,3, 104 and 24 times higher *NAS2* expression than *lhp1*;*alp1* respectively. However, these qRT-PCR results have only been repeated once.

The transcription factor *YABBY5* (At2g26580), the circadian oscillator *CIRCADIAN CLOCK ASSOCIATED 1* (At2g46830) and a pathogenesis-related thionin (At5g36910) all showed a tendency to be up-regulated in *lhp1*;*alp1* compared to the other samples. Only in the first biological replicate the up-regulation towards Col-0 was not clearly visible, revealing a discrepancy between the two biological replicates (Figure 8A). This is possibly resulting from the variation in plant growth, developmental stage at sampling or handling thereafter.

Tested down-regulated genes in *sup3* were the transcription factors *ETHYLENE RESPONSE FACTOR 71* (At2g47520), *WRKY DNA-BINDING PROTEIN 27* (At5g52830) and a cytochrome P450 family enzyme (At4g31940). All three genes were confirmed to be down-regulated with high variation of *ETHYLENE RESPONSE FACTOR 71* expression in the two replicates of *lhp1* (Figure 8B). The expression for the tested P450 family enzyme in the

second biological replicate was below the detection limit and could not be plotted (Figure 8B replicate II).

4.2.5 H3K27me3 decoration of de-regulated genes

De-regulated genes with a FDR < 0,05 and a $\log_2$FC > 2 (Figure 7B) in *lhp1*, *lhp1*;*alp1* and *lhp1*;*elp1* were tested for the direction of de-regulation and their H3K27me3 decoration. In *lhp1*, *lhp1*;*alp1* and *lhp1*;*elp1* approximately 45, 30 and 60 % of the genes were up-regulated respectively (Figure 10). Of the up-regulated genes in each comparison about 60 to 70 % were decorated with H3K27me3 and therefore considered to be direct PRC1 targets. Down-regulated genes were generally less frequently covered with H3K27me3. In *lhp1*;*alp1* and

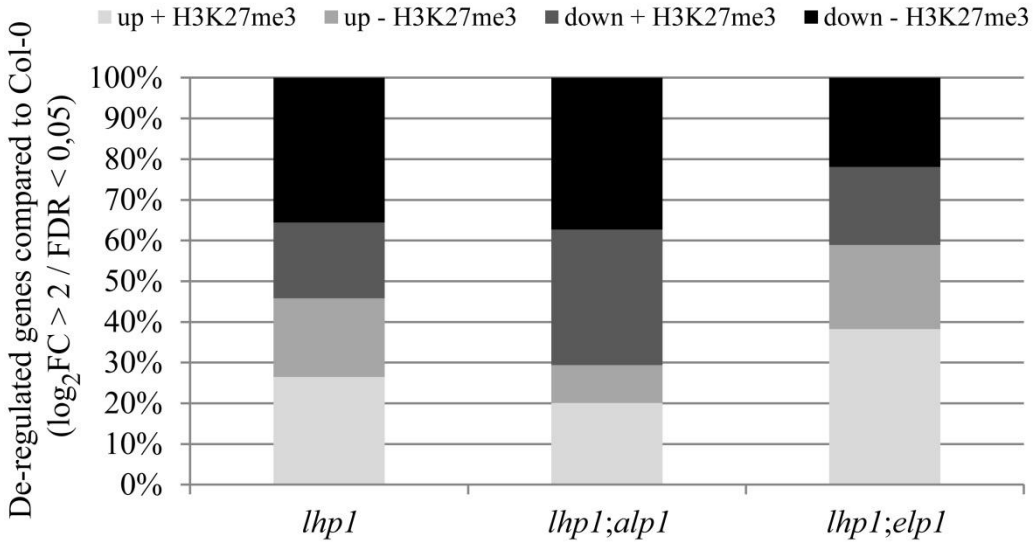


Figure 10: De-regulated genes with or without H3K27me3
De-regulated genes from comparisons to Col-0 were checked for their decoration with H3K27me3 (data analysis by Dr. Julia Engelhorn, MPIPZ; Farrona et al., 2011). Number of de-regulated genes were 624, 279 and 570 for *lhp1*, *lhp1*;*alp1* and *lhp1*;*elp1* respectively (Figure 7B). Genes down-regulated compared to Col-0 are shown in darker shades.

lhp1;*elp1* there were about 50 % and in *lhp1* about 35 % of the down-regulated genes considered to be targeted by PRC1.

Taken together, the percentage of H3K27me3 decorated genes stayed at the same level for up-regulated genes in all comparisons and was slightly altered for down-regulated genes in *lhp1*;*alp1* and *lhp1*;*elp1* compared to *lhp1*. Stronger changes were observed for the number of up- and down-regulated genes. Thus, LHP1, ALP1 and ELP1 influenced transcription and the number of H3K27me3 decorated genes differently.

4.2.6 Active and repressive marks do not change globally

Chromatin regulators such as LHP1 bind chromatin marks and influence the transcriptional state of a gene. In the analysis of transcriptional profiles global changes in transcription were observed for *lhp1*;*alp1* and minor changes for *lhp1*;*elp1* double mutants compared to *lhp1* single mutant (Figure 7; Figure 10). It was tempting to speculate that a global change in the abundance of histone marks in the double mutants could link the candidates more directly to a function in chromatin regulation.

Total protein extracts were used to analyse the abundance of chromatin marks in Col-0, *lhp1*, *lhp1*;*alp1* and *lhp1*;*elp1*. *CLF* and *SWINGER* (*SWN*) mutants were used as positive controls, since it had been shown that at least H3K27me3 marks were altered globally in immunoblot analysis of *clf* (Ws-2) plants (Lafos et al., 2011). However, consistent changes in the abundance of H3K27me3 were not detected in two biological replicates of all samples (Figure 11).

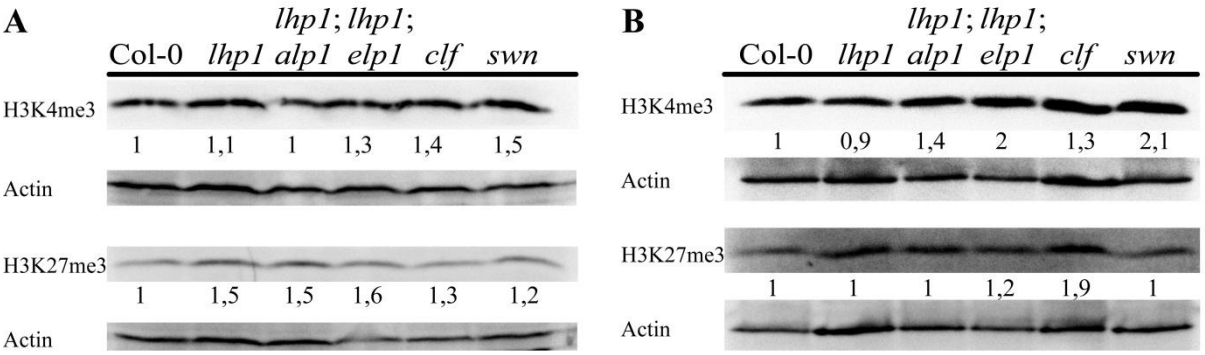


Figure 11: Methylation marks do not change globally in double mutants
(**A**, **B**) Wild-type, *lhp1*, *lhp1*;*alp1*, *lhp1*;*elp1*, *clf* and *swn* protein samples extracted from whole seedlings grown on plates for 10 days in LD conditions were tested for their levels of H3K4me3 and H3K27me3 marks in two biological replicates. Protein was run on a 16 % Polyacrylamide gel under non-native conditions. Quantification was performed in ImageJ by using Actin as a standard protein control.

Changes in H3K4me3 were found for *lhp1*;*elp1* and *swn* samples (Figure 11). Normalized H3K4me3 quantities compared to Col-0 varied from 1,3 to 2,1 for *lhp1*;*elp1* and 1,5 to 2,1 for *swn* in biological replicate 1 and 2 respectively. Variation between the two biological replicates was in general not greater than 0,5 with the exceptions of H3K4me3 levels in *lhp1*;*elp1* and *swn* as well as H3K27me3 levels in *clf*. Those had a slightly higher variation comparing the two replicates. Globally, no consistent changes could be observed for the tested chromatin modifications in *lhp1*;*alp1* and *lhp1*;*elp1* double mutant plants with this method.

4.3 Isogenic mapping of causal mutations by bulked re-sequencing

ALP1 and ELP1 were selected for a mapping-by-sequencing approach. In contrast to conventional mapping, which would make use of two different Arabidopsis accessions carrying a mutation in *lhp1*, an isogenic approach was established. Mapping populations were generated by backcrossing the candidates twice to Col-0 and selecting plants in the F_2 generation (BC_2F_2). The bulked DNA of the plants was then sequenced.

In a primary re-sequencing experiment, leaf samples of 48 *lhp1* plants, 270 *lhp1*;*alp1* plants and 356 *lhp1*;*elp1* plants were pooled and DNA was isolated from the three pools of bulked segregants. Separate libraries were prepared from the three samples, each of which was sequenced on a separate lane on an Illumina Genome Analyzer II platform. Generated sequence data each covered the Col-0 reference genome more than 40 times (Table 11).

High quality reads were aligned to the reference sequence (TAIR 10) using GenomeMapper (Schneeberger et al., 2009a). Sequence differences between the reference sequence (Col-0) and the three sequence sets were independently identified with SHORE (Ossowski et al., 2008; Materials and Methods). The *lhp1* mutant used in this study had been isolated in a previous EMS screen (Larsson et al., 1998). Prior to its application for this screen the *lhp1* mutant had likely been backcrossed several times, since the EMS changes found in the sequenced bulk were not evenly distributed over the genome (personal communication with M.Sc. Geo Velikkakam, MPIPZ).

Due to the heterogeneous characteristics of the *lhp1*;*alp1* and *lhp1*;*elp1* pools, medium allele frequency changes within the analysis of each mutant pool were allowed. All sequence differences that had their origin in the *lhp1* genome were removed from the candidate pools. Thus there were 4284 and 5920 SNPs left for *lhp1*;*alp1* and *lhp1*;*elp1* respectively (Table 11). Those remaining sequence variations were filtered for the canonical EMS changes (G:C and A:T), as the EMS-induced changes were most likely to include the causal mutation. Through this, the SNP set was further reduced to 852 and 652 changes respectively (Table 11). Selection for the lower arm of chromosome 3 for both candidates became apparent through an allele frequency distortion of remaining EMS-induced changes in this region.

An exemplary frequency distribution plot for *lhp1*;*alp1* was created (Figure 12). Frequencies were used in SHORE to create a scale with scores ranging 1 to 40. The SHORE score is influenced by the number of reads and concordance of the reads with respect to the reference sequence (Col-0) and a higher score indicates a higher probability for a change in the input sequence.

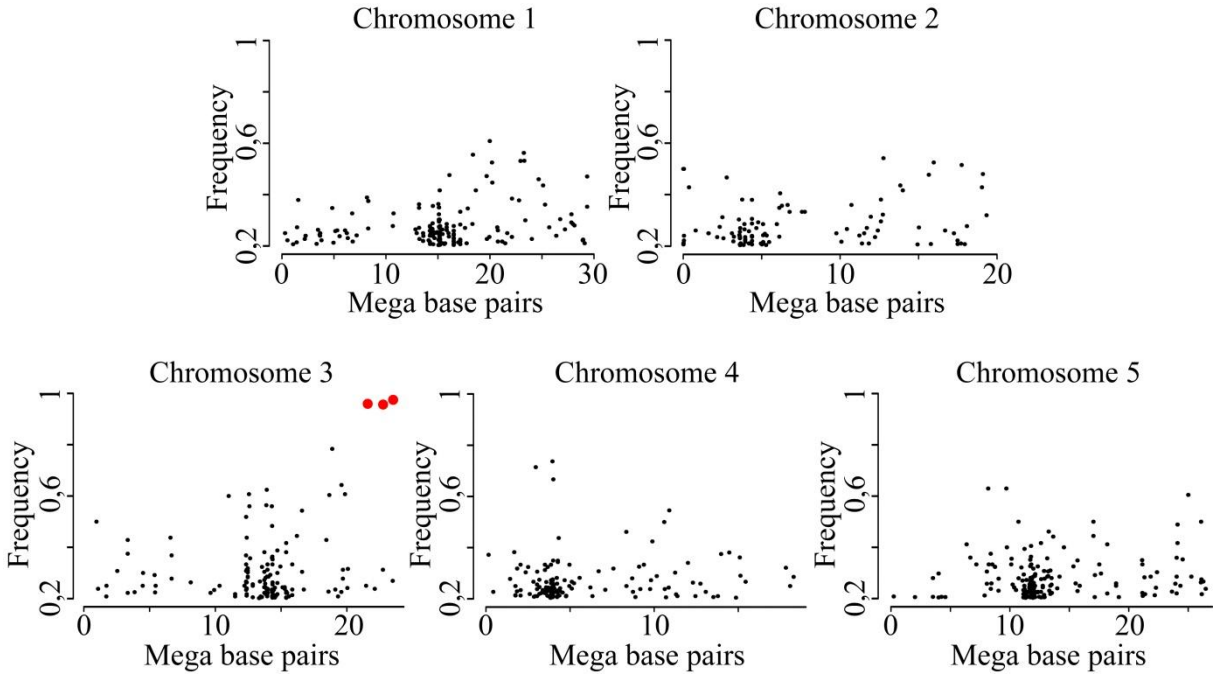


Figure 12: High frequency SNPs of *lhp1*;*alp1* on chromosome 3
Allele frequency distribution of *alp1*;*lhp1*-specific EMS mutations supported by at least seven reads across all five Arabidopsis chromosomes. Each SNP is represented by a black dot. Three SNPs with the highest frequency clustered together on the lower arm of chromosome 3 (red dots). Figure is adapted from M.Sc. Geo Velikkakam and submitted in (Hartwig et al., 2012).

Out of three EMS changes that had a mutant allele frequency higher than 80%, two mutations were found to be located in exons of At3g57940 and At3g63270 and one in an intron of At3g61130. The first two SNPs caused missense mutations leading to amino acid changes of $V_{898}I$ and $G_{273}E$, respectively (Table 11; Figure 13B).

The genes were located at the bottom of chromosome 3 at a distance of about 1,9 Mb to each other. Interestingly, both genes, At3g57940 and At3g63270, were of unknown function, although At3g63270 showed similarities to *Harbinger* transposases when blasted in TAIR.

In the *lhp1*;*elp1* list the sole strongly supported candidate gene with a non-synonymous amino acid change, $W_{75}*$ (* = stop codon), was At3g42660, which codes for a WD40 domain protein of unknown function. Those domains are known interfaces for protein-protein interactions. According to the electronic Fluorescent Pictograph (eFP) browser the gene is highly expressed in the shoot apical meristem and potentially interacting with a number of nuclear proteins (Geisler-Lee et al., 2007; Winter et al., 2007). In fact, the protein with the highest predicted binding probability of At3g42660 is ICU2, a subunit of DNA polymerase alpha and a potential binding partner for LHP1 (Barrero et al., 2007).

A second re-sequencing experiment was carried out with bulked DNA of the double mutants *sup1*, *enh13* and *enh15*. The three DNA libraries were bar-coded, loaded and sequenced on a

Table 11: Re-sequencing data

	lhp1	*lhp1;alp1*	*lhp1;elp1*	*sup1*	*enh13*	*enh15*
No. of bulked plants	48	270	356	308	240	298
Raw reads [Million]	84	84	77	77	60	80
Aligned reads [Million]	79,7	78,1	76,4	75,8	57,5	77,8
Genome coverage	~ 49x	~ 41x	~ 52x	~ 47x	~ 37x	~ 51x
Reliable SNPs vs. *lhp1* [all / EMS induced]	-	4284 / 852	5920 / 655	4467 / 683	4171 / 761	4572 / 625
hom. SNPs reduced [all / EMS induced]	-	6 / 4	1618 / 24	8 / 4	21 / 21	15 / 11
non-synonymous SNPs SHORE score > 35	-	2	1	2	4	4

single lane of an Illumina High Seq flow cell. Two to four non-synonymous SNPs for each of the sequenced candidates were found (Table 11). Most of the SNPs had a SHORE score of 40 and were therefore supported with the highest confidence level. Only the four potentially causal SNPs of *enh15* were scored with a lower concordance coefficient and subsequently a lower SHORE score of 36. The gene with the highest concordance coefficient (0,88) of the four was At3g49850, which codes for TELOMERE REPEAT BINDING FACTOR 3 (TRB3). Interestingly, another member of the same protein family was found to be a putative causal gene of the *enh13* phenotype. This gene was At1g49950, which encodes TRB1 and like TRB3 of *enh15*, had the highest concordance coefficient (1) in the *enh13* list.

The two candidate genes of *sup1*, At5g24500 and At5g25280, were located at the top of chromosome 5 and spaced about 400 kb apart. The SNP found in the exon of At5g24500 resulted in a premature stop codon (protein: W_{639}*), whereas the other candidate SNP caused a proline to serine change ($P_{358}S$) in a different protein.

Results presented here demonstrate a way of efficiently narrowing down the number of target SNPs in an enhancer and suppressor screen for EMS-induced mutants. Mapping by sequencing was applied on singly sequenced mutant plants as well as on bar-coded samples. In each of the five studied mutants not more than four putatively causal, non-synonymous SNPs were found.

4.3.1 Validation of the SNP underlying the *alp1* phenotype

The *lhp1*;*alp1* double mutant was crossed with a Ws-2 *lhp1* mutant as the maternal parent. It was possible to distinguish *lhp1*;*alp1* plants from WT-like or *lhp1*-like plants even though the appearance of the double mutant was altered by the genetic background of the two Arabidopsis accessions. In a rough mapping approach with 61 *lhp1*;*alp1*-like F_2 individuals, linkage of the phenotype to the SSLP marker T20 O10 at the lower arm of chromosome 3 was found (Figure 13A). Only four of the 61 plants were heterozygous at the T20 O10 position and none were homozygous for Ws-2. The marker was lying in between the At3g63270 SNP and the At3g57940 SNP. Another marker, F27K19, on the other side of At3g57940 was heterozygous in eleven of the 61 plants. This result strongly suggested that the causal SNP for the *lhp1*;*alp1* phenotype was located in close proximity to T20 O10.

Two high resolution melting (HRM) markers for the SNPs in At3g57940 and At3g63270 were created (Figure 13B arrows). The markers were used on DNA prepared from 39 F_3 plants with either the *lhp1*;*alp1* or the *lhp1* phenotype derived from different crosses between Ws-2 *lhp1* and Col-0 *lhp1*;*alp1* plants. In the F_2, seven of the used F_3 plants had been homozygous for T20 O10 and heterozygous for F27K19, an SSLP close to the At3g57940 SNP. All other plants had been heterozygous for both SSLP markers in the F_2.

The presence and absence of the nucleic acid change in At3g63270 was correlated with the corresponding phenotype in all 39 cases. However, linkage could not be confirmed for the SNP in At3g57940, where four of the plants with a *lhp1*;*alp1* phenotype were scored as heterozygous and one was scored as homozygous for Ws-2 with the HRM markers (data not shown).

To confirm that the SNP in At3g63270 was causal for the *lhp1*;*alp1* phenotype several measures were taken in addition to the results from outcrossing *alp1* and HRM analysis. First, since *sup2* was found to be allelic to *lhp1*;*alp1* (Figure 4), the two candidate loci identified by re-sequencing were Sanger-sequenced in the two double mutants. Both mutations in *ALP1* could be confirmed. In addition, At3g63270 was disrupted by a G to A change leading to a premature stop codon in *sup2* plants, whereas no mutation was identified for At3g57940 (Figure 13B). Accordingly, the two alleles of ALP1 / At3g63270 were designated *alp1-1* and *alp1*-2 for *sup3* and *sup2* respectively. In further experiments mostly the *alp1-1* allele was used and is hereafter referred to as *alp1*, if not indicated otherwise.

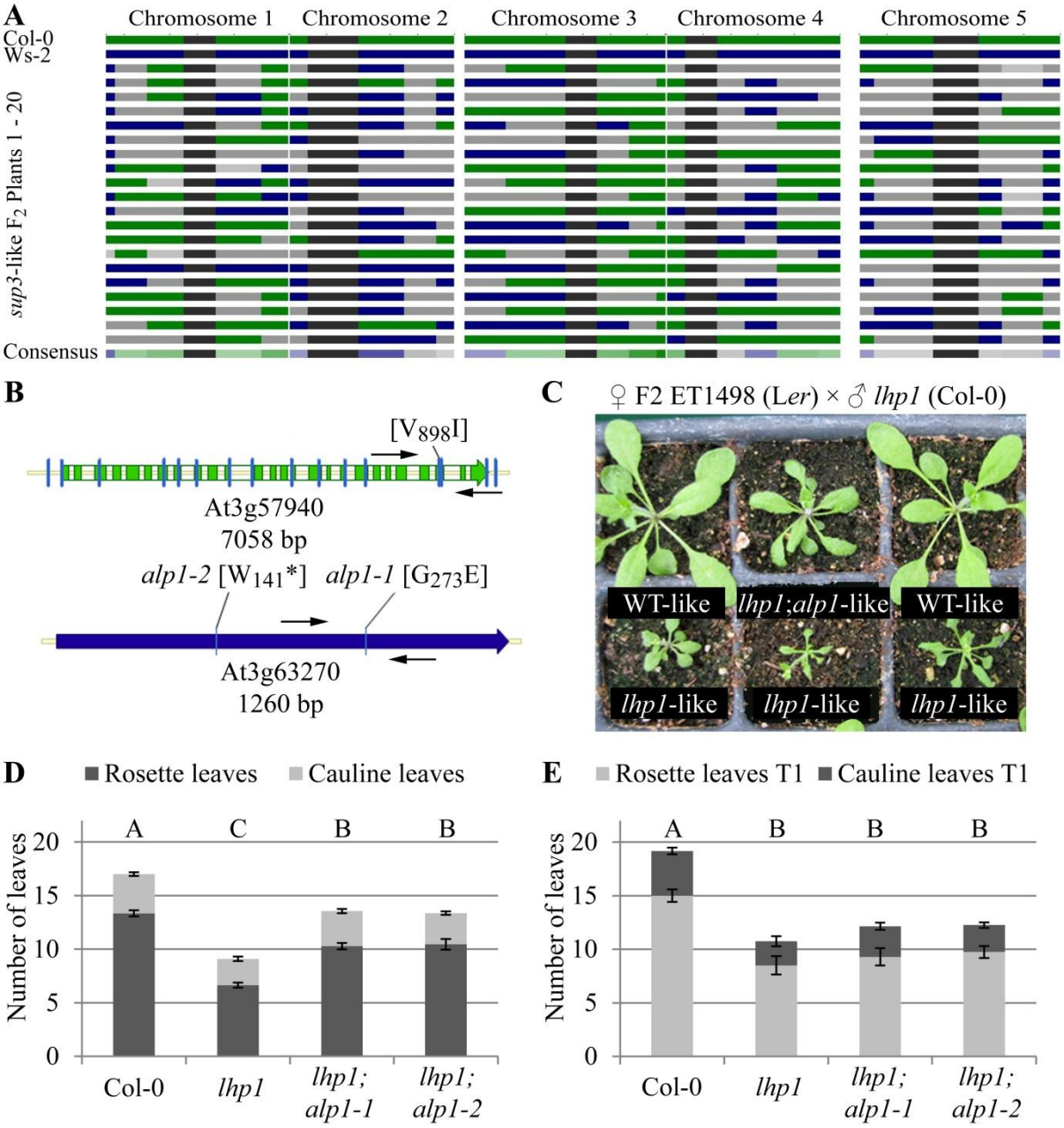


Figure 13: Validation of the causal mutation of *alp1*

(**A**) Rough mapping scheme of 20 F_2 individuals with *alp1*-like phenotypes with a set of 24 SSLP markers distributed evenly over the five chromosomes. Col-0 is shown in green, Ws-2 in blue, heterozygous regions in light grey and centromere regions in dark grey. (**B**, **C**) Identification of a second At3g63270 allele. (**B**) When the two likely target genes with mutations in exons were sequenced in the second allele (*sup2* = *alp1-2*), a second point mutation leading to a premature stop codon could be found in At3g63270. The two candidate genes are displayed as genomic DNA with its Exon-Intron structure (green) and cDNA (blue) respectively. Black arrows indicate primers for high resolution melting. Blue lines indicate sequencing primer positions and point mutations. (**C**) In an F_2 population of a cross between an enhancer trap line with a T-DNA insertion in *ALP1* and *lhp1* some individuals displayed a *alp1*-like phenotype. (**D**, **E**) Complementation of *alp1* plants with a 35S:At3g63270::HA construct could restore flowering time of the two *alp1* alleles to *lhp1* levels. On the y-axis leaf number until bolting is shown. (**D**) Flowering time is shown as an average of 9 to 11 individual plants grown in LD conditions. (**E**) Col-0, *lhp1*, *lhp1*;*alp1-1* and *lhp1*;*alp1-2* were all transformed with the same construct. For the four genotypes 6, 4, 7 and 8 independent T_1 lines grown in LD conditions were considered respectively. (**D**, **E**) Single factor ANOVA followed by an honesty significant difference Tukey test confirmed differences between genotypes with a probability of at least 95 % (letters). Error bars represent the standard error of the mean.

Secondly, a cross between the L*er* line ET1398, containing a T-DNA insertion in *ALP1,* and *lhp1* gave rise of *lhp1*;*alp1*-like individuals in the F_2 generation (Figure 13C). Eighteen of those *lhp1*;*alp1*-like plants were picked and found to be homozygous for the *lhp1* SNP as well as the T-DNA insertion (Supplemental figure S5). When T-DNA lines with insertions in the other two possible candidate genes were crossed to *lhp1* no *lhp1*;*alp1*-like phenotype could be observed among 94 F_2 plants. Thus, the T-DNA allele in ET1398 was designated as *alp1-3*.
Finally, the *lhp1*;*alp1* phenotype could be partially complemented by introducing an ALP1:HA construct driven by a 35S promoter from the CaMV to the double mutant plants comparing at least 4 independent T_1 lines. Transgenic *lhp1*;*alp1-1* and *lhp1*;*alp1-2* plants flowered almost as early as *lhp1* in LDs, when compared to Col-0 and *lhp1* plants carrying the same construct (Figure 13D and 13E).

4.3.2 Phylogenetic analysis of ALP1

In a ClustalW alignment, performed with MEGA5, the amino acid sequence of ALP1 was compared to its closest homologues in Arabidopsis and the 93 most related proteins (sorted by Expect value (E-value)) of all annotated eukaryotic and species available through a Basic Local Alignment Search (BLAST) at the National Center for Biological Information (NCBI). Its amino acid sequence did neither cluster together with the seven other *Harbinger*-like elements from *Arabidopsis*, nor with an out-group of functional *Harbinger*-related transposases, such as the *E. coli* IS5 (Figure 14).
The tree was divided into twelve sub-clades with ALP1 being in sub-clade VI and its closest homologue At3g55350 in sub-clade II. Both sub-clades solely contained plant proteins from different species. Only three of the aligned proteins, apart from the functional transposases, had a function assigned to them. One was a putative RNA binding protein of *Ricinus communis* in sub-clade I, the second was a tryptophan repressor and replication initiator in *Medicago truncata* building the single gene sub-clade III and the third was SALT RESPONSIVE PROTEIN 2 of *Solanum lycopersicum* in sub-clade XII. Two ALP1-related proteins from Arabidopsis, At4g29870 and At5g12010, were also placed into sub-clade XII, together with all mammalian proteins of the alignment.
Sub-clade VI included four additional plant proteins apart from ALP1. They were proteins of unknown function from soy bean, poplar, wine and castor oil. In a ClustalW alignment of the five sequences followed by a protein families (pfam) data base search, a highly conserved helix-turn-helix domain with two DNA binding interfaces (E-value: $1{,}67e^{-03}$) between amino

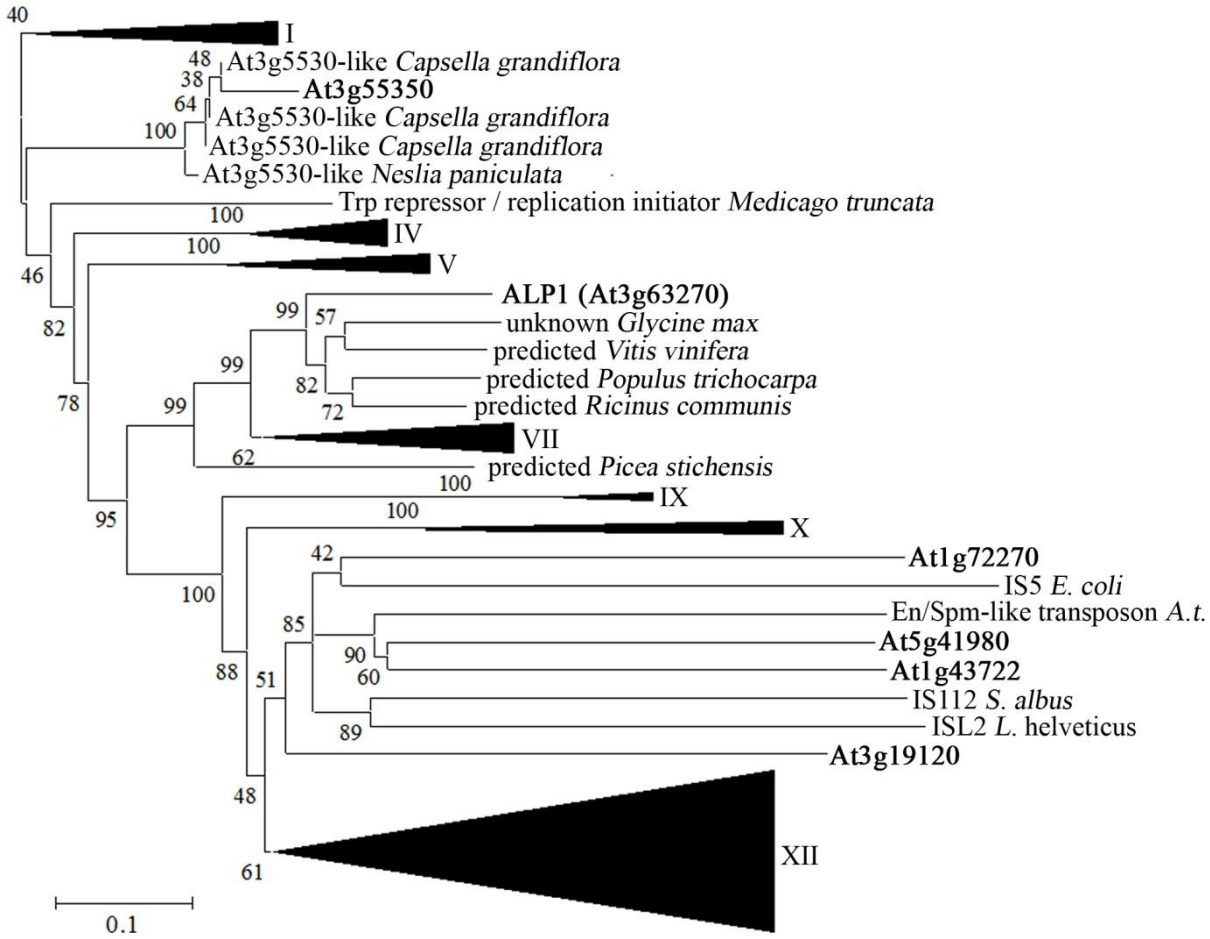


Figure 14: ALP1 clusters apart from active transposases
Most of the twelve sub-clades of the neighbor joining tree have been collapsed and are indicated by roman numerals to emphasize the ALP1, At3g55350 and IS5 sub-clades. The IS5 out-group contained four of the seven ALP1 Arabidopsis homologues (in bold). ALP1 and its closest homologue At3g55350 were assigned to two distinct subgroups. Four other unknown proteins from different plant species clustered together with ALP1 in sub-clade VI.

acids 110 and 141 of ALP1 showed similarity to the DNA binding domain of centromere protein B (Iwahara et al., 1998). This domain was also present in the other proteins of sub-clade VI. In contrast, it was not found in the seven Arabidopsis homologues of ALP1 applying a Conserved Domain Database (CDD) search using standard parameters (Marchler-Bauer et al., 2011).

Amino acid identity of ALP1 paralogues to ALP1 ranged from 23 to 31 % for the six more distantly related proteins. For At3g55350 the amino acid identity was 50 % (E-value $1e^{-121}$). E-values were higher ($1e^{-37}$ to $1e^{-121}$) for the three paralogues that did not associate with the IS5 group ($2e^{-9}$ to $2e^{-25}$).

In a search for conserved domains, the transposase domain of ALP1 was only supported with an E-value of 0,02, which was much less supported than the transposase domain in human

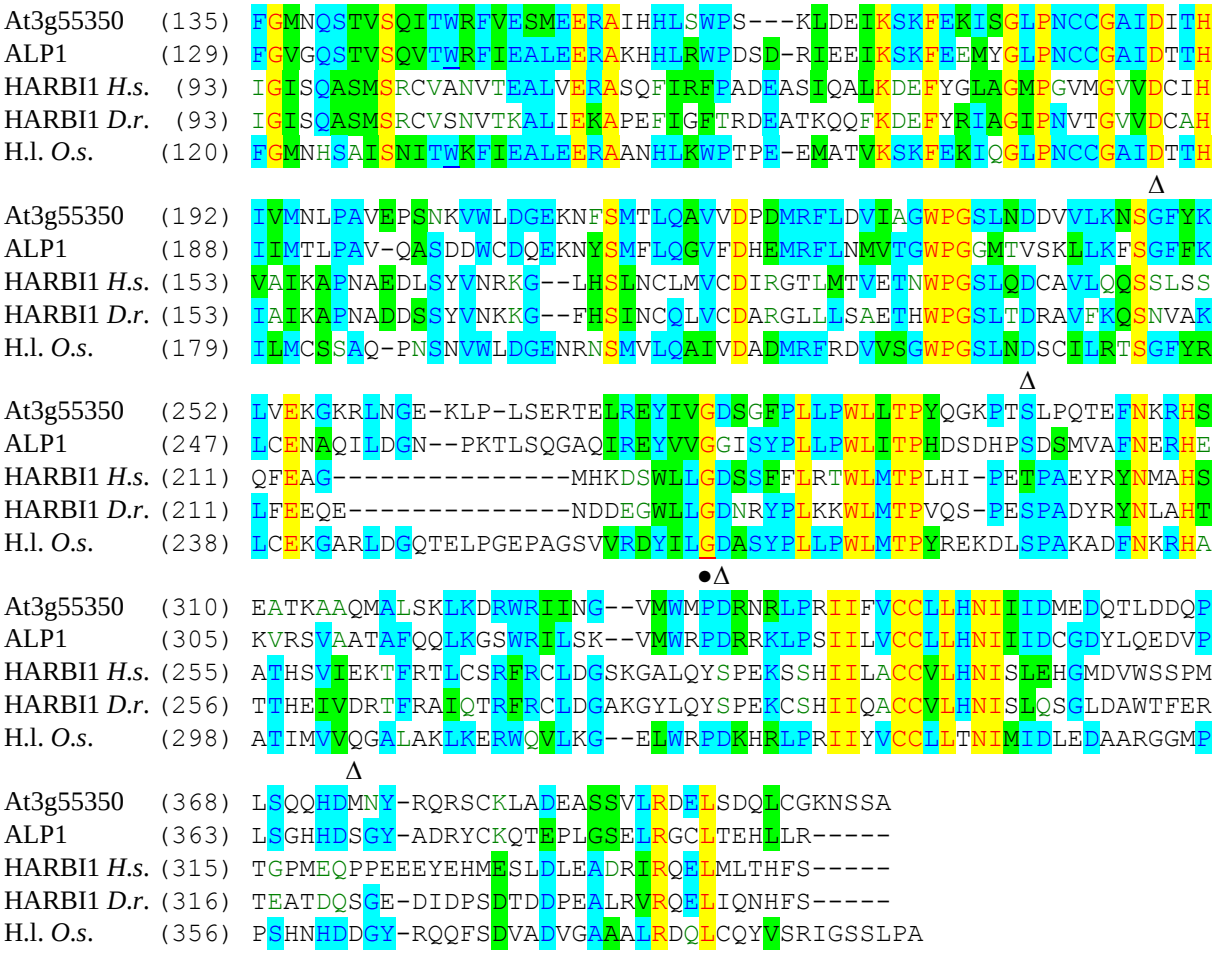


Figure 15: Alignment of *Harbinger* and *Harbinger*-like proteins from different species
ALP1, its closest homologue in Arabidopsis (At3g55350), rice (Os01g0838900), human and zebrafish were aligned using ClustalW in VectorNTI®. The C-terminal part of the alignment is shown. Highly conserved amino acids are in yellow shades. Mutated amino acid of *alp1-1* indicated with a circle below. Amino acids of the "DDE" triad are indicated with triangles below (based on Kapitonov and Jurka, 2004).

Harbinger 1 (HARBI1) ($9.96e^{-5}$) or At3g55350 ($2.93e^{-7}$) (Marchler-Bauer et al., 2011). HARBI1 is the closest human homologue to ALP1 ($3e^{-26}$ / 29 % maximal identity) and predicted to be a domesticated transposase (Kapitonov and Jurka, 2004). HARBI1 is not shown in Figure 14, because all considered proteins had an E-value of $1e^{-34}$ or lower compared to ALP1. When aligning ALP1 with At3g55350, HARBI1 from humans and zebra fish (clade XII) it became evident that the catalytic acidic triad with the conserved amino residues "DDE" was disrupted in ALP1 (Figure 15; Craig, 2002; Kapitonov and Jurka, 2004).

4.3.3 Validation of the SNP underlying the *elp1* phenotype

Only one strongly supported SNP in *lhp1*;*elp1* was found in the re-sequencing analysis (Table 11). When a Gabi-Kat line with a homozygous T-DNA insertion in At3g42660

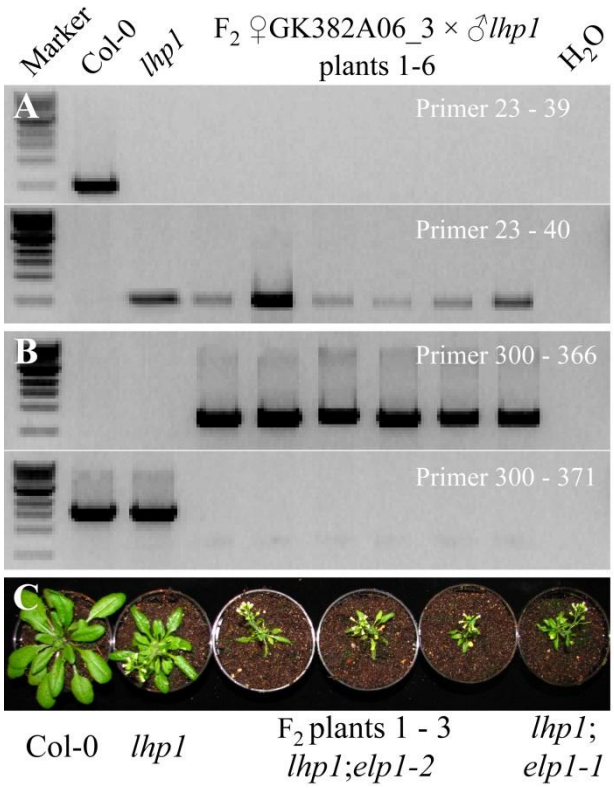


Figure 16: A mutation in At3g42660 causes the phenotype of *lhp1;elp1*
(**A**) Primers were amplifying bands either in WT-like or in *lhp1*-like mutants. (**B**) Primers did give a product when the T-DNA was either present or absent. (**C**) Phenotypes of wild-type, *lhp1*, *lhp1;elp1-2* and *lhp1;elp1-1* double mutant plants were grown at climate chamber conditions.

(GK382A06_3) was crossed with *lhp1*, *elp1*-like individuals were found in the F_2 segregating population (Figure 16C). DNA of six selected individuals was extracted and plants were shown to be homozygous for the *lhp1* point mutation and the T-DNA insertion *via* PCR (Figure 16A and 16B). This confirmed that the mutation of At3g42660 was causing the *lhp1;elp1* phenotype. Consequently, these alleles were designated as *elp1-1* and *elp1-2* for the EMS and T-DNA allele, respectively. In the following experiments only *elp1-1* single and *lhp1;elp1-1* double mutant plants were used. Hence, *elp1-1* is referred to as *elp1* hereafter.

4.3.4 Single mutant analysis

When *lhp1;alp1* and *lhp1;elp1* double mutants were backcrossed to Col-0 no individual mutant phenotype in the F_2 segregating population was observed in LD conditions. Lines with a homozygous SNP in At3g42660 and At3g6370 and no SNP in *LHP1* were selected *via* HRM analysis using Col-0 and double mutant plants as parental controls. Single mutant *elp1-1* lines without an *lhp1* mutation but with a homozygous mutation in At3g42660 could not be distinguished from WT.

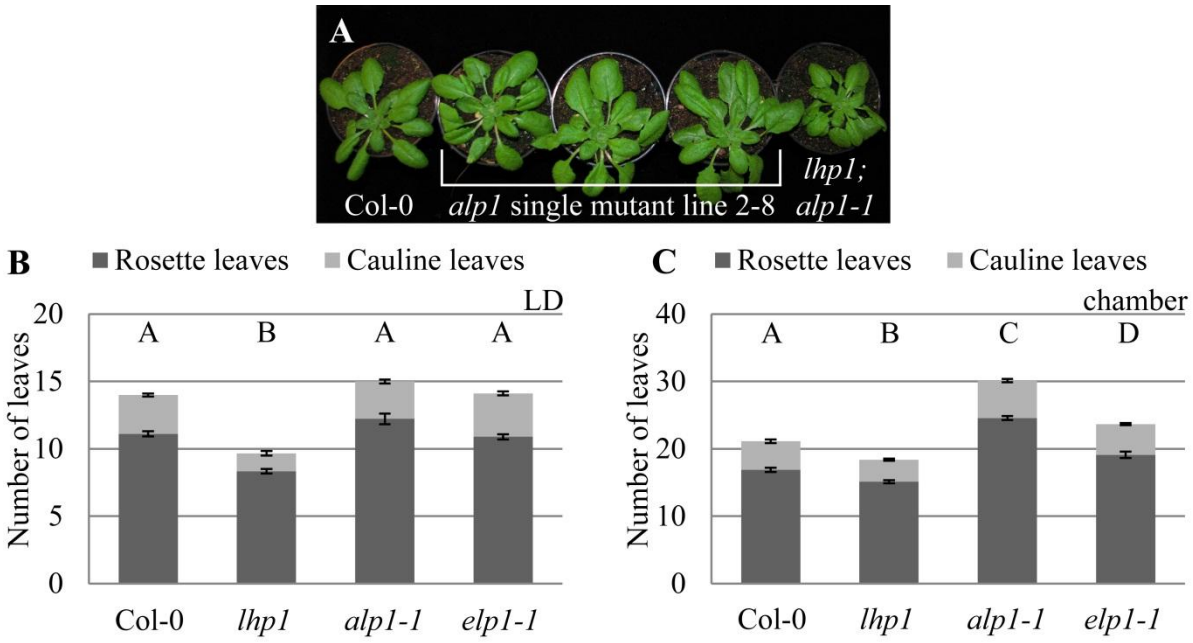


Figure 17: Single mutants of *alp1-1* display a late flowering phenotype
(**A**) Phenotype of one *alp1-1* line (3 representative plants) that had been backcrossed to Col-0 and selected in the F_2 is shown. (**B**, **C**) Flowering time of Col-0, *lhp1*, *alp1-1* and *elp1-1* mutants is plotted for LD and climate chamber conditions repsectively. Single factor ANOVAs followed by Tukey tests with a probability of at least 95 % (letters) have been carried out to compare the genotypes. Error bars represent the standard error of the mean (n = 9).

An aberrant phenotype for *alp1-1* mutants was observed, when plants were grown in a climate chamber (Figure 17A). The rosette leaf architecture distinguished from WT. Leaves seemed slightly bigger with longer internodes and were slightly stronger downwardly curled, which has not been further quantified and will be analyzed in several *alp1-1* lines.

There was no significant difference between the flowering time of WT and single mutant plants observed in LD conditions (Figure 17B). However, when plants were grown at colder temperatures in a growth chamber there were significant differences in flowering time between all genotypes (Figure 17C).

4.3.5 *ALP1* and *ELP1* expression are not influenced by LHP1

In the mRNA-seq analysis *ALP1* was not de-regulated in any of the samples, while *ELP1* was down-regulated in all *lhp1*;*elp1* samples. Both genes were tested for the histone methylation pattern of marks either associated with transcriptionally active (H3K4me3 / H3K36me2) repressed (H3K27me3) or silenced (H3K9me2) loci (Figure 18A and 18B). *ALP1* and *ELP1* displayed a common pattern of actively transcribed genes with H3K4me3 at the TSS and H3K36me2 throughout the gene body and towards the 5'end (Luo and Lam, 2010). The genes

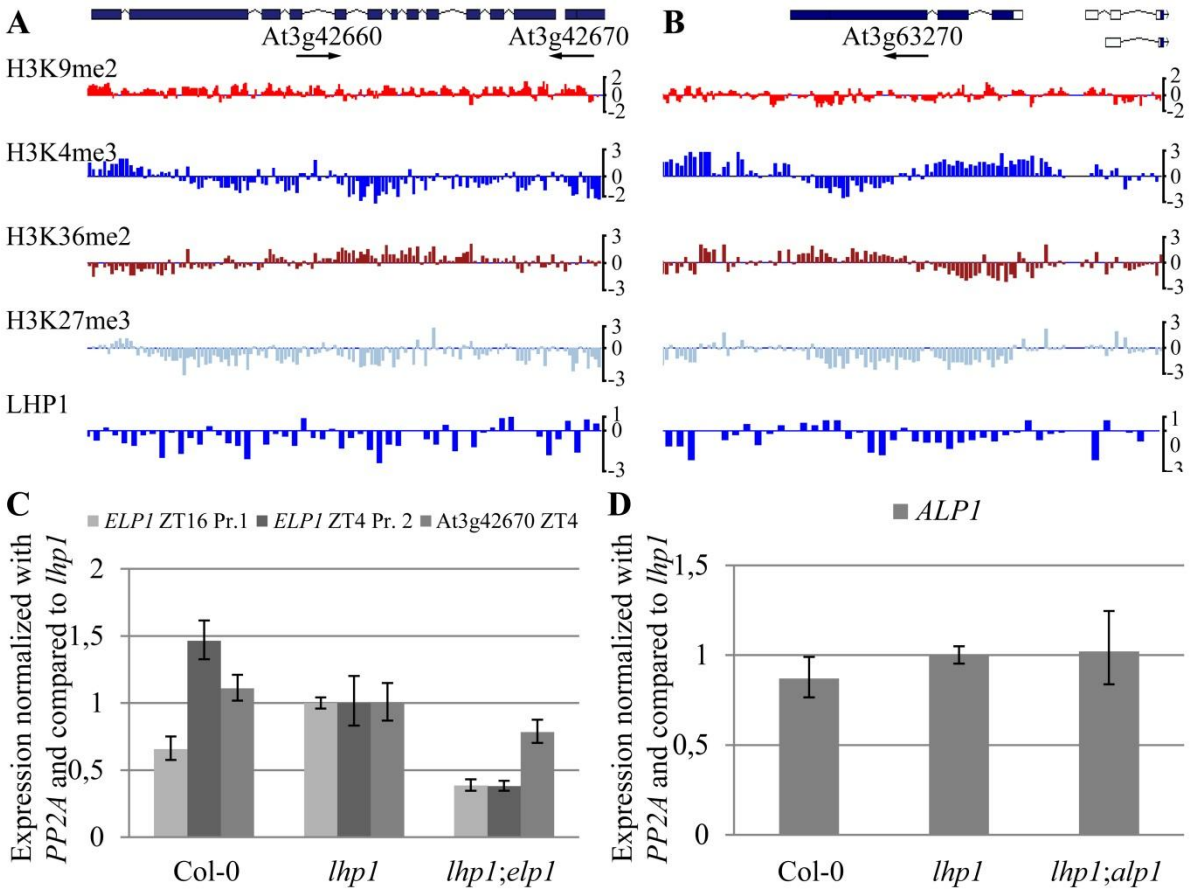


Figure 18: *ALP1* and *ELP1* expression is not influenced by LHP1
(**A**, **B**) Genome browser images of *ELP1* (At3g42660) and *ALP1* (At3g63270) loci with their corresponding histone mark decoration. H3K9me2 ChIP-Chip data was generated by (Rehrauer et al., 2010), H3K4me3 and H3K36me2 by (Oh et al., 2008), H3K27me3 by (Zhang et al., 2007) and LHP1 data by (Turck et al., 2007). 3'-ends of *ELP1* and At3g43670 lie in close proximity to each other. Arrows indicate the direction of transcription. ChIP-Chip data is presented as $\log_2$ (IP/input) (**C**, **D**) Expression levels of *ELP1*, At3g42670 and *ALP1*. (**C**) Expression of *ELP1* was measured with two different primers in Col-0, *lhp1* and *lhp1*;*elp1*. The adjacent gene to *ELP1*, At3g42670, was measured with one primer. (**D**) *ALP1* expression was measured in Col-0, *lhp1* and *lhp1*;*alp1* respectively.

were neither covered by H3K9me2 nor decorated with H3K27me3 and consequently also not bound by LHP1 (Figure 18A and 18B).

In qRT-PCR experiments two different primer-pairs were used to measure *ELP1* and one to measure *ALP1* expression (Figure 18C and 18D). *ELP1* was expressed at lower levels in *lhp1*;*elp1* confirming the mRNA-seq results. Expression levels were at about 40 % compared to Col-0 and *lhp1*, while the neighboring gene's expression, At3g42670, was not significantly affected in *lhp1*;*elp1* (Figure 18C). *ALP1* was expressed at similar levels in Col-0, *lhp1* and *lhp1*;*alp1* (Figure 18D).

These results demonstrate that *ALP1* and *ELP1* are not directly targeted by LHP1. Thus, secondary effects by LHP1 on transcription of these genes can be ruled out.

4.3.6 Analysis of de-regulated MADS domain transcription factors

Transcription factors containing a MADS domain were found to be strongly up-regulated in *lhp1*, *lhp1*;*alp1* and *lhp1*;*elp1* compared to WT (Figure 7E and 7F). The global transcriptional profiles generated could be confirmed for a number of genes in real-time PCR experiments (Figure 8). The same technique was applied to further elucidate the relationship between *lhp1*, *alp1*, *elp1* and MADS domain transcription factors. This class of genes was chosen, due to its high impact on developmental processes, such as flower development or stem cell maintenance in plants (Theissen et al., 2000). Further, only a low number of hits were detected for MADS domain transcription factors in Col-0 samples (zero hits for *AP3*; Supplemental table 1), strengthening the importance of in depth analysis.

All three tested transcription factors, *AP3*, *SEP3* and *AG*, were highly up-regulated in *lhp1* compared to Col-0, *alp1-1* and *elp1-1* (Figure 18A and 18B). *PP2A* or *ACT2* were used as housekeeping genes. Up-regulation was highest for *SEP3* (~ 550 ×), followed by *AP3* (~ 300 ×) and *AG* (~ 60 ×) compared to WT and *alp1-1*. *Elp1-1* mutants had slightly higher levels of *AP3*, *SEP3* and *AG* compared to Col-0 and *alp1-1* (~ 7, 8 and 1,5 × respectively).

In *lhp1*;*alp1* expression levels of *AP3* and *SEP3* dropped to about 20 to 25 %. The drop of expression was not as high for *AG* which was still expressed at about 75 % of *lhp1* levels (Figure 18A and 18B). In *lhp1*;*elp1* on the other hand a tendency for further up-regulation of the transcription factors was observed. The effect was largest on *SEP3*, with expression levels

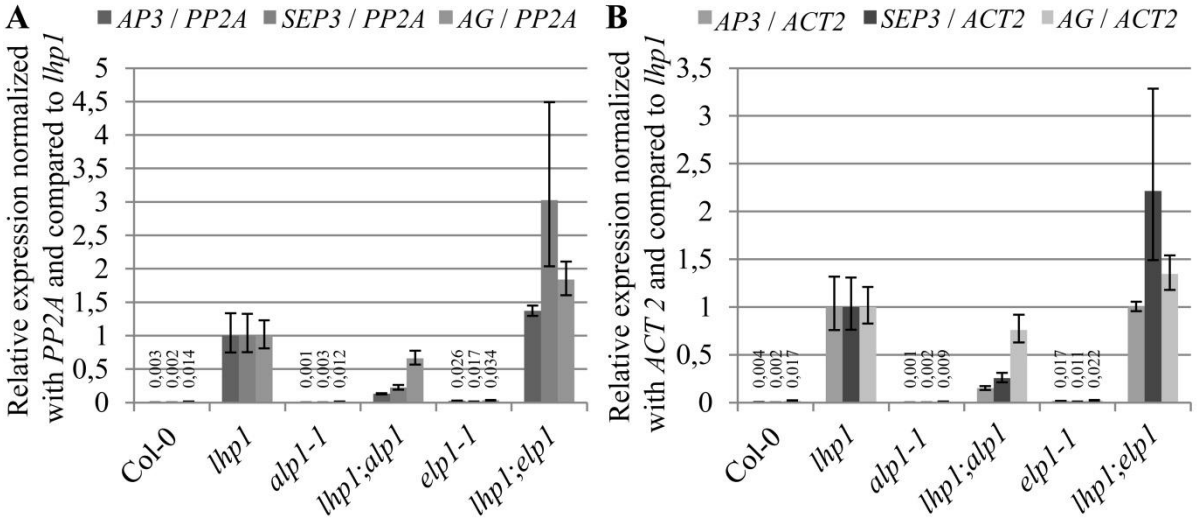


Figure 18: MADS box transcription factors are influenced by *alp1* and *elp1*
(**A**, **B**) Plants were grown on soil in a climate chamber for 14 days before sampling. Error bars represent the standard variation of three technical replicates except for two replicates of *AP3* in *lhp1*;*elp1* samples. All samples were compared to *lhp1* expression values, which were set to 1 and normalized with (**A**) *PP2A* or (**B**) *ACT2*.

rising to ~ 2,5 times of *lhp1* (Figure 18A and 18B). Differences in expression were slightly less when *ACT2* was used as a housekeeping gene. All described tendencies however remained similar.

Considering all of the results presented here, *alp1-1* did have a strong effect on at least a fraction of MADS domain transcription factors. This effect was only present in an *lhp1* mutant background and counteracted the up-regulation in *lhp1*. On the other hand *elp1* seemed to have a slight effect on the up-regulation of MADS domain transcription factors alone. When *lhp1*;*elp1* double mutants were tested the effect of both single mutants appeared to be additive for fold up-regulation (Figure 18).

4.4 Nuclear localization and binding studies

4.4.1 ELP1 is a nuclear protein

Constructs with a 35S promoter, ALP1, ELP1 as well as LHP1 and a Green Fluorescent Protein (GFP) were created to transform *lhp1*;*alp1*, *lhp1*;*elp1* and *lhp1* plants respectively. For the ALP1 construct only three transformants were retrieved in the T_1 and none of them gave any GFP signal (data not shown).

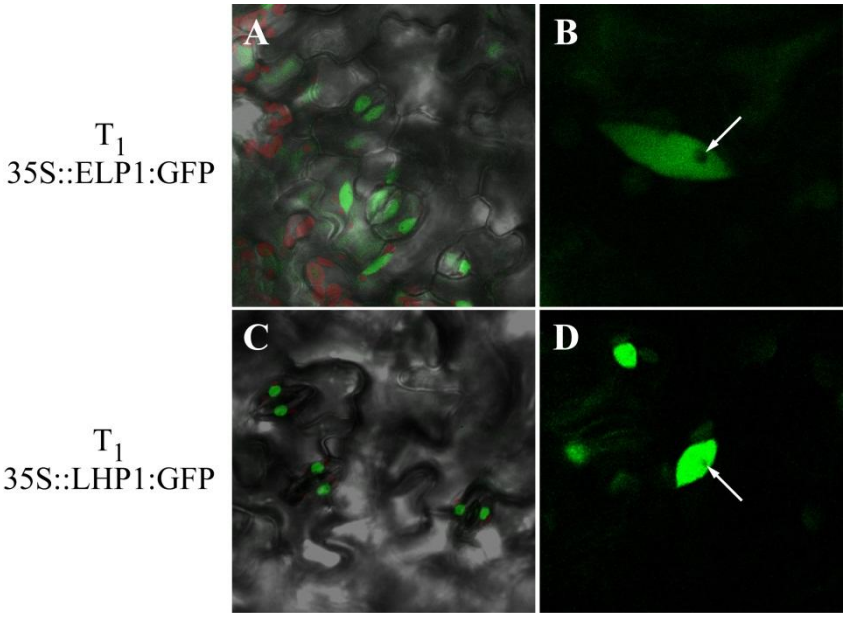
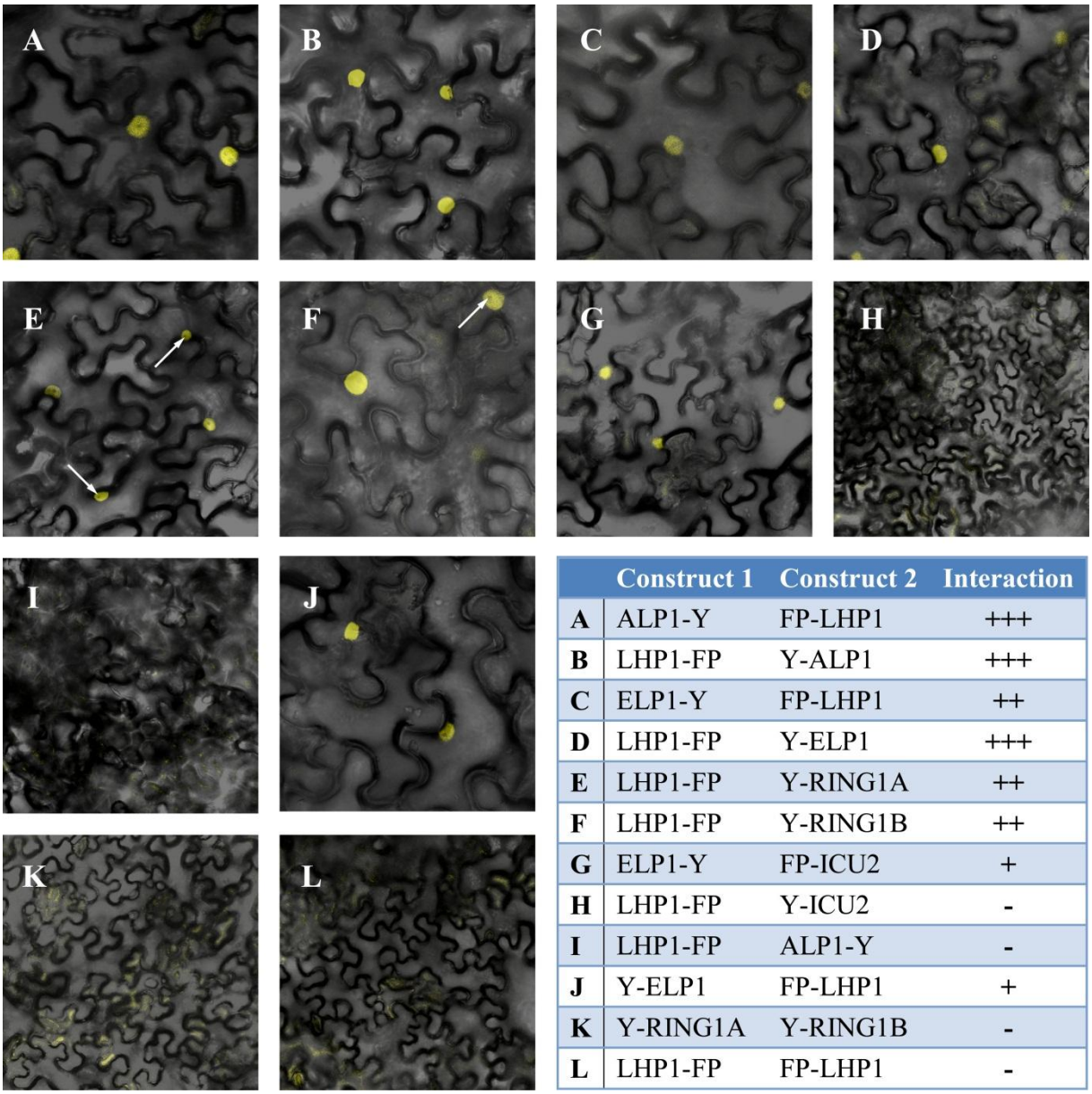


Figure 19: ELP1 is localized at nuclei
(**A**, **B**) One representative Arabidopsis T_1 line transformed with 35S:ELP1::GFP is shown for ELP1 (**C**, **D**) as well as LHP1. (**A**, **C**) Bright field images were set to 50 % transparency and merged with GFP images. Chloroplasts are shown in red color. (**B**, **D**) Protein-fused GFP in close-up images. GFP signal was elusive of nucleoli in both constructs tested (arrows).

One line, complementing *lhp1*;*elp1* partially and two complementing *lhp1* almost fully, were selected in the T_1. Two representative lines of each construct displayed a similar GFP pattern visible in the nuclei of Arabidopsis leaf and stomata cells (Figure 19).
When single cells were closely observed it seemed as though the nucleoli of the cells were not fluorescing with GFP in both lines (Figure 19B and 19D). The ELP1 signal was weaker than the LHP1 signals and its nuclear pattern reminded of the RING1A and RING1B YFP signal in tobacco nuclei (see below). This result confirmes that ELP1 is likely to be an exclusively nuclear protein with a similar distribution pattern as LHP1.

4.4.2 ALP1 and ELP1 interact with LHP1 in tobacco leaf cells

The YFP fluorescence in all assays was scored visually in categories ranging from no signal in three replicates (-) to strong signal in all replicates (+++). When an interaction could be observed in all infiltrated leaves, but was not giving a comparably strong YFP signal in many cells it was scored as a medium (++) interaction (Figure 20C, 20E and 20F). A weak interaction (+) was only giving signal in one or two of three infiltrated plants (Figure 20G) or in only a few cells close to the infiltration site (Figure 20J). Interaction categories for each infiltration experiment were summarized in the table next to the figures (Figure 20).
Parallel infiltration of tobacco leaves with LHP1 fused to the C-terminal part of YFP (FP) and ALP1 as well as ELP1 fused to the N-terminal part of YFP (Y) resulted in nuclear YFP signal in the nuclei of infiltrated cells (Figure 20A - 20D). For ELP1 also the N-terminal Y-fusion gave a weak YFP signal when infiltrated together with N-terminally fused FP to LHP1 (Figure 20J). Little to no signal could be observed in other parts of the cells such as the cytoplasm (Figure 20). To determine the amount of false-positive fluorescence a series of negative controls was used. Neither in parallel infiltration of LHP1 proteins with FP-fusions on either side (Figure 20L) nor the exclusive infiltration of ALP1 and ELP1 fused to Y on either side of the proteins (data not shown) or the N-terminal Y-fusion to ALP1 infiltrated with the C-terminal FP-fusion to LHP1 (Figure 20I), showed any YFP signal in 3 biological replicates of two different infiltration experiments.
RING1A and RING1B are known to interact with LHP1 *in vivo* (Xu and Shen, 2008) and were chosen as positive controls (Figure 20E and 20F). The signal strength was comparable to the interaction observed between ELP1 and LHP1 (Figure 20C) and even stronger in co-infiltrated leaves with LHP1 and ALP1 (Figure 20A and 20B). There was no signal observed when the two RING proteins were fused to Y and infiltrated in parallel (Figure 20K). A weak



	Construct 1	Construct 2	Interaction
A	ALP1-Y	FP-LHP1	+++
B	LHP1-FP	Y-ALP1	+++
C	ELP1-Y	FP-LHP1	++
D	LHP1-FP	Y-ELP1	+++
E	LHP1-FP	Y-RING1A	++
F	LHP1-FP	Y-RING1B	++
G	ELP1-Y	FP-ICU2	+
H	LHP1-FP	Y-ICU2	-
I	LHP1-FP	ALP1-Y	-
J	Y-ELP1	FP-LHP1	+
K	Y-RING1A	Y-RING1B	-
L	LHP1-FP	FP-LHP1	-

Figure 20: LHP1 interacts with ALP1 and ELP1 in tobacco leaf cells
(**A** - **L**) Confocal images were merged with bright field images, while the bright field transparency was set to 50 %. YFP signal is shown in yellow and has only been observed in nuclei of tobacco leaf cells. (**I** - **L**) Smaller images represent the negative controls used. In the table interaction for construct pairs was scored visually, with +++ for strong interaction, ++ for medium interaction, + for weak interaction and – for no interaction between proteins.

interaction could also be found between ELP1 and ICU2. An interaction between ICU2 and LHP1 was not observed (Figure 20H). Interestingly, YFP signals observed for the RING proteins were not seen in nucleoli of cells (arrows in Figure 20E and 20F), which was also the case for ALP1 and ELP1 interactions.

Considering all observations, ALP1 and ELP1 strongly interacted with LHP1. In addition, ELP1 also showed a weak interaction with ICU2 which will be tested in further *in vitro* and eventually *in vivo* assays. A previously reported *in vitro* interaction of ICU2 and LHP1 could not be confirmed using split-YFP constructs (Barrero et al., 2007).

4.4.3 YFP-ALP1 does not co-localize with CFP-LHP1 in tobacco nuclei

Constructs with CFP and YFP fused N-terminally to LHP1 and ALP1 respectively were created to further elucidate the relationship between the two proteins. After infiltration in tobacco of YFP-ALP1 alone, the protein was exclusively localized at nuclei of leaf cells. The YFP signal was spread across the whole nucleus as seen before for split-YFP constructs. However, a denser YFP signal that resembled the form of large speckles or crystals resided within the center of all nuclei with stronger YFP signal (Figure 21A).

Intriguingly, when YFP-ALP1 and CFP-LHP1 were co-infiltrated the CFP signal was weakest at the location of the crystal-like YFP signal (Figure 21B). This separate localization was observed at most nuclei in tobacco leaf cell with only few exceptions found (Figure 21E).

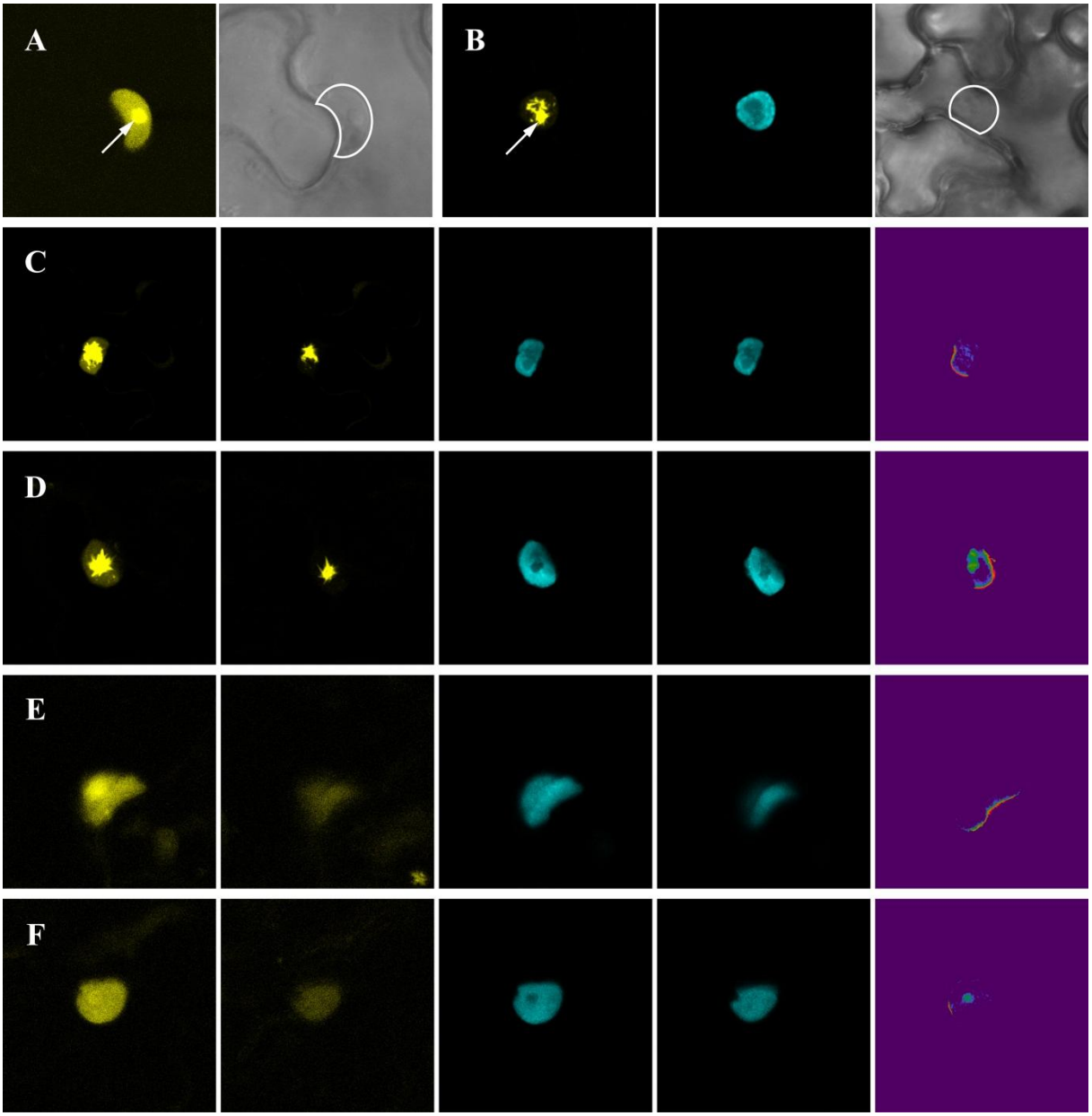


Figure 21: YFP-ALP1 fusions do not interact with LHP1
(**A**) YFP and bright-field image of a nucleus (line) containing YFP-ALP1 singly infiltrated into tobacco leaves. (**B**) YFP-ALP1 and CFP-LHP1 co-infiltrated and shown in the same tobacco nucleus (line). (**A**, **B**) Arrows indicate crystal-like YFP structures in the center of nuclei. (**C** - **D**) Acceptor Photo Bleaching (APB) experiments with co-infiltrated tobacco nuclei. First and second image YFP pre- and post-bleaching. Third and fourth image CFP pre- and post-bleaching. Fifth image calculated gain of CFP fluorescence at region of interest. (**C**, **D**) YFP nuclei with strong accumulation of YFP in the center. (**E**, **F**) YFP nuclei with evenly distributed YFP signal throughout the nucleus.

When nuclei of co-infiltrated cells were subjected to Acceptor Photo Bleaching (APB) experiments, no Fluorescence Resonance Energy Transfer (FRET) was observed (Figure 21C to 21F). However, a few points were striking in those experiments. Firstly, almost all YFP-ALP1 aside the crystal-like YFP structures was bleached at the time of FRET measurement (when 40 % of YFP signal remained), indicating a rigid structure at the nuclear center. Secondly, for half of the nuclei tested, FRET was detected at sub-compartments of nuclei, which might be attributed to a re-localization of CFP-LHP1 when a fraction of YFP-ALP1 had been bleached in these areas (Figure 21D and 21F). Finally, in split-YFP experiments, when only part of YFP was fused N-terminally to ALP1, a crystal-like structure had not been observed. This raises the possibility that the protein structure or function was altered by a longer N-terminal fusion. Alternatively, the crystal-like structure does not interact with LHP1 and the interaction observed in Split-YFP experiments was solely due to the remaining ALP1 distributed throughout the nucleus.
Taken together, ALP1 might be a preferentially nucleolar protein. Consequently, no FRET for YFP-ALP1 and CFP-LHP1 would be measured, since LHP1 is preferentially localized outside the nucleolus in Arabidopsis nuclei (Guan et al., 2011). However, a smaller fraction of YFP-ALP1 is present outside of the nucleolus. This fraction might interact with LHP1 to antagonize its function.

5 Discussion

5.1 Selection of Enhancers and Suppressors of *lhp1*

5.1.1 Enhancers of *lhp1* fall into four categories

Three groups of enhancers (subtle, strong and severe) of *lhp1* have been described in this study. A potential fourth group of enhancers involving gene families with redundant function was not covered. It was unlikely that multiple genes of one family were simultaneously mutagenized in the screen, which had a relatively low mutation rate. Strong and severe enhancers were identified and categorized but not further analyzed. Subtle enhancers were selected for detailed study, when no single mutant phenotype was observed after backcrossing. Selecting those enhancers made it more likely that a novel, direct and, at least in case of ELP1, a specific relationship to LHP1 would be unraveled.

Strong enhancers were attributed to a different class than subtle enhancers. In that group, two enhancers with mutations in *CLF*, one with a mutation in *EMF1* and one with a mutation in *TFL1* were found by Sanger sequencing. Although, the mutation in *EMF1* was not causal for the phenotype of *enh18*, results confirmed that putants with mutations in PcG genes were among the selected candidates in conditions applied in the screen. In fact, an amino acid change in *enh12* was caused by a non-segregating mutation residing within the SET domain of CLF (Goodrich et al., 1997; Kim et al., 1998). This domain is particularly important for the methyl-transferase function of the protein (Doyle and Amasino, 2009; Rea et al., 2000), which is essential for the activity of at least the VRN2- and the EMF2-PRC2 in plants (Schmitges et al., 2011). Further, the phenotype of *enh12* resembled that of a *clf*;*lhp1* double mutant, used as a control for the screen. It is likely that e*nh12* represents a new mutant allele of *CLF*.

In conclusion, Sanger sequencing of known PcG genes has proven to be a useful tool for the removal of contaminants (*enh19*: Ws-2 allele of *CLF*) and for their identification (*enh12* and *enh20*: *TFL1* mutant allele). However, in future experiments this technique is likely to become superfluous, since high-throughput sequencing is a competitive alternative. Prices for genome-sequencing have significantly dropped during the time of this study. Thus, even if some candidates contained causal mutations in previously described genes (1 out of 20 in this study) it would be more time-saving and cost-effective to re-sequence bar-coded DNA-pools than to analyze each candidate individually.

Apart from *CLF* no other novel PcG alleles have been found among tested putants. Notably, M_1 EMS mutants in Arabidopsis are chimeric, meaning that every 7^{th} M_2 plant displays a recessive mutant phenotype (Page and Grossniklaus, 2002). In this study, only 10 seeds were sown of each M_1 seed bulk in the M_2. Probably, there are additional enhancers of *lhp1* yet to be discovered within the screen, of which about 55 % of the M_1 seed bulks were tested. Another explanation to why not many of the known enhancers of *lhp1* have been found is that the original SD screening conditions were biased for the discovery of subtle enhancers, since these were more prone to survive. Last, the low mutation rate of the screen influenced the number of selected putants (see below).

Two severe enhancers of *lhp1* could be retrieved (Figure 3). Remarkably, *enhs1* resembled *clf*;*swn* double mutants (Chanvivattana et al., 2004; Farrona et al., 2011). The double mutant of the PRC1 subunits *AtBMI1A* and *AtBMI1B* display a callus-like structure related to *clf*;*swn* (Chen et al., 2010). These are two of the most severe phenotypes of PRC1 and PRC2 mutant plants that are still viable. Such a severe effect of a mutation in combination with *lhp1* has not been reported previously.

In Drosophila, not all PRC1 components act on the same target genes and the picture starts to emerge that the SCE module, responsible for H2A-K119ub catalysis, might not always be required for the repressive function of the complex (Gutierrez et al., 2012). Given that not all target genes of the Arabidopsis PRC1 components are congruent, a model with multiple active PRCs1 with partially overlapping functions is a likely scenario in plants (Xu and Shen, 2008). If *enhs1* turned out to be a single locus functioning in concert with LHP1, it could be a further indication to a multiplicity of PRCs1. It is thinkable that *ENHs1* encodes a protein which is also able to bind H3K27me3 to anchor PRC1 to target loci. A double knock-out of both anchors would then resemble *Atbmi1a*;*Atbmi1b* double mutants. Alternatively, LHP1 and other PRC1 components might fulfill functions independently of PRC1.

Because of the functional redundancy of AtRING1A and AtRING1B only the triple PRC1 mutant *lhp1*;*Atring1a*;*Atring1b* displays an enhanced *lhp1* single mutant phenotype (Figure 2), (Xu and Shen, 2008; Figure 2). On the same line, *AtBMI1A* is shown to be redundant in function with *AtBMI1B* (Chen et al., 2010), with each mutant displaying a normal growth phenotype. It is likely that other unknown PRC1 components exist in plants that could be organized in gene families with redundant function.

The selection of subtle enhancers of *lhp1* was demonstrated to be successful in finding at least one protein that directly interacts with PRC1 in transient assays. If ELP1 is truly a complex binding partner remains to be confirmed. First, the interaction of ELP1 with LHP1 and ICU2

will be repeated in transient assays. Second, ELP1 interaction to other PRC1 components will be tested in yeast. Finally, transgenic ELP1::RFP lines under the control of the ELP1 promoter have been created to confirm the nuclear ELP1 localization *in vivo* and eventually attempt pull-down or bimolecular fluorescence complementation assays using 35S:LHP1::GFP lines. In conclusion, ELP1 serves as a blueprint of an analysis pipeline for enhancer candidates consisting of four main steps – selection, sequencing, confirmation and functional analysis.

5.1.2 Flowering time of selected enhancers and *FT* regulation

LHP1 has a direct effect on flowering on multiple levels. Early flowering of the *lhp1* mutant is achieved through the up-regulation of *FT* in concert with the up-regulation of multiple MADS domain transcription factors (Adrian et al., 2010; Germann et al., 2006; Kotake et al., 2003). The five selected candidates for re-sequencing did neither have a strong effect on *FT* nor on *FLC* expression (Figure 5), which is also directly regulated by LHP1 (Mylne et al., 2006). However, *lhp1*;*alp1* and *lhp1*;*elp1* plants were demonstrated to flower later and earlier than *lhp1* plants respectively (Figure 6). The other three candidates selected for re-sequencing did display a flowering phenotype, which was not quantified in this study. Hence, all five candidates were pre-selected for an effect on flowering that was not correlated with strongly altered *FT* or *FLC* expression.

Many gene products acting on *FLC* repression or activation have been described (reviewed in He, 2009). Mutations in repressors of *FLC* lead to elevated *FLC* levels accompanied by changes in chromatin at the *FLC* locus and to late flowering plants (Crevillen and Dean, 2011). Especially for suppressors of the *lhp1* phenotype it has been useful to quantify the expression of *FT* and *FLC* in the double mutants. Plants with very high *FLC* levels and consequently mostly low *FT* levels were likely to contain a mutation in a gene participating in the autonomous pathway (Morel et al., 2009; Pazhouhandeh et al., 2011). There was a high probability that those genes were identified in the past 50 years of screening for flowering time mutants (Koornneef et al., 1998; Koornneef et al., 1991; Redei, 1962).

Not all flowering pathways require *FT* as a mobile signal to promote flowering at the SAM. GA induces flowering in SD conditions and at least in roots is also associated with LHP1, which binds SCR (Cui and Benfey, 2009; Wilson et al., 1992). Other emerging flowering pathways are less well described, such as altered flowering time through aging or through ambient temperature changes (Jung et al., 2007; Lee et al., 2007). Hence, some selected

Class Ia putants might be part of another flowering pathway without having a direct effect on LHP1 or MADS domain transcription factors.

5.1.3 Effect of ambient temperature changes

Changes in ambient temperature have an effect on flowering time through alterations in *SVP* and *FT* expression (Blazquez et al., 2003; Lee et al., 2007). Plants growing at lower temperature have a prolonged vegetative phase, lower *FT* expression and consequently flower later. The expression of PcGs is likely not affected by temperatures ranging from 12 °C to 27 °C (NASCarrays-147). Thus, the thermosensory pathway influences flowering independently of PcG mediated-repression.

All candidates of this screen were grown and selected at low temperature conditions (Supplemental Figure 1). In these conditions, the early flowering phenotype of *lhp1* is suppressed and mutants flower with about 20 instead of 10 leaves (compare Figure 6I to Figure 13D). A direct effect of temperature on LHP1 can be ruled out, since *LHP1* expression is not temperature sensitive and the *lhp1* mutant used was probably a null mutant.

Selected double mutants with a mutated temperature-sensitive locus would display the same phenotype independent of temperature changes. Such a phenotype was not observed for any of the re-sequenced candidates when they were grown in LD or SD conditions. Taken together, low temperature conditions acted as a suppressor of *lhp1*, which reduced the severity of enhancer candidates. It was important to dissect a temperature effect from a PcG effect by comparing each candidate phenotype in different conditions.

5.2 Isogenic bulked mapping-by-sequencing

5.2.1 Advantages of isogenic mapping-by-sequencing over conventional approaches

Conventional genetic mapping requires outcrossing to a diverged accession to establish a mapping population. Recognition of the selected mutant in the F_2 of the mapping population is a key point for successful identification of a mutation. Differences in phenotypes between Arabidopsis accessions, such as between Col-0 and Ws-2, are likely to mask phenotypes caused by EMS mutations, especially when subtle enhancers or suppressors are selected (Supplemental Figure 3). Given the pleiotropic nature of the *lhp1* mutant plant (Kotake et al.,

2003) and the limited success in crossing enhancers to *lhp1* Ws-2, a new mapping approach was developed in this study.

Firstly, the phenotype divergence problem has been by-passed by backcrossing double mutant plants to their single mutant parents *e.g*. an isogenic mapping population was generated in the original genetic background. Consequently, conventional markers are absent in the population and cannot be used to distinguish parental alleles. Secondly, a bulk of F_2BC_2 individuals was subjected to genome sequencing. The mutagen-induced changes found in the assembled data sets were then used as markers, as these were specific to the mutant genome and absent in the original. Finally, the SHOREmap platform was used to assemble generated reads and search for causal SNPs (Schneeberger et al., 2009b).

Similar to SHOREmapping, isogenic mapping-by-sequencing is much less labor-intensive and faster than classical mapping approaches. After sequencing, data analysis to produce putative candidate genes took only one day, using automated pipelines (http://Shoremap.org). According to previous work, the number of pooled plants could have been reduced for all sequenced candidate DNA pools (Ashelford et al., 2011). Though, as double mutant phenotypes were easily scored in the isogenic crosses, reducing the number of plants was not a major requirement in this study.

Considering only SNPs with a SHORE score greater than 25, an average of one mutation per 915 kb was found in all five re-sequencing experiments of this study. This is considerably less compared to similar studies describing EMS loads of one mutation per 112-170 kb (Ashelford et al., 2011; Jander et al., 2003) but varying largely among the samples (1 high-quality SNP per 249 kb to 1920 kb). The low mutation rate has helped reducing the number of candidate genes to four at most for each candidate respectively. Possibly, even more important is the depth of sequencing with respect to the success of isogenic mapping. A high fold-coverage has the obvious advantage to distinguish homozygous from near-to-homozygous changes, but comes with additional sequencing costs. It is possible that an increased sequencing depth could have excluded some of the candidates directly. At least in case of *lhp1*;*alp1* segregation of both candidate alleles was observed in the analysis of individuals and thus could have been found by deeper sequencing as well (Figure 13). In addition, there is no need to generate sequencing data in a single run, as seen in three candidates sequenced as bar-coded samples (Table 11). With advancing sequencing technology it would be most cost- and labor-effective to bar-code a high number of pooled candidate DNA samples for sequencing, taking a loss of sequencing depth into account. To find the causal change, several PCR fragments containing high-quality SNPs of one sample could be amplified using the pooled DNA. Those PCR

fragments would finally be sequenced deeply, for example in an Ion Torrent sequencer (Rothberg et al., 2011), to gain a frequency distribution for the SNPs of a candidate pool. The causal EMS change would probably be the one supported by the highest frequency of reads. In practical terms, it is better to start the approach by targeting a relatively low coverage that is only increased if required (Abe et al., 2012). However, the number of bulked plants for each DNA pool has to be high enough and has to be adapted to the type of re-sequencing experiment.

The low mutation rate of this study is a controversial issue, since it might harbor drawbacks in terms of saturation and number of mutants in an EMS screen. Thus, the availability of a second allele with the same phenotype in the same screen proved to be a big advantage in confirming the re-sequencing results for *lhp1*;*alp1* (Figure 13B).

5.2.2 Enhancers of LHP1 reveal a relation to telomere regulation

After having re-sequenced the genomes of *enh13* and *enh15* evidence was gathered that both double mutants might have a mutation in genes of the same family of *TELOMERE REPEAT BINDING FACTOR*s (*TRB*s). TRBs belong to the Single Myb Histone (SMH) group of proteins in Arabidopsis. This family comprises five paralogous members, TRB1 to TRB5, which are plant specific DNA-binding proteins (Byun et al., 2008). Three SMH-like proteins were described in Arabidopsis (TRB1-3) as well as in rice and five are known in maize (Byun et al., 2008; Dvorackova et al., 2010; Marian et al., 2003). TRBs are rather small proteins (30-35 kDa) containing an N-terminal Myb domain, a central H1/H5-like (histone-like) domain and a C-terminal coiled-coil domain.

The Myb domain of TRB proteins is very similar to the telobox of telemore binding proteins in mammals, such as TRF1 and TRF2. Those two proteins are components of the shelterin complex, which is essential for the interaction with telomeric DNA (Hofr et al., 2009). Previously it had been shown that TRB1 to TRB3 can interact with POT1b through their H1/H5-like domain in Arabidopsis. POT1b is a homologue of the telomeric G overhang binding proteins (Schrumpfova et al., 2008). The central domain was also shown to be involved in unspecific DNA binding and formation of homo- and heterodimeric TRB complexes.

TRB1 to TRB3 are preferentially localized in the nucleoli of cells (Dvorackova et al., 2010). The central domain of TRB1 was demonstrated to be sufficient for the nucleolar localization of the protein with contribution of the Myb domain. Using Fish experiments, a GFP antibody

and a telomere marker co-localization of TRB1 and telomeres were found in about 11 % of tested nuclei. The low overlap could be explained by the high mobility of TRB proteins which were able to shuttle rapidly between the nucleus and the nucleolus (Dvorackova et al., 2010). During mitosis most of TRB1 was depleted from DNA, but a small fraction was still binding. Already in the anaphase some of the protein re-associated with chromatin, which made it a very early marker for nucleolar re-assembly. This question, why a protein with DNA binding function was localized mostly in the nucleolus, a compartment associated with RNA processing, ribosome biogenesis and telomere regulation has to be answered yet (Degenhardt and Bonham-Smith, 2008; Dvorackova et al., 2010). LHP1 could possibly influence a telomere binding complex in plants such as the six-protein complex shelterin in mammals or the CST (Cdc13, Stn1 and Ten1) complex in budding yeast (Palm and de Lange, 2008; Price et al., 2010; Watson and Riha, 2010).

Conserved Telomere maintenance Component 1 (CTC1) is likely to be a central component of a CST analogous complex in humans and in Arabidopsis. Interestingly, the C-terminus of ICU2 (DNA Pol alpha subunit) binds CTC1 *in vitro* (Barrero et al., 2007). It is yet unclear how CST would be recruited to telomeres in Arabidopsis and it is speculated that the complex also fulfills non-telomeric functions. It is tempting to speculate that a multi-protein complex containing TRB1 and TRB3 would be involved in telomere control. Further, an epigenetic layer through LHP1 or PRC1 might aid the regulation of C-strand fill in and cell cycle control at chromosome ends. ELP1 might function as an interaction hub between DNA Polα, LHP1 and TRBs. Given that only three enhancers of LHP1 have been sequenced so far, this study provides a fruitful ground towards understanding the role of PRC1 in plants.

5.3 *ALP1* and *ELP1* – two sides of *LHP1*

5.3.1 ALP1 is a domesticated transposase-like gene

On the DNA level, type II (dsDNA) transposases are the most abundant and possibly most essential elements for evolution in viral, bacterial and eukaryotic genomes (Aziz et al., 2010). They can fulfill vital functions for an organism, such as DNA processing (Landweber et al., 2009). The phenomena of domesticated transposases or transposase-derived genes have been described for plant development (Bundock and Hooykaas, 2005) and light signaling (Lin, 2007), where the transposase-like proteins FAR-RED-ELONGATED HYPOCOTYL 3

(FHY3) and FAR-RED-IMPAIRED RESPONSE 1 (FAR1) are shown to be transcription factors to integrate light signals to the circadian clock.

Of the ten different superfamilies for class II elements only five, namely *CACTA*, *hAT*, *Mutator*, *PIF* and *Tc1*/*Mariner*, are present in plants (Benjak et al., 2008). *Harbinger* and *Harbinger*-like transposases have been placed into the P instability factor (*PIF*) family of "cut-and-paste" transposons (Kapitonov and Jurka, 2004). *PIF*-like transposons have been domesticated several times in Drosophila and in mammals (Casola et al., 2007; Kapitonov and Jurka, 2004).

The mammalian *HARBI1* gene is present in several organisms, which do not contain any active *Harbinger* transposases. It is speculated that inactive *Harbinger*-like elements may protect the human genome from active *Harbinger* transposases. If this mechanism exists it is not present in plants where *Harbinger* transposons and *Harbinger*-like elements coexist (Kapitonov and Jurka, 2004). Of the seven close homologues to the *Harbinger*-like element At3g63270 in Arabidopsis four clustered together with active, bacterial transposases in a Neighbour Joining analysis. Two of the others, ALP1 and At3g55350, were present in distinct clades of the tree. These sub-clades contained only plant proteins from other species, supporting the idea of coexistence between active and inactive *Harbingers* in Arabidopsis (Figure 14 sub-clades II and VI).

The catalytic acidic triad "DDE" that was found to be disrupted in ALP1 is characteristic of the DDE transposase / integrase supergroup and is essential in coordinating metal ions involved in the "cut and paste" mechanism (Casola et al., 2007; Craig, 2002). The fact that important amino acid residues were not conserved in ALP1, but in At3g55350, further supported the finding that the protein does no longer fulfill a transposase function (Figure 15). ALP1 might be able to bind DNA through one or two helix-turn-helix motifs towards the N-terminal end of the protein (Iwahara et al., 1998). This function could well be exclusive for ALP1, since all of the Arabidopsis homologues did not have the same domain structure in a CDD search. However, all other proteins in sub-clade VI were predicted to also contain helix-turn-helix motifs. It would be interesting to confirm a DNA binding function of the *Harbinger*-like transposase and explore a possible link to the repression of target genes by LHP1.

Combining results from qRT-PCR, phylogenetic analysis and the alignment with other *Harbinger*-like transposases *ALP1* was demonstrated to be an actively transcribed gene. In fact, the 5'-end of the gene is decorated with H3K4me3 and the gene body 3'-end with H3K36me2 as it is expected for active transcription (Figure 18). It is likely to have lost its

ability to transpose and might have been co-opted by the plant host to suppress the Polycomb pathway of gene silencing.

5.3.2 Localization and binding properties of ALP1

In transient assays ALP1 protein localized exclusively to nuclei of tobacco leaf cells (Figure 20 and Figure 21). The likely nuclear localization of ALP1 goes with its potential function in DNA binding. Further, ALP1 was demonstrated to bind LHP1 repetitively in transient assays (Figure 20). However, the physical interaction between ALP1 and LHP1 could not be demonstrated in Acceptor Photo Bleaching assays. In most nuclei YFP-tagged ALP1 formed a rigid, crystal-like structure that did not fit the distribution pattern in Split-YFP experiments (Figure 21).

Possibly, the YFP tag altered the structure of ALP1 and might have changed its binding properties. In some nuclei a partial, stronger fluorescence of CFP-LHP1 protein was observed after bleaching of YFP-ALP1 (Figure 21D and 21F). Therefore, LHP1 did either only interact with ALP1 in parts of the nucleus, or not at all. In the first case, the increase in CFP fluorescence measured in some nuclei could have been due to shuttling of CFP-LHP1 to the nucleolus after bleaching of YFP-ALP1 (Figure 21D and 21F). However, normally LHP1 is not localized at the nucleolus in Arabidopsis (Guan et al., 2011; Kotake et al., 2003).

Alternatively, the interaction observed in split-YFP assays was only due to a smaller fraction of ALP1 that was distributed throughout the nucleus and overlapped with the localization of LHP1. To test this hypothesis *in vivo*, transgenic lines with ALP1 fused to RFP and GFP are being created. Additionally, it would be interesting to compare ALP1 and ALP1mu (mutated ALP1 protein / transcribed and translated *alp1-1*) localization in *lhp1* and WT plants, to elucidate if LHP1 has influence on ALP1 distribution or potential shuttling between nuclear compartments. Further, ALP1mu will be used in split-YFP assays together with LHP1, to test if the binding properties are affected by the mutation.

5.3.3 ALP1 does not alter selected histone marks globally

ALP1 is not covered by H3K27me3 and consequently not a target of LHP1 (Figure 18; Zhang et al., 2007). The expression of *ALP1* was not influenced in *lhp1* mutants compared to Col-0 (Figure 18D). Hence, *ALP1* was found to be an actively transcribed gene that is not directly influenced by LHP1. If ALP1 functions as an antagonist to the PcG pathway, it is of interest to clarify if ALP1 targets primarily PRC1 function or also affects the function of PRC2.

H3K27me3 is globally less abundant in *clf* mutant plants but remains unaltered in *lhp1* mutants (Jiang et al., 2008; Lafos et al., 2011; Turck et al., 2007). The effect of other PRC1 mutations on global H3K27me3 levels has not been elucidated. An increase of H3K27me3 in *lhp1*;*alp1* and eventually *clf*;*alp1* seedlings could indicate that ALP1 acts at the level of PRC2. However, lower H3K27me3 levels in *clf* mutants were not observed in this study (Figure 11), which argues against the suitability of the assay. Compared to published results, the nature of different *clf* alleles used in this study possibly influences the impact on H3K27me3 abundance. Alternatively, the freeze-thawing method for protein extraction, applied in this study, had an influence on the composition of nucleosomes. Last, the developmental stage of the plant, ambient temperature and culture conditions in general could explain the results, which have not been reproduced using changed parameters. A difference in the abundance of H3K4me3, as a mark associated with active transcription, was not found in *lhp1*;*alp1*, despite its large influence on the transcriptome (Figure 11 and Figure 7).

When the de-regulated genes in *lhp1*;*alp1* were analyzed, a more genes were down-regulated compared to the de-regulated genes in *lhp1* (Figure 10). This would be expected for a transcriptional regulator that normally activates genes. However it is not yet clear if ALP1 is a functional DNA-binding protein that is able to regulate transcription directly. The percentage of genes covered with H3K27me3 in *lhp1*;*alp1* was different for down-regulated genes and nearly the same for up-regulated genes compared to *lhp1* (Figure 10). These results are difficult to interpret, since it is not clear how many, if any, of the de-regulated genes are directly targeted by ALP1. Only 50 % of the de-regulated genes in *lhp1* are PRC1 targets, which illustrates that the secondary effect on transcription can be high for mutants involved in epigenetic regulation. However, a smaller number of up-regulated genes were PRC1 target genes in *lhp1*;*alp1* compared to *lhp1*. ChIP studies are planned to test ALP1 binding of MADS-domain transcription factors and other potential target genes. Additionally, the distribution of H3K27me3 and H3K4me3 on these loci will be elucidated.

5.3.4 A dual function of ALP1 in gene activation

Most genes de-regulated in *lhp1* mutant plants displayed expression patterns similar to the WT state in *lhp1*;*alp1* (Figure 7C). In the mRNA-seq study, MADS domain transcription factors seemed to be up-regulated to the same extend in *lhp1* and *lhp1*;*alp1* samples, but a greater number were classified as outliers with an insufficient number of reads to draw statistically significant conclusions. Unaltered expression between *lhp1* and *lhp1*;*alp1* was

confirmed for *AG*, but disproved for *AP3* and *SEP3* by quantitative real-time PCR (Figure 18). The varying results on transcription of MADS domain genes could also be due to the different developmental stages of the plants e.g. ten-day old seedlings grown in LD conditions and 14 day old seedlings grown in climate chamber conditions. Hence, ALP1 has an effect on the transcription of nearly all commonly de-regulated genes in *lhp1* and *lhp1*;*alp1* mutants, which explains the intermediate phenotype of *lhp1*;*alp1* plants.

Without *lhp1*, *alp1* plants displayed a subtle, late flowering phenotype that was detected in climate chamber growth conditions and not further analyzed (Figure 17C). This effect could be due to temperature affecting *ALP1* directly or enhancing PcG repression. However, also in LD conditions there was a tendency for late flowering (bars in Figure 17B), which argues for a temperature independent effect of ALP1 on flowering that will be further elucidated.

On the same line, a set of genes was exclusively de-regulated in *lhp1*;*alp1* plants. The down-regulated genes in this set showed enrichment for defense responsive genes, such as *GRIM REAPER* (*GRI*), involved in programmed cell death or At4g09420, which encodes a disease resistance protein of the Toll-interleukin receptor nuclear binding signal leucine-rich repeat (TIR-NBS-LRR) class (Meyers et al., 2002; Wrzaczek et al., 2009). These data indicate a dual function of ALP1. On one hand, ALP1 is binding LHP1 to possibly counterbalance its repressive function on a subset of targets. On the other hand ALP1 activates genes related to defense response.

Preliminary results from a yeast two hybrid screen, to find binding partners of ALP1, strengthens the association of ALP1 as a positive regulator of defense response. The screen might also help to understand the role of ALP1 in PcG repression, especially when the interaction with LHP1 can be confirmed (Liang and Hartwig in preparation).

If ALP1 associates with PRC1 *in vivo* and has retained its DNA binding function it is tempting to speculate that ALP1 is involved in the propagation of histone marks associated with the activation of transcription. Alternatively, ALP1 acts in the repression of H3K27me3, binds to transcription factors, or is a transcription factor itself. An example for a transcription factor that is binding LHP1 *in vivo* and also has a DNA-binding activity is SVP. This MADS box transcription factor is involved in the repression of *SEP3* in combination with LHP1 (Liu et al., 2009). An attractive model would therefore include PRC1 as the yin and ALP1 as the yang in PRC1 mediated gene repression for a subset of target genes (Figure 22B).

In this model ALP1 would bind DNA as well as LHP1 of PRC1 to prevent binding of H3K27me3 and subsequent H2A mono-ubiquitylation. Recently it has been demonstrated that the demethylase RELATIVE OF EARLY FLOWERING 6 (REF6) actively removes

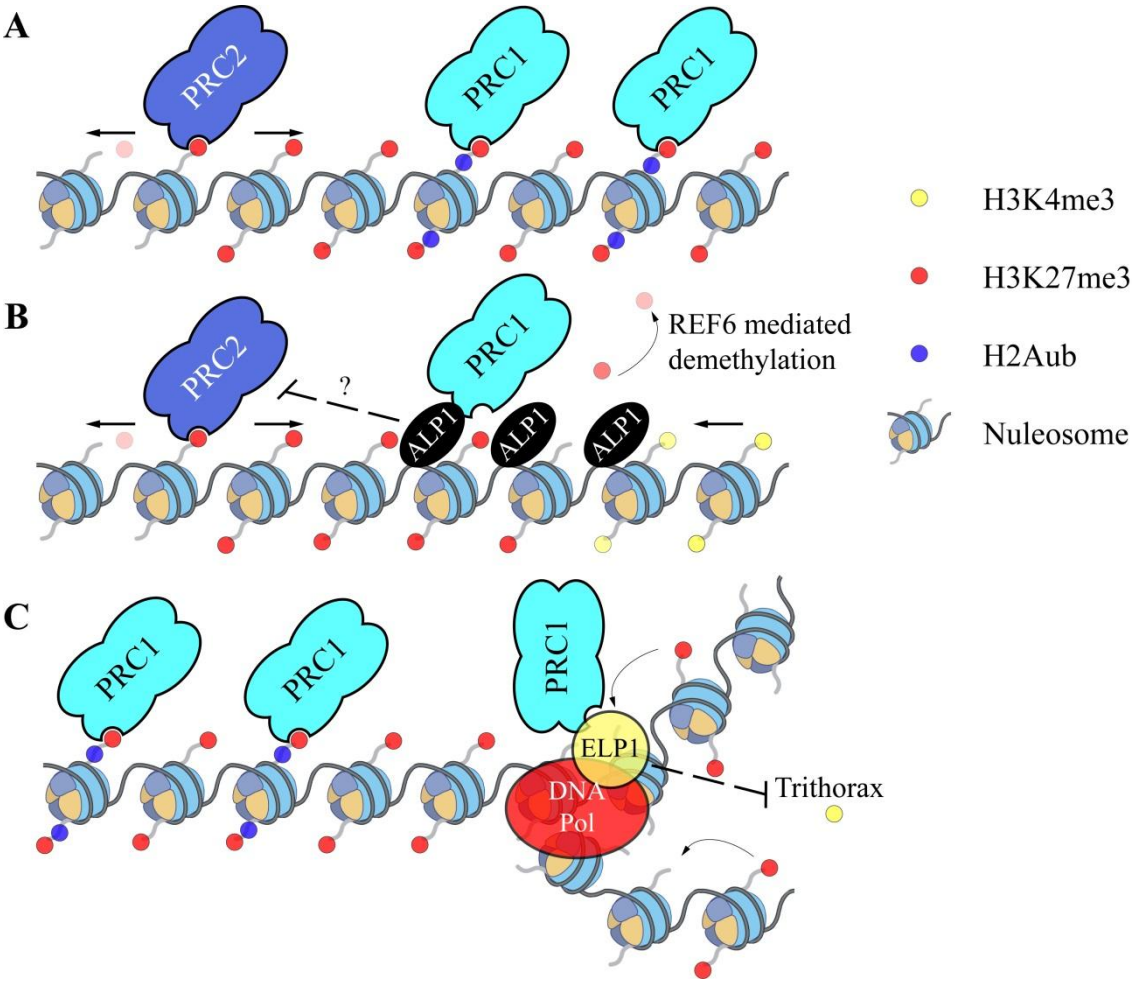


Figure 22: Model for the function of ALP1 and ELP1
(**A**) Simplified model for Polycomb mediated gene expression. PRC2 sets the mark which spreads by an unknown mechanism and is subsequently bound by PRC1 which mediates H2A mono-ubiquitylation. (**B**) Working model for ALP1 mediated repression of PRC1 and possibly PRC2. (**C**) Working model for ELP1 mediated enhancement of PRC1 repression possibly by connecting PRC1 to DNA-dependent DNA Polymerases. Figure in parts adapted from (Margueron and Reinberg, 2011) and (Beisel and Paro, 2011).

H3K27me3 in Arabidopsis (Lu et al., 2011). This could pave the way for the initiation and spreading of active histone marks such as H3K4me3 (Figure 22A and 22B).

5.3.5 *ELP1* is neither regulated by LHP1 nor by its neighboring gene

The *ELP1* locus is not decorated with H3K27me3 and therefore not likely to be regulated directly by LHP1 (Figure 18A) (Turck et al., 2007; Zhang et al., 2007). The expression of *ELP1* in *lhp1*;*elp1* was about 60 % lower than in WT and *lhp1* plants (Figure 18C). In *lhp1*;*elp1* ELP1 is truncated and only 75 amino acids long. Quantitative real-time primers were located downstream of the premature stop codon of the truncated protein. However, a small amount of transcript was still detectable in *lhp1*;*elp1* which might mean that *elp1-1* is

not a null allele. However, early transcripts of the gene are likely to be subjected to the nonsense-mediated decay pathway for degradation (Chang et al., 2007).

ELP1 lies in close proximity to At3g42670. The 3'-ends of the genes are facing each other and might share regulatory elements (Figure 18A). At3g42670 encodes a nuclear protein named CHROMATIN REMODELING 38 or CLASSY1 (CLSY1). It is an ATP-dependent DNA binding protein that resides in the nucleus but not in the nucleolus with a similar distribution pattern as LHP1 (Smith et al., 2007). *CLSY1* expression was measured to rule out that the SNP in *ELP1* had a secondary effect on *CLSY1*. Expression of *CLSY1* was not different in *lhp1*;*elp1*, *lhp1* or WT (Figure 18C).

5.3.6 ELP1 is an interaction hub binding LHP1 in the nucleus

ELP1 encodes a protein with a WD40 domain towards the N-terminal end. Those proteins are prominent for their vast number of protein-protein interactions and *ELP1* has been predicted to interact with a high number of nuclear proteins (Geisler-Lee et al., 2007). Interestingly, the group of MSI proteins, of which MSI1 resides within PRC2, also contains WD40 domains (Hennig et al., 2005). Especially MSI4 is connected to other complexes and developmental processes apart from flowering (Morel et al., 2009; Pazhouhandeh et al., 2011). These findings allow speculations on ELP1 being associated with or part of the PRC1 complex. Four results of this study support this hypothesis.

Firstly, ELP1 protein resides within the nuclei of Arabidopsis cells and interacts with LHP1 in transient expression assays (Figure 19A, Figure 20C and 20D). Secondly, commonly de-regulated genes in *lhp1* and *lhp1*;*elp1* plants are mis-regulated in the same way (Figure 7B and 7F). Thirdly, MADS domain transcription factors are up-regulated in *elp1* and *lhp1*;*elp1*. The up-regulation in *lhp1*;*elp1* is higher than in *lhp1* alone which favors an additive effect of both mutations (Figure 18). Hence, ELP1 does act neither up- nor down-stream of LHP1 at least in the regulation of *SEP3* and *AG*, which might also account for the earlier flowering phenotype of *lhp1*;*elp1* compared to *lhp1* (Figure 6I). Finally, more genes are highly up-regulated in *lhp1*;*elp1* compared to *lhp1*. The percentage of up-regulated PRC1 targets remains the same for *lhp1*;*elp1* compared to *lhp1*, meaning that an overall higher number of PRC1 targets were found among that group of genes (Figure 10).

Before ELP1 can be placed in PRC1 the interaction with LHP1 has to be experimentally verified *in vivo* and association with other PRC1 components has to be clarified. Studying H3K27me3 and H3K4me3 levels on potential ELP1 targets and known PRC1 targets will

provide further details for the functional dissection of ELP1. With its potentially high number of interactions to other nuclear proteins ELP1 would provide numerous novel insights into PRC1 regulation.

5.3.7 Interaction with ICU2 and influence on chromatin remodeling

The interaction between ICU2 and ELP1 was predicted *in silico* (Interactions Viewer / Geisler-Lee et al., 2007) and confirmed in this study by Split-YFP assays (Figure 20G). It was demonstrated that ICU2 is also binding LHP1 in vitro, which could not be confirmed here using transient assays (Figure 20H). The interaction between ELP1 and ICU2 was comparatively weak. Fewer fluorescing cells that were also closer to the point of bacterial injection were found compared to other positive interactions. Thus, the ICU2 protein could have been altered by the YFP tag. It remains to be shown if ICU2 and ELP1 interact *in vivo*. ICU2 has been suggested to be involved in mitotic memory and inheritance of repressive histone marks. Hence, ELP1 could fulfill the function of an interaction hub between DNA Polymerase α, PRC1 and PRC2 (Barrero et al., 2007 and Figure D2 C).
In this model DNA strands are unraveled from histones and their marks to be replicated. Later the marks are distributed evenly to daughter, like in human cells, or are re-established by PRC2, as demonstrated in yeast (reviewed in Beisel and Paro, 2011). To re-establish H2Aub, PRC1 would have to be re-introduced to each daughter strand. Fine-tuning of the re-introduction process would increase the flexibility of repression. Interactive proteins, such as ELP1, present a valuable platform for the implementation for such a process (Figure 22C).
Global H3K4me3 levels were slightly elevated in *lhp1*;*elp1* and *swn* mutants (Figure 11). This effect has been observed previously for *clf-28*;*swn-7* double mutant plants (Lafos et al., 2011). H3K4me3 is placed at histones by the catalytic activity of the Trithorax complex in Arabidopsis (Tamada et al., 2009). It has been shown that at least at bivalent loci *e.g.* genes that are decorated by active and repressive histone marks, such as *FLC* one mark is dominant, when activators for the other are missing (Carles and Fletcher, 2009; Jiang et al., 2008). Hence, in a model ELP1 as well as SWN might be involved in the repression of either the initiation or the spreading of H3K4me3 (Figure 22C).

6 Conclusions and Perspectives

This study is one of the first to apply next-generation-sequencing for the identification of EMS-induced mutants in Arabidopsis. A robust pipeline has been established for the fast identification of enhancers and suppressors of LHP1. The selection process for successful mapping-by-sequencing has so far taken the largest amount of time, since candidates were backcrossed twice, the number of pooled plants was comparably large and molecular methods were applied to ensure mapping of novel loci.

In future experiments the bar-coding technology and more powerful in-depth sequencing will lead to further improvement of the pipeline. It is even thinkable that a number of bar-coded individual plants will be sequenced. This would lead to an EMS profile for each candidate which withholds valuable information about all homozygous EMS-induced changes in each plant. These profiles will be particularly important for Class Ib putants that contain multiple mutated loci to rule out secondary effects and understand each mutations contribution to the mutant phenotype.

ALP1 was the first mapped candidate of this study. As a repressor of LHP1, ALP1 could also be associated with TrxG proteins (Figure 22). Trithorax complexes are functioning as epigenetic transcriptional activators, for instance, by setting H3K4me3 at target loci. Excellent work on understanding those complexes has been published over recent years in Arabidopsis (Carles and Fletcher, 2009; Schuettengruber et al., 2011; Tamada et al., 2009). However, since ChIP experiments of ALP1 lines or H3K4me3 on potential ALP1 targets have not been performed, there is no direct link between ALP1 and TrxG proteins yet. Hence, they have not been discussed in detail in this study, but will be considered in future experiments. ELP1 could be an important interactive protein in explaining a likely multiplicity or the multiple functions of PRCs1 and their components in Arabidopsis. Since, repression of loci is thought to be a flexible state, interaction hubs could be key for the re-localization or targeting of PRCs.

Notably, *alp1* and WT plants have been transformed with a 35S:AG construct and *lhp1*;*elp1* has been crossed to *icu2* and various other PcG mutant plants. The study of the resulting phenotypes may shed light on the interplay between ALP1, ELP1 and PRCs1. Placing ALP1 and ELP1 into the Polycomb pathway will be a crucial step to gain additional insight on their functions.

6 Bibliography

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Mitsuoka, C., Tamiru, M., Innan, H., *et al.* (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol.

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.

Adrian, J., Torti, S., and Turck, F. (2009). From decision to commitment: the molecular memory of flowering. Mol Plant *2*, 628-642.

Ahmad, A., Zhang, Y., and Cao, X.F. (2010). Decoding the epigenetic language of plant development. Mol Plant *3*, 719-728.

Alonso, A.G.D., Gutierrez, L., Fritsch, C., Papp, B., Beuchle, D., and Muller, J. (2007). A genetic screen identifies novel polycomb group genes in Drosophila. Genetics *176*, 2099-2108.

Ashelford, K., Eriksson, M.E., Allen, C.M., D'Amore, R., Johansson, M., Gould, P., Kay, S., Millar, A.J., Hall, N., and Hall, A. (2011). Full genome re-sequencing reveals a novel circadian clock mutation in Arabidopsis. Genome biology *12*.

Aubert, D., Chen, L.J., 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.

Austin, R.S., Vidaurre, D., Stamatiou, G., Breit, R., Provart, N.J., Bonetta, D., Zhang, J.F., Fung, P., Gong, Y.C., Wang, P.W., *et al.* (2011). Next-generation mapping of Arabidopsis genes. Plant Journal *67*, 715-725.

Aziz, R.K., Breitbart, M., and Edwards, R.A. (2010). Transposases are the most abundant, most ubiquitous genes in nature. Nucleic acids research *38*, 4207-4217.

Bantignies, F., and Cavalli, G. (2011). Polycomb group proteins: repression in 3D. Trends Genet *27*, 454-464.

Barrero, J.M., Gonzalez-Bayon, R., del Pozo, J.C., Ponce, M.R., and Micol, J.L. (2007). INCURVATA2 encodes the catalytic subunit of DNA polymerase alpha and interacts with genes involved in chromatin-mediated cellular memory in Arabidopsis thaliana. Plant Cell *19*, 2822-2838.

Bedford, M.T., and Clarke, S.G. (2009). Protein Arginine Methylation in Mammals: Who, What, and Why. Molecular Cell *33*, 1-13.

Beisel, C., and Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. Nature Reviews Genetics *12*, 123-135.

Benjak, A., Forneck, A., and Casacuberta, J.M. (2008). Genome-Wide Analysis of the "Cut-and-Paste" Transposons of Grapevine. PLoS ONE *3*.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met *57*, 289-300.

Berger, N., Dubreucq, B., Roudier, F., Dubos, C., and Lepiniec, L. (2011). Transcriptional Regulation of Arabidopsis LEAFY COTYLEDON2 Involves RLE, a cis-Element That Regulates Trimethylation of Histone H3 at Lysine-27. The Plant cell.

Berger, S.L. (2007). The complex language of chromatin regulation during transcription. Nature *447*, 407-412.

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol *26*, 2560-2569.

Blazquez, M.A., Ahn, J.H., and Weigel, D. (2003). A thermosensory pathway controlling flowering time in Arabidopsis thaliana. Nat Genet *33*, 168-171.

Boss, P.K., Bastow, R.M., Mylne, J.S., and Dean, C. (2004). Multiple pathways in the decision to flower: Enabling, promoting, and resetting. Plant Cell *16*, 18-31.

Bratzel, F., Lopez-Torrejon, 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.

Bundock, P., and Hooykaas, P. (2005). An Arabidopsis hAT-like transposase is essential for plant development. Nature *436*, 282-284.

Buzas, D.M., Tamada, Y., and Kurata, T. (2011). FLC: A Hidden Polycomb Response Element Shows Up in Silence. Plant & cell physiology.

Byun, M.Y., Hong, J.P., and Kim, W.T. (2008). Identification and characterization of three telomere repeat-binding factors in rice. Biochem Bioph Res Co *372*, 85-90.

Calonje, M., Sanchez, R., Chen, L.J., and Sung, Z.R. (2008). EMBRYONIC FLOWER1 participates in Polycomb group-mediated AG gene silencing in Arabidopsis. Plant Cell *20*, 277-291.

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. Genes Dev *23*, 2723-2728.

Casola, C., Lawing, A.M., Betran, E., and Feschotte, C. (2007). PIF-like Transposons are common in Drosophila and have been repeatedly domesticated to generate new host genes. Mol Biol Evol *24*, 1872-1888.

Chang, Y.F., Imam, J.S., and Wilkinson, M.E. (2007). The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem *76*, 51-74.

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, D., Molitor, A., Liu, C., and Shen, W.H. (2010). The Arabidopsis PRC1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth. Cell research *20*, 1332-1344.

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

Corbesier, L., Vincent, C., Jang, S.H., Fornara, F., Fan, Q.Z., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., *et al.* (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science (New York, NY *316*, 1030-1033.

Craig, N.L. (2002). Mobile DNA II (Washington, D.C., ASM Press).

Crevillen, P., and Dean, C. (2011). Regulation of the floral repressor gene FLC: the complexity of transcription in a chromatin context. Current opinion in plant biology *14*, 38-44.

Cui, H.C., and Benfey, P.N. (2009). Interplay between SCARECROW, GA and LIKE HETEROCHROMATIN PROTEIN 1 in ground tissue patterning in the Arabidopsis root. Plant Journal *58*, 1016-1027.

Cuperus, J.T., Montgomery, T.A., Fahlgren, N., Burke, R.T., Townsend, T., Sullivan, C.M., and Carrington, J.C. (2010). Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing. Proceedings of the National Academy of Sciences of the United States of America *107*, 466-471.

Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-Wide Kinetics of Nucleosome Turnover Determined by Metabolic Labeling of Histones. Science (New York, NY *328*, 1161-1164.

Degenhardt, R.F., and Bonham-Smith, P.C. (2008). Arabidopsis ribosomal proteins RPL23aA and RPL23aB are differentially targeted to the nucleolus and are disparately required for normal development. Plant physiology *147*, 128-142.

del Olmo, I., Lopez-Gonzalez, L., Martin-Trillo, M.M., Martinez-Zapater, J.M., Pineiro, M., and Jarillo, J.A. (2010). EARLY IN SHORT DAYS 7 (ESD7) encodes the catalytic subunit of DNA polymerase epsilon and is required for flowering repression through a mechanism involving epigenetic gene silencing. Plant Journal *61*, 623-636.

Doyle, M.R., and Amasino, R.M. (2009). A Single Amino Acid Change in the Enhancer of Zeste Ortholog CURLY LEAF Results in Vernalization-Independent, Rapid Flowering in Arabidopsis. Plant physiology *151*, 1688-1697.

Du, Z., Zhou, X., Ling, Y., Zhang, Z.H., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. Nucleic acids research *38*, W64-W70.

Dvorackova, M., Rossignol, P., Shaw, P.J., Koroleva, O.A., Doonan, J.H., and Fajkus, J. (2010). AtTRB1, a telomeric DNA-binding protein from Arabidopsis, is concentrated in the nucleolus and shows highly dynamic association with chromatin. Plant Journal *61*, 637-649.

Exner, V., Aichinger, E., Shu, H., Wildhaber, T., Alfarano, P., Caflisch, A., Gruissem, W., Kohler, 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.

Farrona, S., Thorpe, F.L., Engelhorn, J., Adrian, J., Dong, X., Sarid-Krebs, L., Goodrich, J., and Turck, F. (2011). Tissue-Specific Expression of FLOWERING LOCUS T in Arabidopsis Is Maintained Independently of Polycomb Group Protein Repression. Plant Cell *23*, 3204-3214.

Felsenstein, J. (1985). Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. Evolution *39*, 783-791.

Ficz, G., Heintzmann, R., and Arndt-Jovin, D.J. (2005). Polycomb group protein complexes exchange rapidly in living Drosophila. Development *132*, 3963-3976.

Fornara, F., de Montaigu, A., and Coupland, G. (2010). SnapShot: Control of Flowering in Arabidopsis. Cell *141*.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science (New York, NY *306*, 1574-1577.

Franklin, K.A. (2010). Plant chromatin feels the heat. Cell *140*, 26-28.

Gaudin, V., Libault, M., Pouteau, S., Juul, T., Zhao, G.C., Lefebvre, D., and Grandjean, O. (2001). Mutations in LIKE HETEROCHROMATIN PROTEIN 1 affect flowering time and plant architecture in Arabidopsis. Development *128*, 4847-4858.

Geisler-Lee, J., O'Toole, N., Ammar, R., Provart, N.J., Millar, A.H., and Geisler, M. (2007). A predicted interactome for Arabidopsis. Plant physiology *145*, 317-329.

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 (New York, NY *297*, 1871-1873.

Germann, S., Juul-Jensen, T., Letarnec, B., and Gaudin, V. (2006). DamID, a new tool for studying plant chromatin profiling in vivo, and its use to identify putative LHP1 target loci. Plant Journal *48*, 153-163.

Gildea, J.J., Lopez, R., and Shearn, A. (2000). A screen for new trithorax group genes identified little imaginal discs, the Drosophila melanogaster homologue of human retinoblastoma binding protein 2. Genetics *156*, 645-663.

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.

Gorfinkiel, N., Fanti, L., Melgar, T., Garcia, E., Pimpinelli, S., Guerrero, I., and Vidal, M. (2004). The Drosophila Polycomb group gene Sex combs extra encodes the ortholog of mammalian Ring1 proteins. Mech Develop *121*, 449-462.

Grewal, S.I.S., and Elgin, S.C.R. (2007). Transcription and RNA interference in the formation of heterochromatin. Nature *447*, 399-406.

Guan, H., Zheng, Z., Grey, P.H., Li, Y., and Oppenheimer, D.G. (2011). Conservation and divergence of plant LHP1 protein sequences and expression patterns in angiosperms and gymnosperms. Mol Genet Genomics *285*, 357-373.

Gutierrez, L., Oktaba, K., Scheuermann, J.C., Gambetta, M.C., Ly-Hartig, N., and Muller, J. (2012). The role of the histone H2A ubiquitinase Sce in Polycomb repression. Development *139*, 117-127.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nature Cell Biology *10*, 1291-U1289.

Hariganeya, N., Kikuchi, A., Mizoguchi, T., and Kamada, H. (2007). Roles of CURLY LEAF (CLF), SWINGER (SWN) and LEAFY COTYLEDON1 (LEC1) in the somatic embryogenesis of Arabidopsis. Plant and Cell Physiology *48*, S51-S51.

Hartwig, B., Velikkakam, G., konrad, K., Schneeberger, K., and Turck, F. (2012). Fast isogenic mapping-by-sequencing of EMS-induced mutant bulks. Plant physiology *Submitted*.

Haughn, G.W., Davin, L., Giblin, M., and Underhill, E.W. (1991). Biochemical Genetics of Plant Secondary Metabolites in Arabidopsis-Thaliana - the Glucosinolates. Plant physiology *97*, 217-226.

He, Y.H. (2009). Control of the Transition to Flowering by Chromatin Modifications. Molecular Plant *2*, 554-564.

Helliwell, C.A., Wood, C.C., Robertson, M., Peacock, W.J., and Dennis, E.S. (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 Journal *46*, 183-192.

Hennig, L., Bouveret, R., and Gruissem, W. (2005). MSI1-like proteins: an escort service for chromatin assembly and remodeling complexes. Trends in Cell Biology *15*, 295-302.

Hennig, L., and Derkacheva, M. (2009). Diversity of Polycomb group complexes in plants: same rules, different players? Trends Genet *25*, 414-423.

Heo, J.B., and Sung, S. (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science (New York, NY *331*, 76-79.

Hofr, C., Sultesova, P., Zimmermann, M., Mozgova, I., Schrumpfova, P.P., Wimmerova, M., and Fajkus, J. (2009). Single-Myb-histone proteins from Arabidopsis thaliana: a quantitative study of telomere-binding specificity and kinetics. Biochem J *419*, 221-228.

Iwahara, J., Kigawa, T., Kitagawa, K., Masumoto, H., Okazaki, T., and Yokoyama, S. (1998). A helix-turn-helix structure unit in human centromere protein B (CENP-B). Embo J *17*, 827-837.

Jander, G., Baerson, S.R., Hudak, J.A., Gonzalez, K.A., Gruys, K.J., and Last, R.L. (2003). Ethylmethanesulfonate saturation mutagenesis in Arabidopsis to determine frequency of herbicide resistance. Plant physiology *131*, 139-146.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science (New York, NY *293*, 1074-1080.

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.

Jung, J.H., Seo, Y.H., Seo, P.J., Reyes, J.L., Yun, J., Chua, N.H., and Park, C.M. (2007). The GIGANTEA-regulated MicroRNA172 mediates photoperiodic flowering independent of CONSTANS in Arabidopsis. Plant Cell *19*, 2736-2748.

Kanhere, A., Viiri, K., Araujo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., *et al.* (2010). Short RNAs Are Transcribed from Repressed Polycomb Target Genes and Interact with Polycomb Repressive Complex-2. Molecular Cell *38*, 675-688.

Kapitonov, V.V., and Jurka, J. (2004). Harbinger transposons and an ancient HARBI1 gene derived from a transposase. DNA and Cell Biology *23*, 311-324.

Kaul, S., Koo, H.L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L.J., Feldblyum, T., Nierman, W., Benito, M.I., Lin, X.Y., *et al.* (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796-815.

Kennison, J.A. (1995). The Polycomb and trithorax group proteins of Drosophila: Trans-regulators of homeotic gene function. Annu Rev Genet *29*, 289-303.

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

Kim, J.H., Durrett, T.P., Last, R.L., and Jander, G. (2004). Characterization of the Arabidopsis TU8 glucosinolate mutation, an allele of TERMINAL FLOWER2. Plant Mol Biol *54*, 671-682.

King, I.F.G., Emmons, R.B., Francis, N.J., Wild, B., Muller, 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.

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.

Klymenko, T., Papp, B., Fischle, W., Kocher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Muller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes & Development *20*, 1110-1122.

Koncz, C., and Schell, J. (1986). The Promoter of Tl-DNA Gene 5 Controls the Tissue-Specific Expression of Chimeric Genes Carried by a Novel Type of Agrobacterium Binary Vector. Mol Gen Genet *204*, 383-396.

Koornneef, M., Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C.J., and Peeters, A.J.M. (1998). Genetic interactions among late-flowering mutants of Arabidopsis. Genetics *148*, 885-892.

Koornneef, M., Hanhart, C.J., and Vanderveen, J.H. (1991). A Genetic and Physiological Analysis of Late Flowering Mutants in Arabidopsis-Thaliana. Mol Gen Genet *229*, 57-66.

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 physiology *44*, 555-564.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X.H., Chi, A.S., *et al.* (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. Plos Genetics *4*.

Kumar, S.V., and Wigge, P.A. (2010). H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in Arabidopsis. Cell *140*, 136-147.

Lafos, M., Kroll, P., Hohenstatt, M.L., Thorpe, F.L., Clarenz, O., and Schubert, D. (2011). Dynamic Regulation of H3K27 Trimethylation during Arabidopsis Differentiation. Plos Genetics *7*.

Lafos, M., and Schubert, D. (2009). Balance of power - dynamic regulation of chromatin in plant development. Biol Chem *390*, 1113-1123.

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

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jorgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., *et al.* (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nature Cell Biology *12*, 618-U214.

Landweber, L.F., Nowacki, M., Higgins, B.P., Maquilan, G.M., Swart, E.C., and Doak, T.G. (2009). A Functional Role for Transposases in a Large Eukaryotic Genome. Science (New York, NY *324*, 935-938.

Larsson, A.S., Landberg, K., and Meeks-Wagner, D.R. (1998). The TERMINAL FLOWER2 (TFL2) gene controls the reproductive transition and meristem identity in Arabidopsis thaliana. Genetics *149*, 597-605.

Latrasse, D., Germann, S., Houba-Herin, N., Dubois, E., Bui-Prodhomme, D., Hourcade, D., Juul-Jensen, T., Le Roux, C., Majira, A., Simoncello, N., *et al.* (2011). Control of flowering and cell fate by LIF2, an RNA binding partner of the polycomb complex component LHP1. PLoS ONE *6*, e16592.

Lee, J.H., Yoo, S.J., Park, S.H., Hwang, I., Lee, J.S., and Ahn, J.H. (2007). Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes & Development *21*, 397-402.

Li, H., and Luan, S. (2011). The Cyclophilin AtCYP71 Interacts with CAF-1 and LHP1 and Functions in Multiple Chromatin Remodeling Processes. Molecular Plant *4*, 748-758.

Li, Y.H., Kirschmann, D.A., and Wallrath, L.L. (2002). Does heterochromatin protein 1 always follow code? Proceedings of the National Academy of Sciences of the United States of America *99*, 16462-16469.

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. (2007). Transposase-derived transcription factors regulate light signaling in Arabidopsis (vol 318, pg 1302, 2007). Science (New York, NY *318*, 1866-1866.

Liu, C., Xi, W.Y., Shen, L.S., Tan, C.P., and Yu, H. (2009). Regulation of Floral Patterning by Flowering Time Genes. Developmental Cell *16*, 711-722.

Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. Genes & Development *21*, 1530-1545.

Lu, F.L., Cui, X., Zhang, S.B., Jenuwein, T., and Cao, X.F. (2011). Arabidopsis REF6 is a histone H3 lysine 27 demethylase. Nature Genetics *43*, 715-U144.

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

Luo, C., and Lam, E. (2010). ANCORP: a high-resolution approach that generates distinct chromatin state models from multiple genome-wide datasets. Plant J *63*, 339-351.

Maison, C., and Almouzni, G. (2004). HP1 and the dynamics of heterochromatin maintenance. Nat Rev Mol Cell Bio *5*, 296-304.

Marchler-Bauer, A., Lu, S.N., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., *et al.* (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic acids research *39*, D225-D229.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., *et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-U711.

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nature Reviews Genetics *11*, 285-296.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343-349.

Marian, C.O., Bordoli, S.J., Goltz, M., Santarella, R.A., Jackson, L.P., Danilevskaya, O., Beckstette, M., Meeley, R., and Bass, H.W. (2003). The maize Single myb histone 1 gene, Smh1, belongs to a novel gene family and encodes a protein that binds telomere DNA repeats in vitro. Plant physiology *133*, 1336-1350.

Meyers, B.C., Morgante, M., and Michelmore, R.W. (2002). TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes. Plant Journal *32*, 77-92.

Morel, P., Trehin, C., Breuil-Broyer, S., and Negrutiu, I. (2009). Altering FVE/MSI4 results in a substantial increase of biomass in Arabidopsis-the functional analysis of an ontogenesis accelerator. Mol Breed *23*, 239-257.

Morey, L., and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. Trends Biochem Sci *35*, 323-332.

Mylne, J.S., Barrett, L., Tessadori, F., Mesnage, S., Johnson, L., Bernatavichute, Y.V., Jacobsen, S.E., Fransz, P., and Dean, C. (2006). LHP1, the Arabidopsis homologue of HETEROCHROMATIN PROTEIN1, is required for epigenetic silencing of FLC. Proceedings of the National Academy of Sciences of the United States of America *103*, 5012-5017.

Nei, M., and Kumar, S. (2000). Molecular evolution and phylogenetics (Oxford ; New York, Oxford University Press).

Oh, S., Park, S., and van Nocker, S. (2008). Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. PLoS Genet *4*, e1000077.

Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N., and Weigel, D. (2008). Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res *18*, 2024-2033.

Page, D.R., and Grossniklaus, L. (2002). The art and design of genetic screens: Arabidopsis thaliana. Nature Reviews Genetics *3*, 124-136.

Palm, W., and de Lange, T. (2008). How Shelterin Protects Mammalian Telomeres. Annu Rev Genet *42*, 301-334.

Pasini, D., Cloos, P.A.C., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306-U193.

Pazhouhandeh, M., Molinier, J., Berr, A., and Genschik, P. (2011). MSI4/FVE interacts with CUL4-DDB1 and a PRC2-like complex to control epigenetic regulation of flowering time in Arabidopsis. Proc Natl Acad Sci U S A *108*, 3430-3435.

Penin, A.A., Chub, V.V., and Ezhova, T.A. (2005). Basic principles of terminal flower formation. Ontogenez *36*, 90-95.

Piacentini, L., Fanti, L., Berloco, M., Perrini, B., and Pimpinelli, S. (2003). Heterochromatin protein 1 (HP1) is associated with induced gene expression in Drosophila euchromatin. J Cell Biol *161*, 707-714.

Price, C.M., Boltz, K.A., Chaiken, M.F., Stewart, J.A., Beilstein, M.A., and Shippen, D.E. (2010). Evolution of CST function in telomere maintenance. Cell Cycle *9*, 3157-3165.

Rea, S., Eisenhaber, F., O'Carroll, N., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., *et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Redei, G.P. (1962). Supervital Mutants of Arabidopsis. Genetics *47*, 443-&.

Rehrauer, H., Aquino, C., Gruissem, W., Henz, S.R., Hilson, P., Laubinger, S., Naouar, N., Patrignani, A., Rombauts, S., Shu, H., *et al.* (2010). AGRONOMICS1: a new resource for Arabidopsis transcriptome profiling. Plant physiology *152*, 487-499.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins. Annu Rev Genet *38*, 413-443.

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development *134*, 223-232.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., *et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by Noncoding RNAs. Cell *129*, 1311-1323.

Rizzardi, K., Landberg, K., Nilsson, L., Ljung, K., and Sundas-Larsson, A. (2011). TFL2/LHP1 is involved in auxin biosynthesis through positive regulation of YUCCA genes. Plant Journal *65*, 897-906.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome biology *11*.

Rojas-Pierce, M., and Springer, P.S. (2003). Gene and enhancer traps for gene discovery. Methods Mol Biol *236*, 221-240.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol Biol *53*, 247-259.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., *et al.* (2011). An integrated semiconductor device enabling non-optical genome sequencing. Nature *475*, 348-352.

Roudier, F., Teixeira, F.K., and Colot, V. (2009). Chromatin indexing in Arabidopsis: an epigenomic tale of tails and more. Trends Genet *25*, 511-517.

Saitou, N., and Nei, M. (1987). The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. Mol Biol Evol *4*, 406-425.

Sanchez-Pulido, L., Devos, D., Sung, Z.R., 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*, -.

Sanchez, R., Kim, M.Y., Calonje, M., Moon, Y.H., and Sung, Z.R. (2009). Temporal and Spatial Requirement of EMF1 Activity for Arabidopsis Vegetative and Reproductive Development. Molecular Plant *2*, 643-653.

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., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S.L., *et al.* (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Molecular Cell *42*, 330-341.

Schneeberger, K., Hagmann, J., Ossowski, S., Warthmann, N., Gesing, S., Kohlbacher, O., and Weigel, D. (2009a). Simultaneous alignment of short reads against multiple genomes. Genome biology *10*, R98.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jorgensen, J.E., Weigel, D., and Andersen, S.U. (2009b). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nature methods *6*, 550-551.

Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H.H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. Embo J *25*, 3110-3122.

Schrumpfova, P.P., Kuchar, M., Palecek, J., and Fajkus, J. (2008). Mapping of interaction domains of putative telomere-binding proteins AtTRB1 and AtPOT1b from Arabidopsis thaliana. Febs Lett *582*, 1400-1406.

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 Bio *12*, 799-814.

Scofield, S., Dewitte, W., and Murray, J.A. (2008). A model for Arabidopsis class-1 KNOX gene function. Plant Signal Behav *3*, 257-259.

Searle, I., He, Y.H., 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 & Development *20*, 898-912.

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., *et al.* (2002). A high-throughput Arabidopsis reverse genetics system. Plant Cell *14*, 2985-2994.

Shao, Z.H., 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.

Shaver, S., Casas-Mollano, J.A., Cerny, R.L., and Cerutti, H. (2010). Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. Epigenetics *5*, 301-312.

Shi, C.L., Stenvik, G.E., Vie, A.K., Bones, A.M., Pautot, V., Proveniers, M., Aalen, R.B., and Butenko, M.A. (2011). Arabidopsis Class I KNOTTED-Like Homeobox Proteins Act Downstream in the IDA-HAE/HSL2 Floral Abscission Signaling Pathway. Plant Cell *23*, 2553-2567.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of Polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Bio *10*, 697-708.

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H.D., and Cordes, S.P. (2009). A Vertebrate Polycomb Response Element Governs Segmentation of the Posterior Hindbrain. Cell *138*, 885-897.

Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S., and Baulcombe, D.C. (2007). An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. Plant Cell *19*, 1507-1521.

Sprenger-Haussels, M., and Weisshaar, B. (2000). Transactivation properties of parsley proline-rich bZIP transcription factors. Plant Journal *22*, 1-8.

Sridhar, V.V., Kapoor, A., Zhang, K.L., Zhu, J.J., Zhou, T., Hasegawa, P.M., Bressan, R.A., and Zhu, J.K. (2007). Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. Nature *447*, 735-U718.

Strutt, H., and Paro, R. (1997). The polycomb group protein complex of Drosophila melanogaster has different compositions at different target genes. Mol Cell Biol *17*, 6773-6783.

Sung, S.B., He, Y.H., Eshoo, T.W., Tamada, Y., Johnson, L., Nakahigashi, K., Goto, K., Jacobsen, S.E., and Amasino, R.M. (2006). Epigenetic maintenance of the vernalized state in Arabidopsis thaliana requires LIKE HETEROCHROMATIN PROTEIN 1. Nature Genetics *38*, 706-710.

Swiezewski, S., Liu, F.Q., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature *462*, 799-U122.

Takada, S., and Goto, K. (2003). TERMINAL FLOWER2, an Arabidopsis homolog of HETEROCHROMATIN PROTEIN1, counteracts the activation of FLOWERING LOCUS T by CONSTANS in the vascular tissues of leaves to regulate flowering time. Plant Cell *15*, 2856-2865.

Tamada, Y., Yun, J.Y., Woo, S.C., and Amasino, R.M. (2009). ARABIDOPSIS TRITHORAX-RELATED7 Is Required for Methylation of Lysine 4 of Histone H3 and for Transcriptional Activation of FLOWERING LOCUS C. Plant Cell *21*, 3257-3269.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol *28*, 2731-2739.

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

Turck, F., Fornara, F., and Coupland, G. (2008). Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annual Review of Plant Biology *59*, 573-594.

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 Genetics *3*, 855-866.

Uchida, N., Sakamoto, T., Kurata, T., and Tasaka, M. (2011). Identification of EMS-induced causal mutations in a non-reference Arabidopsis thaliana accession by whole genome sequencing. Plant & cell physiology *52*, 716-722.

Vakoc, C.R., Mandat, S.A., Olenchock, B.A., and Blobel, G.A. (2005). Histone H3 lysine 9 methylation and HP1 gamma are associated with transcription elongation through mammalian chromatin. Molecular Cell *19*, 381-391.

van Berloo, R. (2008). GGT 2.0: Versatile software for visualization and analysis of genetic data. J Hered *99*, 232-236.

van Steensel, B. (2011). Chromatin: constructing the big picture. Embo J *30*, 1885-1895.

Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant Journal *33*, 949-956.

Watson, J.M., and Riha, K. (2010). Comparative biology of telomeres: Where plants stand. Febs Lett *584*, 3752-3759.

Wilson, R.N., Heckman, J.W., and Somerville, C.R. (1992). Gibberellin Is Required for Flowering in Arabidopsis-Thaliana under Short Days. Plant physiology *100*, 403-408.

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*.

Woo, C.J., Kharchenko, P.V., Daheron, L., Park, P.J., and Kingston, R.E. (2010). A Region of the Human HOXD Cluster that Confers Polycomb-Group Responsiveness. Cell *140*, 99-110.

Wrzaczek, M., Brosche, M., Kollist, H., and Kangasjarvi, J. (2009). Arabidopsis GRI is involved in the regulation of cell death induced by extracellular ROS. Proceedings of the National Academy of Sciences of the United States of America *106*, 5412-5417.

Xu, L., and Shen, W.H. (2008). Polycomb silencing of KNOX genes confines shoot stem cell niches in Arabidopsis. Curr Biol *18*, 1966-1971.

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. Nature Structural & Molecular Biology *14*, 869-871.

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. Molecular Cell *29*, 69-80.

7 Appendix

7.1 Supplemental Figures

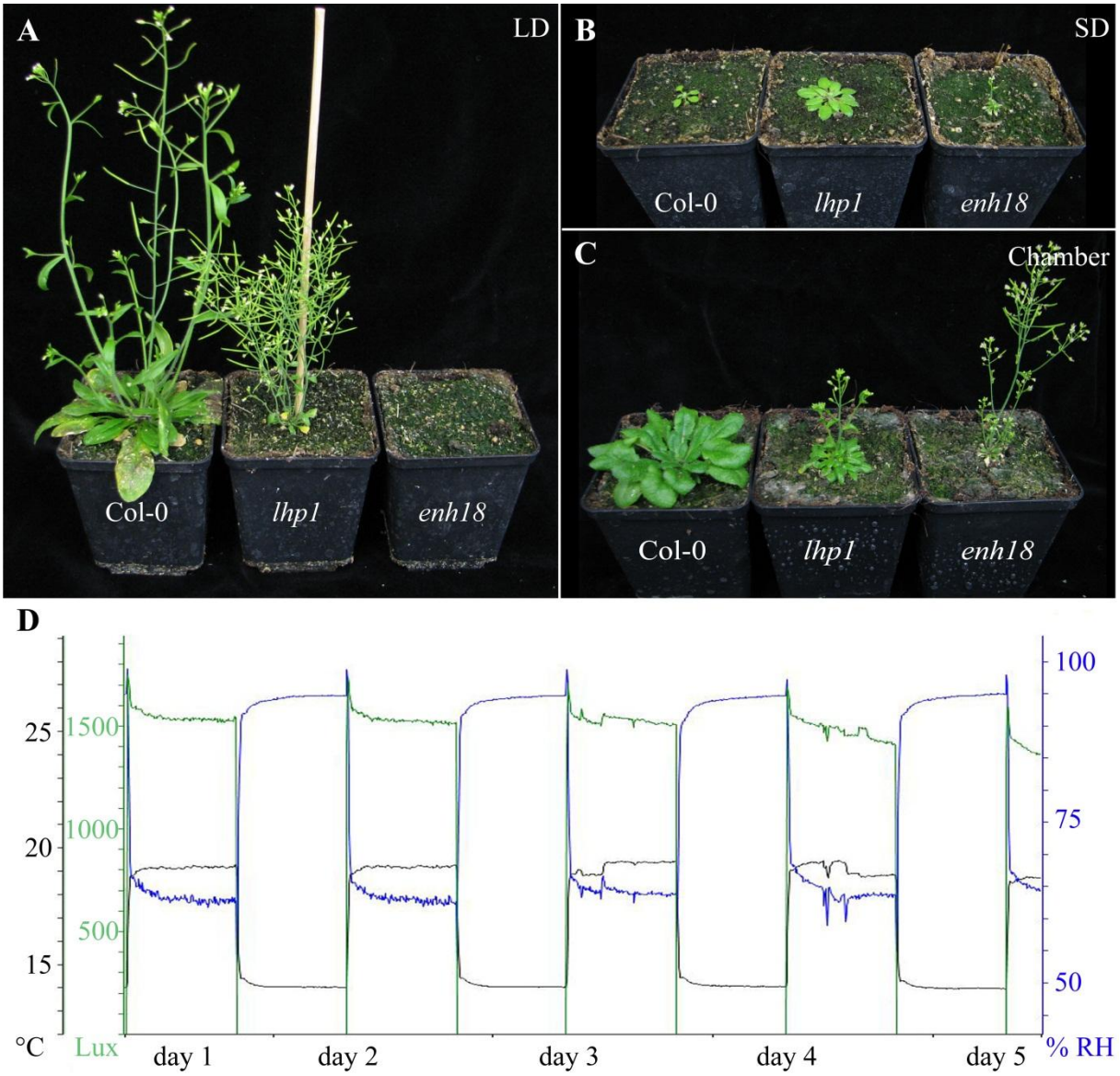


Figure S1: Ambient temperature has a developmental effect on *lhp1* and enhancer mutants
(**A**) Plants were grown for 44 days in long day, (**B**) short day in the greenhouse and (**C**) at 12 h days controlled climate chamber conditions. (**D**) Temperature, light intensity and humidity in the chamber were monitored over several days. Though conditions varied through day-night cycles, temperature was set to 16 °C and humidity to 60 %.

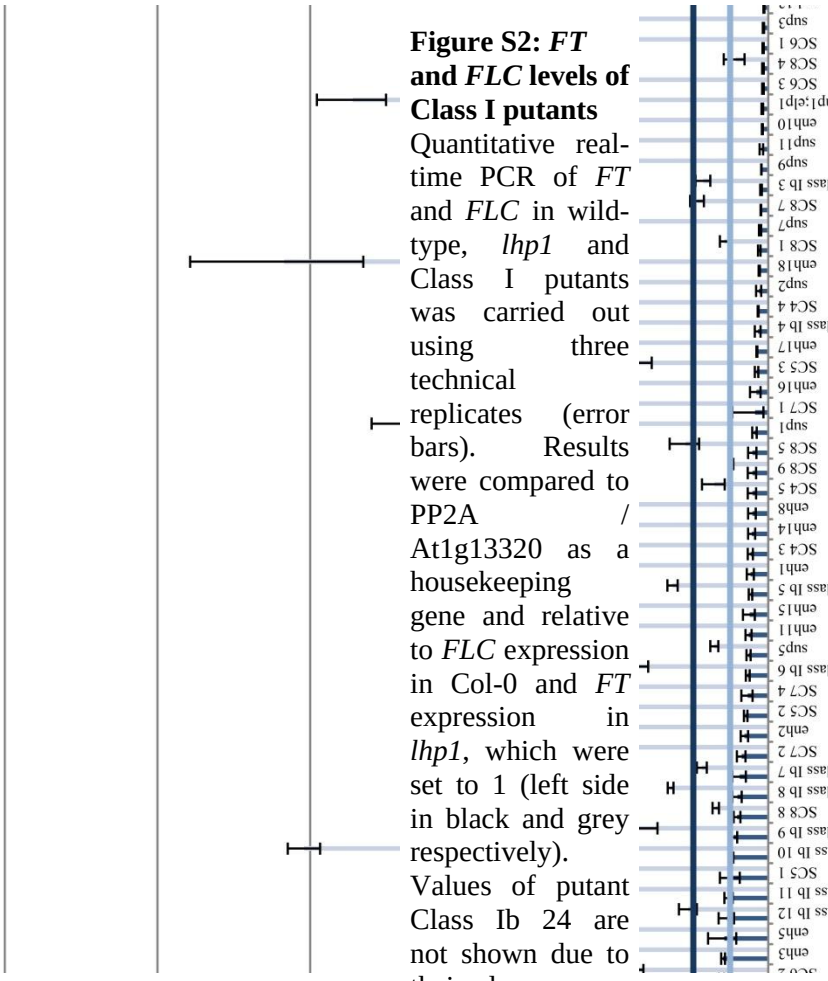


Figure S2: *FT* and *FLC* levels of Class I putants
Quantitative real-time PCR of *FT* and *FLC* in wild-type, *lhp1* and Class I putants was carried out using three technical replicates (error bars). Results were compared to PP2A / At1g13320 as a housekeeping gene and relative to *FLC* expression in Col-0 and *FT* expression in *lhp1*, which were set to 1 (left side in black and grey respectively). Values of putant Class Ib 24 are not shown due to their large error bars (average values 91,1 for *FLC* and 6,3 for *FT*). The graph was ordered by descending *FLC* levels of each putant which are represented by the left y-Axis (dark blue). On the right y-Axis *FT* levels are shown (light blue). Thresholds for *FLC* and *FT* were set to 5 and 0,25 and are displayed as a dark blue and a light blue line respectively.

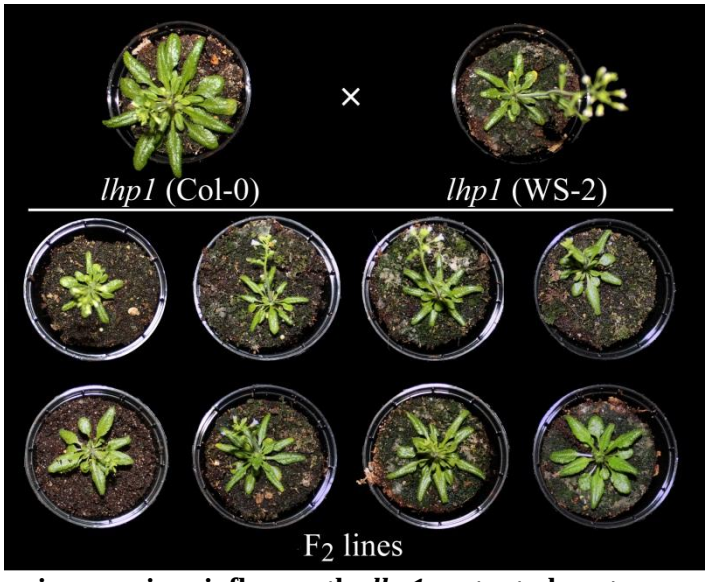


Figure S3: Differences in accessions influence the *lhp1* mutant phenotype
Plants were grown for 40 days in a climate chamber. Eight different F_2 lines of a c
(Col-0) and *lhp1-1* (Ws-2) have been sorted from *lhp1-1* (Ws-2) like to *lhp1-3* (Co

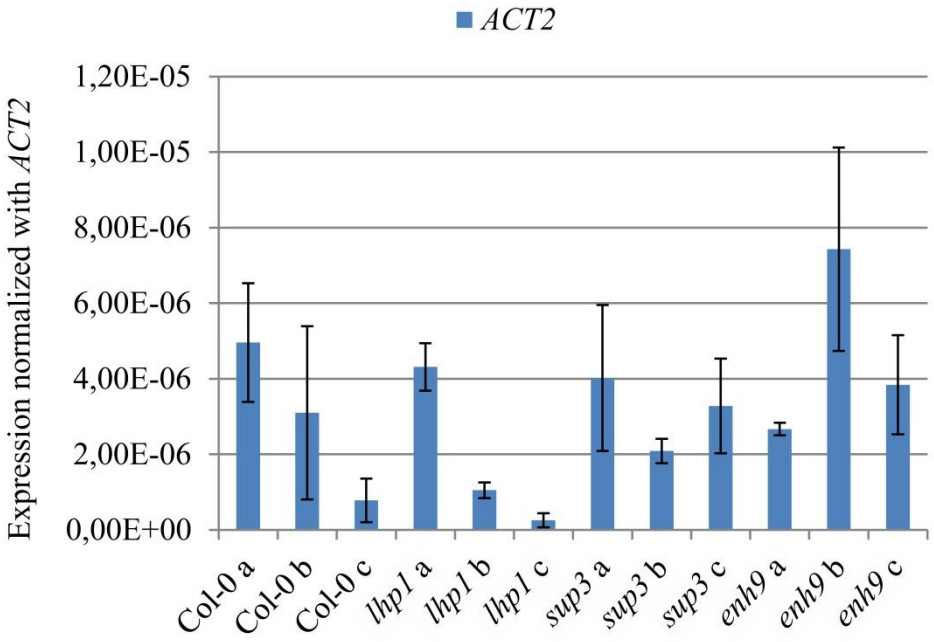


Figure S4: Differences in *Actin 2* expression of mRNA-seq libraries
Eleven libraries, except for Col-0 c, have been prepared at the same time and amplified for 17 cycles prior to measurement of *ACT2*. The three biological replicates of *sup3* and *enh9* show comparable *ACT2* levels while these differ in Col-0 and *lhp1*. Col-0 c and *lhp1* c had to be removed from the analysis after high-throughput sequencing. Error bars represent the standard deviation of three technical replicates.

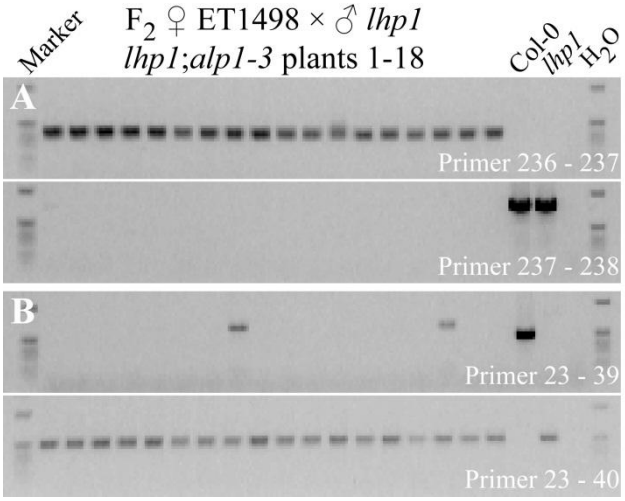


Figure S5: A mutation in At3g63270 causes the phenotype of *lhp1*;*alp1*
(**A**, **B**) Leave samples of 18 *lhp1*;*alp1*-like plants were collected for DNA extraction from 35 day old plants grown in LD conditions. (**A**) Primers did give a product when the T-DNA was either present or absent. (**B**) Primers were amplifying bands either in WT-like or in *lhp1*-like mutants.

7.2 Supplemental Table

Table S1: Outliers of mRNA-Seq analysis with non-normalized raw hits (Discarded replicates in grey / MADS-box in black)

AGI	Col-0 a	Col-0 b	Col-0 c	*lhp1* a	*lhp1* b	*lhp1* c	*sup3* a	*sup3* b	*sup3* c	*enh9* a
AT1G01250	14	17	5	0	0	0	0	18	16	1
AT1G03710	0	0	0	23	27	0	19	35	34	55
AT1G04110	12	1	4	0	0	0	6	8	7	8
AT1G09350	0	0	9	2	10	0	4	0	8	10
AT1G10770	4	6	0	0	0	0	0	2	0	0
AT1G24577	2	8	0	0	0	0	5	4	2	6
AT1G30160	0	0	0	4	19	0	0	0	2	11
AT1G31690	29	10	18	0	0	0	0	2	2	2
AT1G33940	8	4	0	0	0	0	11	6	0	2
AT1G34245	8	5	2	0	0	0	4	0	2	0
AT1G49640	0	0	0	10	7	0	3	0	4	9
AT1G51040	2	7	0	0	0	0	2	3	7	0
AT1G52790	1	0	2	0	20	0	0	0	0	7
AT1G53480	48,5	37,5	44,5	1	0	0	0	0	0	1,5
AT1G53940	0	0	0	5	10	0	0	0	0	4
AT1G58160	11	10	2	1	1	0	0	0	0	5
AT1G58643	0	0	0	14	10	0	0	1	0	3
AT1G58936	0	0	0	14	10	0	0	1	0	3
AT1G59312	0	0	0	14	10	0	0	1	0	3
AT1G62262	2	0	4	12	12	0	0	0	0	8
AT1G63055	0	0	0	8	2	0	13	2	10	8
AT1G63600	0	0	4	6	11	0	3	11	7	5
AT1G69120 / *AP1*	0	0	0	7	12	0	0	10	0	11
AT1G71420	5	6	4	0	0	0	0	1	0	2
AT1G78360	0	1	0	0	14	0	0	0	0	2
AT1G79770	20	7	2	0	0	0	16	7	4	2
AT2G14760	0	0	0	6	4	0	0	0	2	0
AT2G15042	15	2	3	0	0	0	0	0	0	7
AT2G15780	0	0	4	3	8	0	0	0	0	2
AT2G17770	7	5	0	0	0	0	0	8	9	3
AT2G20870	0	0	0	18	26	4	18	20	11	43
AT2G28340	0	0	0	4	5	2	10	38	17	3

AGI	**Col-0 a**	**Col-0 b**	**Col-0 c**	***lhp1* a**	***lhp1* b**	***lhp1* c**	***sup3* a**	***sup3* b**	***sup3* c**	***enh9***
AT2G29940	13	4	1	0	0	1	6	9	2	
AT2G37070	9	6	4	0	0	0	0	5	0	
AT2G39030	0	0	0	2	10	0	0	4	6	1
AT2G45135	0	0	0	2	12	0	11	12	8	
AT2G46880	2	11	0	0	0	0	0	14	10	
AT2G47520	0	0	11	13	8	0	0	0	0	
AT3G01345	1131	3	0	0	0	0	2	0	0	
AT3G03830	2	10	4	0	0	0	2	1	5	
AT3G16670	24	25	1	0	0	0	0	22	0	1
AT3G18010	8	8	12	0	0	0	0	10	2	
AT3G21040	0	0	0	2	12	0	3	0	5	
AT3G22100	17	2	0	0	2	2	7	6	2	
AT3G22231	13	2	0	0	0	0	0	0	0	
AT3G23480	19	23	6	0	0	2	4	12	5	
AT3G25670	8	6	12	0	0	0	13	5	4	
AT3G29590	11	11	8	0	0	0	0	0	2	
AT3G30260 / *AGL79*	0	0	0	10	4	0	4	0	0	
AT3G30720	215	5	4	0	0	0	2	4	4	2
AT3G30737	8	8	6	0	0	0	2	3	7	
AT3G46760	0	0	0	5	8	0	2	4	2	
AT3G46800	0	0	0	2	8	0	0	0	0	
AT3G48290	21	0	0	0	0	0	4	5	3	
AT3G48940	0	0	0	0	15	0	5	3	4	
AT3G49270	0	0	0	2	20	0	0	0	0	
AT3G49900	20	0	4	0	0	0	3	13	3	
AT3G52115	8	4	7	0	0	4	15	11	4	
AT3G54340 / *AP3*	0	0	0	6	18	2	12	6	2	3
AT3G56350	1	2	22	14	8	0	0	0	0	
AT3G57500	8	8	4	0	0	0	4	6	3	
AT3G61250	13	10	0	0	0	0	6	1	4	1
AT3G62930	7	6	10	0	0	0	13	14	3	
AT4G00970	8	3	0	0	0	0	0	4	0	
AT4G01130	9	8	0	0	0	0	10	19	14	
AT4G05470	0	0	0	3	9	0	10	18	15	
AT4G05475	0	0	0	15	28	0	22	33	11	3
AT4G09110	3	0	0	10	10	0	0	0	0	

AGI	**Col-0 a**	**Col-0 b**	**Col-0 c**	***lhp1* a**	***lhp1* b**	***lhp1* c**	***sup3* a**	***sup3* b**	***sup3* c**	***enh9* a**	***enh9***
AT4G14819	12	0	4	0	0	0	0	0	0	0	
AT4G19590	10	4	0	0	0	0	0	0	4	1	
AT4G20160	0	0	4	5	4	0	0	0	1	1	
AT4G23950	8	10	12	0	0	0	4	6	12	4	
AT4G26255	4	6	1	0	0	0	4	6	3	2	
AT4G26950	11	2	0	0	0	2	0	0	2	0	
AT4G31950	2	4	37	30	7	0	0	0	0	22	
AT4G31970	88	22	40	99	202	2	0	2	0	81	
AT4G33970	0	0	3	5	3	0	4	3	0	4	
AT4G35810	18	9	21	0	0	0	3	6	0	2	
AT5G06490	0	0	11	0	12	0	0	0	0	2	
AT5G06900	0	0	0	10	10	0	0	0	0	6	
AT5G09470	0	0	1	6	4	0	0	0	0	1	
AT5G13210	12	13	21	41	25	2	0	0	0	17	
AT5G19700	0	0	18	19	0	0	0	0	0	4	
AT5G24640	0	0	6	16	6	0	5	0	0	10	
AT5G27780	2	11	2	0	0	0	2	2	0	0	
AT5G35935	71	3	4	0	0	0	2	2	0	0	
AT5G37478	0	0	0	0	12	0	6	4	4	0	
AT5G44420	6	11	21	0	0	0	4	2	13	4	
AT5G45150	0	11	2	0	0	0	0	0	2	2	
AT5G46830	21	5	0	0	0	0	2	1	1	6	
AT5G50335	17	6	6	0	0	0	0	6	0	3	
AT5G55460	7	6	1	0	0	0	2	0	2	0	
AT5G59990	2	0	2	5	8	0	0	0	0	2	

7.3 Abbreviations

7.3.1 General abbreviations

-	minus / not present
%	percentage
:	under the control of (promoter-gene constructs)
::	fused to (gene fusion constructs)
+	plus / present
°C	degrees Celsius
µ	micro
3’	three prime end of DNA fragment
35S	promoter of the Cauliflower Mosaic virus
5’	five prime end of DNA fragment
A (base)	adenine of nucleic acid
A.t.	*Arabidopsis thaliana*
Agrobacterium	*Agrobacterium tumefaciens*
ANOVA	analysis of variance
APB	acceptor photo bleaching
Arabidopsis	*Arabidopsis thaliana*
BC	backcross
BLAST	Basic Local Alignment Search
Bp	base pair
C (base)	cytosine of nucleic acid
C-	carboxy-terminal
$CaCl_2$	calcium chloride
CAF-1	Chromatin Assembly Factor-1
CaMV	Cauliflower Mosaic Virus
CAPS	cleaved amplified polymorphic sequence
CCG	Cologne Center for Genomics
Cd	chromodomain
CDD	Conserved Domain Database
cDNA	complementary DNA
ChIP	Chromatin Immunoprecipitation
Col-0	Columbia
CSD	chromo-shadow domain
CST	Cdc13 / Stn1 / Ten1
D (amino acid)	aspartate
D.m.	*Drosophila melanogaster*
D.r.	*Danio rerio*
dH_2O	distilled water
DMSO	dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
Drosophila	*Drosophila melanogaster*
dT18 primer	18-mer oligonucleotide consisting of 18 thymine bases
DTT	Dithiothreitol
E (amino acid)	glutamic acid
E.coli	*Escherichia coli*
EDTA	Ehylenediaminetetraacetic acid

eFP	electronic Fluorescent Pictograph
EMS	Ethyl methanesulfonate
enh	encancer
et al.	*et alii* / *et aliae* [Lat.] and others
ET	enhancer trap
E-value	Expect value
F_2	Second generation (*filius* 2)
FDR	false discovery rate
FRET	Fluorescence Resonance Energy Transfer
G (base)	guanine of nucleic acid
G (amino acid)	glycine
g	gram
GA	gibberellic acid
Gabi	Genomanalyse im biologischen System Pflanze
GAF	GAGA factor
GGT	graphical genotype
GM	growth medium / ½ strength Murashige & Skoog
GO	Gene Ontology
GW	Gateway™
h	hour(s)
H.s.	*Homo sapiens*
H2A-Kub	mono-ubiquitylated lysine at histone 2
H3K27me3	tri-methylated lysine 27 at histone 3
H3K4me3	tri-methylated lysine 4 at histone 3
H3K9me2	di-methylated lysine 9 at histone 3
H4K20	lysine 20 at histone 4
HCl	hydrogen chloride
HRM	high resolution melting
INRA	Institute National de la Recherche Agronomique
KCl	potassium chloride
KOH	potassium hydroxide
l	liter
LD	long day
L*er*	Landsberg *erecta*
lncRNA	long non-coding RNA
log_2FC	logarithmic fold change (to the basis of 2)
m	milli
M	mol / liter
M_1, M_2 and M_3	First, second and third generation after application of EMS
MADS	MCM1, AGAMOUS, DEFICIENS, and SRF
MCS	multiple cloning site
MEGA	molecular evolutionary genetics analysis
MES	Sodium 4-morpholin-1-ylethylsulphonate; 4-Morpholineethanesulfonic acid sodium salt; 2-(N-Morpholino)ethanesulfonic acid sodium salt
$MgCl_2$	magnesium chloride
min	minute(s)
miRNA	micro RNA
MPIPZ	Max Planck Institute for Plant Breeding Research
mRNA	messenger RNA
n.d.	not defined

NaCl	sodium chloride
NaOH	sodium hydroxide
NASC	Nottingham Arabidopsis Stock Centre
NCBI	National Center for Biological Information
NEB	New England Biolabs
NOS	nopaline synthase terminator
O.s.	*Oryza sativa*
OD	optical density
P (amino acid)	proline
PcG	Polycomb group
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfam	protein families
PHORC	PHO-repressive complex
PNK	polynucleotid kinase
Pol II	RNA Polymerase II
PRC1 and PRC2	Polycomb Repressive Complex(es) 1 and 2
PR-DUB	Polycomb repressive deubiquitinase
PRE	PcG response element
Putant	potential mutant
PVDF	polyvinyl difluoride
qRT-PCR	quantitative real-time polymerase chain reaction
RFP	red fluorescent protein
RNA	Ribonucleic acid
rpm	rotations per minute
rRNA	ribosomal RNA
s	second(s)
S (amino acid)	serine
SC	subclass
SD	short day
SET	SUVAR3-9/E(Z)/Trithorax
SHORE	Short Read
SIGnAL	Salk Institute Genomic Analysis Laboratory
SMH	single myb histone
SNP	single nucleotide polymorphism
SSLP	single-strand length polymorphism
sup	suppressor
T (base)	thymine of nucleic acid
T_1	First generation after transformation
TAIR	The Arabidopsis Information Resource
Taq	*Thermus aquaticus*
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
trxG	trithorax Group
TSS	transcriptional start site
W (amino acid)	tryptophan
Ws	Wassilewskija
WT	wild-type
ZT	Zeitgeber (each number represents an hour of the day)

7.3.2 Gene and protein names

ACT2	Actin 2
AG	AGAMOUS
ALP1	ANTAGONIST OF LHP1
AP3	APETALA 3
BPC	BASIC PENTACYSTEINE
CBX	Chomobox protein homologue
CCA1	CIRCADIAN CLOCK ASSOCIATED 1
CFP	cyan fluorescent protein
CLF	CURLY LEAF
CLSY1 (CHR38)	CLASSY 1 (CHROMATIN REMODELING 38)
CO	CONSTANS
CTC1	Conserved Telomere maintenance Component 1
CYP82C4	CYTOCHROME P450 subunit
E(Z)	Enhancer of Zeste
EMF1 and EMF2	EMBRYONIC FLOWER 1 and 2
ERF71	ETHYLENE RESPONSE FACTOR 71
ESC	Extra Sex Combs
ESD7	EARLY IN SHORT DAYS 7
F15D2_21	DSL-like lipase
FAR1	FAR-RED-IMPAIRED RESPONSE 1
FD	FLOWERING LOCUS D
FHY3	FAR-RED-ELONGATED HYPOCOTYL 3
FIE	FERTILIZATION INDEPENDENT ENDOSPERM
FIS	FERTILIZATION INDEPENDENT SEED
FLC	FLOWERING LOCUS C
FT	FLOWERING LOCUS T
GFP	green fluorescent protein
H1	histone 1
H2A(.Z)	histone 2A(.Z)
H2B	histone 2B
H3	histone 3
H4	histone 4
HA	hemagglutinin
HARBI1	Harbinger 1
ICU2	INCURVATA 2
LEC2	LEAFY COTYLEDON 2
LHP1	LIKE HETEROCHROMATIN PROTEIN 1
LIF2	LHP1-Interacting Factor 2
MEA	MEDEA
MSI	MULTICOPY SUPPRESSOR OF IRA
NAS2	NICOTIANAMINE SYNTHASE 2
Nurf55	Nucleosome Remodeling Factor 55
PC	Polycomb
PCL	Polycomblike
PH	Polyhomeotic
PI	PISTILLATA
PIF	P instability factor
PP2A	Protein Phosphatase 2
PSC	Posterior Sex Combs

SCE	Sex Comb Extra
SCR	SCARECROW
SEP1 – SEP3	SEPALLATA 1 - 3
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
SU(Z)12	Suppressor of Zeste 12
SVP	SHORT VEGETATIVE PHASE
SWN	SWINGER
THI2.2	THIONIN 2.2
TRB1 – TRB3	TELOMERE REPEAT BINDING FACTOR 1 - 3
VIN	VERNALIZATION INDEPENDENT
VRN1 and VRN2	REDUCED VERNALIZATION RESPONSE 1 and 2
WRKY27	WRKY DNA-BINDING PROTEIN 27
YAB5	YABBY 5
YFP	yellow fluorescent protein
YUC	YUCCA
YY1	yin-yang 1

Acknowledgements

I thank all of those who have supported me and believed in me during the time of my studies. This work would not have been accomplished without the help of many people. I want to thank…

Dr. Franziska Turck for her ideas, her support, her supervision and her general interest in science that has been a great example to me

Prof. George Coupland for the opportunity to work in his department and for his beautiful mind.

Prof. Ute Höcker for being my second examiner and for her ability to ask the right question at the right time.

Prof. Martin Hülskamp for being the head of my thesis committee.

Prof. Klaus Theres for being my second supervisor and his helpful input when it was needed most.

Everyone in the Turck group for good times, a great working atmosphere and delicious cakes. Especially I want to thank **Sara Farrona**, for her great advice during the first years of my PhD, **Julia Reimer** and **Julia Engelhorn** for their support and critical reading of my manuscript and **Jesús López Corrales** for his crossing abilities and hard work.

Korbinian Schneeberger and **Geo Velikkakam James** for a great collaboration and their help in improving this study.

Dr. Justin Goodrich and **Shih Chieh Liang** for a wonderful collaboration.

Elmon Schmelzer for introducing me to electron microscopy.

Dr. Seth Davis for providing equipment and sharing his thoughts.

The gardeners for taking care of my plants.

Everybody in the Coupland department for a great atmosphere.

Jarod Rollins for critical reading, making the best coffee in the world and his friendship.

Lena Försch, **Barbara Kagon**, **Patrick Fürst** and **Anke Zellmer** for their love and support.

Gabi Deeg for critical reading of the manuscript.

The stage for always being there.

My family for making me, shaping me and loving me.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. George Coupland betreut worden.

Köln, den 21. Februar 2012 _____

Benjamin Hartwig

Teilpublikationen

Hartwig B., Velikkakam G. , Konrad K., Schneeberger K. and Turck F. (2012): Fast isogenic mapping-by-sequencing of EMS-induced mutant bulks. *Plant Phys*, submitted

Lebenslauf

Benjamin Hartwig

Ilmenaudeich 50
21423 Winsen/Luhe
Tel.: 04179-646
E-mail: benhartwig@gmail.com

geboren am 25.06.83 in Winsen/Luhe

Ausbildung

10/2008 – gepl. 04/2012	**Doktorand** der International Max-Planck Research School (IMPRS) am **Max-Planck Institut** für Züchtunsforschung (MPIZ) in Köln
2012	**Doktorarbeit**: „ALP1 und ELP1 – neue genetische Modifikatoren von LIKE HETEROCHROMATIN PROTEIN 1 partizipieren in Polycomb vermittelter Repression von Genen in Arabidopsis“
2008	**Masterarbeit:** „Molekulare Charakterisierung von resistance gene analogues (RGAs) verschiedener Rosenarten“
08/2006 – 09/2008	Aufbaustudium der **Pflanzenbiotechnologie** an der Universität Hannover mit dem Abschluss **Master of Science**
08/2006 – 12/2006	**Auslandssemester an der Purdue University (Indiana/USA) mit dem Hortecus Programm**
2006	**Bachelorarbeit: „**In-vitro-Vermehrung von *Oncidium* ‘Sweet Sugar’ in temporären Immersionssystemen“
10/2003 – 08/2006	Studium der **Pflanzenbiotechnologie** an der Universität Hannover mit dem Abschluss **Bachelor of Science**
2002	**Abitur** am Gymnasium Winsen/Luhe

Zivildienst

08/2002 – 06/2003	Kindergarten der Lebenshilfe e.V. in Winsen/Luhe Betreuung körperlich und geistig behinderter Kinder

Anstellungen

07/2007 – 12/2007	Anstellung als Hilfswissenschaftler im Institut für Pflanzengenetik der Universität Hannover

Köln, 21. Februar 2012: _____ / Benjamin Hartwig