University of South Bohemia in České Budějovice
Faculty of Science

Diversity of Polycomb complexes and their function

Bachelor thesis

Luboš Říha

Supervisor: Mgr. Iva Mozgová, Ph.D.

České Budějovice 2021

Říha, L., 2021: Diversity of Polycomb complexes and their function. Bc. Thesis, in English. – 54 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

The aim of this bachelor thesis is testing and further development of vector system that should help clarify alleged functional redundancy of Polycomb repressive complex 2 (PRC2) subunits. The theoretical part introduces the field of epigenetics and the role of Polycomb group complexes (PcGs) in *Arabidopsis thaliana* (mouse-ear cress) is explained. The issue of PRC2 subunits SWN and CLF redundancy is set in context and the tested hypothesis is explained. Genetic engineering tools relevant for this study are presented. Finally, the background of the promoter and marker vectors developed in the practical part is explained. In the practical part vectors with markers and promoters were developed and transgenic plants were grown on selection and genotyped. Results are presented and discussed.

Plagiarism statement

I certify that I am the author of this thesis and that I completed it only with the use of sources acknowledged in the cited literature.

In České Budějovice 14th of April 2021 Luboš Říha

Acknowledgements

I want to thank my supervisor Mgr. Iva Mozgová, Ph.D. for the opportunity and for her exemplary mentorship. I also want to thank her and members of her lab for friendly environment and for all the advice and help I have received from them. My gratitude also goes to my family, friends, and partner for giving me emotional and nutritional support.

Content

1. INTRODUCTION AND BACKGROUND

1.1. Introduction

Polycomb Repressive Complex 2 (PRC2) is one of the key epigenetic regulators of gene expression found in most eukaryotic organisms. In *Arabidopsis thaliana* (Arabidopsis) this complex targets thousands of developmental and environmentally responsive genes and its function is necessary for correct plant development. The central focus of this thesis is testing and further development of a tool that allows researchers to examine the seemingly redundant function of CURLY LEAF (CLF) and SWINGER (SWN) subunits of PRC2 and their effect on the transcription of specific developmental and light signalling-related genes (*HY5*, *HFR1*, *ELIP1*, *ELIP2*, *CHL27*, *FLU*, *GUN5*, *FUS3*, *LEC1*, *LEC2*, *ABI3*, *ABI4*, *NPQ4*, *RBSC1B*) in specific tissues and stages of plant development. In this thesis, the whole process of generation of transgenic plants, starting with cloning design and ending with a selection of transformants was carried out. It is thoroughly explained and shown on examples.

1.2. Background

1.2.1. The importance of studying plants

As Peter H. Raven (2021) describes in his article, based on the works presented at the National Geographic Society's symposium "A World of Plants" held in 2019, plants and other photosynthetic organisms are, without a doubt, the backbone of life on the planet Earth. There are many proofs that plants are essential for the existence of the world as we know it. It is the photosynthetic activity that produces O_2 and carbon-containing organic compounds that enabled the existence and survival of animals and humans. The CO_2-binding capabilities have kept the atmosphere in stable cycles for millennia. The abilities of creation and maintenance of the environment including soil management and water management provided and provide stable habitats for fauna's life and evolution. Most of the reasons are known for quite some time, nevertheless, it is substantial to highlight that complex understanding of basic mechanisms of plant growth has been, is and will be of significant importance.

Now, in times of rapid climate change and extreme world-population growth, it is crucial more than ever to have a deep knowledge of inner mechanisms that govern plants' growth and development. An example of the impact of climate change on plants can be the change of their phenology, the periodic events in the life cycles of plants. It has been proven

that both spring and autumn phenology has been altered (Piao et al., 2019). For example, in the measured time period 1982-2011 the spring phenology of 61 species in China was 5.5 days/decade earlier (Ge et al., 2015).

However, population growth is perhaps the biggest drive behind plant research. The world population is above 7,5 billion and rising. According to the Action Against Hunger organisation (https://www.actionagainsthunger.org) around 690 million people have suffered from hunger in 2020. Understanding the plant development, physiology, genetics and epigenetics is essential for fighting the world hunger. One of the keystones of this fight might be the development of genetically modified organisms (GMOs) with traits surpassing the original ones (Oliver, 2014).

1.2.2. Arabidopsis as a plant model organism

Arabidopsis was not the first candidate for plant model organism. Until the 1980s, the used model plants were usually those relevant for agriculture - maize, tomato, rice, barley etc. Nevertheless, these plants did not have traits that would make them easily accessible for studying fundamental molecular processes and researchers were unable to answer some of the core questions about development, growth, hormonal response and environmental response (Meinke et al., 1998).

Therefore, a new model organism had to be selected and that was Arabidopsis. This little weed was selected because it combines all the essential characteristics of a model organism. One of them is that it is, in many aspects, representative of flowering plants. Additionally, certain features of Arabidopsis such as genome maintenance can be representative for the whole eukaryotic group (Hays, 2002; Leonelli & Ankeny, 2013; Meinke et al., 1998).

Furthermore, Arabidopsis is a good model organism because it is only 10-20 cm tall and its rosette is around 2-10 cm in diameter which means that many plants can be grown in a small area. The next important feature is the high speed of the life cycle with a short generation time. The plant can germinate, grow, flower and have mature seeds in 8-12 weeks. A very high number of seeds that reach up to 5000 seeds per plant is also a valued quality (Leonelli & Ankeny, 2013; Meinke et al., 1998).

1.2.3. The history of epigenetics and the discovery of Polycomb

The field of epigenetics was undefined until half of the 20th century. At that time, the embryologist Conrad Waddington established the term “epigenetics” for the first time and he thought of it as a field that connects development and genetics (Waddington, 2012, reprinted).

The unexplained problems such as stem-cell differentiation were proofs that some mechanisms additional to genetics are involved. But it wasn't until 1969 when Griffith and Mahler suggested that DNA methylation could be the mechanism underlying the changes of gene expression. This idea was further developed in 1975 when several independent works concluded that methylation could be the mechanism responsible for on/off switching of gene expression or, as in the case of the chromosome X, even for repression and condensation of whole chromosomes. However, these were just theories and the experimental proof came several years later when researchers identified restriction endonucleases that allowed differential restriction of methylated and unmethylated sequence (reviewed in Doerfler, 1981; Holliday, 2006, Holliday & Pugh, 1975). By that time researchers also discovered that the methylation is not distributed equally all over the genome but there are often clusters of methylated DNA. Also, that methylation often occurs in certain regions such as repetitive elements (Doerfler, 1981).

At approximately the same time, Ed Lewis discovered that mutation of *Polycomb* (Pc) affects embryogenesis of *Drosophila* due to failure of repression of transcription factor genes that are responsible for the identity of body segments along the anterior-posterior axis, the Homeobox genes (HOX) (Lewis, 1978, Schuettengruber et al., 2017). Yet, the mechanism of the regulation of gene expression by Pc and other PcGs remained unknown for the next 25 years. It wasn't even technically possible to discover that effect until the 1990s when the regulatory function of chromatin compaction was discovered (Grunstein, 1992; King et al., 2006). Chromatin is a complex of nucleic acids and proteins in the nucleus of a eukaryotic cell, comprising the basic units nucleosomes, which are octamers of basic histone proteins binding approximately 150 bp of DNA, and accessory regulatory proteins (Spiess, 2006).

Until the end of the 20th century, it was believed that histones serve purely for the compaction of DNA (Grunstein, 1992). Discoveries revealing that changes of the composition of functional groups (methyl, acetyl, ubiquitin…) at the N-terminal part of histones (for closer explanation see next chapter 1.2.4) have a significant effect on gene regulation are actually relatively new. As reviewed by Kouzarides (2007), the works that examine the effect of these compositional changes, called post-translational modifications (PTMs), on gene expression come from the beginning of the 21st century. At the same time PcG's ability to control the PTMs was discovered (reviewed by Simon & Kingston, 2013) and the research exploring mechanisms and extent of PcG protein influence on gene expression and development of organisms is very active ever since.

As was indicated, multiple PcG proteins have been discovered since the discovery of Pc. The most well-described proteins from this group are proteins that are part of Polycomb repressive complex 1 (PRC1) and PRC2 (reviewed by Mozgova & Hennig, 2015a; Schuettengruber et al., 2017) Both complexes were first described in *Drosophila* and will be further characterised in chapter 1.2.4.1. devoted to it.

1.2.4. The importance of Polycomb-group proteins

The importance of Polycomb-group (PcG) protein complexes resides in their histone-modifying abilities. The modifications then promote chromatin compaction, which makes PcGs important components of gene transcriptional silencing (Calonje, 2014; Mozgova & Hennig, 2015a; Schuettengruber, 2017). There are, generally speaking, two states in which chromatin can be found – heterochromatin and euchromatin. Heterochromatin is a compact state whereby the residing genes are transcriptionally inactivated. While constitutive heterochromatin relates to stably repressed loci, facultative heterochromatin is found at sites of changing transcription level. On the contrary, euchromatin comprises transcriptionally active genes (King, 2015; Kouzarides, 2007). Chromatin states are dynamic, owing to the dynamic presence of histone post-translational modifications (PTMs). Histones are octamers composed of the subunits H2A and H2B, forming 2 dimers, and of H4 and H3, forming one tetramer. Histones are globular proteins that are highly structured but the N-terminal amino acid "tails" protrude from the nucleosome and are accessible for histone-modifying enzymes that catalyse the PTMs. Different PTMs of histone N-terminal tails are instructive for chromatin remodelling (Kouzarides, 2007). In the case of PRC1, the PTM mediated is ubiquitination of lysine (118 in *Drosophila*, 119 in Arabidopsis) on histone H2A (H2AK118ub/ H2AK119ub). PRC2 functions as a methyltransferase and its activity results in tri-methylation of lysine 27 on histone H3 (H3K27me3) (reviewed in Tamburri et al., 2020). For a long time, scientists believed a model that suggested that placement of H3K27me3 is a prerequisite for docking of PRC1 and thus the placement of ubiquitin. However, in recent years this model was challenged by discoveries of independent activities of both complexes and some studies even suggest that PRC1 activity is important for the recruitment of PRC2 at some loci (reviewed in Yang et al., 2017).

Through the modifications of histones followed by gene transcriptional repression, PcG proteins are involved in the establishment of body plan in *Drosophila* m. (Lewis, 1978), in the regulation of stem cell differentiation, in senescence and cancer formation, in X chromosome inactivation and genomic imprinting in humans (Völkel et al., 2012) and

development regulation in plants (Mozgova & Hennig, 2015a; Yang et al., 2017). These are not all the aspects in which PcG proteins are involved. Nevertheless, as is shown here, PcG proteins are essential for a broad range of different eukaryotic organisms and thus deep knowledge of their structure and function is highly desirable.

1.2.4.1. PcG of *Drosophila*

PRC1 of *Drosophila* contains a conserved core of 2 RING-finger proteins: dRing (also known as Sex combs extra - Sce) + Posterior sex combs [Psc; also known as Supressor of zeste 2 - Suz(2)]. These domains have ubiquitin ligase activity, mediating the monoubiquitylation. Psc, in addition, promotes chromatin compaction and enhances dRing function. Besides the catalytic subunits, the core of PRC1 in *Drosophila* contains two additional subunits: the previously described Pc and Polyhomeotic (Ph). Pc is essential for binding with H3K27me3 mark placed by PRC2. Ph's sterile alpha motif (SAM) domain is responsible for oligomerization of proteins (Calonje, 2014; Mozgova & Hennig, 2015a; Völkel et al., 2012; Yang et al., 2017).

PRC2 is composed of 4 subunits: Enhancer of zeste [E(z)], Suppressor of zeste [Su(z)], Extra sex combs (Esc) and Nurf 55 (p55). E(z) catalyses H3K27me3 via its catalytic domain SET [Su(var)3-9, E(z), Trithorax (Trx)]. E(z) cooperates with Su(z)12, which is a Cys2-His2 zinc-finger and VEFS [VRN2-EMF2-FIS2-Su(z)12]–domain protein responsible for the stability of the complex. Esc is necessary for H3K27me3 binding and Nurf 55 is a histone-binding, nucleosome-remodelling factor. Both Esc and Nurf55 mediate their activities via their WD40 domains (Kasinath et al., 2018; Mozgova & Hennig, 2015a; Schuettengruber et al., 2017; Yang et al., 2017).

1.2.4.2. PcG diversity of Arabidopsis

The core of PRC1 in Arabidopsis is a protein dimer that contains two domains: RING-finger domain and RAWUL-domain (RING-finger and WD-associated ubiquitin-like) (Sanchez-Pulido et al., 2008). One of the proteins is either of two homologues AtRING1A or AtRING1B, which have E3 ubiquitin ligase activity. The other is represented by one of three homologues AtBMI1A, AtBMI1B or AtBMI1C, which influence chromatin compaction. Additionally, AtBMI1/2/3 have E3 ubiquitin ligase activity as well (Calonje, 2014; Gómez-Zambrano et al., 2019; Mozgova & Hennig, 2015a; Yang et al., 2017). The accessory subunits of PRC1 in plants are less well known. LHP1 (LIKE-HETEROCHROMATIN PROTEIN 1,

also known as TFL2- TERMINAL FLOWER 2) is a functional homologue of Pc and an orthologue of animal HP1 (HETEROCHROMATIN PROTEIN 1) that binds H3K9me (Sanchez-Pulido et al., 2008). Unlike HP1 however, LHP1 recognizes H3K27me3 (Exner et al., 2009). Studies have shown that LHP1 physically connects to EMF, VRN and other subunits of PRC2, suggesting, that LHP1 is actually the bridge between PRC1 and PRC2 (Derkacheva & Hennig, 2014). Another plant-specific protein that interacts with more than one PcG complex is EMBRYONIC FLOWER 1 (EMF1) (Aubert et al., 2001).

PRC2 is composed of homologues of *Drosophila* PRC2 subunits, however, the number of these homologues is larger. The E(z) homologues are CURLY LEAF (CLF), MEDEA [MEA, also known as FERTILIZATION INDEPENDENT SEED 1 (FIS1)] and SWINGER (SWN). All of these show histone methyltransferase activities. EMBRYONIC FLOWER 2 (EMF2), VERNALIZATION 2 (VRN2) and FERTILIZATION INDEPENDENT SEED 2 (FIS2) are homologues of Su(z)12 and are essential for the stability of the complex. MULTIPLE SUPPRESSOR OF INHIBITORY REGULATOR PROTEIN (IRA) 1 (MSI1) is a homologue of Nurf55 and FERTILIZATION INDEPENDENT ENDOSPERM (FIE) is a homologue of Esc (Calonje, 2014; Mozgova & Hennig, 2015a; J. Shu et al., 2019; Yang et al., 2017). Both MSI and FIE have WD40 repeats and serve for protein-protein interaction (FIE for H3 binding and MSI1 for nucleosome binding) (Mozgova & Hennig, 2015a). It is important to note that in Arabidopsis, PRC2 can be found in at least three forms, depending on which homologue of Su(z)12 is present: VRN-PRC2, FIS-PRC2, EMF-PRC2 (Mozgova & Hennig, 2015a; Yang et al., 2017). Evolutionarily, subunits of these complexes are not equally old. Both MEA and FIS2 (both a part of the FIS-PRC2) have only been found in *Brassicaceae* so far. CLF and SWN share functions and seem to act partly redundantly in EMF-PRC2 and VRN-PRC2 (Mozgova et al., 2015b; Mozgova & Hennig, 2015a).

1.2.4.3. PRC2's function in development of Arabidopsis

The role of PRC2 in development is relatively highly conserved throughout distinct eukaryotes such as Arabidopsis and *Caenorhabditis elegans* (Mozgova et al., 2015b; Patel et al., 2012; Chanvivattana et al., 2004). In Arabidopsis, PRC2 is partially responsible for transitions between developmental stages as well as maintenance of the identity of the cells in those stages (Chanvivattana et al., 2004).

PRC2 mediates the transition from the seed to the seedling. That is accomplished by repression of genes that promote late embryo maturation [*ABA INSENSITIVE 3* (*ABI3*), *LEAFY COTYLEDON 1* (*LEC1*), *LEAFY COTYLEDON 2* (*LEC2*) and *FUSCA3* (*FUS3*)], as

well as genes that are responsible for the seed dormancy (Bouyer et al., 2011; Mozgova et al., 2015b). In case of severe PRC2 phenotypes like in the double mutant *clf swn* (the severity will be explained in chapter1.2.4.3.1.), the breakage of dormancy is delayed (Bouyer et al., 2011).

Secondly, PRC2 is responsible for flowering induction following vernalization. EMF-PRC2 complex initiates repression of FLOWERING LOCUS C (FLC), a repressor of floral-meristem identity genes, which enforces the recruitment of VRN-PRC2 with accessory subunits, that furthermore maintains the repression of *FLC* via H3K27me3 (reviewed in Mozgova et al., 2015b; Derkacheva et al., 2013, Bastow et al., 2004; Feng et al., 2010; Gendall et al., 2001).

PRC2 is also important for cell and tissue differentiation. As Lafos et al. (2011) have shown, the H3K27me3 marks are variously distributed across different types of tissues such as the shoot apical meristem (SAM) or leaves. The different distribution of H3K27 can be seen on KNOX-class transcription factor (TF) genes (homeodomain TFs) that are targeted in the leaves but not in the apical meristem tissue. Furthermore, the difference between SAM and specific organs is marked by repression of meristem-identity genes by PRC2 and promotion of the genes specific for different organs by other factors (Gan et al., 2013). A similar situation is in the root where PRC2 suppresses the meristematic genes and thus assists with the differentiation of the tissue (Aichinger et al., 2011).

For all those reasons, PRC2 is an indispensable part of the developmental regulation and mutations of its subunits have often severe phenotype effects. In the next chapter, mutations in the E(z) homologues CLF and SWN will be discussed as this problematics is the core of this study.

1.2.4.4. Specific activity of FIS- VRN- EMF-PRC2 complexes and involvement of CLF, SWN and MEA in these complexes.

The diversity of E(z) homologues CLF, SWN and MEA is closely related to the function of PRC2 type they can be found in.

MEA is only found in the FIS complex. FIS-PRC2 plays a major role in female gametophyte and endosperm development in connection to fertilisation. That is based on experiments revealing that the endosperm over-proliferates in plants with a mutation in FIS-PRC2 regardless of fertilisation. It is also proposed that it is essential for the repression of paternally expressed imprinted genes (PEGs) (Wolff et al., 2011). Before the seed is fertilized the FIS-PRC2 that contains MEA or SWN is responsible for repression of cell division and signalling pathways in the female gametophyte central cell nucleus (2n). At the same time,

VRN- and EMF-PRC2 with CLF or SWN subunit are responsible for the repression of the cell elongation and differentiation in the maternal integuments (2n). After fertilisation, repression is terminated in the newly formed endosperm (3n) (originating from the central cell) and in in the seed coat (2n) (originating from the integuments) (reviewed by Derkacheva & Hennig, 2014; Mozgova et al., 2015b; Wang et al., 2006).

The sporophyte, starting with the embryo-to-seedling transition, is controlled by the EMF- and VRN-PRC2 (Bouyer et al., 2011). By description, EMF-PRC2 is responsible for the transition from the vegetative phase to flowering phase and VRN-PRC2 is affecting the vernalisation process but mutations *emf2* and *vrn2* are genetically additive, suggesting at least partial redundancy. Both complexes can contain CLF or SWN (reviewed in Derkacheva & Hennig, 2014). Their function also appears to be partially redundant. Despite higher expression of the *SWN* gene compared to *CLF* (Iva Mozgova – personal communication), *swn* mutant does not have severe developmental phenotypes but *clf* does (Fig. 1). Chanvivattana et al., (2004) show, that *clf* cannot be complemented with $35S_{pro}$*::SWN* or $35S_{pro}$*::MEA* ($35S_{pro}$ or *CaMV* $35S_{pro}$ is a strong promoter driving the *35S* gene in cauliflower mosaic virus (Somssich, 2018)). The absence of phenotype in *swn* does not mean that all SWN targets are also targeted by CLF. The genes that are affected only by SWN are mostly related to lipid localization and storage and cell-wall modifications and development (Shu et al., 2019). The severity of the phenotype is closely related to the H3K27me3 distribution. *swn* has almost identical H3K27me3 distribution as wild type (WT). On the contrary, *clf* loses a lot of the marks which correlates with the more severe phenotype. Finally, the *swn clf* homozygotic double mutant (*ccss*) shows almost no sign of H3K27 marks (Shu et al., 2019; Wang et al., 2016). Thus, it is not surprising that the double mutant has the most severe phenotype, which is visible soon after germination. The root in a proportion of the mutant plants is short and swollen, accumulating embryonic storage lipids (develops so-called pickle-root phenotype) (Fig. 1). The cotyledons are deformed. SAM does not elongate nor generates organs and eventually forms callus-like structures, that accumulate embryonic lipids (Aichinger et al., 2009; Chanvivattana et al., 2004). This evidence clearly shows that at least one of either CLF or SWN has to be functional in order for the plant to develop into a seedling. However, that appears to be contradicted by the fact that *swn* shows a wild-type-like developmental phenotype. It is possible that there could be a similar mechanism as in the mammalian cells where the E(z) homologues EZH1 and EZH2 have different functions. EZH2 is active in pluripotent cells that actively divide, whereas EZH1 is active in fully differentiated cells (reviewed in Mozgova et al., 2015b).

While the best-well described functions of the PRC2 relate to plant development by targeting developmentally-regulated genes, recent evidence suggests that PRC2 targets also environmentally, metabolically or stress-responsive genes (Bellegarde et al., 2018; Chica et al., 2017; Shu et al., 2019). Recent work in our lab has revealed targeting of PRC2 to light-responsive genes and function in light-regulated plant development (Konečný, Mozgová et al. – manuscript in preparation and personal communication). This thesis builds on this recent work to establish the specific functions of CLF and SWN and to develop a reporter system to differentiate PRC2 activity affecting developmental and light-responsive genes.

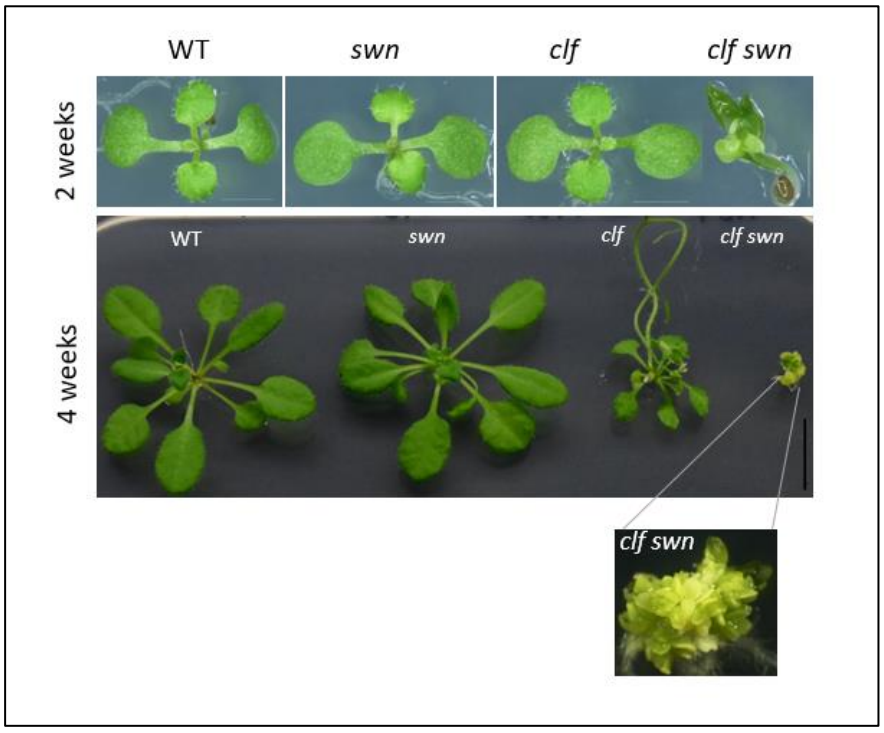

Fig. 1: Comparison of phenotypes of WT, *swn*, *clf* and *clf swn* throughout the post-embryonal development.

1.2.5. The project's hypothesis and questions

1.2.5.1. Development of reporter lines for visualisation of PRC2 activity affecting developmental and light-signalling genes

The first part of the thesis aims to develop binary constructs that will serve to generate transgenic plants used as reporters of PRC2 activity targeting developmental and light-signalling genes. To reach this goal, the task of the thesis work was to identify promoters of targeted genes, design primers and isolate the sequences from genomic DNA.

In addition, primers for amplification of reported genes from available plasmids were to be designed. The amplified fragments were to be subcloned into Gateway entry vectors and verified by Sanger sequencing.

1.2.5.2. Redundancy and tissue-specificity of CLF and SWN

The question of SWN and CLF redundancy and tissue-specific function is the core of this project's aim. As was suggested in the chapter 1.2.4.4., it is possible that the activity of the two homologues could be dependent on the identity of the cell (less differentiated actively dividing cell /differentiated, non-dividing cell), respectively the tissue the cell can be found in. To test this hypothesis, Tomáš Konečný, a PhD student in the lab, developed plasmid vectors for plant transformation, where *CLF* and *SWN* are driven by different promoters. The promoters are either native to one of the two genes, or they are tissue-specific. The selected tissue-specific promoters are *ARABIDOPSIS PUMILIO 10* promoter (*APUM10*$_{pro}$, shortly *APUM*$_{pro}$) active in SAM and *LIGHT-HARVESTING CHLOROPHYLL B-BINDING 2.1* (*LHCB2.1*$_{pro}$, shortly *LHCB*$_{pro}$) active in leaf mesophyll (Klepikova et al., 2016). Tab.1 shows all the constructs designed by Tomáš Konečný that have been tested in this study. Each plasmid [= binary vector (BV)] contains one construct A and one construct B, that are inserted into the plant simultaneously.

Tab.1: Constructs designed by Tomáš Konečný in binary vector R4pGWB6650-MD8 (Aboulela et al., 2017) .

ID	Construct A	Construct B
BV1	*CLF*$_{pro}$*::CLF-tagRFP*	*CLF*$_{pro}$*::CLF-G3GFP*
BV2	*APUM*$_{pro}$*::CLF- tagRFP*	*APUM*$_{pro}$*::CLF- G3GFP*
BV3	*LHCB*$_{pro}$*::CLF- tagRFP*	*LHCB*$_{pro}$*::CLF- G3GFP*
BV4	*SWN*$_{pro}$*::SWN- tagRFP*	*SWN*$_{pro}$*::SWN- G3GFP*
BV5	*APUM*$_{pro}$*::SWN- tagRFP*	*APUM*$_{pro}$*::SWN- G3GFP*
BV6	*LHCB*$_{pro}$*::SWN- tagRFP*	*LHCB*$_{pro}$*::SWN- G3GFP*
BV7	*SWN*$_{pro}$*::CLF- tagRFP*	*CLF*$_{pro}$*::SWN- G3GFP*

BV8	*APUM$_{pro}$::CLF- tagRFP*	*APUM$_{pro}$::SWN- G3GFP*
BV9	*LHCB$_{pro}$::CLF- tagRFP*	*LHCB$_{pro}$::SWN- G3GFP*
BV10	*CLF$_{pro}$::CLF- tagRFP*	*SWN$_{pro}$::SWN- G3GFP*
BV11	*APUMpro::CLF- tagRFP*	*LHCBpro::SWN- G3GFP*
BV12	*LHCBpro::CLF- tagRFP*	*APUMpro::SWN- G3GFP*

The questions asked here are: If *SWN* is driven by *CLF$_{pro}$* will it be able to complement *ccss* and vice versa? In other words: Are the *SWN$_{pro}$* and *CLF$_{pro}$* responsible for the difference between *swn* and *clf* and are the proteins CLF and SWN interchangeable?

For the tissue-specific promoters, we are asking, whether CLF can still complement *ccss* in case it is only expressed in certain tissue. Or if SWN can complement *ccss* if it is expressed in certain tissue. And there is of course the potential of the emergence of new phenotypes that would occur due to partial compensation of *ccss*.

As the binary constructs were made and used for plant transformation by Tomáš Konečný, the goal of this part of the thesis was to select transgenic plant lines using antibiotics resistance and provide their basic characterisation by PCR and phenotype description.

1.2.6. Genetic engineering tools

1.2.6.1. MultiSite Gateway Technology

MultiSite Gateway Technology (MSGT) is one of the most widely used cloning methods today. This technology uses recombination mechanisms naturally used by bacteriophage lambda. After the bacteriophage recognises specific att sequences, it recombines its own DNA into *Escherichia coli* and thus creates a novel sequence. The att sites are partially flanking, which means that when the recombination occurs the att sites are hybrids of the original bacteriophage and *E. coli* sequences. This reaction is mediated by the proteins of both participants. *E. coli* supplies the reaction with Integration Host Factor (IHF) protein. The bacteriophage's protein setup is variable. If the phage produces only Integrase (Int) then it will enter the lysogenic pathway. If it additionally produces Excisionase (Xis) then it will change its pathway to lytic (Chiew Foan Chin, 2015; Invitrogen, 2010; Landy, 1989). This may be bad news for *E. coli*, however, the two pathways allowed researchers to develop the complex genetic engineering tool that MSGT is.

The feature that makes MSGT so useful and efficient is its accuracy given by the specificity of the att sites. The att sites are divided into four categories: attB, attP, attL and attR. The dogma here is, that attB can only pair with attP and attL can only pair with attR. Additionally, the att sequences are specified and tagged with a number. And only the atts with same number can couple. For instance, attB1 only pairs with attP1 (Invitrogen, 2010).

The attB(1) and attP(1) reaction is called BP reaction (Fig. 2) and the outcome of recombination of these two is the attL(1) sequence. Therefore, the BP reaction is directly preceding the LR reaction, a reaction between attL and attR. If the desired outcome of the BP reaction is attR, then the attB(1) and attP(1) must be modified and changed into attB(1)r and attP(1)r (Invitrogen, 2010).

The BP reaction is mediated by *BP Clonase® II Enzyme Mix*, which includes the IHP and Int proteins. The *LR Clonase® II Plus Enzyme Mix* contains additionally Xis. The proteins form a complex that binds both of the strands that are about to be recombined. It cuts the strands; recombines them; and finally, releases them (Invitrogen, 2010).

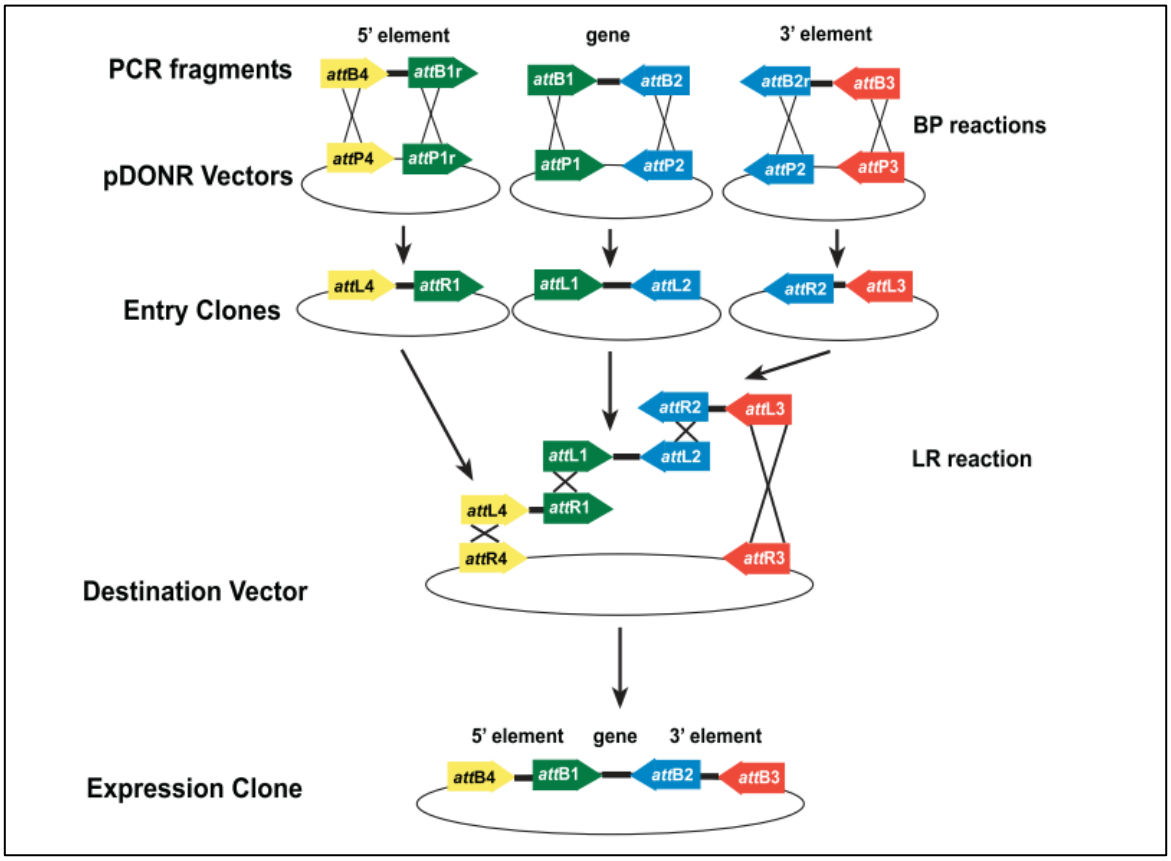

Fig. 2: Overview of MultiSite Gateway Technology's mechanism (Invitrogen, 2010).

The BP reaction is classically a reaction between an attB-flanked PCR product and a donor vector (pDONR) that contains two attP sites. The result of their reaction is called entry

clone. The preparation of attB-flanked PCR product is explained in more details in the chapters 2.2.1.1.-2.2.2.1.2. The pDONR plasmid is predesigned and besides attP sites, it must contain several other necessary sequences. The part that is going to be recombined out during the BP reaction contains the *ccdB* gene that encodes DNA gyrase toxin. This toxin should later kill every cell that contains the non-recombined pDONR. Another important part is a selective marker, Kanamycin resistance (Kan), that should prevent the survival of non-transformed cells on Kanamycin-rich substrate (for more detail see chapter 1.2.6.4.) (Invitrogen, 2010; Miki & McHugh, 2004). In addition, very useful parts of the vector are the sequences complementary to M13 primers. These sequences are placed in the proximity of the att sites but from the outer side of the Gateway cassette. Thus, they do not undergo recombination and can be used for both genotyping and sequencing of the plasmid.

The next step is the LR reaction (Fig. 2). The mechanism is essentially the same as the one of the BP reaction but the two counterparts that enter the reaction are the entry clone and the destination vector. The product of this reaction is called expression vector (Invitrogen, 2010).

This approach traditionally allows the researchers to easily assemble 4 fragments into one expression vector, so the outcome can be for example Promotor-Gene-Marker-Terminator. However, to make more complex assembly such as Promotor1-Gene1-Marker1-Terminator1-Promotor2-Gene2-Marker2-Terminator2 requires a more laborious process and thus is both time and money consuming (Aboulela et al., 2017). It is important to explain that this more complex assembly is highly desired for experiments like the ones in this study where the researchers want to introduce two constructs into plants without having to establish two single-construct plant lines and carry out genetic crosses to bring the two constructs together in one transgenic plant line.

1.2.6.2. Gateway technology-compatible binary vector system

For the purpose described above, Aboulela et al. (2017) designed Gateway technology-compatible binary vector system. The two vectors designed here are R4 Destination Donor (R4DD) and its destination vector R4 Dual-Site vector (R4DS). The binary vectors already contain reporter genes and terminators, which means that only the promoter and gene of interest must be added (in the simplest form of use).

The beginning of the cloning process is the same as described in the chapter about MultiSite Gateway Technology. Two vectors, one with the promoter and one with the gene of interest, that are created by BP reaction are recombined into R4DD vector in the 1st LR reaction

(Fig. 3A). The newly created vector then proceeds into the 2nd LR reaction alongside another two vectors, one with the promoter one with the gene of interest. These three vectors are recombined into the expression vector R4DS (Fig. 3B).

The newly formed vector then contains promotor1-gene1-marker1-terminator1-gap sequence-promotor2-gene2-marker2-terminator2. That all is a part of a one T-DNA that will be integrated into the plant genome (closely described in the next chapter) and will be as well inherited together as one allele (Aboulela et al., 2017).

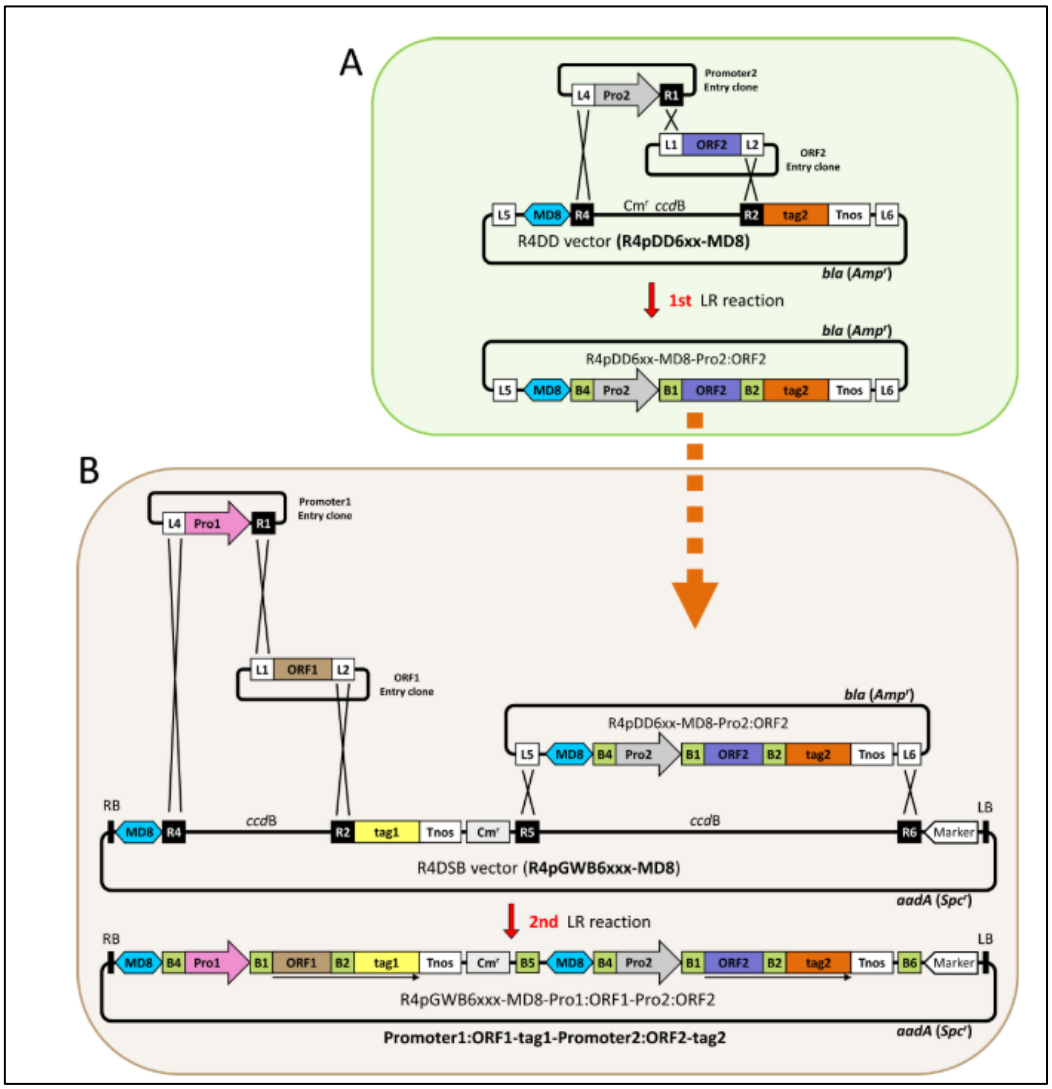

Fig. 3: Overview of Gateway technology-compatible binary vector system (Aboulela et al., 2017). A is the scheme of the 1st LR reaction and B of the 2nd one.

1.2.6.3. *Agrobacterium*-mediated plant transformation

Agrobacterium tumefaciens-mediated plant transformation is a commonly used tool in plant genetic engineering. The reason why *Agrobacterium* is fit for genetic engineering is that it has a natural mechanism for the transfer of genetic material (DNA) between the bacterial and plant cells. As a result, the bacterium transforms the plant cells to start over-proliferating, causing the crown gall disease and enabling the bacteria to amplify in a nutrition-rich environment (Hwang et al., 2017).

A. tumefaciens lives in soil and it responds by chemotaxis to acetosyringone molecules that are released from a wound on a plant that is close to the ground and bind to a periplasmic dimer virA. Furthermore, the bacterium activates virG which then activates transcription of the rest of the vir-box, a segment of a tumour-inducing plasmid (Ti plasmid). One of the translated proteins, virD, then nicks the T-DNA. The T-DNA is a region of the plasmid that is defined by two boundaries, left (LB) and right (RB), that are approx. 24-28 bp long. In nature, the T-DNA carries genes that the plant itself incorporates into its own DNA. The incorporation is randomly placed and occurs because the plant's repair mechanisms recognise microhomologies between the T-DNA and its own DNA. The important genes are those encoding proteins required for the synthesis of plant hormones and opines, amino acid-related substances, which are *Agrobacterium*'s source of nutrition. The plant hormones auxin and cytokinin are responsible for the induction of the crown gall tumour. This tumour is composed of cells that synthesize an opine - octopine, nopaline, succinamopine or agropine, thus becoming a factory for bacterial nutrition. (Frame et al., 2002; Hwang et al., 2017).

In the strains used for research, the natural T-DNA is removed and is replaced with DNA sequence of interest. This new T-DNA must also contain LB and RB. Vector with the T-DNA is introduced into the *A. tumefaciens* which then uses its natural mechanisms described above to introduce this T-DNA into the plant. There are two categories of transformation. The first is transient transformation. Here the T-DNA is not integrated into the genome, so the expression is not permanent (usually only lasts a few days). The second category is stable transformation, which has been also used in this thesis. In this case, the *A. tumefaciens* is introduced into germinal line cells via a method called floral dip. The plants grown from seeds formed from infected germinal cells are then transgenic. That means that the T-DNA is present in all cells of the plant (Hwang et al., 2017). The R4DSB vector (Aboulela et al., 2017) is an example of such a vector and its T-DNA framed by LB and RB can be seen in Fig. 3.

1.2.6.4. Marker genes

Marker genes are an important tool in experimental biology. They allow examining the outcome of the experiment quickly and accurately. Marker genes have variable use.

They can be used for the selection of transformants either as positive or negative selectors. Positive selection means that the transformed organism has a beneficial trait that helps it survive. Negative selection is the opposite - the transformed organism has a disability that eventually terminates its life. Both can be conditional, dependent on a substance in the environment, or non-conditional, independent of the environment. The selection markers that were used in this study are described in Tab.2 (Miki & McHugh, 2004).

Tab.2: Selective markers used in this study and their features.

Selective substance	**Type of substance**	**Type of selection**	**Enzyme**	**Gene**	**Origin**
Phosphinotricin (PPT)	Herbicide	Conditional-Positive	Phosphinotricin acetyltransferase	*bar*	*Streptomyces hygroscopicus, Streptomyces viridochromogenes,* Tu494
Kanamycin (Kan)	Antibiotics	Conditional-Positive	Phosphotransferases	*aphA2*	*Escherichia coli* Tn5

The second important group are non-selective genes, usually called reporter genes. These genes do not affect the survival of the transformed organism, but they help to identify survivors that escaped the selection (false positives) and, perhaps more importantly, they help to localize the transformed cells or visualize promoter activity or protein presence in cells or tissues. Again, the reporter gene function may be conditional or non-conditional. The need for the presence of an external substance is an important feature of the selection of a reporter gene. While the visualization of some reporters (e.g. GUS) is destructive for the plant, other reporters can be used for in-vivo live imaging (e.g. GFP). Tab. 3 shows all the reporter genes used in this study and their characteristics in more detail (Miki & McHugh, 2004).

Tab.3: Reporter markers and their features.

Reporter (Aboulela et al., 2017)	**Type**	**Substrate**	**Function**	**Origin**	**Ex / Em λ (nm) ***	**Recognition of transformed tissue**	**References**
Beta-glucuronidase (GUS)	Conditional Destructive	4-methyl umbelliferyl glucuronide (MUG)	Measurement of specific activity	uidA - *Escherichia coli*	-	Blue pigmentation	(Miki & McHugh, 2004)
		5-bromo-4-chloro-3-indolyl glucuronide (X-gluc)	Histological location				
G3 green fluorescent protein (G3GFP)	Non-conditional Non-destructive	None	Histological and intracellular location; Selection	cFP484 - *Clavularia sp.*	498/515	Green fluorescence	(Lambert, 2019; Miki & McHugh, 2004)
Tag red fluorescent protein (TagRFP)	Non-conditional Non-destructive	None	Histological and intracellular location; Especially localization of proteins (due to increased longevity)	eqFP578 - *Entacmaea quadricolor*	555/584	Red fluorescence	(Lambert, 2019; Merzlyak et al., 2007)
Enhanced yellow florescent protein (EYFP)	Non-conditional Non-destructive	None	Histological and intracellular location; Selection	avGFP - *Aequorea victoria*	513/527	Yellow fluorescence	(Lambert, 2019; Miki & McHugh, 2004)
Luciferase (LUC)	Conditional Non-destructive	Luciferin	Continuous monitoring of gene activity during development	LUC - *Photinus Pyralis*	-	Yellow/green glow in dark	(Miki & McHugh, 2004)

* Feature of the fluorescent markers only.

Tab. 3 shows all the important attributes of each of the reporter gene. However, note that the functions can be more extensive, and the presented functions are only the most valuable ones according to Milky & McHugh (2004).

In this study, the marker genes are used in two variants with slightly different purposes. The first is PROMOTER-MARKER (transcriptional fusion) and the second is PROMOTER-GENE-MARKER (translational fusion). The less complex of the two options is designed to report the location and extent of the activity of a particular promoter. This construct should not affect the phenotype of a plant. The second variant with three components may additionally affect the phenotype and the marker is here to (i) control for the expression of the transgene and (ii) visualize the protein of interest.

1.2.7. The promoter-marker vector design

To further develop the tested project designed by Tomáš Konečný, 14 promoters of genes that are upregulated in PRC2 mutant plants (*clf swn*) and some of which are directly targeted by PRC2 were selected (Mozgova et al. personal communication). These genes are either related to the embryo-to-seedling transition during plant development: *ABI3, LEC1, LEC2* and *FUS3*. Or to light-signalling and photosynthesis *ELONGATED HYPOCOTYL 5* (*HY5*), *LONG HYPOCOTYL IN FAR-RED* (*HFR1*), *EARLY LIGHT-INDUCIBLE PROTEIN* (*ELIP1*), *EARLY LIGHT-INDUCIBLE PROTEIN 2* (*ELIP2*), *FLUORESCENT IN BLUE LIGHT* (*FLU*), *GENOMES UNCOUPLED 5* (*GUN5*), *ABA INSENSITIVE 4* (*ABI4*), *NONPHOTOCHEMICAL QUENCHING 4* (*NPQ4*), *RUBISCO SMALL SUBUNIT 1B* (*RBCS1B*) and *COPPER RESPONSE DEFECT 1* (*CHL27*). Functions of the selected genes as well as the locations with the highest rates of mRNA of each of these genes are described in Tab. 4.

The aim of this part was to create a set of promoter and marker constructs that can be arbitrarily combined and once inserted into the plant they would show the location and the level of transcription from the promoters of the native genes, but without changing the phenotype.

Tab.4: Selection of genes affected by PRC2 activity.

Gene	Type	Functions (related to light-signalling or/and development)	High occurrence of its mRNA
HY5	Light signalling and development	Transcription factor of basic leucine zipper (bZIP) which affects photomorphogenesis (reviewed in Chen et al., 2021). Response to light induced abiotic stress. (Ulm et al., 2004; reviewed in Chen et al., 2021)	Germinating seed, Pod of 1st silique, Mature flower (Klepikova et al., 2016).
HFR1	Light signalling and development	Transcription factor that recruits HAC1, a co-activator of AG, and thus affects flower development. Connects circadian rhythm and flower development (Duren et al., 2019)	Mature Flower, Mature leaf, Silique, 1st internode (Klepikova et al., 2016).
ELIP1	Light response	Chloroplast protein that affects germination in stress conditions (salinity, temperature, light) (Rizza et al., 2011).	Germinating seed, Mature flower Klepikova et al., 2016).
ELIP2	Light response	Chloroplast protein that affects germination in stress conditions (salinity, temperature, light) (Rizza et al., 2011).	Germinating seed, 8th flower abscission Klepikova et al., 2016).
CHL27	Photosynthetic pigment biogenesis	Accumulates substrate for formation of protochlorophyllide, a chlorophyll precursor (Tottey et al., 2003).	Young leaf, Cotyledon, Silique, Young flower, Germinating seed, Pedicel (Klepikova et al., 2016).
FLU	Photosynthetic pigment biogenesis, light-signalling pathway	Repression of 5-aminolevulinic acid (ALA), a precursor of the tetrapyrroles, in the dark. Regulation of the glutamil-tRNA reductase (GluTR), a tetrapyrrole biosynthesis limiting factor (Hou et al., 2019)	Cotyledons, Young leaf, Intermediate leaf 1, Young flower (Klepikova et al., 2016)
GUN5	Photosynthetic pigment biogenesis, retrograde signalling	Affects biosynthesis of tetrapyrroles. Affects transcription of Acetyl-CoA Carboxylase-2 (ACC2), a factor that influences plant growth (rewieved in Wang et al., 2018).	Young leaf, Cotyledon, Silique, Young flower, Germinating seed, Pedicel (Klepikova et al., 2016).
LEC1	Embryo development	A key inducer of somatic embryogenesis (Stone et al., 2001).	8th flower abscission, Silique (Klepikova et al., 2016).
LEC2	Embryo development	A key inducer of somatic embryogenesis (Stone et al., 2001).	8th flower abscission, Silique (Klepikova et al., 2016).
FUS3	Embryo development	Affects morphogenesis and maturation during seed development (Stone et al., 2001).	8th flower abscission, Silique (Klepikova et al., 2016).
ABI3	Embryo development	Transcription factor that affects the seed maturation (Stone et al., 2001).	Early germinating seed, Silique, Seeds of 1st silique, Dry seeds
ABI4	ABA response	Positively regulates dormancy of the seeds and negatively greening of cotyledon (Shu et al., 2013)	Early germinating seed (Klepikova et al., 2016).
NPQ4	Photosynthesis	necessary for correct function of Eq, the most important part of non-photochemical quenching (Dall'Osto et al., 2014).	Young leaf, Cotyledons, Young flower, Silique, Germinating seed (Klepikova et al., 2016).
RBCS1B	Photosynthesis	Partially responsible for RUBISCO holoenzyme accumulation and thus sufficient CO_2 assimilation (Izumi et al., 2012).	Cotyledon, Early leaf, Silique, Early flower (Klepikova et al., 2016).

2. MATERIALS AND METHODS

2.1. MATERIALS

2.1.1. Organisms and plasmids

Organisms:

Arabidopsis thaliana:
wild type (WT): ecotype *Columbia* (*Col-0*)
mutant: *clf-29/+ swn-3/-; CLFpro::CLF-GR* (CLF protein fused to the glucocorticoid receptor "GR") -*BAR* control*; clf-29; swn-3*;
Growing conditions: long-day (16/8 h day/night, 100-120 $\mu mol.m^{-2}.s^{-1}$, 21°C)
Agrobacterium tumefaciens – strain GV3101
Escherichia coli (*E.coli*) – strain TOP10

Plasmids:

pDONR-P4P1r
pDONR221 (containing P1P2)

2.1.2. Chemicals

water; EDTA (#39761.02, Serva); TRIS (#252859, Sigma Aldrich); NaCl (#3957.1, Roth); Sodium dodecyl sulfate (=SDS) (#11667289001, Sigma Aldrich); DreamTaq green buffer (#14966123, Thermo Fisher Scientific); GreenTaq polymerase (#EP0701, Thermo Fisher Scientific); dNTPs (#R0191, Thermo Fisher Scientific); Phusion polymerase (#F530S, Thermo Scientific); Phusion HF buffer (#F518L, Thermo Fisher Scientific); TAE buffer (#42549.01, Serva); Agarose (#11404.03, Serva); GelRed (#41003, Biotium); GeneRuler 1kb DNA ladder (#SM0311, Thermo Fisher Scientific); DNA Gel Loading Dye 6x (#R0611,Thermo Scientific; #B7024S, BioLabs); Ethanol 96% (#70390, Penta); Glufosinate-Ammonium (=Phosphinotricin, PPT) (#45520, Riedel-de Haen); Kanamycin (#K0126.0010, Duchefa); HCl (#10033-A35, Lachner); NaOH (#740609, Macherey-Nagel); Gateway™ BP Clonase™ Enzyme mix (=BP Clonase) (#11789013, Thermo Fisher Scientific); Proteinase K (#AM2546, Thermo Fisher Scientific); Silwet L-77 (AgroBio Opava); Substral Careo ultra (SCOTTS); Plant agar (#P1001, Duchefa); LB Broth (#L1703, Duchefa); Murashige & Skoog

medium including vitamins (=MS) (#M0255, Duchefa); sucrose; LB Broth with agar (#L2897, Sigma Aldrich)

2.1.3. Kits

Thermo Scientific™ MagMAX™ Plant Genomic DNA Kit (#15650969, Thermo Fisher Scientific); GeneJET Plasmid Miniprep Kit (#K0503, Thermo Fisher Scientific); NucleoSpin® Gel and PCR Clean-up XS (#740611, Macherey-Nagel)

2.1.4. Prepared buffers and media

TE buffer – 12,114 g of TRIS (121 g/mol) were dissolved in 100 ml of water and the pH was adjusted to 8 by HCl. 18.6 g of EDTA (186 g/mol) were dissolved in 200 ml of water and the pH was adjusted to 8 by NaOH. 5 ml of 1 M TRIS and 1ml of 0,5 M EDTA were mixed with 496 ml of water and stored at -20 °C (Aitken, 2016).

DNA-extraction buffer based on Edwards et al. (1991) (Edwards buffer) – 0.93 g of EDTA (186 g/mol), 2.42 g of TRIS (121 g/mol) and 1.45 g of NaCl (58.44 g/mol) were dissolved in 100 ml and stored at room temperature.

Liquid LB – 6.25 g of LB Broth were dissolved in water in 0.5 l flask, autoclaved and stored at 8 °C.

Solid LB – 12.5 g of LB Broth with agar were dissolved in water in 0.5 l flask, autoclaved and stored at 8 °C.

1/2MS + 1% sucrose – 4 g of plant agar, 4.9 g of MS and 20 g of saccharose are mixed in 2 l of water, adjusted to 8,3 pH, autoclaved and stored at 8 °C.

2.1.5. Tools and machines

Pipettes [(20 µl, 200 µl, 1000 µl), Eppendorf Research Plus; 10 µl, DISCOVERY Comfort]; Laminar hood (SCS 1-5, MERCI; HB2448, Holten LaminAir; OSN-5, MERCI); dentist machine (ivoclar vivadent, SILAMAT); Thermo-Shaker (TS 100C, BioSan); Thermal Cycler (T100™ Thermal Cycler, BIO RAD); scales (IMB, KERN; 770, KERN); microwave (SAMSUNG); electrophoresis tanks (Cleaver Scientist); electrophoresis power supply (Power Pac Basic, BIO RAD); ChemidDoc™MP Imaging System (BIO RAD); Spectrophotometer (NanoDrop™ ND-1000, Thermo Scientific); minishaker (MS1; IKA); shaking incubator (NB-205; N-BIOTECH. INC); biological thermostat (BT120, Laboratorní přístroje Praha); centrifuge (Z 216 MK, HERMLE; UNIVERSAL 320 R, Hettich ZENTRIFUGEN);

minicentrifuge (M-6, BOECO), falcon tubes (50ml; 15ml); Eppendorf tubes (1.5 ml; 2 ml); tweezers; stratification box; Petri plates; glass beads (ROTH)

2.1.5.1. Computer tools

AmplifX 2.0.7 (Jullien, 2020); Primer3 4.1.0 (Untergasser; Koressaar ; Koressaar); Geneious 2020.2.5.; IDT OligoAnalyzer™ Tool

2.2. METHODS

2.2.1. Primer design

The primers were developed according to the common rules described in Appendix 1. Sometimes the primer length and CG content had to be sub-optimal because primers for cloning must be placed in a specific region and if this region does not contain the optimal amounts of CG, the only two solutions are to make the primer longer or to leave the lower CG content. The tools used for primer development and quality assessment were Primer3 and IDT OligoAnalyzer™ Tool.

Different primer design approaches were applied in the case of primer for amplifying the promoter or the marker gene. For promoter, the source DNA sequence information was the TAIR10 assembly of the *Arabidopsis thaliana* genome (Berardini et al., 2015). The optimal reverse primer would start immediately in front of the gene start codon and the forward primer would be placed 1.5-2 kb upstream of the start codon. However, this was not always possible, so the longest possible region had to suffice. The length was limited by the proximity of upstream genes.

In the case of the marker genes, the sources of sequences were R4DD and R4DS vectors (Aboulela et al., 2017) that include the desired marker gene. Primers had to be designed to target the start codon and the stop codon of the marker gene (Dieffenbach et al., 1993; Onodera & Melcher, 2004).

2.2.1.1. Primer adjustment

Primers were enriched with part of attB sequence. Promoter primers were designed to be recombined into the pDONR-P4P1r plasmid and thus contained part of attB4 on the 5'-end of the forward primers and with part of attB1r on the 5'-end of the reverse primer. The primers of the markers that were designed to be recombined into pDONR221 (plasmid that contains

attP1 and attP2 sites) contained part of attB1 on the 5'-end of the forward primer and with part of attB2 on the 5'-end of the reverse primer (Invitrogen, 2010).

2.2.2. Polymerase chain reaction (PCR)

2.2.2.1. PCR for cloning

This PCR had to be divided into two separate PCRs (PCR1 and following PCR2). The reason for two consecutive PCR reactions is that a primer comprised of both the promoter/gene-specific primer and the full att site would be very long, increasing the possibilities of mutations during primer manufacturing and possibly affecting the efficiency of first PCR cycles.

2.2.2.1.1. PCR1

The PCR that precedes cloning must be highly accurate and high processivity is also desirable. That is why Phusion (Phu) DNA polymerase with proof-reading activity and its High-fidelity buffer (HF) were used (for further information about PCR and the criteria of choice see Appendix 2). 50 µl reaction using Phu DNA polymerase was mixed according to the manufacturer´s instructions (Thermo Scientific, 2020) (Tab. 5). The reaction mix was prepared on ice. DMSO was not added. PCR cycling conditions followed the manufacturer´s recommendations (Thermo Scientific, 2020) (Tab. 6).

Tab. 5: Master mix preparation for PCR with Phusion Polymerase.

Component	**50 µl Reaction**
Nuclease-free water	to 50 µl
5X Phusion HF or GC Buffer	10 µl
10 mM dNTPs	1 µl
10 µM Forward Primer	2.5 µl
10 µM Reverse Primer	2.5 µl
Template DNA	variable
DMSO (optional)	(1.5 µl)
Phusion DNA Polymerase	0.5 µl

Tab. 6: PCR with Phu polymerase program.

Step	Temperature (°C)	Time	N of cycles
Initial denaturation	98	30''	-
Denaturation	98	10''	35
Annealing	Primer dependent	20''	
Extension	72	20''/kb	
Final extension	72	8'	-
Hold	4	Infinite	-

The annealing temperatures were adjusted based on the individual primer melting temperatures as calculated by the AmplifX 2.0.7. (Jullien, 2020) program. The efficiency of the amplification conditions was analysed using agarose electrophoresis as described in chapter 2.2.3. The products of PCR1 were stored at 4 °C or at -20 °C.

2.2.2.1.2. PCR 2

The second PCR (PCR2) served to attach the remaining parts of the att sites to the product of PCR1. PCR2 was performed in the same way as PCR1. However, this time PCR1 product served as the template and primers carrying the sequence of att sites were used. The amplification results were analysed via agarose electrophoresis as described in chapter 2.2.3.

2.2.2.2. PCR for genotyping and screening

This PCR was run with GreenTaq Polymerase, which does not have a proof-reading activity and has lower processivity than Phu Pol. Master mix and PCR program were prepared according to the user manual (Thermo Scientific, 2016) (Tab. 7 and Tab. 8).

Tab. 7: Master mix preparation for PCR with DreamTaq.

Component	**20 µl Reaction**
Nuclease-free Water	to 20 µl
10X DreamTaq Green Buffer	2 µl
10 mM dNTPs	0.4 µl
10 µM Forward Primer	0.5 µl
10 µM Reverse Primer	0.5 µl
Template DNA	variable
DreamTaq Polymerase	0.1 µl

Tab.8: PCR with DreamTaq polymerase program.

Step	Temperature (°C)	Time	N of cycles
Initial denaturation	98	3’	-
Denaturation	98	30’’	35
Annealing	Primer dependent	30’’	
Extension	72	1’/kb	
Final extension	72	8’	-
Hold	4	Infinite	-

2.2.3. Agarose gel electrophoresis

A 1% agarose gel was prepared from a mixture of agarose and 1xTris-acetate-EDTA (TAE) buffer and GelRed (2.5 µl/50 ml). 5 µl of the PCR1 product were mixed with 1 µl of DNA 6x Gel Loading Dye, loaded into the wells of the agarose gel placed in an electrophoresis tank (Cleaver Scientific, n.d.) and separated at 5-10 V/cm for 45 min. The GeneRuler 1kb DNA Ladder was used as a DNA fragment size marker. The result was documented using the bottom UV illumination mode on the ChemiDoc gel documenting system (Bio-Rad).

2.2.4. DNA isolation

2.2.4.1. DNA isolation from leaf (for genotyping)

A fresh sample of a leaf was immediately after collection placed into an Eppendorf tube containing 250 µl of Edwards buffer and 2-4 glass beads. It was placed into the bead-beater (homogeniser) and shaken vigorously for 15-25 sec. Further 250 µl of Edwards buffer were added and the tube was spun for 5 min at 13 500 rpm. 400 µl of supernatant without leaf residues were transferred into a new tube and 400 µl of isopropanol were added. The tube was again spun at 13 500 rpm for 5 min The supernatant was decanted, and the pellet was left to dry for 45 min with the tube placed upside down on a lab bench. 100 µl of water were added after all isopropanol evaporated. The sample was either immediately used as a template for PCR or stored at -20 °C (laboratory protocol based on Edwards et al., 1991).

2.2.4.2. DNA isolation from leaf (for cloning)

WT plant sample was frozen with liquid nitrogen and crushed to a fine powder in a mortar. 100 mg of sample was transferred into an Eppendorf tube with 350 µl of Lysis Buffer A from MagMAX™ Plant DNA Kit. User guide for manual purification (Thermo Fisher, 2016), where the DNA is first bound to magnetic beads, then RNAse A-treated, purified and eluted into TE buffer, was followed exactly without any modifications. The DNA samples were either used immediately as a template for PCR or were stored at -20 °C.

2.2.4.3. DNA isolation from agarose gel

A segment of the agarose gel that contains a product of required length was cut out. The sample was weighed and mixed with Buffer NTI from NucleoSpin® Gel and PCR Clean-up XS in ratio 1 mg: 2 µl. The manufacturers manual was followed (Macherey-Nagel, 2021). The purified product was stored at -20 °C.

2.2.5. Gateway cloning (Invitrogen)

2.2.5.1. BP reaction

The concentrations (ng/µl) of fresh products of PCR2 were measured using NanoDrop™ ND-1000 (Thermo Fisher Scientific) spectrophotometer. Then the weight (ng) of 50 fmol was calculated for each of the amplicons. The amount of 50 fmol is recommended by the manufacturer for BP reaction. It is calculated using the following equation where *N* represents the length of the att-flanked product in bp:

$$ng = (fmol)(N)(\frac{660\ fg}{fmol})(\frac{1\ ng}{10^6\ fg})$$

2 µl of the 25 fmol of the att-flanked product were mixed with 0.5 µl of the pDONR plasmid (150 ng/µl) (later the concentration of att-flanked product was increased due to low efficiency of the reaction as described in Results – chapter 3.2.). 1.5 µl of TE buffer was added to the reaction. A 5 µl aliquot of BP Clonase Enzyme Mix was taken from -80 °C and placed on ice to thaw for less than 2 minutes. It was quickly 2x vortexed and spun down. 1 µl was added into the reaction mix shortly and vortexed 2 times, spun down and placed in the thermoshaker set to 25 °C. The BP reaction incubated for different times - 1 hour, 3 hours and overnight, depending on the efficiency of the BP reaction (for closer information, see results

chapter 3.2.). The BP reaction was terminated by addition of 0.5 µL of proteinase K, vortexed, and left for 10 min at 37 °C. The reaction was stored at 4 °C until the next step (Invitrogen, 2010).

2.2.5.2. *Escherichia coli* transformation, screening and culture preparation for plasmid isolation

1 µl of BP reaction was added to 50 µl of TOP10 competent cells placed on ice and the mixture was left to thaw. After 1 h, the sample was incubated in the thermoshaker for 1 min at 42 °C. Then the sample was transferred back on the ice and 500 µl of cooled liquid LB was added. The sample was incubated for 2-3 h at 37 °C. In the next step, 75 µl and 150 µl of the sample were spread on plates containing 25 ml of solid LB medium and Kanamycin (c = 50 µg/ml) at room temperature. The samples were incubated at 37 °C overnight.

The next day, samples from several colonies from each plate were collected and thoroughly mixed with 10 µl of water. 1 µl from each was used as a template for PCR for genotyping using M13 primers and DreamTaq polymerase described in chapter 2.2.2.2. (Thermo Scientific, 2016).

Positively tested colonies were spread on a new solid LB with Kan (c = 50 µg/ml) and incubated at 37 °C overnight. The next day a part of each colony was transferred into 5 ml of liquid LB with Kan (c = 50 µg/ml) and incubated overnight at 37 °C and 150-250 rpm.

2.2.5.3. Plasmid isolation

Each of the cultures grown in liquid LB was centrifuged at 8000 rpm for 2 min at room temperature. LB was decanted and cells were resuspended with 250 µl of the Resuspension solution from GeneJET Plasmid Miniprep Kit (Thermo Scientific). The user manual for the kit was followed exactly. The only modification was done at the end of the purification process where the elution was performed twice with 25 µl of the Elution buffer instead of once with 50 µl (Thermo Scientific, 2012).

2.2.5.4. Sanger sequencing

The concentration of the purified plasmid sample was measured using the NanoDrop™ ND-1000 (Thermo Fisher Scientific) spectrophotometer. The concentration of the sample was adjusted to 80 – 100 ng/µl. 5µl of the sample were mixed with 5 µl of one of the M13 primers. The samples were sent for Sanger sequencing:

Eurofins Genomics: https://eurofinsgenomics.eu/en/custom-dna-sequencing/. Results of the sequencing were analysed using multiple alignment analysis tool in Geneious 2020.2.5.

2.2.5.5. Floral dip

The floral dip was performed according to (Logemann et al., 2006). Cells of transformed *Agrobacterium t.* were cultivated in Petri plates and resuspended in 30 ml of liquid LB. The bacterial solution was poured inside a plastic bag with 120 ml of freshly prepared 5% sucrose solution supplemented with 0.03% Silwet L-77. Flowers were dipped into the plastic bag. The plants were removed from the plastic bag after 1 min. Transformed plants were grown at long-day conditions (16/8 h day/night, 100-120 $\mu mol.m^{-2}.s^{-1}$, 21°C) in a growth chamber. Dry seeds were collected and stored at room temperature.

2.2.5.6. Seed sterilization and growth on selection substrate

Seeds of transformants and WT control were surface sterilized using 70% Ethanol (EtOH) for 15 min, followed by 96% EtOH for 10 min. After EtOH removal, seeds were left to dry inside a laminar hood. 900 ml of 1/2 MS supplemented with 1% sucrose were melted and after cooling down mixed with Phosphinotricin (PPT) (final c = 10µg/ml). The gel was distributed into Petri plates - 25 ml to each. After the gel solidified, the dry sterile seeds were spread on the plate. The Petri plates were sealed and put to stratify at 8°C for 4 nights. For the controls, the selection antibiotics were not used. Next, the Petri plates were placed in the growth chamber in long-day conditions (16/8 h day/night, 100-120 $\mu mol.m^{-2}.s^{-1}$, 21 °C). After two weeks the Petri plates were evaluated for the presence of the survivors. Representative plates of all categories (germinated and died / germinated with survivors / didn't germinate) that were found within each transgenic line were photographed. Survivors were replanted into watered, gently sterilized soil and placed into the growth chamber in long-day conditions (16/8 h day/night, 100-120 $\mu mol.m^{-2}.s^{-1}$, 21 °C). After 2-3 days the plants were sprayed with systematic insecticide Substral Careo ultra.

2.2.5.7. Statistical methods for determination of randomness of survivors' distribution

Each dataset analysed comprised all survivors grown from seeds generated by a single maternal plant and all the plates these seeds were sown on.

2.2.5.7.1. Comparison of variation and median value

Mean value ($\bar{x}$) of each dataset was calculated. $\sum_{i=1}^{n} x_i$ is the sum of all survivors and n is the number of plates.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Variation (σ^2) was calculated.

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

Median (μ) was calculated. The results were analysed. If $\sigma^2 < \mu$ then the survivors were distributed evenly across all the plates. If $\sigma^2 = \mu$ then the distribution was random. If $\sigma^2 > \mu$ then the survivors were distributed non-randomly and there was tendency for grouping. The bigger the difference between σ^2 and μ the less random the distribution was (Lepš & Šmilauer, 2016).

2.2.5.7.2. Poisson probability

Probability (P) of occurrence of the exact number of survivors ($x = X$) on each selective plate was calculated. λ represents an average number of survivors (Lepš & Šmilauer, 2016).

$$P_{(x=X)} = \frac{e^{-\lambda} \lambda^X}{X!}$$

3. RESULTS

3.1. Development of reporter lines for visualisation of PRC2 activity affecting developmental and light-signalling genes

Primers for the 14 selected promoters and 5 marker genes were designed and part of att-site was added at the 5'-end of each of the primers (Tab. 9). The annealing temperatures suggested by the AmplifX software were tested (Tab. 9).

Tab. 9: The promoter/marker-specific primers.

Gene name	Gene ID*	Type	F/R	Sequence (5'-3') of promoter primer fused with part of att-site (in capitals)	Primer ID	Annealing T(°C)
HY5$_{pro}$	*AT5G11260*	Promoter	F	*AGAAAAGTTGCTtcctcacctgatccaagtc*	AT_498	64.4
			R	*ACAAACTTGCttttcttactctttgaagatcgatc*	AT_499	
HFR1$_{pro}$	*AT1G02340*	Promoter	F	*AGAAAAGTTGCTtgtagtggctgtggttttg*	AT_500	64.4
			R	*ACAAACTTGCgttagttaaagagatatcggagatg*	AT_501	
ELIP1$_{pro}$	*AT3G22840*	Promoter	F	*AGAAAAGTTGCTgggaacaaaatctaaagaa*	AT_502	63
			R	*ACAAACTTGCttctaaagcttagaactactagtg*	AT_503	
ELIP2$_{pro}$	*AT4G14690*	Promoter	F	*AGAAAAGTTGCTgcaaattcggccaaaactaa*	AT_504	64.4
			R	*ACAAACTTGCttctgattaggttttctaaaagccg*	AT_505	
CHL27$_{pro}$	*AT3G56940*	Promoter	F	*AGAAAAGTTGCTgttgatgagagtggaaagag*	AT_506	64.4
			R	*ACAAACTTGCtgactgtgagttgcagaag*	AT_507	
FLU$_{pro}$	*AT3G14110*	Promoter	F	*AGAAAAGTTGCTcaaaagctgagccttctgt*	AT_508	64.4
			R	*ACAAACTTGCcactaaagaaagctctctgagag*	AT_509	
GUN5$_{pro}$	*AT5G13630*	Promoter	F	*AGAAAAGTTGCTtcctcatcgtgcaccacctc*	AT_510	70
			R	*ACAAACTTGCtttgcggctgctggattctccaaac*	AT_511	
LEC1$_{pro}$	*AT1G21970*	Promoter	F	*AGAAAAGTTGCTaccaattcaccgcctccta*	AT_512	64.4
			R	*ACAAACTTGCtgtttctctgccgtctttttt*	AT_513	
LEC2$_{pro}$	*AT1G28300*	Promoter	F	*AGAAAAGTTGCTtccatcaacccatgcctc*	AT_514	67.3
			R	*ACAAACTTGCttttcccggagagagagaga*	AT_515	
FUS3$_{pro}$	*AT3G26790*	Promoter	F	*AGAAAAGTTGCTgccacttgtccatgcaaaga*	AT_516	67.3
			R	*ACAAACTTGCttttctctctcaattggttaacactgc*	AT_517	
ABI3$_{pro}$	*AT3G24650*	Promoter	F	*AGAAAAGTTGCTtgccatctcgacgtacaa*	AT_518	64.4
			R	*ACAAACTTGCcgttgaagtggaaatgaaaca*	AT_519	
ABI4$_{pro}$	*AT2G40220*	Promoter	F	*AGAAAAGTTGCTcatttagatcttttactagggttg*	AT_520	64.4
			R	*ACAAACTTGCagatgaagaagaagaagaagaag*	AT_521	
NPQ4$_{pro}$	*AT1G44575*	Promoter	F	*AGAAAAGTTGCTcgactggttgagcgtttgat*	AT_522	67.3
			R	*ACAAACTTGCtctttctgaggatgagagaagga*	AT_523	
RBCS1B$_{pro}$	*AT1G67090*	Promoter	F	*AGAAAAGTTGCTtcccacatcgcttaaaaa*	AT_524	63
			R	*ACAAACTTGCtacttcttcttcttcttttgc*	AT_525	
G3GFP	-	Marker	F	*AAAAAGCAGGCTTAatgagtaa--aggagaagaacttttcact*	UP_29	62
			R	*AGAAAGCTGGGTAttatttgtatagttcatccatgcca*	UP_30	
TagRFP	-	Marker	F	*AAAAAGCAGGCTTAatggtgtctaagggcgaag*	UP_27	63
			R	*AGAAAGCTGGGTAtcaattaagtttgtgcccca*	UP_28	
EYFP	-	Marker	F	*AAAAAGCAGGCTTAatggtgagcaagggcga*	UP_25	65
			R	*AGAAAGCTGGGTActaagccttgtacagctcgt*	UP_26	
GUS	-	Marker	F	*AAAAAGCAGGCTTAatgttacgtcctgtagaaacc*	UP_23	62
			R	*AGAAAGCTGGGTAtcattgtttgcctccctg*	UP_24	
LUC	-	Marker	F	*AAAAAGCAGGCTTAatggaagacgccaaaaacat*	UP_21	63
			R	*AGAAAGCTGGGTAttacacggcgatctttccg*	UP_22	

*Only for promoters.

To establish suitable annealing temperatures for some primer pairs, a gradient PCR was used. Here, the same reaction mix is aliquoted and exposed to different annealing temperatures and the best-performing conditions are selected. In the case, where several PCR products of different lengths were produced, the product of correct size was isolated from the agarose gel before amplification by PCR2. Tab. 10 shows the conditions used when each of the amplicons was successfully amplified. Note that $NPQ4_{pro}$ was not successfully amplified and was not included in further steps.

Tab. 10: Specific conditions and results of PCR1.

Promotor/marker	**Specific parameters**	**Length of the expected product**	**Band**
	Primers; annealing T(°C); elongation time; N of cycles	bp	(Single*/More than one**/None)
$HY5_{pro}$	AT_498+499; 64,4 °C; 2'; 35X	1181	Single
$HFR1_{pro}$	AT_500+501; 64,4 °C; 2'; 35X	1736	Single
$ELIP1_{pro}$	AT_502+503; 63 °C; 2'; 35X	1462	Single
$ELIP2_{pro}$	AT_504+505; 64,4 °C; 2'; 35X	1971	Single
$CHL27_{pro}$	AT_506+507; 64,5 °C; 2'; 35X	1144	More than one
FLU_{pro}	AT_508+509; 64,5 °C; 2'; 35X	1534	More than one
$GUN5_{pro}$	AT_510+511; 66,4-72 °C; 2'; 35X	1545	Single
$LEC1_{pro}$	AT_512+513; 64,5 °C; 2'; 35X	1211	More than one
$LEC2_{pro}$	AT_514+515; 66,4-70,3 °C; 2'; 35X	1710	More than one
$FUS3_{pro}$	AT_516+517; 66,4-72 °C; 2'; 35X	1614	More than one
$ABI3_{pro}$	AT_518+519; 64,5 °C; 2'; 35X	1902	Single
$ABI4_{pro}$	AT_520+521; 64,5 °C; 2'; 35X	1020	More than one
$NPQ4_{pro}$	AT_522+523; 59-70 °C; 1'45"; 35X	1540	None
$RBSC1B_{pro}$	AT_524+525; 63-66,4 °C; 2'; 35X	1682	Single
	Primers; annealing T(°C); elongation t; N of cycles; source of DNA – plasmid ID***		
GUS	UP_23+24; 62-63 °C; 1'20"; 35X; IMP22	1839	Single
LUC	UP_21+22; 62 °C; 1'20"; 35X; IMP23	1680	Single
tagRFP	UP_27+28; 62-65 °C; 1'20"; 35X; IMP24	741	Single
G3GFP	UP_29+30; 55-65 °C; 1'20"; 35X; IMP14	744	Single
EYFP	UP_25+26; 58-65 °C; 1'20"; 35X; IMP14	750	Single

*of a correct length **one of them had a correct length ***IMP22 = R4pDD633-MD8, IMP23 = R4pDD635-MD8, IMP24 = R4pDD659-MD8, IMP14 = R4pGWB6540-MD8 (Aboulela et al., 2017)

The specific conditions used during the successful PCR of each promoter/marker are in the Tab. 11. The products with more differently sized bands were not purified anymore.

Some of the elongation times (Tab. 10; Tab. 11) were unnecessarily long because the author was not aware of the high processivity of the PhuPol.

Tab. 11: Specific conditions and results of PCR2.

Promotor/marker	PCR2	Single band
	Primers; polymerase; annealing T(°C); elongation t; N of cycles	(YES/NO)
HY5$_{pro}$	AT_94+95; 57 °C; 2'20"; 30X	YES
HFR1$_{pro}$	AT_94+95; 57 °C; 2'20"; 30X	YES
ELIP1$_{pro}$	AT_94+95; 57 °C; 1'20"; 30X	YES
ELIP2$_{pro}$	AT_94+95; 57 °C; 2'20"; 30X	YES
CHL27$_{pro}$	AT_94+95; 57 °C; 1'45"; 35X	YES
FLU$_{pro}$	AT_94+95; 57 °C; 1'20"; 30X	YES
GUN5$_{pro}$	AT_94+95; 57 °C; 1'20"; 30X	YES
LEC1$_{pro}$	AT_94+95; 57 °C; 1'45"; 30X	NO
LEC2$_{pro}$	AT_94+95; 57 °C; 1'45"; 30X	NO
FUS3$_{pro}$	AT_94+95; 57 °C; 1'20"; 30X	NO
ABI3$_{pro}$	AT_94+95; 57 °C; 1'20"; 30X	YES
ABI4$_{pro}$	AT_94+95; 57 °C; 1'20"; 30X	NO
RBSC1B$_{pro}$	AT_94+95; 57 °C; 1'45"; 30X	YES
GUS	AT_102+103; 61 °C; 1'; 35X	YES
LUC	AT_102+103; 61 °C; 1'; 35X	YES
tagRFP	AT_102+103; 61 °C; 1'; 35X	YES
G3GFP	AT_102+103; 61 °C; 1'; 35X	YES
EYFP	AT_102+103; 61 °C; 1'; 35X	YES

The recommended weight (ng) of each product was calculated according to the manufacturer's manual (Tab.12). The att-flanked PCR product with optimized concentration was then mixed with the pDONR plasmid: pDONR-P4P1r - promoter; pDONR221 – marker gene. The BP reaction was carried out according to protocol described in methods. At first the BP reaction time was 1 h, as recommended by manual. However, in some cases the manufacturer's optimal conditions did not work, therefore the amount (ng) of att-flanked product was increased and the BP reaction was prolonged. The final amounts (ng) and reaction times are in Tab. 13.

Tab. 12: amounts of att-flanked products in 50fmol.

Promoter	***HY5*$_{pro}$**	***HFR1*$_{pro}$**	***ELIP1*$_{pro}$**	***ELIP2*$_{pro}$**	***CHL27*$_{pro}$**	***FLU*$_{pro}$**	***GUN5*$_{pro}$**
ng/50 fmol	40	58	49	66	39	52	52
Promoter	***LEC1*$_{pro}$**	***LEC2*$_{pro}$**	***FUS3*$_{pro}$**	***ABI3*$_{pro}$**	***ABI4*$_{pro}$**	***NPQ4*$_{pro}$**	***RBCS1B*$_{pro}$**
ng/50 fmol	41	58	54	64	35	52	57
Marker	***G3GFP***	***tagRFP***	***EYFP***	***LUC***	***GUS***	-	-
ng/50 fmol	27	26	26	57	62	-	-

The *E. coli* transformation, cultivation, screening and sequencing was performed according to the protocol described in Methods. Constructs positively tested with M13 primers were sent for sequencing and compared with expectation. In all of the sequenced samples no relevant mutation was indicated (Tab. 13), so they were stored at -80°C to be used for the LR reaction in the future. In some cases, there was an unusually high rate of colonies that successfully grew on Kan but did not test positive with M13 primers. But when screened with cloning primers used for PCR1 some of the colonies tested positive. Few of the plasmids isolated form these colonies were sent for sequencing with M13, which showed that there is no insertion present (Tab. 13).

Tab. 13: Results summary of cloned constructs. Constructs in light grey were successfully obtained. Constructs in darker grey didn't undergo successful recombination into a vector. Construct in dark grey were not obtained, but it is unknown in which phase the problem occurred.

Construct	**Last used amount**	**Successfully recombined***	***E. coli* growth on selection**
promoter in pDONR-P4P1r; Marker in pDONR221	PCR product (ng); t of BP		
HY5$_{pro}$	40ng;1h	Yes	Yes
HFR1$_{pro}$	58ng; 1h	Yes	Yes*
ELIP1$_{pro}$	49ng; 1h	Yes	Yes
ELIP2$_{pro}$	63ng; 3h	Yes	Yes
CHL27$_{pro}$	39ng; 75ng; 2h	Yes	Yes
FLU$_{pro}$	241,6ng; overnight	Yes	Yes
ABI4$_{pro}$	231ng; overnight	Yes	Yes
GUN5$_{pro}$	285,7ng; overnight	No	Yes
LEC1$_{pro}$	41ng; 2h	No	Yes
LEC2$_{pro}$	57ng; 2h	No	Yes
FUS3$_{pro}$	241ng; overnight	No	Yes
ABI3$_{pro}$	272,5ng; overnight	No	Yes
RBCS1B$_{pro}$	57ng; 2h	?	No
NPQ4$_{pro}$	-	-	-
GUS	61ng; 2h	?	No
LUC	57ng; 2h	No	Yes
tagRFP	25ng; 2h	No	Yes
G3GFP	26ng; 2h	Yes	Yes
EYFP	25ng; 2h	Yes	Yes

*this construct has a point mutation, but it is 1460bp far from the start codon of the gene and hypothetically is not likely to influence the promoter activity.

3.2. Redundancy and tissue-specificity of CLF and SWN

Binary vector 1 (BV1) and BV12 (see Tab. 1) were not successfully cloned and BV4-BV11 contained mutation in the coding region of *SWN*, so the potentially transgenic seeds/plants were eliminated and were not carried further in the process. The only vectors left were BV2 and BV3. Their seeds of T1 generation were sterilized and sown on selection plates according to the protocol described in Methods. After approximately 3 weeks the plates were analysed (Tab. 14). Five phenotypes were identified (Fig. 4)

Tab. 14: Results of T1 generation.

ID	Constructs		N of plates	N of plates without germination	N of survivors	N of survivors positively PCR-tested for PPT resistance gene
BV2	***APUMpro::CLF-tagRFP***	***APUMpro::CLF-G3GFP***	**62**	**0**	**71**	**0**
BV3	***LHCBpro::CLF-tagRFP***	***LHCBpro::CLF-G3GFP***	**77**	**0**	**17**	**0**

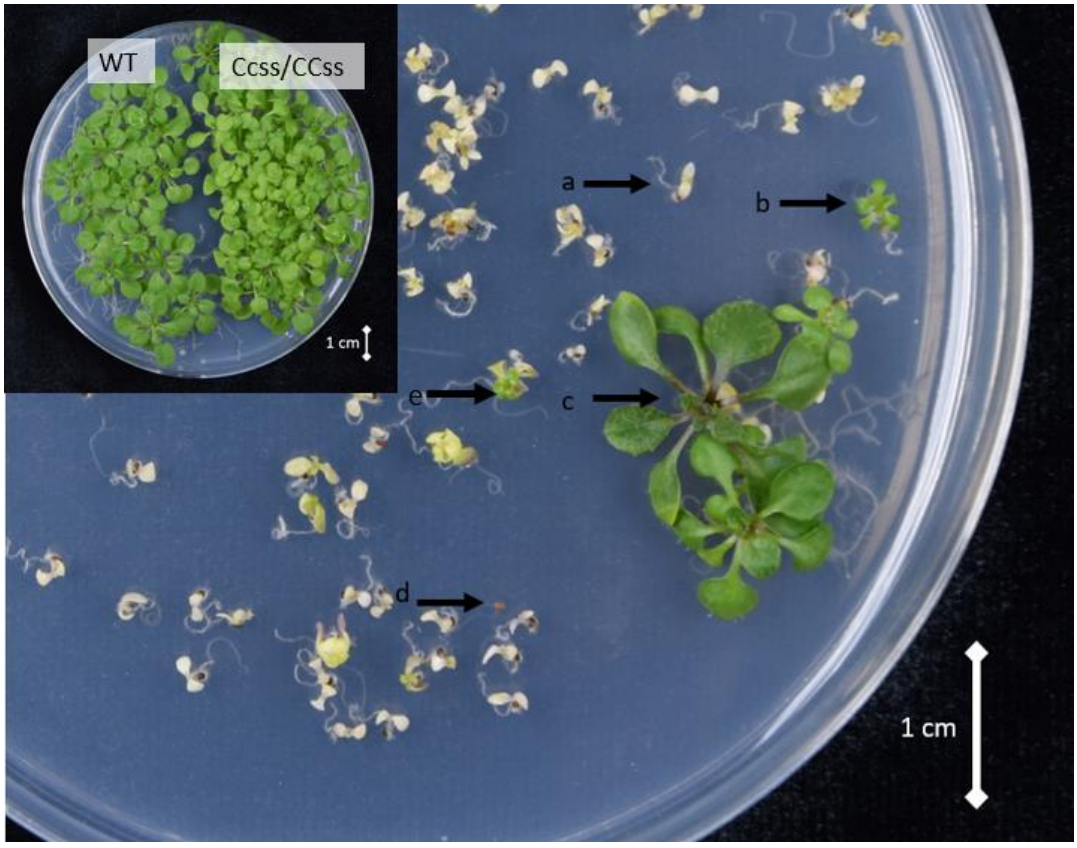

Fig. 4: Phenotypes **a** and **e** were both considered negative transformants and **c** and **b** positive transformants. Although phenotype **b** seemed less developed than phenotype **c**, it showed no signs of sensitivity to PPT, unlike phenotype **e**. Phenotype **d** was not categorised as sensitive or insensitive. It was a result either of a seed collection done too early or of too long sterilization. However, the occurrence was not high.

The presumably positive transformants (PPT-resistant) were replanted into soil and the plants were grown according to protocol. DNA was isolated according to the protocol for DNA isolation for genotyping.

The PCR was run according to the protocol for Green Taq Polymerase. Controls used were WT- *Col-0* (*swn-3* negative (-)/*G3GFP-/tagRFP-/PPT-resistance gene (PPT-R)-*), plasmid IMP109 (*swn-3-/G3GFP positive (+)/tagRFP+/PPT-R+*) and eventually *CLFpro::CLF-GR* (*swn*-3-/*G3GFP-/tagRFP-/PPT-R+*). Each DNA sample was tested with several primers (Tab. 15).

Tab. 15: List of primers used for genotyping.

Gene	Primers's ID (F+R)	Primers' sequences	T(°C) of annealing	Size of amplicon	Works
swn-3	AT_31+AT_33	F*: TGGTTCACGTAGTGGGCCATCG* R*: TGGAACTTTTGAGTGGCTAGA--GGTG*	59	721	Yes
PPT-R	AT_42+AT_43	F*: ATCTACCATGAGCCCAGAAC* R: *GTCATCAGATCTCGGTGACG*	53	563	Yes
PPT-R	UP_31+UP_33	F*: ATGAGCCCAGAACGAC* R*: TCAGATCTCGGTGACGG*	52	552	No
G3GFP	AT_474+AT_475	F*: AGAGGGTGAAGGTGATGCAA* R*: CCATGTGTAATCCCAGCAGC*	55	603	No
tagRFP	AT_476+AT_477	F*: CTTCAAGTGCACATCCGAGG* R*: GTTTGTGCCCCAGTTTGCTA*	55	626	No
G3GFP	UP_15+UP_16	F*: TTTCACTGGAGTTGTCCCAAT* R*: AAAGGGCAGATTGTGTGGAC*	53	592	No
tagRFP	UP_27+UP_28	F: *AAAAAGCAGGCTTAAT--GGTGTCTAAGGGCGAAG* R*: AGAAAGCTGGGTATCA--ATTAAGTTTGTGCCCCA*	65	714	No
G3GFP	UP_29+UP_30	F*: AAAAAGCAGGCTTAATGAGTA--AAGGAGAAGAACTTTTCACT* R*: AGAAAGCTGGGTATTATT--TGTATAGTTCATCCATGCCA*	65	717	No

The *swn-3* amplicon was a reliable indicator of the quality of isolated DNA (Fig. 5G). But the reliability of other primers was unclear. The *G3GFP* primers AT_474, AT_475, UP_15 and UP_16 had unspecific activity. So, they were tested using different annealing temperaures (53-65°C) and in all possible combinations, but their activity remained unspecific with false-positive WT controls. Example of unspecific activity of UP_15 and UP_16 is in Fig. 5A. Additionally, the primers used for cloning UP_29 and UP_30 were used but the result was also unspecific and produced false-positive amplicons (Fig. 5B). The *tagRFP* primers AT_476 and AT_477 (Fig. 5E) and *tagRFP* cloning primers UP_27 and UP_28 (Fig. 5F) were also tested but they acted non-specifically, similarly to the G3GFP primers. Two sets of *PPT-R* primers were also tested. UP_31 and UP_32 were highly unspecific (Fig. 5D). However, the second set of primers, AT_42 and AT_43, seemed as a reliable indicator of transgene presence, because it amplified/didn't amplify three different controls according to expectation (Fig. 5C).

Note that there were long term technical problems with correct distribution of the samples in agarose gel during electrophoresis.

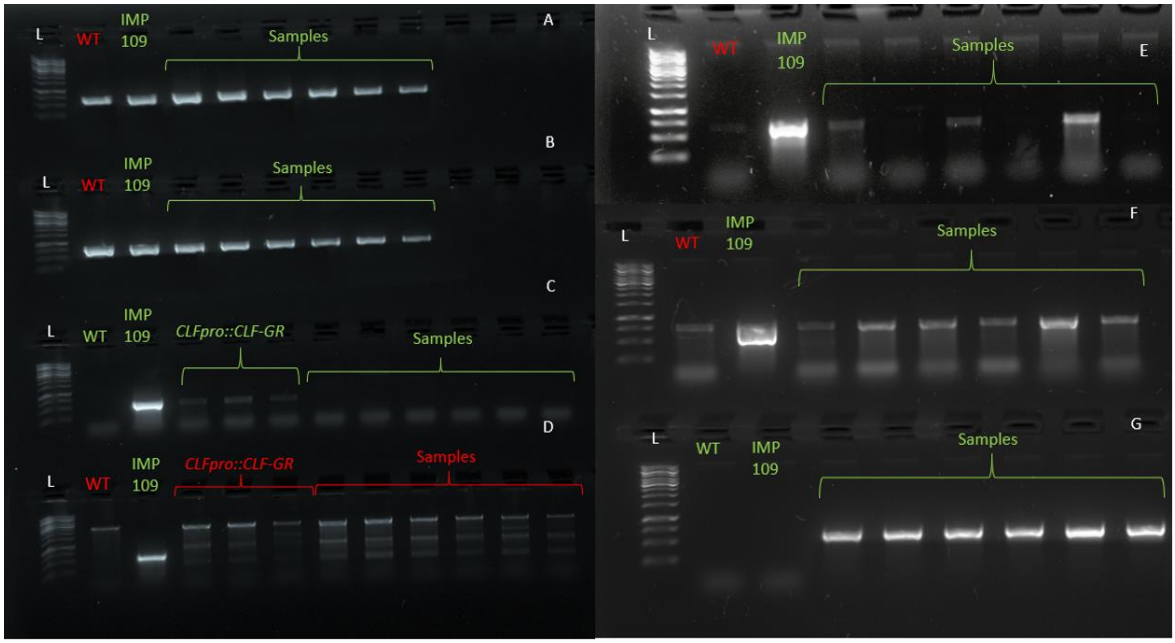

Fig. 5: Results of genotyping of samples of TKA3. Green colour stands for expected result (working PCR reaction) in controls and correct product length in case of Samples. The red colour indicates other than expected PCR outcome. The bands are results of a PCR reaction performed with: A: *G3GFP* primers UP_15 and UP_16. B: *G3GFP* primers UP_29 and UP_30. C: *PPT-R* primers AT_42 and AT_43. D: *PPT-R* primers UP_31 and UP_32. E: *tagRFP* primers AT_476 and AT_477. F: *tagRFP* primers UP_27 and UP_28. G: *swn-3* primers AT_31 and AT_33.

After the analysis of the T1 generation it was therefore unclear whether the results are positive or not. *PPT-R* gene (primers AT_42 and AT_43) indicated that the transformation wasn't successful (Fig. 5C). In order to clarify the situation more, it was necessary to grow the T2 generation and see whether it will be resistant to PPT. The seeds were sterilized according to protocol, but because a subset of the seeds failed to germinate (Fig. 6A), the next batches of seeds were dried on a filter paper instead because the 90% EtOH then evaporates faster. Additionally, the stratification was prolonged by 1 day. The amount of herbicide was increased to 30 µl of PPT (final c=12 µg/ml) since there was a possibility that the survivors in the T1 generation escaped the selection somehow. The increased amount of herbicide was first tested

on a plate with WT and *CLFpro::CLF-GR*, a transgenic line that has previously been tested as PPT-resistant (Fig. 6B).

The plants were grown in the same conditions as T1 generation. After 3 weeks the dishes were analysed in the same manner as T1 generation (Tab.16).

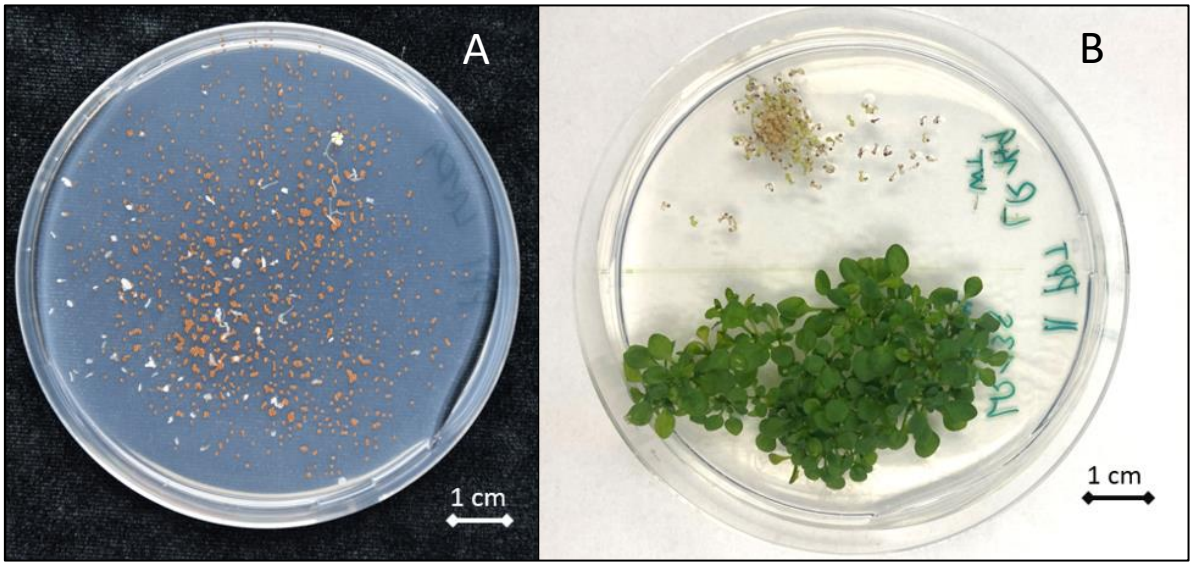

Fig. 6: A) Example of a plate without germination (after 1 month). B) control plate with WT (upper part) and *CLFpro::CLF-GR* (lower part) on PPT c=12 µg/ml.

Tab. 16: Results of T2 generation.

ID	Constructs		N of plates	N of plates without germination	N of survivors	N of survivors positively PCR-tested for PPT resistance
BV2	***APUMpro::CLF-tagRFP***	***APUMpro::CLF-G3GFP***	**10**	**0**	**0**	**0**
BV3	***LHCBpro::CLF-tagRFP***	***LHCBpro::CLF-G3GFP***	**119**	**9**	**7**	**0**

The very low number of survivors again indicated that the plants weren't transgenic. However, they were genotyped for complete assurance and the results were negative.

For better understanding of the situation, T1 generation data was statistically analysed to discover whether the distribution of survivors was random. Hypothetically, the seeds of each plant should have been distributed on the plates randomly. Therefore, the distribution of survivors should be coherent with models of random distribution. Variation and mean of each

dataset (all plates with seeds from the same plant) were compared and Poisson probability was calculated (Tab.17). Only datasets where all seeds from one maternal plant produced at least one survivor were analysed. T2 wasn't analysed because the dataset was too small.

Tab. 17: Analysis of randomness of survivors' distribution in T1 generation.

Maternal plant ID	**N of plates**	**N of positive plates**	**N of survivors per positive plate**	6^2	μ	λ	$6^2 / \mu$	**Poisson probability**
BV2-1	12	2	5; 1	1.9	0	0.5	$6^2 > \mu$	P(5)=0.0002 P(1)=0.3033 P(0)=0.6065
BV2-2	1	1	1	0	1	1	$6^2 < \mu$	P(1)=0.3679
BV2-3	2	1	4	4	2	2	$6^2 > \mu$	P(4)=0.0902 P(0)=0.1353
BV2-4	2	1	5	6.25	2.5	2.5	$6^2 > \mu$	P(5)=0.0668 P(0)=0.0821
BV2-5	2	1	1	0.25	0.5	0.5	$6^2 < \mu$	P(1)=0.3033 P(0)=0.6065
BV3-1	9	3	17; 1; 4	28.01	0	2.44	$6^2 > \mu$	P(17)=7.4168E−10 P(1)=0.2127 P(4)=0.1287 P(0)=0.0872
BV3-2	18	3	12; 6; 8	11.47	0	1.44	$6^2 > \mu$	P(12)=3.9321E−8 P(6)=0.0029 P(8)=0.0001 P(0)=0.2369
BV3-3	23	7	9; 3; 1; 1; 4; 3; 2	4.26	0	1	$6^2 > \mu$	P(9)=1.0138E−6 P(3)=0.0613 P(1)=0.3679 P(4)=0.0153 P(2)=0.1839 P(0)=0.3679

4. DISCUSSION

Although the aim of the study was only partially fulfilled it will make a good base for the continuation of the project. 7 of 14 promoter constructs (*HY5*$_{pro}$, *HFR1*$_{pro}$, *ELIP1*$_{pro}$, *ELIP2*$_{pro}$, *CHL27*$_{pro}$, *FLU*$_{pro}$, *ABI4*$_{pro}$) and 2 of 5 marker gene constructs (*G3GFP*, *EYFP*) were successfully recombined into vectors and are now ready to proceed in the LR reaction.

6 promoter constructs (*GUN5*$_{pro}$, *FUS3*$_{pro}$, *LEC1*$_{pro}$, *LEC2*$_{pro}$, *ABI3*$_{pro}$, *RBSC1B*$_{pro}$) and 3 marker gene constructs (*tagRFP*, *LUC*, *GUS*) were successfully amplified in PCR2 but weren't successfully recombined into pDONR vectors. The factor that affected the success of

the recombination appears to be time-dependent because it was more difficult to obtain successfully recombined vectors as the project proceeded. BP Clonase is known for its intolerance to repeated thawing and is guaranteed to be stable for only 6 months. Therefore, it is possible that, even if the BP Clonase was aliquoted to 5µl aliquots and stored in the -80°C freezer for only a few months, its efficiency decreased. For future experiments, it would be beneficial to make smaller aliquots (2-3 µl) that will be used at once. And plan experiments in the way that the BP Clonase is used as fast as possible. It is unlikely that the problem here would be in the design of att-flanked primers because all of the primers were designed in the same way.

Another problem that could have occurred with some of the promoter constructs is that non-specific (other than targeted) amplicon might have been recombined into the pDONR plasmid. Even though the products of PCR1 were purified there were in some cases small amounts of short amplicons after PCR2. These shorter amplicons probably have att-B sites, so there is a chance that they would get recombined into the vector. However, this would be recognised through genotyping.

The last promoter construct ($NPQ4_{pro}$) was not successfully amplified by PCR, therefore it would be good to test the primers in gradient PCR and depending on the result either proceed with new annealing T (°C) or design new primers.

As mentioned earlier, in the second part a mutation in *SWN* in most of the vectors occurred, thus we did not proceed with these constructs further. The constructs that were left, BV2 ($APUM_{pro}$*::CLF-tagRFP +* $APUM_{pro}$*::CLF-G3GFP*) and BV3 ($LHCB_{pro}$*::CLF-tagRFP +* $LHCB_{pro}$*::CLF-G3GFP*), were probably not inserted into the plant genome, indicating problems with plant transformation.

That is suggested based on the selection of transformants as well as the genotyping. The T1 generations had overall 88 survivors from which none tested positive for *PPT-R.* This gene's amplification with primers AT_42 and AT_43 appeared to show results that were coherent with three different controls. Thus, it is unlikely, however not impossible, that these primers would not reveal PPT-resistant transformants. The T2 generations had overall only 7 survivors on 129 plates and none of these tested positive for *PPT-R.* Based on this low number of survivors and the negative *PPT-R* test it can be said with a high level of confidence that the plants were not transgenic. Therefore, the question is why the plants survived. In the first generation locally decreased or depleted concentration of PPT, a mistake in pipetting or PPT inactivated due to too high temperature of the ½ MS+ 1% sucrose could be the reason. In the second generation, the concentration of PPT was increased from c=10µg/ml to c=12µg/ml and

the number of survivors quite significantly decreased. The higher concentration possibly prevented the local insufficiency of PPT amount.

Alternatively, it is also possible, that the plants that survived the selection were transgenic, and the inserted genes were not detected with any of the primers used in genotyping. The low rate of survivors in T2 generation could then be a result of *PPT-R* being silenced by RNA interference (RNAi) due to its high transcription.

However, the data analysis of T1 generation also hints that the distribution of survivors on selective plates was not random and that that there was another factor that could influence surviving/dying, besides transgenes. The variation and mean comparison indicated that there was a tendency for grouping of the survivors on certain plates ($\sigma^2 > \mu$) in 6 out of 8 datasets. The last two datasets, BV2-2 and BV2-5 were showing the opposite tendency, but their size was extremely small and therefore the analysis was probably distorted. The Poisson probability test revealed that some of the datasets, especially the bigger ones, tended to have a much larger percentage of plates without any survivors than the probability suggests. For example, any BV3-2 plate should theoretically have a 24% chance of having 0 survivors. But in reality, 83% of the plates were without survivors. On the contrary, one BV3-2 plate had 12 survivors and the probability of that is only 3.9321E−8. It is not impossible, but it strongly suggests that the distribution of survivors was not random. Similar tendencies are apparent in all BV3 datasets. Most of the BV2 datasets do not show the same tendency, but they are significantly smaller and thus probably show distorted results. Both analytical methods suggest that the distribution of survivors was not random. The seeds from each maternal plant were sterilised and kept together in one tube. The tube was thoroughly shaken before the distribution of seeds on selective plates. Therefore, it is unlikely that transgenic seeds would be somehow separated from the non-transgenic seeds when they were sowed. That again suggests that the surviving was influenced by an external factor. And the previously suggested locally depleted/lowered concentration of PPT is the most likely the cause. That doesn't mean that some of the plants could not be transgenic, undetected with PCR and their T2 generation was terminated by RNAi. But it indicates that the survival rate was at least partially dependent on another factor than the presence of PPT-R.

Overall, the evidence suggests that the transformation of Arabidopsis itself could be the key problem. There is potential for improvement in multiple ways. The *A. tumefaciens* colony density may be increased to ensure a successful transformation. The floral dip can be repeated after few days to increase the chance of transformation. It may also be suggested to select younger plants that have fresh flowers.

The dual marker system generated within project one can be used separately or in combination with the *CLF* and *SWN* transgenic system. If used separately, it can be used in WT or PRC2 mutant (*clf swn*) genetic backgrounds.

The marker constructs can be inserted into WT and function as a dual-marker for a forward genetic screen of PRC2 disability after mutagenesis. This use may help discover new genes that are involved in the PRC2 regulatory system. An example of such use might be a line in which $LEC1_{pro}$*::G3GFP* and $ELIP2_{pro}$*::tagRFP* has been inserted. The mutagenesis here might target a gene that is involved in the repression of both developmental and response genes and therefore the seedlings would have a strong G3GFP and tagRFP signal. But it may also uncover a gene that is only involved in the repression of developmental genes (G3GFP ectopic upregulation) or in repression of response genes (tagRFP ectopic upregulation). If the line shows ectopic upregulation of tagRFP then another dual-marker including promoters of response genes (for example $ELIP1_{pro}$*::LUC* and FLU_{pro}*::GUS*) could be introduced to discover, whether the newly identified gene only plays role in the repression of *ELIP2* or if it is a common regulator for more response genes. This method could as well potentially help to identify some differences between SWN and CLF. The fluorescent signal could be in some tissues weaker than in *clf swn* that also contains $LEC1_{pro}$*::G3GFP* and $ELIP2_{pro}$*::tagRFP*. But it could be stronger than in WT that did not undergo mutagenesis but also contains $LEC1_{pro}$*::G3GFP* and $ELIP2_{pro}$*::tagRFP*. That would indicate that PRC2 is still functioning, but less than in WT and one of the possible explanations could be that the gene is only linked with one of the two homologues.

If the dual-marker was introduced to *clf swn* that underwent mutagenesis then we would screen for seedling with a lowered expression of the marker and hence carry out a repressor genetic screen. The lack of expression could indicate that the gene that has been turned off was part of the pathway that is responsible for ectopic activation of expression of the development/response genes in the mutant genetic background. Additionally, in *clf swn* (without mutagenesis) the dual-marker can be used for screen of environmental conditions that influence ectopic overexpression of developmental/response genes. In an experiment of this kind, *clf swn* with markers under control of response genes' promoters could be grown in stress-inducing conditions such as prolonged periods of darkness or prolonged periods of very bright light. The outcome could be that in the light-deficient conditions the markers would be even more expressed and new response genes' TF could be identified. In the case of an over-lit environment, the possible result would be that at some point the expression of the marker could be down-regulated and again new TF of response genes could be identified.

As described earlier in this work, the binary vectors that contain *CLF* and/or *SWN* under tissue-specific promoters inserted in *clf swn* will allow us to observe complementation of CLF/SWN function on developmental/responsive genes in certain tissue. For example, transgenic *clf swn* plant with *CLF* controlled by $LHCB_{pro}$ has theoretically a good chance to successfully transit into the vegetative phase. This could for example indicate that CLF is required in differentiated cells to suppress the reversal of vegetative to embryonic cells, but it is dispensable in the SAM. In combination with the promoter-marker constructs, this system introduced to *clf swn* can test the importance of CLF/SWN in specific tissues and reveal how they affect expression of specific developmental/response genes. For instance, the intensity of EYFP signal in transgenic plant with $LHCB_{pro}$*::CLF-tagRFP* and $HFR1_{pro}$*::EYFP* can be compared with a transgenic plant with $APUM_{pro}$*::CLF-tagRFP* and $HFR1_{pro}$*::EYFP*. The results of this experiment could show whether CLF expressed in only one type of cells can efficiently repress the expression of *HFR1* or whether the repression is dependent on the abundance of CLF in both tissues. This system could also clarify the presumed SWN/CLF redundancy in specific tissues. Similarly to the previous example, the intensity of EYFP signal in transgenic a line with $LHCB_{pro}$*::CLF-tagRFP* and $HFR1_{pro}$*::EYFP* can be put up for comparison with $LHCB_{pro}$*::SWN-tagRFP* and $HFR1_{pro}$*::EYFP* transgenic line. The results may show that EYFP is not expressed in the seedling that contains *CLF*. It could be expressed in the one with SWN, but maybe less than in *clf swn* control with $HFR1_{pro}$*::EYFP*. This would indicate that CLF is essential for the repression of *HFR1* in the vegetative phase and SWN has only a supportive function.

Nevertheless, the combinations can be much more complex. There are 12 dual vectors with *SWN/CLF* fused with *tagRFP* and *G3GFP*. The promoter-marker constructs can then contain either *GUS*, *LUC* or *EYFP* (since *tagRFP* and *G3GFP* are already used) and any of the 14 promoters. Each promoter-marker construct can be paired with any of the other 13 promoter-marker constructs in binary vector. Additionally, any of the other 12 promoter-marker constructs can be inserted into separate vector. In total there are 26 208 possible combinations in case all the named constructs with the five different markers are used. It is without a doubt that not all of these combinations will be necessary or interesting for the research. However, the potential of the tool is large, and it will be interesting to see all the findings that it will help to discover.

5. APPENDIXES

5.1. APPENDIX 1: DNA primer design generally:

The primers must have certain features for the required outcome of the PCR reaction. The two main features are *specificity* and *efficiency*. Both of these features are affected by length of the primer, its cytosine-guanine content and homology-based annealing site and the temperature required for annealing. The optimal length of the primer is a compromise between the specificity and efficiency. The longer the primer is, the more specific it is. But it also takes more time to anneal fully than shorter primer, which may result in lower amount of the PCR product, thus lower efficiency. 18-24 bp is generally considered being the optimal length of a primer because it is both specific and efficient (Dieffenbach, 1993).

Cytosine (C) and guanin (G) share triple hydrogen-bond, which is much stronger in comparison to the double-bond between adenine (A) and thymine (T). This stronger bond of C and G is important for the correct annealing because primers with low ratios of CG tend to disconnect from the template DNA strand (double strand melting). The optimal ratio is around 50% (50% AT). The content of CG also greatly increases the annealing temperature (T_m) (Dieffenbach, 1993). It is often argued that at least one C or G should be located amongst last five bases on the 3' end of the primer. That is because this part of primer (sometimes unlike 5' end) must be perfectly annealed so the polymerase has a stable binding spot. However, many scientists also say that there should not be GC or CG because of the danger of self-complementarity leading in creation of hairpin and primer dimer structures (Dieffenbach, 1993).

T_m of a primer that is 18-24 bp and has around 50% of CG content is usually between 56°C and 62°C. However, exact T_m for each primer can be calculated with use of software described below. It is important to perform PCR with optimal T_m because too low T_m can result in unspecific annealing. The lower the temperature is, the less of the bases need to anneal to form a stable connection. Too high T_m will result in no primer-template duplex formation. That is because, if the T_m is too high the primer will not have enough binding energy to remain inert and it will disconnect from the template DNA (Dieffenbach, 1993).

6. LITERATURE

A. Simon, J., & E. Kingston, R. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell*, *49*(5), 808–824. https://doi.org/10.1016/j.molcel.2013.02.013.Occupying

Aboulela, M., Tanaka, Y., Nishimura, K., Mano, S., Nishimura, M., Ishiguro, S., Kimura, T., & Nakagawa, T. (2017). Development of an R4 dual-site (R4DS) gateway cloning system enabling the efficient simultaneous cloning of two desired sets of promoters and open reading frames in a binary vector for plant research. *PLoS ONE*, *12*(5), 1–18. https://doi.org/10.1371/journal.pone.0177889

Aichinger, E., Villar, C. B. R., di Mambro, R., Sabatini, S., & Köhler, C. (2011). The CHD3 chromatin remodeler PICKLE and polycomb group proteins antagonistically regulate meristem activity in the Arabidopsis root. *Plant Cell*, *23*(3), 1047–1060. https://doi.org/10.1105/tpc.111.083352

Aichinger, E., Villar, C. B. R., Farrona, S., Reyes, J. C., Hennig, L., & Köhler, C. (2009). CHD3 proteins and polycomb group proteins antagonistically determine cell identity in Arabidopsis. *PLoS Genetics*, *5*(8). https://doi.org/10.1371/journal.pgen.1000605

Aitken, A. (2016) TE Buffer. Accessed October 2020. https://www.protocols.io/view/TE-Buffer-c5uy6v

AmplifX 2.0.7 by Nicolas Jullien ; Aix-Marseille Univ, CNRS, INP, Inst Neurophysiopathol, Marseille, France - https://inp.univ-amu.fr/en/amplifx-manage-test-and-design-your-primers-for-pcr

Aubert, D., Chen, L., Moon, Y.-H., Martin, D., Castle, L. A., Yang, C.-H., & Sung, Z. R. (2001). EMF1, A Novel Protein Involved in the Control of Shoot Architecture and Flowering in Arabidopsis. *The Plant Cell*, *13*(8), 1865. https://doi.org/10.2307/3871324

Bastow, R., Mylne, J. S., Lister, C., Lippman, Z., Martienssen, R. A., & Dean, C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature*, *427*(6970), 164–167. https://doi.org/10.1038/nature02269

Bellegarde, F., Herbert, L., Séré, D., Caillieux, E., Boucherez, J., Fizames, C., Roudier, F., Gojon, A., & Martin, A. (2018). Polycomb Repressive Complex 2 attenuates the very high expression of the Arabidopsis gene NRT2.1. *Scientific Reports*, *8*(1), 1–9.

https://doi.org/10.1038/s41598-018-26349-w

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis*, *53*(8), 474–485. https://doi.org/10.1002/dvg.22877

Bouyer, D., Roudier, F., Heese, M., Andersen, E. D., Gey, D., Moritz, K., Goodrich, J., Renou, J., Grini, P. E., Colot, V., & Schnittger, A. (2011). *Polycomb Repressive Complex 2 Controls the Embryo-to- Seedling Phase Transition*. *7*(3). https://doi.org/10.1371/journal.pgen.1002014

Calonje, M. (2014). Prc1 marks the difference in plant pcg repression. *Molecular Plant*, *7*(3), 459–471. https://doi.org/10.1093/mp/sst150

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y., Sung, Z. R., & Goodrich, J. (2004). *Interaction of Polycomb-group proteins controlling flowering in Arabidopsis*. 5263–5276. https://doi.org/10.1242/dev.01400

Chen, S., Ma, T., Song, S., Li, X., Fu, P., Wu, W., Liu, J., Gao, Y., Ye, W., Dry, I. B., & Lu, J. (2021). Arabidopsis downy mildew effector HaRxLL470 suppresses plant immunity by attenuating the DNA-binding activity of bZIP transcription factor HY5. *New Phytologist*, *February*, 0–1. https://doi.org/10.1111/nph.17280

Chica, C., Louis, A., Roest Crollius, H., Colot, V., & Roudier, F. (2017). Comparative epigenomics in the Brassicaceae reveals two evolutionarily conserved modes of PRC2-mediated gene regulation. *Genome Biology*, *18*(1), 1–15. https://doi.org/10.1186/s13059-017-1333-9

Chiew Foan Chin, J. Y. C. (2015). Gateway Cloning Technology: Advantages and Drawbacks. *Cloning & Transgenesis*, *04*(01), 10–12. https://doi.org/10.4172/2168-9849.1000138

Cleaver Scientific. Agarose gel electrophoresis of DNA. Accessed April 2021. https://www.cleaverscientific.com/applications/agarose-gel-electrophoresis-of-dna/

Dall'Osto, L., Cazzaniga, S., Wada, M., & Bassi, R. (2014). On the origin of a slowly reversible fluorescence decay component in the Arabidopsis npq4 mutant. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1640). https://doi.org/10.1098/rstb.2013.0221

Derkacheva, M., & Hennig, L. (2014). Variations on a theme: Polycomb group proteins in plants. *Journal of Experimental Botany*, *65*(10), 2769–2784. https://doi.org/10.1093/jxb/ert410

Derkacheva, M., Steinbach, Y., Wildhaber, T., Mozgová, I., Mahrez, W., Nanni, P., Bischof, S., Gruissem, W., & Hennig, L. (2013). Arabidopsis MSI1 connects LHP1 to PRC2 complexes. *EMBO Journal*, *32*(14), 2073–2085. https://doi.org/10.1038/emboj.2013.145

Dieffenbach, C. W., Lowe, T. M. J., & Dveksler, G. S. (1993). General concepts for PCR primer design.. *Genome Research*, *23*(4), 566–573. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.965.1253

Doerfler, W. (1981). DNA Methylation-A Regulatory Signal in Eukaryotie Gene Expression. *Journal of genetic Virology*, 1–20

Duren, Z., Wang, Y., Wang, J., Zhao, X. M., Lv, L., Li, X., Liu, J., Zhu, X. G., Chen, L., & Wang, Y. (2019). Hierarchical graphical model reveals HFR1 bridging circadian rhythm and flower development in Arabidopsis thaliana. *Npj Systems Biology and Applications*, *5*(1), 1–11. https://doi.org/10.1038/s41540-019-0106-3

Edwards, K., Johnstone, C., & Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, *19*(6), 1349. https://doi.org/10.13031/aim.20141900217

Eurofins. (n.d.). *LIGHTRUN Tube*. https://www.eurofinsgenomics.eu/en/custom-dna-sequencing/gatc-services/lightrun-tube/

Exner, V., Aichinger, E., Shu, H., Wildhaber, T., Alfarano, P., Caflisch, A., Gruissem, W., Köhler, C., & Hennig, L. (2009). The chromodomain of LIKE HETEROCHROMATIN PROTEIN 1 is essential for H3K27me3 binding and function during arabidopsis development. *PLoS ONE*, *4*(4), 1–10. https://doi.org/10.1371/journal.pone.0005335

Feng, S., Jacobsen, S. E., & Reik, W. (2010). Epigenetic reprogramming in plant and animal development. *Science*, *330*(6004), 622–627. https://doi.org/10.1126/science.1190614

Frame, B. R., Shou, H., Chikwamba, R. K., Zhang, Z., Xiang, C., Fonger, T. M., Pegg, S. E. K., Li, B., Nettleton, D. S., Pei, D., & Wang, K. (2002). Agrobacterium tumefaciens-mediated transformation of maize embryos using a standard binary vector system. *Plant Physiology*, *129*(1), 13–22. https://doi.org/10.1104/pp.000653

Gan, E. S., Huang, J., & Ito, T. (2013). Functional roles of histone modification, chromatin remodeling and microRNAs in arabidopsis flower development. In *International Review of Cell and Molecular Biology* (1st ed., Vol. 305). Elsevier Inc. https://doi.org/10.1016/B978-0-12-407695-2.00003-2

Ge, Q., Wang, H., Rutishauser, T., & Dai, J. (2015). Phenological response to climate change in China: A meta-analysis. *Global Change Biology*, *21*(1), 265–274. https://doi.org/10.1111/gcb.12648

Gendall, A. R., Levy, Y. Y., Wilson, A., & Dean, C. (2001). The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. *Cell*, *107*(4), 525–535. https://doi.org/10.1016/S0092-8674(01)00573-6

Geneious 2020.2.5 (https://www.geneious.com)

Gómez-Zambrano, Á., Merini, W., & Calonje, M. (2019). The repressive role of Arabidopsis H2A.Z in transcriptional regulation depends on AtBMI1 activity. *Nature Communications*, *10*(1), 1–12. https://doi.org/10.1038/s41467-019-10773-1

Grunstein, M. (1992). Histones as regulators of genes. *Scientific American*, *267*(4), 68–75. https://doi.org/10.1038/scientificamerican1092-68

Hays, J. B. (2002). Arabidopsis thaliana, a versatile model system for study of eukaryotic genome-maintenance functions. *DNA Repair*, *1*(8), 579–600. https://doi.org/10.1016/S1568-7864(02)00093-9

Holliday, R. (2006). Epigenetics: A historical overview. *Epigenetics*, *1*(2), 76–80. https://doi.org/10.4161/epi.1.2.2762

Hou, Z., Yang, Y., Hedtke, B., & Grimm, B. (2019). Fluorescence in blue light (FLU) is involved in inactivation and localization of glutamyl-tRNA reductase during light exposure. *Plant Journal*, *97*(3), 517–529. https://doi.org/10.1111/tpj.14138

Hwang, H.-H., Yu, M., & Lai, E.-M. (2017). Agrobacterium -Mediated Plant Transformation: Biology and Applications . *The Arabidopsis Book*, *15*(15), e0186. https://doi.org/10.1199/tab.0186

Invitrogen, G. (2010). MultiSite Gateway ® Three- Fragment Vector Construction Kit (Catalog no. 12537-023). *Protocols and Product Manuals*, *H*.

Izumi, M., Tsunoda, H., Suzuki, Y., Makino, A., & Ishida, H. (2012). RBCS1A and RBCS3B, two major members within the Arabidopsis RBCS multigene family, function to yield sufficient Rubisco content for leaf photosynthetic capacity. *Journal of Experimental Botany*, *63*(5), 2159–2170. https://doi.org/10.1093/jxb/err434

Kasinath, V., Faini, M., Poepsel, S., Reif, D., Feng, X. A., Stjepanovic, G., Aebersold, R., & Nogales, E. (2018). Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science*, *359*(6378), 940–944. https://doi.org/10.1126/science.aar5700

King, G. J. (2015). Crop epigenetics and the molecular hardware of genotype × environment interactions. *Frontiers in Plant Science*, *6*(NOVEMBER), 1–19. https://doi.org/10.3389/fpls.2015.00968

Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., & Penin, A. A. (2016). A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant Journal*, *88*(6), 1058–1070. https://doi.org/10.1111/tpj.13312

Koressaar T, Lepamets M, Kaplinski L, Raime K, Andreson R and Remm M. Primer3_masker: integrating masking of template sequence with primer design software. Bioinformatics 2018;34(11):1937-1938.

Koressaar T and Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics 23(10):1289-91.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell*, *128*(4), 693–705. https://doi.org/10.1016/j.cell.2007.02.005

Lafos, M., Kroll, P., Hohenstatt, M. L., Thorpe, F. L., Clarenz, O., & Schubert, D. (2011). Dynamic regulation of H3K27 trimethylation during arabidopsis differentiation. *PLoS Genetics*, *7*(4). https://doi.org/10.1371/journal.pgen.1002040

Lambert, T. J. (2019). FPbase: a community-editable fluorescent protein database. *Nature Methods*, *16*(4), 277–278. https://doi.org/10.1038/s41592-019-0352-8

Landy, A. (1989). Dynamic, structural, and regulatory aspects of λ site-specific recombination. *Annual Review of Biochemistry*, *58*(1), 913–941. https://doi.org/10.1146/annurev.bi.58.070189.004405

Leonelli, S., & Ankeny, R. A. (2013). What makes a model organism? *Endeavour*, *37*(4),

209–212. https://doi.org/10.1016/j.endeavour.2013.06.001

Lepš, J., & Šmilauer, P. (2016). *Biostatistika*.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature*, *276*(5688), 565–570. https://doi.org/10.1038/276565a0

Logemann, E., Birkenbihl, R. P., Ülker, B., & Somssich, I. E. (2006). An improved method for preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol. *Plant Methods*, *2*(1), 1–5. https://doi.org/10.1186/1746-4811-2-16

Macherey-Nagel. (2021). *NucleoSpin ® Gel and PCR Clean-up XS*. *February*, 5–6.

Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., & Koornneef, M. (1998). Arabidopsis thaliana: A model plant for genome analysis. *Science*, *282*(5389). https://doi.org/10.1126/science.282.5389.662

Merzlyak, E. M., Goedhart, J., Shcherbo, D., Bulina, M. E., Shcheglov, A. S., Fradkov, A. F., Gaintzeva, A., Lukyanov, K. A., Lukyanov, S., Gadella, T. W. J., & Chudakov, D. M. (2007). Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature Methods*, *4*(7), 555–557. https://doi.org/10.1038/nmeth1062

Miki, B., & McHugh, S. (2004). Selectable marker genes in transgenic plants: Applications, alternatives and biosafety. *Journal of Biotechnology*, *107*(3), 193–232. https://doi.org/10.1016/j.jbiotec.2003.10.011

Mozgova, I., & Hennig, L. (2015a). The Polycomb Group Protein Regulatory Network. *Annual Review of Plant Biology*, *66*(1), 269–296. https://doi.org/10.1146/annurev-arplant-043014-115627

Mozgova, I., Köhler, C., & Hennig, L. (2015b). Keeping the gate closed: Functions of the polycomb repressive complex PRC2 in development. *Plant Journal*, *83*(1), 121–132. https://doi.org/10.1111/tpj.12828

Oliver, M. J. (2014). Why we need GMO crops in agriculture. *Missouri Medicine*, *111*(6), 492–507.

Onodera, K., & Melcher, U. (2004). Selection for 3′ end triplets for polymerase chain reaction primers. *Molecular and Cellular Probes*, *18*(6), 369–372. https://doi.org/10.1016/j.mcp.2004.05.007

Piao, S., Liu, Q., Chen, A., Janssens, I. A., Fu, Y., Dai, J., Liu, L., Lian, X., Shen, M., & Zhu, X. (2019). Plant phenology and global climate change: Current progresses and challenges. *Global Change Biology*, *25*(6), 1922–1940. https://doi.org/10.1111/gcb.14619

Raven, P. H. (2021). Plants make our existence possible. *Plants, People, Planet*, *3*(1), 2–6. https://doi.org/10.1002/ppp3.10173

Rizza, A., Boccaccini, A., Lopez-Vidriero, I., Costantino, P., & Vittorioso, P. (2011). Inactivation of the ELIP1 and ELIP2 genes affects Arabidopsis seed germination. *New Phytologist*, *190*(4), 896–905. https://doi.org/10.1111/j.1469-8137.2010.03637.x

Sanchez-Pulido, L., Devos, D., Sung, Z. R., & 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*, 1–11. https://doi.org/10.1186/1471-2164-9-308

Schuettengruber, B., Bourbon, H. M., Di Croce, L., & Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell*, *171*(1), 34–57. https://doi.org/10.1016/j.cell.2017.08.002

Shu, J., Chen, C., Thapa, R. K., Bian, S., Nguyen, V., Yu, K., Yuan, Z. C., Liu, J., Kohalmi, S. E., Li, C., & Cui, Y. (2019). Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in Arabidopsis seedlings. *Plant Direct*, *3*(1). https://doi.org/10.1002/pld3.100

Shu, K., Zhang, H., Wang, S., Chen, M., Wu, Y., Tang, S., Liu, C., Feng, Y., Cao, X., & Xie, Q. (2013). ABI4 Regulates Primary Seed Dormancy by Regulating the Biogenesis of Abscisic Acid and Gibberellins in Arabidopsis. *PLoS Genetics*, *9*(6). https://doi.org/10.1371/journal.pgen.1003577

Somssich, M. (2018). *A short history of the CaMV 35S promoter*. 1–21. https://doi.org/10.7287/peerj.preprints.27096

Spiess, E. B. (2006). A dictionary of genetics and cytogenetics. In *Journal of Heredity* (Vol. 52, Issue 6). https://doi.org/10.1093/oxfordjournals.jhered.a107088

Stone, S. L., Kwong, L. W., Yee, K. M., Pelletier, J., Lepiniec, L., Fischer, R. L., Goldberg, R. B., & Harada, J. J. (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proceedings of the National*

Academy of Sciences of the United States of America, *98*(20), 11806–11811. https://doi.org/10.1073/pnas.201413498

Tamburri, S., Lavarone, E., Fernández-Pérez, D., Conway, E., Zanotti, M., Manganaro, D., & Pasini, D. (2020). Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. *Molecular Cell*, *77*(4), 840-856.e5. https://doi.org/10.1016/j.molcel.2019.11.021

Thermo Scientific. (2012). *GeneJET Plasmid Miniprep*. 503.

Thermo Scientific. (2016). *MagMAX Plant DNA Kit User Guide (Pub. no. MAN0015954 Rev. A.0)*. https://www.thermofisher.com/mx/es/home/life-science/dna-rna-purification-analysis/automated-purification-extraction/kingfisher-automated-purification-kits-selection-guide.html?gclid=Cj0KCQiAxNnfBRDwARIsAJlH29BcBGPJXhjvVDsKM955EbWToIFQgBVPkB5QpxgcNhenuaTG

Thermo Scientific. (2020). *Product Information Sheet - Phusion ™ High–Fidelity DNA Polymerase*. MAN0012393.

Thermo Scientific. (2016). *Thermo Scientific DreamTaq DNA Polymerase*. *2016* (December), 2–3.

Tottey, S., Block, M. A., Allen, M., Westergren, T., Albrieux, C., Scheller, H. V., Merchant, S., & Jensen, P. E. (2003). Arabidopsis CHL27, located in both envelope and thylakoid membranes, is required for the synthesis of protochlorophyllide. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 16119–16124. https://doi.org/10.1073/pnas.2136793100

Tulsi, Patel; Baris, Tursun; Dylan, P. Rahe; Oliver, H. (2012). *Removal of Polycomb Repressive Complex 2 makes C. elegansgerm cells susceptible to direct conversion into specific somaticcell types*. *2*(5), 1178–1186. https://doi.org/10.1016/j.celrep.2012.09.020.Removal

Ulm, R., Baumann, A., Oravecz, A., Máté, Z., Ádám, É., Oakeley, E. J., Schäfer, E., & Nagy, F. (2004). Genome-wide analysis of gene expression reveals function of the bZIP transcription factor HY5 in the UV-B response of Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(5), 1397–1402. https://doi.org/10.1073/pnas.0308044100

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M and Rozen SG. Primer3--new capabilities and interfaces. Nucleic Acids Res. 2012 Aug 1;40(15):e115.

Völkel, P., le Faou, P., Vandamme, J., Pira, D., & Angrand, P. O. (2012). A human polycomb isoform lacking the Pc box does not participate to PRC1 complexes BUT forms protein assemblies and represses transcription. *Epigenetics*, *7*(5), 482–491. https://doi.org/10.4161/epi.19741

Waddington, C. H. (2012). The epigenotype. 1942. *International Journal of Epidemiology*, *41*(1), 10–13. https://doi.org/10.1093/ije/dyr184

Wang, D., Tyson, M. D., Jackson, S. S., & Yadegari, R. (2006). Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(35), 13244–13249. https://doi.org/10.1073/pnas.0605551103

Wang, H., Liu, C., Cheng, J., Liu, J., Zhang, L., He, C., Shen, W. H., Jin, H., Xu, L., & Zhang, Y. (2016). Arabidopsis Flower and Embryo Developmental Genes are Repressed in Seedlings by Different Combinations of Polycomb Group Proteins in Association with Distinct Sets of Cis-regulatory Elements. *PLoS Genetics*, *12*(1), 1–25. https://doi.org/10.1371/journal.pgen.1005771

Wang, J., Xia, H., Zhao, S. Z., Hou, L., Zhao, C. Z., Ma, C. Le, Wang, X. J., & Li, P. C. (2018). A role of GUNs-Involved retrograde signaling in regulating Acetyl-CoA carboxylase 2 in Arabidopsis. *Biochemical and Biophysical Research Communications*, *505*(3), 712–719. https://doi.org/10.1016/j.bbrc.2018.09.144

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M. T. A., Spillane, C., Nordborg, M., Rehmsmeier, M., & Köhler, C. (2011). High-Resolution analysis of parent-of-origin allelic expression in the Arabidopsis endosperm. *PLoS Genetics*, *7*(6). https://doi.org/10.1371/journal.pgen.1002126

Yang, X., Tong, A., Yan, B., & Wang, X. (2017). Governing the silencing state of chromatin: The roles of polycomb repressive complex 1 in arabidopsis. *Plant and Cell Physiology*, *58*(2), 198–206. https://doi.org/10.1093/pcp/pcw209