Unequal genetic redundancy between the PcG proteins CLF and SWN has created distinct biochemical properties

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

ALEXANDER FÖRDERER

aus Sankt Ingbert

Köln, 2019

Max Planck Institute for Plant Breeding Research

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

Prüfungsvorsitzender:

Prof. Dr. Maria Albani

Berichterstatter:

Prof. Dr. George Coupland

Prof. Dr. Ute Höcker

MAX-PLANCK-GESELLSCHAFT

Tag der Disputation: 11 September 2018

Abstract

Individual cells of multicellular organisms display distinct usage of identical genetic information. Gene expression states often correlate with posttranslational modifications (PTMs) on histones. Methylation of histone H3 on its lysine residue in position 27 (H3K27me) is such a signature epigenetic mark and its trimethylated form, H3K27me3, strongly correlates with transcriptional repression. In the model plant *Arabidopsis thaliana* (Arabidopsis), H3K27me3 in the sporophyte is exclusively catalyzed by the highly conserved histone methyltransferases CURLY LEAF (CLF) and SWINGER (SWN), which act in a protein complex called Polycomb repressive complex 2 (PRC2). Arabidopsis *swn* mutant plants do not have an obvious morphological phenotype, while *clf* mutant plants have a mild phenotype compared to the *clf swn* double mutant plants, which are full knockouts of sporophytic H3K27me3. However, a comparative biochemical analysis of CLF and SWN proteins in a PRC2 context has been lacking until now and it was uncertain if distinct differences are encoded in the coding or non-coding parts of *CLF* and *SWN* genes. This work shows that CLF and SWN share the enzymatic activity in a PRC2 oligomeric context to catalyze H3K27 methylation, but that their contribution to H3K27me3 is unequal due to an individually different specificity (K_m/kcat) to H3K27 methylated forms (H3K27me0/1/2). Given their overlapping expression in meristems and their interchangeability of the respective non-coding parts, however, *CLF* and *SWN* are genetically redundant *in planta*. These findings are consistent with previous H3K27me3 ChIP-seq data in *clf* and *swn* mutants and strengthen the argument, that CLF protein is able to hypermethylate its target genes, classified as CLF-dependent genes, in the absence of SWN protein; as is the case in the *swn* mutant. Conversely, SWN protein relies on the presence of CLF protein to achieve H3K27me3 at these CLF-dependent target genes; as is the case in the *clf* mutant. My results demonstrate how the duplication of the common ancestral gene of *CLF* and *SWN* at the base of Angiosperm phylogeny has led to an unequal genetic redundancy in Arabidopsis. The results further imply a strong divergence of the coding sequences of Arabidopsis *CLF* and *SWN* from their green-lineage orthologues outside the core Brassicaceae. I anticipate my study to be a starting point to gain a better understanding of the PRC2 oligomeric composition in Arabidopsis and to further characterize such PRC2 variants.

Zusammenfassung

Einzelne Zellen mehrzelliger Organismen verwenden die identische genetische Information unterschiedlich. Genexpressionszustände korreliert häufig mit dem lokalen Vorliegen von posttranslationalen Modifikationen (PTMs) an Histonen. Die Methylierung von Histon H3 an seinem Lysinrest in Position 27 (H3K27me) ist eine solche charakteristische epigenetische Markierung und seine trimethylierte Form, H3K27me3, korreliert stark mit transkriptioneller Repression. In der Modellpflanze *Arabidopsis thaliana* (Arabidopsis) wird H3K27me3 im Sporophyt ausschließlich durch die hochkonservierten Histon-Methyltransferasen CURLY LEAF (CLF) und SWINGER (SWN) katalysiert, welche in einem Proteinkomplex namens Polycomb repressive complex 2 (PRC2) vorkommen. *Swn*-mutierte Arabidopsis-Pflanzen haben keinen offensichtlichen morphologischen Phänotyp, wohingegen *clf*-Mutanten einen schwächeren Phänotyp als die *clf-swn*-Doppelmutante haben, welche durch den vollständigen Verlust sporophytischer H3K27-Trimethlyierung gekennzeichnet ist. Bisher fehlte jedoch eine vergleichende biochemische Analyse der CLF- und SWN-Proteine in einem PRC2-Kontext und es war ungewiss, ob deutliche Unterschiede in den kodierenden oder nicht-kodierenden Teilen von *CLF*- und *SWN*-Genen kodiert werden. Hier zeige ich, dass CLF und SWN sich in ihrer enzymatischen Aktivität im PRC2 Oligomer ähneln, insofern dass sie alle Schritte der H3K27-Methylierung katalysieren. Ihr jeweiliger Beitrag zur H3K27 Trimethlyierung ist jedoch ungleich wegen einer individuell unterschiedlichen Spezifitätskonstante ($K_m/kcat$) im Bezug auf die zu methylkierende H3K27-Form (H3K27me0/1/2). Außerdem überlappt die Expression der Gene *CLF* und *SWN* in Meristemen und sie verhalten sich genetisch redundant *in planta,* da ihre nicht-kodierenden Teile gentechnisch austauschbar sind. Diese Befunde bestätigen und erweitern frühere H3K27me3 ChIP-seq Daten in *clf* und *swn* Mutanten, welche implizierten, dass das CLF Protein seine Zielgene in Abwesenheit des SWN Proteins hypermethylieren kann (wie in der *swn* Mutante), und umgekehrt, dass das SWN Protein abhängig ist vom Vorhandensein des CLF Proteins, um H3K27me3 bei den gleichen Zielgenen zu realisieren (wie in der *clf* Mutante). Meine Ergebnisse zeigen, dass die Duplikation des gemeinsamen Vorläufergens von *CLF* und *SWN* an der Basis der Stammesgeschichte von Angiospermen zu einer ungleichen genetischen Redundanz in Arabidopsis geführt hat. Die Ergebnisse implizieren ferner eine starke Divergenz der codierenden Sequenzen von Arabidopsis CLF und SWN von ihren Orthologen in anderen Grünpflanzen außerhalb der ‚core Brassicaceae'. Es ist davon auszugehen, dass meine Studie ein Ausgangspunkt für die Aufklärung der Zusammensetzung von PRC2-Oligomeren in Arabidopsis sein wird und eine tiefergehende biochemische Analyse dieser PRC2 Varianten ermöglicht.

Abbreviations

^{14}C SAM – ^{14}C isotope-labelled SAM

3C – chromatin conformation capture

Ab – antibody

Aethionema arabicum – *A. arabicum*, Ae

AG – AGAMOUS (Arabidopsis)

Arabidopsis thaliana – Arabidopsis, At

Arabis alpina – *A. alpina*, Aa

Asx – Calypso or Additional sex combs (Drosophila)

AtBMIA/B/C – ARABIDOSPSIS BMI HOMOLOG A, B and C (Arabidopsis)

AtRING1A/B – ARABIDOSPSIS RING1 HOMOLOG A and B (Arabidopsis)

ATRX5 – ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 (Arabidopsis)

ATRX6 – ARABIDOPSIS TRITHORAX-RELATED PROTEIN 6 (Arabidopsis)

BAH – protein domain; stands for 'bromo adjacent homology domain'

BAM – protein domain; stands for 'ß-addition motif'

Bar – BASTA resistance

BN PAGE – blue native polyacrylamide gel electrophoresis

Brachypodium distachyon – *B. distachyon*, Bd

BRCT – protein domain, named after 'BRCA1 C-terminus'

C – carboxy

C5/MCSS – protein domain, interacting with VEFS protein domain

CAF1 – CHROMATIN ASSEMBLY FACTOR 1 (Arabidopsis)

CDS – coding sequence

ChIP – chromatin immunoprecipitation

CLF – CURLY LEAF (Arabidopsis)

CLF-PRC2 – Polycomb repressive complex 2 composed of CLF, EMF2, FIE and MSI1

CoIP-MS – Co-immunoprecipitation followed by mass spectrometry

CUL4-DBB1 – CULLIN4 –DNA DAMAGE BINDING-PROTEIN1 (Arabidopsis)

CV – column volume

CW-Zf – protein domain, named after its zinc finger and CW motif

CXC – protein domain, named after its C-X(6)-C-X(3)-C-X-C motif

dKDM2 – Lysine demethylase 2 (Drosophila)

Drosophila – *Drosophila melanogaster*, Dm

E(z) – Enhancer of zeste (Drosophila)

EBD – protein domain; stands for 'EED-binding domain'

EED – EMBRYONIC ECTODERM DEVELOPMENT (human)

ELF6 – EARLY FLOWERING 6

EMF1 – EMBRYONIC FLOWER 1 (Arabidopsis)

EMF2 – EMBRYONIC FLOWER 2 (Arabidopsis)

Esc – Extra sex comb (Drosophila)

EST – expressed sequence tag

EZH1 – ENHANCER OF ZESTE HOMOLOG 1 (human)

EZH2 – ENHANCER OF ZESTE HOMOLOG 2 (human)

FIE – FERTILIZATION INDEPENDENT ENDOSPERM (Arabidopsis)

FIS2 – FERTILIZATION INDEPENDENT SEED 2 (Arabidopsis)

FISH – fluorescence in situ hybridisation

FLC – FLWOERING LOCUS C (Arabidopsis)

FPLC – fast protein liquid chromatography

FRET – Förster resonance energy transfer

FT – FLOWERING LOCUS T (Arabidopsis)

FUS3 – FUSCA 3(Arabidopsis)

gCLF – genomic *CLF* sequence

GFP – green fluorescent protein

gSWN – genomic *SWN* sequence

H3 – histone H3

H3K27me3 – Histone H3 trimethylation on residue lysine 27

HMTase – histone methyltransferase

HMTase X – unknown H3K27 dimethylase (Arabidopsis)

HSP70 – heatshock protein 70

HTRF – homogeneous time-resolved fluorescence

JMJ – protein domain; stands for 'Jumonji'

kcat – turnover number

K_d – dissociation constant

K_m – Michaelis-Menten constant

K_m/kcat – specificity constant

LD – long days

LHP1 – LIKE HETEROCHROMATIN 1 (Arabidopsis)

mCpG – methylation of CpG dinucleotide

MEA – MEDEA (Arabidopsis)

minigene – minimal gene consisting of promoter, coding sequence and 3'untranslated region, e.g. *miniSWN:SWN*

MIR156 – microRNA 156

MOI – multiplicity of infection

mRNA – messenger RNA

MSI1 – MULTISUPPRESSOR OF IRA 1(Arabidopsis)

N – amino

n – Hill coefficient

N – number of replicates

Nicotiana benthamiana – N. benthamiana, tobacco

Nurf55 – nuclear remodeling factor 55 (Drosophila)

NURF55 – NUCLEAR REMODELLING FACTOR 55 HOMOLOG (human)

OD600 – optical density at 600 nm

Pc – Polycomb

PcG – Polycomb Group

PCL – POLYCOMB LIKE (human)

PEV – position effect variegation

Ph – Polyhomeotic (Drosophila)

PHD – protein domain; stands for 'plant homeodomain'

Pho – Pleiohomeotic (Drosophila)

Physcomitrella patens – P. patens, Physcomitrella, Pp

PID – percentage identity

PRC1 – Polycomb repressive complex 1

PRC2 – Polycomb repressive complex 2

Psc – Posterior sex combs (Drosophila)

PTM – posttranslational modification

RAWUL – protein domain, stands for 'Ring-finger And WD40 associated Ubiquitin-Like'

RB – RETINOBLASTOMA (Arabidopsis)

RBR – RETINOBLASTOMA RELATED (Arabidopsis)

rDNA – ribosomal DNA

REF6 – RELATIVE OF EARLY FLOWERING

SAH – S-adenosylhomocysteine

SAL – protein domain, stands for 'SET activation loop'; derived from structural annotation

SAM – S-adenosylmethionine

SANT – protein domain, named after its occurrence in Swi3, Ada2, N-Cor, and TFIIIB

SBD – protein domain, stands for 'SANT1-binding domain'; derived from structural annotation

SD – short days

SDS PAGE – SDS polyacrylamide gel electrophoresis

SEP3 – *SEPALLATA* 3 (Arabidopsis)

SET – protein domain, named after its occurrence in Su(var)3-9, Enhancer of zeste and Trithorax proteins

Sf21 – insect cells line recovered from *Spodoptera frugiperda*

SlEZ1 – *SOLANUM LYCOPERSICUM* ENHANCER OF ZESTE HOMOLOG 1; homolog of SWN (tomato)

SlEZ2 – SOLANUM LYCOPERSICUM ENHANCER OF ZESTE HOMOLOG 2; homolog of CLF (tomato)

SN – supernatant

Solanum lycopersicum – *S. lycopersicum*, Sl, tomato

SRM – protein domain, stands for 'stimulation response motif'; derived from structural annotation

Su(z)12 – Suppressor of zeste 12 (Drosophila)

SWN – SWINGER (Arabidopsis)

SWN-PRC2 – PRC2 complex composed of SWN, EMF2, FIE and MSI1

TE – transposable element

UCL1 – UPWARD CURLY LEAF 1 (Arabidopsis)

UTR – untranslated region

VEFS – protein domain, named after its occurrence VRN2, EMF2, FIS2 and Su(z)12

VEL1 – VERNALIZATION-LIKE 1 (Arabidopsis)

V_{max} – maximal velocity

VRN2 – VERNALIZATION 2 (Arabidopsis)

VRN5 – VERNALIZATION 5 (Arabidopsis)

WD40 – protein domain, named after its repeats of W and D

WGD – whole-genome duplication

WT – wild-type

αSAH Ab – specific antibody towards S-adenosylhomocysteine

Table of Contents

1 Introduction

1.1 Epigenetics – chromatin states correlate with gene expression

Multicellular organisms possess a diverse range of cell types, each with disparate use of identical genetic information. Historically, cellular differentiation has been considered an epigenetic phenomenon orchestrated by non-genetic heritable changes forming the 'epigenetic landscape' described by Waddington (Waddington 1957). Today it is widely accepted that posttranslational modifications (PTMs) on both DNA and DNA-associated proteins embody a significant portion of these non-genetic changes (Goldberg, Allis et al., 2007).

Two major discoveries established some important aspects about the transcriptional state of genes and their genomic location: (1) When chromatin is visualised with stains such as Giemsa or Feulgen, it reveals two highly contrasting states defined as euchromatin and heterochromatin (Heitz 1928). (2) A specific gene, called *white*, is either silent or active when located in heterochromatin or euchromatin respectively. This phenomenon was termed 'position effect variegation' (PEV) (Müller 1930, Demerec and Slizynska 1937). These two findings (1 and 2) inspired the idea that euchromatin tends to be transcriptionally active, while heterochromatin is transcriptionally idle (Grewal & Moazed, 2003). However, such a simplistic binary model does not capture the dynamic changes occurring during cellular differentiation that have been identified more recently. Modern epigenetics further differentiates between facultative heterochromatin, which may switch between heterochromatic and euchromatic states, and constitutive heterochromatin present at regions such as telomeres, centromeres and additional repetitive and noncoding regions of higher eukaryotes (Margueron, Trojer et al., 2005).

Modern techniques such as fluorescence *in situ* hybridisation (FISH), super-resolution microscopy and chromatin conformation capture (3C and related methods e.g. 4C, 5C and HiC) are instrumental in understanding the dynamic nature of nuclear organisation (Fransz & de Jong, 2011, Grob & Grossniklaus, 2017, Pombo & Dillon, 2015, Ricci, Cosma et al., 2017). According to the consensus model of nuclear chromatin, single chromosomes also occupy defined regions or territories within the nucleus (Baroux, Pecinka et al., 2007, Cremer, Cremer et al., 1982).

At a molecular scale nuclear DNA is packaged into chromatin through hydrogen-bonding with small positively charged proteins called histones (Luger, Mäder et al., 1997). The DNA connects consecutive histones in a beads-on-a-string-like fashion (Luger et al., 1997). The reoccurring motif of this 10 nm fiber is the nucleosome: a linker DNA of varying length (commonly 55 bp) and a ~147 bp-long DNA, which wraps itself around an octamer of core histones in a left-handed super-helix (Fyodorov, Zhou et al., 2017). Core histones are generally composed of two copies each of H2A, H2B, H3 and H4. In addition, specific isoforms of core histones can replace canonical histone isoforms. In

plants and animals. For instance, metazoan H3.3 can replace H3.1 at specific genomic locations upon active transcription of a gene.

According to the textbook model, nuclear DNA is further compacted from the 10 nm into a 30 nm fiber by the action on linker histone H1. Indeed, H1 causes the formation of a 30 nm fiber *in vitro* (Finch & Klug, 1976). However, the long-standing paradigm that the 30 nm fiber is a fundamental building block of higher-order native chromatin is now being questioned (Efroni, Duttagupta et al., 2008, Fussner, Strauss et al., 2012, Joti, Hikima et al., 2012, Nishino, Eltsov et al., 2012, Ricci et al., 2017). Current consensus is that histone H1 facultatively binds the exit/entry point of the nucleosomal DNA (Fyodorov et al., 2017). H1 is enriched in heterochromatin and a recent study showed that *in vivo*, H1 is found in nucleosome clusters containing only 4-8 nucleosomes rather than in a continuous 30 nm fiber (Ricci, Manzo et al., 2015). The core histones possess a DNA-binding globular domain, while exposing their amino (N)-terminal tails (Luger et al., 1997). The tails, as well as the globular domain, are a target for a multitude of PTMs, for example phosphorylation of serine and threonine residues (S and T), acetylation of lysine (K), methylation of arginine (R) and of K (Margueron et al., 2005). Although these PTM might be the most abundant (Zhang, Cooper et al., 2015), the list of novel modifications is frequently updated (Arnaudo & Garcia, 2013). Many histone marks correlate with transcriptional state of individual genes and signify cell identity (Mikkelsen, Ku et al., 2007). Histone modifications are annotated as follows: histone variant (H3) – amino acid (K) - number (27) - modification (me) - count (3) – e.g. H3K27me3

1.2 The histone code – readers, writer and erasers

In analogy to the genetic code, scientists developed the concept of a 'histone code' (Jenuwein & Allis, 2001). In the genomic era modification-specific antibodies in combination with chromatin immunoprecipitation (ChIP) allowed the identification of genetic regions enriched with specific chromatin marks and histone isoforms in many model organisms (Kouzarides, 2010). The advent of the genomic era also brought with it a looser usage of the terms euchromatin and heterochromatin (see Infobox I 'definition: heterochromatin and euchromatin').

Epigenetic 'readers', 'writers' and 'erasers' are proteins that assess and edit the 'histone code' by scanning, depositing and removing chromatin marks, respectively. Indeed, mutants of these proteins tend to show defects in heterochromatin or euchromatin formation and also global misexpression of developmental genes.

Infobox I

Definition: Heterochromatin and Euchromatin

Today, according to the idea 'guilt by association' (Weblink [1]), the association of certain chromatin marks with certain DNA sequences inside the genome usually satisfies the definition of heterochromatin or euchromatin without the need of microscopic confirmation. For example in Arabidopsis, DNA methylation and H3K9me9 are enriched at centromeres, telomeres, transposable elements (TEs) and repetitive regions and are therefore called 'heterochromatic marks', while H3K4me2 and H3K36me3 dominate in coding genes of the chromosome arms and are called 'euchromatic marks' (Jiang & Berger, 2017). In this study, I will use the terms 'heterochromatic region' and 'euchromatic region' in accordance with this modern definition, while 'heterochromatin' and 'euchromatin' are used according to Heitz's, microscopy-based, definition (Heitz 1928).

1.3 PRC2 and H3K27me3

Polycomb (*Pc*) is a gene coding for an epigenetic 'writer'. *Pc* mutant fruit flies of *Drosophila melanogaster* (Drosophila) show misregulation of homeotic genes, which govern body segmentation (Grimaud, Nègre et al., 2006).The phenotype of *Pc* mutant flies is similar to *Suppressor of zeste* 12 (*Su(z)12*)and *Enhancer of zeste* (*E(z)*) mutant flies. Therefore, the underlying genes in these mutants were classed as Polycomb group (PcG) genes (Jürgens, 1985).

E(z) encodes a histone code 'writer' belonging to the protein family of SET domain methyltransferases (HMTase) and E(z) in particular was shown to methylate H3K27 (Czermin, Melfi et al., 2002). Homologs of E(Z) exist in many species, including plants. Arabidopsis encodes three E(Z) homologous proteins, named MEDEA (MEA), CURLY LEAF (CLF) and SWINGER (SWN). While MEA is expressed in the gametophytic life stage of Arabidopsis, CLF and SWN are responsible for H3K27 methylation in the sporophytic life stage (Mozgova & Hennig, 2015b). As indicated by their mutant phenotypes, the PcG also governs developmental genes in plants. For example, double mutants of *clf swn* show formation of an amorphous cell cluster that randomly starts forming embryo and leaf-like structures without ever completing organ development (PcG-callus) (Forderer, Zhou et al., 2016). Eventually, this progressive loss of cell identity in *clf swn* double mutants leads to somatic embryogenesis (Ikeuchi, Iwase et al., 2015).

E(z)-type proteins associate in a complex named Polycomb repressive complex 2 (PRC2) (Simon & Kingston, 2009). Drosophila PRC2 comprises E(z), Su(z)12, Extra sex comb (Esc) and Nuclear remodeling factor 55 (Nurf55). The vertebrate homologs of tetrameric PRC2 are ENHANCER OF ZESTE HOMOLOG 1 or 2 (EZH1 or EZH2), SUPPRESSOR OF ZESTE 12 HOMOLOG (SUZ12), EED – EMBRYONIC ECTODERM DEVELOPMENT HOMOLOG (EED, present in different isoforms) and NUCLEAR

REMODELLING FACTOR 55 HOMOLOG (NURF55) (listed in the same order like Drosophila homologs). Arabidopsis PRC2 homologs are: CLF, SWN or MEA; EMBRYONIC FLOWER 2 (EMF2), VERNALIZATION 2 (VRN2) or FERTILIZATION INDEPENDENT SEED 2 (FIS2); FERTILIZATION INDEPENDENT ENDOSPERM (FIE); MULTISUPPRESSOR OF IRA 1 (MSI1) (see Table 1 for overview of PRC2 nomenclature). Mutations of PRC2 components can lead to similar phenotypes. For example, double mutants of *emf2 vrn2* (knockout alleles) or mutants of the single copy gene *fie* (gametophytic rescue allele) develop into a PcG callus similar to that of *clf swn* double mutants (Nowack, Shirzadi et al., 2007, Schubert, Clarenz et al., 2005).

Table 1 'Nomenclature overview of PRC2' components of drosophila, Human and Arabidopsis

Droso-phila	**Human**	**Arabidopsis**
Nurf55	NURF55	MSI1
Esc	EED	FIE
Su(z)12	SUZ12	EMF2, VRN2, FIS2
E(z)	EZH1, EZH2	CLF, SWN, MEA

1.3.1 Genome-wide distribution of H3K27 methylation states in Arabidopsis

H3K27me is the signature mark of PRC2 and can generally occur in three states: mono- (H3K27me1), di- (H3K27me2) and trimethylation (H3K27me3).

In Arabidopsis seedlings H3K27me3 marks about 20% of coding genes and is predominantly found in euchromatic regions. These H3K27me3 target genes are typically developmental genes with cell type-specific function and are located in euchromatic regions (Fuchs, Demidov et al., 2006, Lafos, Kroll et al., 2011, Lippman, Gendrel et al., 2004, Oh, Park et al., 2008, Roudier, Ahmed et al., 2011, Tanurdzic, Vaughn et al., 2008, Turck, Roudier et al., 2007, Zhang, Clarenz et al., 2007). From animal models we know that H3K27me3-marked genes show reduced transcription(Brookes, de Santiago et al., 2012), decreased H3K27 acetylation (Pasini, Malatesta et al., 2010) and increased chromatin compaction(Deaton, Gómez-Rodríguez et al., 2016, Eskeland, Freyer et al., 2010). A study in human cancer cells showed that the linker histone variant H1.2 recognises H3K27me3 and that an increased incorporation of H1.2 at H3K27me3 regions mediates chromatin compaction and transcriptional silencing (Kim, Kim et al., 2015). Such a direct link between plant linker histone variants H1.3 or H1.1, and H3K27me3 was not addressed and the transferability awaits verification.

H3K27me2 and H3K27me1 are less studied, especially in plants, as only two genome-wide ChIP experiments are published (Park, Oh et al., 2012, Roudier et al., 2011), and a comprehensive study of transcriptional correlation is missing. The two studies highlight that H3K27me1 is a signature mark

of heterochromatic regions, which have high TE content, high methylation at CpG dinucleotides (mCpG) and high H3K9me2 (Jacob, Feng et al., 2009, Jacob, Stroud et al., 2010, Jiang & Berger, 2017, Roudier et al., 2011, Sequeira-Mendes & Gutierrez, 2015, Stroud, Hale et al., 2012). Unlike H3K27me3, which favorably marks coding genes, H3K27me1 and H3K27me2 mark both heterochromatic regions and coding genes (Roudier et al., 2011) Both genome-wide and at coding gene level, H3K27me2 correlates with both H3K27me1 and H3K27me3, while H3K27me3 and H3K27me1 anticorrelate at genes with diverse cellular function (Roudier et al., 2011).

For the most part, FISH experiments corroborate the genome-wide distribution of all three H3K27me marks. While H3K27me3 disperses more widely over the nucleus, both H3K27me1 and H3K27me2 concentrate in heterochromatic chromocenters, which harbor rDNA and other repetitive regions of the Arabidopsis pericentromer(Baroux et al., 2007, Fuchs et al., 2006, Lindroth, Shultis et al., 2004, Mathieu, Probst et al., 2005).

1.3.2 Writers of H3K27 methylation

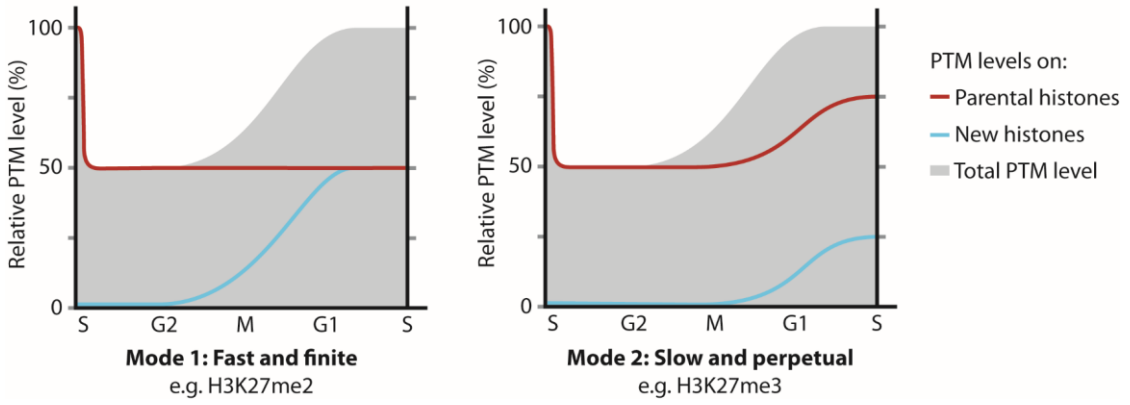
CLF and SWN are 'writers' of sporophytic H3K27me. However, western blots using specific antibodies (Ab) against all three kinds of H3K27 methylation in *clf swn* double mutants showed that they are not the only 'writers' of H3K27me1 (Lafos et al., 2011). In *clf swn* double mutants, H3K27me3 is absent, whereas H3K27me1 is unaltered (Lafos et al., 2011). The redundant ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 and 6 (ATRX5, ATRX6) proteins were shown *in vivo* and *in vitro* to be functional H3K27 monomethylases (Jacob et al., 2009). Double mutants of *atrx5 atrx6* show 22% decreased H3K27me1, unaltered H3K27me2/3 levels, partial heterochromatin decondensation and transcriptional activation of heterochromatic ribosomal DNA (rDNA) (Jacob et al., 2009, Pontvianne, Blevins et al., 2012, Stroud et al., 2012). Interestingly, ATRX5 and ATRX6 are shown to only methylate H3.1, but not H3.2 (Jacob, Bergamin et al., 2014). This indicates that CLF and SWN do not act in redundancy with ATRX5 and ATRX6 in the deposition of H3K27me1 on the H3 variant H3.2. Mass spectrometry analysis quantified K27 methylation in wild-type (WT) Arabidopsis and found that: (1) H3.1 carries 60% K27me1, 16% K27me2 and 5% K27me3. (2) H3.2 carries overall less K27 methylation and has 36% K27me1 and 6% K27me2 and no detectable K27me3 (Johnson, Mollah et al., 2004). This is relevant, because H3.1 is by default incorporated during DNA replication, whereas H3.2 functions as a replacement histone during transcription (Johnson et al., 2004) (see Infobox II 'chromatin marks and the cell cycle'). At last, H3K27me2 might also not exclusively stem from CLF and SWN activity. In fact, western blot analysis showed even increased H3K27me2 levels in *clf swn* double mutants(Lafos et al., 2011). Hypomorphic *fie* mutants and other strong PcG mutants also retain H3K27me2 at heterochromatin (Lindroth et al., 2004, Schubert et al., 2005). However, the specificity of the H3K27me2 antibodies (Ab) used in those studies is somewhat controversial (Lafos et al., 2011, Lindroth et al., 2004, Schubert et al., 2005).

The effect of H3K27me3 on plant gene transcription is probably best studied at the floral repressor gene *FLOWERING LOCUS C* (*FLC*). Prolonged cold in winter triggers flowering in some Arabidopsis accessions after return to warmer temperatures, a process known as vernalisation (Yang, Berry et al., 2017). During cold exposure, *FLC* acquires a repressive chromatin state, which involves PcG activity at the FLC locus. During cold temperatures, CLF and SWN initiate H3K27me3 deposition or 'nucleation' close to the FLC transcription start site. After return to warm temperatures, CLF, but not SWN, is necessary for H3K27me3-spreading over the entire FLC locus and robust transcriptional shutdown that is heritable through cell divisions is realised (Yang et al., 2017). Similar to H3K27me3, H3K27me2 might also be involved in transcriptional shutdown of FLC (Bastow, Mylne et al., 2004, Sung & Amasino, 2004), but due to the lack of a specific Ab against H3K27me2, at the time, a reconfirmation of this result is needed. In addition, no genome-wide study has made a correlation between H3K27me2 and the transcriptional state of all genes. A mathematical model using quantitative fit to experimental data was made of epigenetic memory (Berry, Dean et al., 2017). In this model two preconditions were slow H3K27 methylation by PRC2 and, importantly, a similarly repressive role of H3K27me2 and H3K27me3 (Berry et al., 2017).

Infobox II

Chromatin marks and the cell cycle

The DNA replication machinery disrupts histone–DNA interaction during S-phase. Newly synthesised DNA strands are decorated with new histone by two mechanisms: (1) recycling of old (modified) histones (2) incorporation of new (unmodified) histones. This leads to a two-fold dilution of chromatin marks. With an elegant combination of nascent chromatin capture and triple SILAC applied in a human cell line, Alabert and colleagues shed light on chromatin mark restoration after DNA replication. Two modes are postulated for complete chromatin mark restoration: (mode 1) modification of new histones (most chromatin marks e.g. H3K27me1/2) and (mode 2) modification of both new and old histones (H3K27me3, H3K9me3). See figure below, modified from (Alabert, Barth et al., 2015).



1.3.3 Erasers of H3K27 methylation

Besides the 'writers' of H3K27 methylation, 'erasers' might play an equally fundamental role. In Arabidopsis four K27 demethylases of the Jumonji (JMJ) family are identified: RELATIVE OF EARLY

FLOWERING (REF6), EARLY FLOWERING 6 (ELF6), JMJ30 and JMJ32 (Gan, Xu et al., 2015). REF6, JMJ30 and JMJ32 have H3K27m2/3, but not H3K27me1, demethylation activity *in vitro (Gan et al., 2015)* and all four contribute to H3K27me2/3 demethylation *in vivo* (Crevillén, Yang et al., 2014, Gan, Xu et al., 2014, Lu, Cui et al., 2011). During floral transition, REF6, ELF6, JMJ30 and JMJ32 act as H3K27me3 demethylases at *FLC* to achieve transcriptional derepression *vivo* (Crevillén et al., 2014, Gan et al., 2014, Lu et al., 2011). With regard to their genome-wide effect on H3K27me3, only *ref6* mutants were studied and showed a global increase in H3K27me3 levels (Lu et al., 2011).

1.3.4 Readers of H3K27 methylation

'Reader' proteins encompass many protein classes and exist in all eukaryotes (Goldberg et al., 2007). In animal models three single H3K27me3 readers were found within the following three, vastly different protein classes: (1) Chromo domain proteins, (2) WD40 domain proteins and (3) Bromo-adjacent homology (BAH) domain proteins (Zhao, Zhang et al., 2016).

In Drosophila the chromo domain protein Pc plays a pivotal role in the classic model of PcG repression. Found within the WD40 class of proteins is the mammalian PRC2 component EED. It specifically binds to H3K27me3 and triggers allosteric activation of PRC2 (Margueron, Justin et al., 2009). Found within the BAH class of proteins is the mammalian BAHD1, which was only recently described to be a H3K27me3 reader *in vitro* (Zhao et al., 2016), but its relevance in H3K27me3 dynamics *in vivo* is not understood. It is noteworthy, that no H3K27me1/2 readers are identified in any model species so far.

Plants possess a vast range of putative histone mark readers belonging to nine protein families grouped by their specific domains (Zhao, Zhang et al., 2018). These domains are: Plant homeodomain (PHD), Tudor, Bromo-adjacent homology (BAH), Chromo, Proline-tryptophan-tryptophan-proline (PWWP), Bromo, BRCA1 C-terminus (BRCT), 14-3-3 and zinc finger CW (CW-Zf) domains (Zhao et al., 2018).

To date, the only H3K27me3 reader identified in plants is LIKE HETEROCHROMATIN 1 (LHP1) (Turck et al., 2007, Zhang et al., 2007). Despite homology to Drosophila HETEROCHOMATIN PROTEIN 1 (HP1), which binds to heterochromatin and H3K9me *in vitro*, LHP1 localises to euchromatin and behaves as a functional homolog of Pc (Turck et al., 2007, Zhang et al., 2007). Although LHP1 was shown to bind H3K27me3 and H3K9me2 *in vitro*, it localises to H3K27me3 *in vivo* (Turck et al., 2007) (Exner, Aichinger et al., 2009, Zhang et al., 2007). Like Pc, LHP1 is necessary for H3K27me3 maintenance *in vivo* and acts at the interface of both PRCs by direct binding of PRC1 and PRC2 components (Derkacheva, Steinbach et al., 2013). However, *lhp1* single mutants show a mild phenotype and a small decrease in H3K27me3 levels compared with the full PRC2 knockout *clf swn* of Arabidopsis and the lethal *pc* mutant of Drosophila. This either indicates that, in plants, (1)

additional H3K27me3 readers exist, or that (2) the classic linear model for PcG recruitment is incomplete. Both hypotheses are currently being explored.

1.4 PRC1 and the hierarchal model

According to the classic hierarchical model of PcG gene repression, PRC1 is recruited *via* the interaction of Pc as a 'reader' of H3K27me3, which was previously deposited by PRC2. RING1 and BMI1 type proteins of PRC1 then deposit H2A118ub1 (H2A119ub1 in vertebrates, H2A121ub1 in Arabidopsis) and gene repression is achieved.

In Drosophila the PRC1 core module is a heterodimer between two RING domain proteins. It comprises dRing1 and either Posterior sex combs (Psc), Su(z)2 or potentially L(3)73 AH and acts as E3 ligase to catalyze H2A ubiquitination (Kahn, Dorafshan et al., 2016, Lee, Kahn et al., 2015)). Variant PRC1 can have different activities such as chromatin compaction, heterochromatin condensation, DNA-binding and H2A deubiquitination. Which one of these activities occurs depends on the modular composition of PRC1, which can encompass Polyhomeotic (Ph) and Pc, Lysine demethylase 2 (dKDM2), Pleiohomeotic (Pho), Calypso or Additional sex combs (Asx) with or without the RING heterodimer. Vertebrate homologs are identified for most these components. Arabidopsis PRC1 composition is less clear. The RING domain core module includes one of the AtBMI1s (AtBMI1A/B/C) and AtRING1A or AtRING1B. Similar to *clf swn* double mutants, triple mutants *atbmi1/2/3* and double mutants *atring1a/b* develop into PcG calli (Bouyer, Roudier et al., 2011, Bratzel, Lopez-Torrejon et al., 2010, Bratzel, Yang et al., 2012, Chen, Molitor et al., 2010, Lindroth et al., 2004, Nowack et al., 2007).

LHP1 and EMBRYONIC FLOWER 1 (EMF1) are additional proteins that are specific to plants. Similarly to LHP1, EMF1 interacts with both PRCs. Although a single copy gene, EMF1 might not be an obligatory component of PRC1 as its mutant phenotype is less severe than PRC1 null mutant *atbmi1/2/3* or PRC2 null mutant *clf swn*. Biochemical experiments showed that EMF1 binds DNA, hinders transcription, inhibits chromatin remodeling and facilitates chromatin compaction (Beh, Colwell et al., 2012, Calonje, Sanchez et al., 2008, Kim, Lee et al., 2012). *In vivo* experiments showed that EMF1 mediates heterochromatin formation and co-distributes widely with H3K27me3 (Kim et al., 2012).

Recently, the hierarchical model of PRC recruitment has been revised and new models of PcG recruitment are emerging in most model species (Calonje, 2014). A new model in Arabidopsis postulates that PRC1 can act upstream or in parallel to PRC2 (Calonje, 2014, Zhou, Romero-Campero et al., 2017a). This model is supported by genome-wide ChIP data from PRC1 knockdown mutants *atmbi1a/b/c*, PRC2 knockout mutants *clf swn* and the *lhp1* single mutant (Zhou et al., 2017a): (1) H2A121ub1 deposition occurs independent of H3K27me3 marking at most genes in *lhp1* mutant. (2) H3K27me3 depends on H2Aub at most genes in *atbmi1a/b/c*. This indicates that LHP1 is mostly

inessential for H2A119ub1 deposition and that feedback loops exist directly from PRC1 to PRC2. Indeed, EMF1 and LHP1 associate with PRC2 (Calonje et al., 2008, Derkacheva et al., 2013, Liang, Hartwig et al., 2015).

The complexity of PcG recruitment is reflected in the architecture of the underlying molecular machineries.

1.5 Biochemical analysis of PRC2

Both its sheer size of 260 kDa and the tetrameric nature of PRC2 hampered the resolution of its molecular structure(Tan, Yan et al., 2013). Due to collective effort, individual atomic structures were available for the two WD40 domain proteins, EED and NURF55, and a truncated, catalytically inactive CXC-SET twin-domain of EZH2 (Tan et al., 2013). The inactivity of the isolated SET domain was later explained due to occlusion by the CXC domain and might be a conformation also found *in vivo* (Antonysamy, Condon et al., 2013, Wu, Zeng et al., 2013).

1.5.1 Electron microscopy of PRC2

Technological progress in electron microscopy made it eventually possible to readdress the challenge of resolving PRC2 structure: In 2012, Ciferri *et al.* proposed a three dimensional model of the molecular architecture of human PRC2 (Ciferri, Lander et al., 2012). Briefly, PRC2, being a heterotetramer, forms a four-lobed dumbbell interacting with two adjacent nucleosomes. While EED and EZH2 build one double lobe and NURF55the other; SUZ12 is the principal component of the junction region and the second double lobe. In the catalytic double lobe EED faces one nucleosome where it fits H3K27me3 tails into the center of its WD40-repeat aromatic cage, while EZH2 faces the other nucleosome allowing binding to the H3 substrate tail. Although Ciferri *et al.* give an intimate view of unprecedented clarity of PRC2 architecture, the low resolution of 21 Å left much unresolved, particularly the enigmatic allosteric mechanisms of PRC2.

1.5.2 Crystallisation of partial PRC2 and single domains

Eventually, Jiao, Brooun and Justin were able to crystallise large parts of PRC2 from yeast *Chaetomium thermophilum* (≥ 2.3 Å), lizard *Anolis carolinensis* (≥ 2.62 Å) and human *Homo sapiens* (≥ 2.95 Å), respectively (Brooun, Gajiwala et al., 2016, Jiao & Liu, 2015, Justin, Zhang et al., 2016). These studies focused on four principle components of PRC2: namely, EZH2, EED, the VEFS domain of SUZ12 and the histone H3 tail.

EZH2 received much attention and its structure was subdivided into different moieties. In the C-terminal part, EZH2 carries a catalytic moiety, which includes the SET and CXC domain, a middle lobe, which comprises the VEFS-binding MCSS (or C5) and the SANT2 domain, and, finally, the regulatory N-terminus, which embraces EED in a belt-like structure. In more detail, the N-terminus harbors six

structural domains named SANT1-binding domain (SBD), EED-binding domain (EBD), ß-addition motif (BAM), SET activation loop (SAL), stimulation response motif (SRM) and SANT1 domain (Jiao *et al.* 2015). Specifically, the domains can be sub-grouped by function: (1) SANT1 and 2, together with SBD, seem to warrant the overall structure and are binding platforms for RNAs, for instance. (2) C5 tethers EZH2 to SUZ12 and transmits allosteric changes. (3) SAL, SRM, CXC and SET are in close contact and form a catalytic compartment sensitive to allosteric shifts in the remaining complex. (4) EBD and BAM are in close contact with EED, where EBD passages along a ravine of EED's WD40 propeller, and BAM adds three ß-sheets to EED. The activity of the SET domain depends on conformational shifts in the flexible and structurally disordered N-terminus (Justin et al., 2016).

To date, plant PRC2 structure remains unresolved. Figure 1 shows a cartoon of plant PRC2 based on Ciferri's model.

Complementary to molecular structure resolution, biochemical studies were used in animal models, to study substrate specificity and drug/co-factor inhibition or activation.

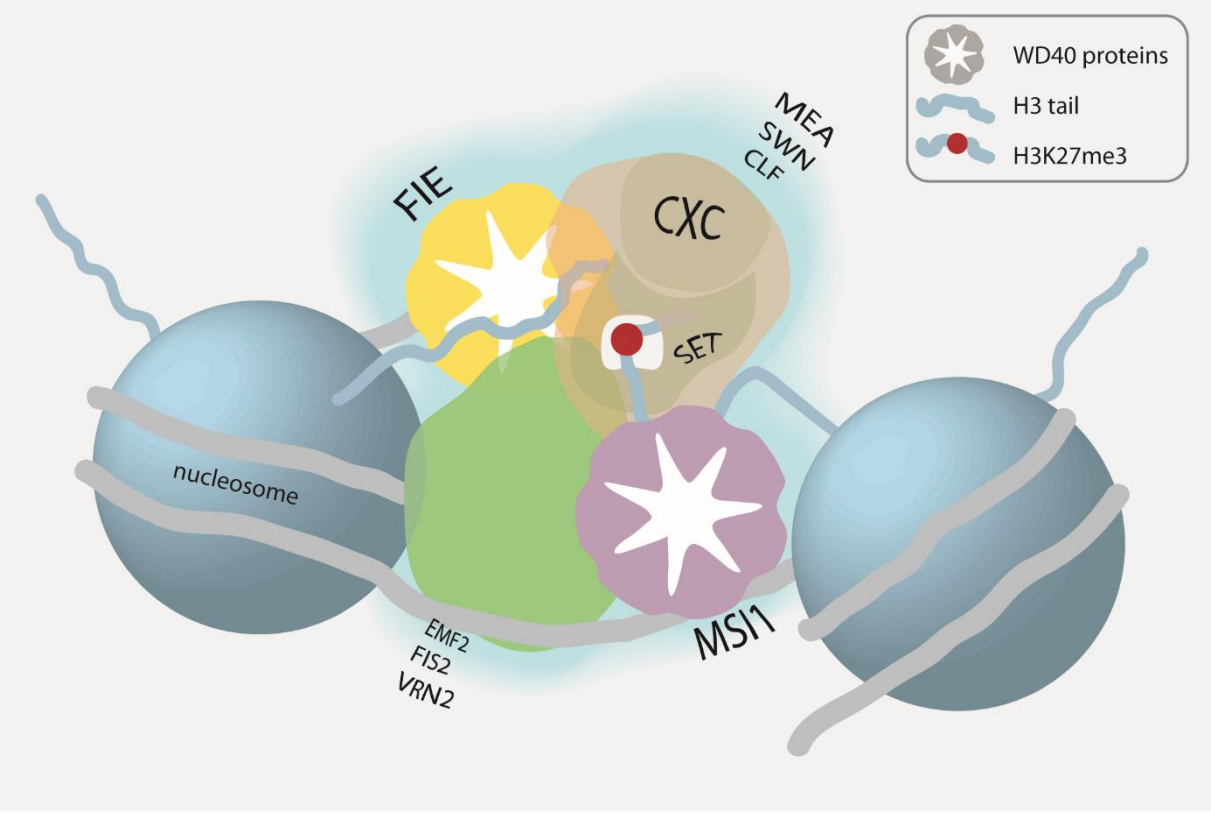


Figure 1 Cartoon of plant PRC2 based on high resolution electron microscopy model of human PRC2 by Ciferri et al. 2012. While the two WD40 propeller proteins FIE and MSI1 face one histone H3 tail each, EMF2/VRN2/FIS2 stabilise the complex and are sensitive to allosteric modulation originating on H3 tails. CLF/SWN/MEA are the catalytic SET domain components depositing H3K27 methylation. (image modified from Forderer et al. 2016).

1.5.3 PRC2 activity *in vitro*

Enzymatic activity of E(z) homologs can be studied using HMTase assays . In a histone methylation reaction, histone substrate and *S*-adenosylmethionine (SAM) co-substrate will react into the R/K-methylated histone and the byproduct *S*-adenosylhomocysteine (SAH), respectively (Luo 2013). Due to the slow enzymatic turnover, product or byproduct detection is favourable over detection of co-substrate depletion (Luo, 2012).

E(Z) on its own has poor activity in HMTase assays (Nekrasov, Wild et al., 2005). Instead, it depends on allosteric activation by ESC and SU(Z)12 (Nekrasov et al., 2005); similarly, human EZH2 depends on EED and SUZ12 (Margueron et al., 2009). PRC2 binding to nucleosome depends on SU(Z)12 and NURF55 subunits (Nekrasov et al., 2005).

In vitro, PRC2 methylates H3K27 in a non-processive manner meaning that it rather monomethylates H3K27, H3K27me1 or H3K27me2 substrate, than it trimethylates H3K27 in one (processive) reaction (Berry et al., 2017). ESC homologs have the ability to allosterically translate pre-existing H3K27me3, not H3K27me2 or H3K27me1, into complex activation (Margueron et al., 2009). *In vivo*, H3K27me3 is mostly formed from existing H3K27me2 substrates, and H3K27me2 is formed from H3K27me1 (Zee, Britton et al., 2012). Comparing the specificity constant $kcat/K_m$ as an indicator of catalytic efficiency, EZH2 loses activity when progressively more methyl groups are incorporated into H3K27 (H3K27me0 > H3K27me1 > H3K27me2), with the ratio of catalytic efficiencies for H3K27me0, H3K27me1, and H3K27me2 between 14:4:1 (Sneeringer, Scott et al., 2010) and 9:6:1 (McCabe, Graves et al., 2012). In addition, PRC2 association with other proteins, such as H1 or POLYCOMB-LIKE (PCL), can further boost activity (Jacob et al., 2009, Martin, Cao et al., 2006), while presence of other chromatin marks such as H3K4me3 or H3K36me2/3 can inhibit (Schmitges, Prusty et al., 2011). Comparing the two human homologs EZH1 and EZH2, EZH2 exceeds EZH1 as an HMTase, but EZH1 is able to repress transcription from chromatinised templates and compacts chromatin in the absence of SAM (Margueron, Li et al., 2008). Experiments *in vivo* support the idea, that EZH2 deposits H3K27me3 *de novo* in dividing cells, while EZH1 is required for H3K27me3 maintenance in resting cells (Margueron et al., 2009)

While biochemical studies of PcG proteins in plants are scarce, many evolutionary studies in the green lineage exist due to the availability of many sequenced plant genomes. However, PcG proteins in the green lineage are significantly diverged from animals and thus the transferability is not always clear.

1.6 PcG evolution in the green lineage

Whole genome duplication (WGD) have massively shaped plant genome evolution (Panchy, Lehti-Shiu et al., 2016, Van de Peer, Fawcett et al.). Five WGD events took place in the evolutionally history of Arabidopsis and At-α refers to the most recent WGD (47 mio years ago) (Hohmann, Wolf et al., 2015, Vision, Brown et al., 2000). Gene-duplications are common in the green lineage and WGD especially favour duplication of protein complexes since WGD ensure dosage balance between complex members (Panchy et al., 2016) (see Infobox III 'Protein Evolution').

Infobox III

Protein evolution – the birth and fate of homologous pairs

Gene duplications can either be the product of whole genome duplication or individual gene duplication. Duplicates can be either retained because they are adaptive, or they can be lost/pseudogenised because they are mal-adaptive. If retained, two paralogs are positively selected because of an existing function, or because of a novel function.

Paralog retention because of an existing function can be adaptive for an organism because of:
1. an increased gene dosage.
2. the preservation of the ancestral function of one paralog, while the other paralog is free to sub-functionalise (e.g. shift in temporal/spatial expression)
3. the maintenance of the stoichiometric balance in a protein complex (especially favored by WGD).
4. the prevention of paralog interference.

Paralog retention because of a novel function can be adaptive for an organism because of:
1. a gain of function.
2. the escapement from an adaptive conflict.

The plant PcG family has expanded widely. Conservation of PRC1 between animals and plants was doubtful until the identification of the RAWUL domain (Sanchez-Pulido, Devos et al., 2008), which is part of PcG ubiquitin E3-ligases. Due poor sequence identity, EMF1 conservation was also obscure, but functional conservation was shown with the Psc C-terminus, which has a characteristic, high charge and is intrinsically disordered (Beh et al., 2012). Within the green lineage, AtRING1, AtBMI1 and also LHP1 have obvious conservation on the primary sequence level with homologs in mosses, lycophytes and monilophytes (Berke & Snel, 2015).

The presence of homologs of E(z), Su(z)12, Esc and Nurf55 in basal unicellular algae, basal heterokonta and basal metazoa suggests that the last eukaryotic common ancestor had functional histone methylation (Shaver, Casas-Mollano et al., 2010). Studies in the moss *Physcomitrella patens (*Pp, Physcomitrella*)* and the red alga *Cyanidioschizon merolae* confirmed evolutionary conserved function of H3K27me3 in gene repression (Pereman 2016, Mikulski 2017). In Physcomitrella, PpCLF and PpFIE were shown to suppress heterochronic expression of sporophytic developmental program in the gametopohyte (Mosquna, Katz et al., 2009, Okano, Aono et al., 2009).

Together these data suggest, that both PRC1 and PRC2 components might be of monophyletic origin in plants and animals (Shaver et al., 2010). Within the green lineage, PRC2 components have individual evolutionary histories.

The homologous pairs SWN and MEA and VRN2 and FIS2 originate from At-α (Qiu, Liu et al., 2017, Spillane, Schmid et al., 2007)). MEA and FIS2 underwent a concerted divergence and subfunctionalised in gametophytic expression (Qiu et al., 2017). Gametophytic FIS-PRC2 formed by FIS2, MEA/SWN, FIE and MSI1 prevents seed development in the absence of fertilisation and guides normal seed development. Deleterious mutations in FIS2, MEA or FIE lead to embryo abortion caused by overproliferation and delayed/abolished cellularisation of the endosperm (Guitton, Page et al., 2004, Kohler, Hennig et al., 2003, Luo, Bilodeau et al., 1999).

Although, MSI1 has four additional homologs in Arabidopsis, namely MSI2-5, only MSI1 takes part in PcG mechanism (Derkacheva et al., 2013). However, MSI1 does not only associate with PRC2, but is also moonlighting with the chromatin assembly factor (CAF) complex, the ubiquitination CULLIN4 – DNA DAMAGE BINDING-PROTEIN1 (CUL4-DDB1) complex, the cell cycle regulator RETINOBLASTOMA RELATED (RBR) complex and a RETINOBLASTOMA (RB)-MSI4 complex (Dumbliauskas, Lechner et al., 2011, Gu, Jiang et al., 2011). The five MSI homologs of Arabidopsis group in three distinct clades and evolved before the divergence of monocots and dicots (Hennig, Bouveret et al., 2005).

Like Physcomitrella and Arabidopsis, few plant species encode more than one FIE homolog with the exception of cereals, indicating that the single copy gene status is ancestral (Tonosaki & Kinoshita, 2015). VRN2, however, was formed from a duplication of EMF2 at the base of the rosids clade (Chen, Diao et al., 2009)

CLF and SWN are present in all monocots and dicots, whereas only one ancestral homolog was identified in Selaginella and Physcomitrella (Zografou, dissertation, 2013). However, Zografou's study left a distinct resolution gap between diverged angiosperms and mosses. SWN evolved under purifying selection in the Brassicaceae family, while MEA is under diversifying selection (Spillane et al., 2007). Accordingly, duplication of the common ancestral E(z) homolog might date back at least 170 mio years (Panchy et al., 2016).

Few functional studies exist of CLF and SWN homologs in other plant species than Arabidopsis, with the exception of tomato. The tomato homolog of SWN, SlEZ1, is involved in stamen development and carpel number determination. The CLF homolog, SlEZ2, plays a special role in fruit development. Similar to Arabidopsis *fie* mutants, mutation of the tomato FIE homolog results in more severe phenotype than either of the single E(z) mutants (Boureau, How-Kit et al., 2016, Gallusci, Hodgman et al., 2016, How Kit, Boureau et al., 2010).

1.7 CLF and SWN – truly redundant?

Arabidopsis mutants of *CLF* and *SWN* have distinct phenotypes rather different from tomato E(z) mutants. Arabidopsis null mutants of *clf* show leaf curling, early flowering under 8 hours light and 16 hours dark conditions (short days, SD), increased chance of floral homeotic transformation and mildly delayed aging when grown under 16 hours light and 8 hours dark conditions (long days, LD) (Lopez-Vernaza, Yang et al., 2012, Xu, Guo et al., 2016). Null mutants of *swn* lack a relatively obvious phenotype. However, *swn* mutants of the allele *swn7* have shorter roots and mutants of the allele *swn3* show a mild delay of aging in LD (Chanvivattana, Bishopp et al., 2004, de Lucas, Pu et al., 2016, Xu et al., 2016). Notably, a PcG callus phenotype can be obtained in case of both *swn3* and *swn7* alleles (*swn7 clf28 or swn3 clf50)* (Chanvivattana et al., 2004, Lafos et al., 2011, Liang et al., 2015). Heterochronic upregulation of the developmental genes *AGAMOUS* (*AG*), *FLC*, *FLOWERING LOCUS T* (FT), *SEPALLATA 3* (*SEP3*), *FUS3* and *MIR156* due to decreased H3K27me3 levels have been shown to cause the *clf* null mutant phenotype (Goodrich, Puangsomlee et al., 1997, Lopez-Vernaza et al., 2012, Xu, Hu et al., 2015, Xu et al., 2016). Conversely, *MIR156* is the only described target gene of SWN and is expressed prematurely in *swn3* mutants (Xu et al., 2015, Xu et al., 2016). Notably, tomato homologs of CLF and SWN do not regulate any of the above target genes (Boureau et al., 2016, How Kit et al., 2010)

SWN redundancy with CLF is implied by their overlapping expression domains (Chanvivattana et al., 2004, Goodrich et al., 1997). Similarly, SWN redundancy with MEA is suggested by an overlapping expression domain in the endosperm and gametophyte, the genetic relationship of the mutants and the ability of SWN to interact with FIS2 in yeast-two-hybrid (Derkacheva et al., 2013). Notably, SWN and MEA fail to complement the *clf* mutation, what is indicative for distinct molecular functions between the Arabidopsis E(z) homologs(Chanvivattana et al., 2004). In the sporophyte, SWN is more abundant than CLF, as indicated by transcriptome and proteome data, and dominates in EMF-and VRN-complex purification (Derkacheva et al., 2013). Co-immunoprecipitation followed by mass spectrometry (CoIP-MS) suggests that CLF and SWN are mutually exclusive in PRC2 (Liang et al., 2015)

Genome-wide H3K27me3 distribution was addressed by Wang et al. in *clf* mutants (Wang, Liu et al., 2016) and by Zografou in *clf* and *swn* single mutants (Zografou, 2013). Both studies found substantially more regions with decreased rather than increased enrichment of H3K27me3 in *clf* mutants. Many of these CLF-dependent genes have gene ontology terms such as flowering and floral development. Both findings are consistent with western blot analysis, where overall H3K27me3 mildly decreases in *clf* mutants, and the *clf* morphological and misregulation phenotype. However, Wang and colleagues also found hundreds of genes that show an increase in H3K27me3 enrichment and decreased expression in *clf* mutants, suggesting an involvement of SWN at these genes. Full redundancy of CLF and SWN is found at 164 genes that, in concert with AtBMI1A/B/C and

AtRING1A/B, are specifically involved in embryo development (Wang et al., 2016). Zografou found that CLF-dependent genes could gain H3K27me3 in *swn* mutants compared to WT suggesting hypermethylation of CLF in the absence of SWN. Together these data indicate that CLF and SWN have dynamic relationship at a common set of target genes.

2 Aim of this study

Swn single mutants do not have an obvious phenotype. The retention of SWN in all Angiosperms, the higher protein abundance of SWN protein and messenger RNA (mRNA) in Arabidopsis, the hypermethylation by CLF in the *swn* mutant plants and the residual H3K27me3 in *clf* mutant plants, all together suggest that CLF- and SWN-PRC2 variants play distinct roles in Arabidopsis and other species of the green lineage. This study tries to answer the following questions:

(1) What are the biochemical differences of SWN-PRC2 compared to CLF-PRC2 *in vitro*?

(2) Are CLF and SWN fully redundant *in vivo*?

(3) What are the protein domains that underlie CLF and SWN functional divergence?

(4) How general is the functional divergence of CLF and SWN in the green lineage?

3 Materials and Methods

3.1 Sequence acquisition

3.1.1 Cross-species complementation

To identify clonable CDS of CLF and SWN across the plant phylogeny, the phytozome platform v 9.1 was used and identified *B. distachyo*n, *S. lycopersicum* and *P. patens*. The *CLF* and *SWN* orthologues of *A. arabicum* were identified using the method described by Lyons (2008). Orthologues of *A. alpina* were identified by blast in the *A. alpina* genome assembly (Willing, Rawat et al., 2015).

3.1.2 Phylogenetic tree

To construct a phylogenetic tree, orthologous protein sequences in other plant species were searched for most species on Phytozome v10.3 (Weblink [2]) by using Arabidopsis protein sequences of CLF and SWN as a query. For a handful of orthologues, no complete protein sequence was available. These remaining orthologues sequences were searched by using the *PpCLF* CDS as a query and to extract EST tracks. EST tracks were extracted of *Picea albea*, *Picea glauca*, *Pinus taeda*, *Pseudotsuga menziesii* and *Podocarpus macrophyllus* from Gbrowse tool of the Congenie website (Weblink [3]); sequences of *Amborella trichopoda* were extracted from the Amborella genome database (Weblink [4]); and sequences of *Ginkgo biloba* were extracted from the medicinal plants genomics resource (Weblink [5]). The extracted sequences were translated using the Expasy translate tool (Weblink [6]). All protein sequences were aligned using the ClustalOmega algorithm (Weblink [7]) and trimmed to the conserved region of Arabidopsis CLF corresponding to amino acids 742-868 (SET domain).

3.1.3 Arabidopsis sequences

Sequences of Arabidopsis were acquired from the Arabidopsis Information Resource (TAIR; Weblink [8], on www.arabidopsis.org, 2014-2018, version TAIR9 and TAIR10).

3.2 Plasmid constructions

3.2.1 PCR

PCR reactions were performed using Phusion Polymerase (NEB) according to manufacturer's instructions. Typical 50µL reaction mix was is shown in Table 2. DNA concentrations was adjusted if cDNA or Plasmid was used. DMSO was added for difficult to amplify templates. Typical reaction conditions are shown in Table 3.

Table 2 Typical PCR reaction mixture using Phusion polymerase (NEB)

Component	50 µL reaction	Final concentration
H_2O	to 50 µL	
5x phusion HF buffer	10 µL	1 x
primer A (10 mM)	2.5 µL	50 µM
primer B (10 mM)	2.5 µL	50 µM
dNTP (10 mM)	1 µL	200 µM
DMSO (optional)	(1.5 µL)	(3%)
DNA		50-100ng (total inflorescnce cDNA), 1-10 ng (total plasmid)
Phusion polymerase	1 µL	1 unit/50 µL

Table 3 PCR conditions used for Phusion Polymerase (NEB)

Cycles	Step	Temperature	Duration
	1	98°C	5 min
x34	2	98°C	30 sec
	3	50-70°C (depending on primers)	30 sec
	4	72°C	30 sec /1kb
	5	72°C	5 min (product<5kb) 10 min (product>5kb)
	6	4°C	hold

3.2.2 Tobacco expression

Sequences were amplified of Arabidopsis *CLF*, *SWN*, *FIE*, *MSI1* and *EMF2* CDS from Arabidopsis cDNA which was made from young inflorescence tissue using Phusion polymerase (NEB) and gene specific primers (Table 4). Silent mutations were introduced in *CLF* and *MSI1* to eliminate endogenous BsaI restriction sites by using primers AF77/AF76 for CLF and AF73/AF72 for MSI1 (Table 4). The PCR products were incubated in a golden gate reaction (Engler & Marillonnet, 2014). During the reaction the 100 ng of the vector pICH31070 (Icon genetics) was used in combination with an annealing product of the primers AF57 and AF58 (Table 5) to introduce the N-terminal tags 6xHis and HRV3C and. These constructs lacked a start codon, since N-terminal apoplastic localization sequence was provided by the vector pICH20111 (Icon genetics). If cytoplasmic expression was done, pICH20155 was used during co-transfection in combination with *CLF*, *SWN*, *FIE*, *MSI1* and *EMF2* cloned into pICH31070 including their respective start codon. Using a PCR and Gibson assembly (NEB) strategy

including the primers of Table 6, the start codons were individually introduced into the recombinant PcG plCH31070 plasmids. Here and in subsequent experiments, Gibson assembly (NEB) was done according to manufacter's instructions and by using DpnI digested (NEB) purified PCR products (gel extraction kit, Machery Nagel). Usually, 100ng of the largest fragment was combined with insert(s) in a 1:3 molar ratio (largest fragment:insert) using the NEBioCalculator online tool (Weblink [9]).

Table 4 Primers used to clone *CLF*, *SWN*, *EMF2*, *MSI1* and *FIE* CDS into *pICH31070*.

primer	ID	sequence
1	AF80-SWNrev	CGGTCTCCCAAGATGAGATTGGTGCTTTCTGGCTC
2	AF79-SWNfwd	GGGTCTCGAGGTGTGACGGACGATAGCAACTCCTC
3	AF78-CLFrev	CGGTCTCCCAAGAGCAAGCTTCTTGGGTCTACCAAC
4	AF77-CLFmfd	GGGTCTCGTGAGAAAAGCCTTTTTGATAAAGGTG
5	AF76-CLFmrv	CGGTCTCCCTCAAGTGGCCTCCACAACTTATTTG
6	AF75-CLFfwd	GGGTCTCGAGGTGCGTCAGAAGCTTCGCCTTCTTC
7	AF74-MSI1rev	CGGTCTCCCAAGAGAAGCTTTTGATGGTTCTTCCCC
8	AF73-MSImfwd	GGGTCTCGTCAAGCAGGGTCATTTGCTTAGTGGCTCTG
9	AF72-MSImrev	CGGTCTCCTTGAACTTACTCCAGCTGAGCCCATATCCC
10	AF71-MSI1fwd	GGGTCTCGAGGTGGGAAAGACGAAGAGGAAATGCG
11	AF70-EMF2rev	CGGTCTCCCAAGAATTTGGAGCTGTTCGAGAAAGG
12	AF69-EM2fwd	GGGTCTCGAGGTCCAGGCATTCCTCTTGTTAGTCG
13	AF68-FIErev	CGGTCTCCCAAGCTTGGTAATCACGTCCCAGC
14	AF67-FIEfwd	GGGTCTCGAGGTTCGAAGATAACCTTAGGGAACG

Table 5 Primers used to introduce the N-terminal tag 6xHis-HRV3C-.

Primer	ID	Sequence
1	AF58-HRV3C_6HIS_rev	CGGTCTCCAAGCCTAGTGATGGTGGTG GTGATGAGCCGCGGAGGGTCCCTGAA AGAGGACTTCAAGCGAGACCC
2	AF57-HRV3C_6HIS_fwd	GGGTCTCGCTTGAAGTCCTCTTTCAGG GACCCTCCGCGGCTCATCACCACCACC ATCACTAGGCTTGGAGACCG

Table 6 Primers used to introduce start codon ATG into PcG *pICH31070* vectors for cytoplasmic expression

Primer	ID	Sequence
1	AF163-B_fwd_CLF_ATG	TGA GAA AAG CCT TTT TGA TAA AGG TG
2	AF164-B_rev_CLF_ATG	CTT CTG ACG CCA TAC CTG CAA CA
3	AF167-B_fwd_MSI1_ATG	TGT TGC AGG TAT GGG GAA AGA C

4	AF168-B_rev_MSI1_ATG	GTC TTT CCC CAT ACC TGC AAC A
5	AF170-B_fwd_SWN_ATG	TGT TGC AGG TAT GGT GAC GG
6	AF171-B_rev_SWN_ATG	CCG TCA CCA TAC CTG CAA CA
7	AF173-B_fwd_FIE_ATG	TGT TGC AGG TAT GTC GAA GAT AAC C
8	AF174-B_rev_FIE_ATG	GGT TAT CTT CGA CAT ACC TGC AAC A
9	AF176-B_fwd_EMF2_ATG	TGT TGC AGG TAT GCC AGG CAT T
10	AF177-B_rev_EMF2_ATG	AAT GCC TGG CAT ACC TGC AAC A

3.2.3 Insect expression

Donor plasmids designed for the Bac-to-Bac system (Invitrogen), named *pNT74i* (His tag) and *pNT79i* (StepII-tag), which contained *CLF*, *EMF2*, *MSI1* and *FIE* CDS expression were acquired from Schmitges et al. (2011). *SWN* CDS was amplified by Phusion (NEB) PCR from Arabidopsis cDNA. The *SWN* PCR product was excised from Agarose gel and purified using a gel purification kit (Machery Nagel) according to manufacturer's instructions. The PCR product was digested according to manufacturer's instructions using DpnI (NEB) to remove template DNA and NotI and KpnI to create 5' and 3' overhangs. Simultaneously, the vector *pNT74i* was linearized using NotI and KpnI restriction enzymes (NEB). Digested insert and digested vector were ligated using T4 DNA ligase (NEB). Sequenced donor plasmids were transformed into Dh10bac bacterial stain harboring recombinant baculovirus DNA. White colonies were selected from plates containing X-Gal. Theoretically, these white clones are identified by successful TE transposition of the gene of interest including its regulatory sequences into the acceptor bacmid. The extracted, recombinant bacmid DNA was tested by PCR using M13fwd and M13rev standard primers (Sigma) to confirm by product size for the successful transposition of the full-length expression cassette into the baculovirus bacmid.

3.2.4 GUS reporter lines

CLF and *SWN* promoters were amplified by Phusion PCR from genomic constructs (Zografou, 2013) using gateway primers (**Error! Reference source not found.**). Entry clones were recombined by LR reaction using Gateway protocol (Invitrogen) into *pFASTG04* (Shimada, Shimada et al., 2010).

3.2.5 GFP and mCherry reporter lines

To make a binary vector version of *pFASTG01*, which in addition to the seed coat green-fluorescence protein (GFP) marker also harbors a BASTA (Bayer) resistance marker (Bar), *pFASTG01* and *pFASTG02* (Shimada et al., 2010) were recombined. The two vectors *pFASTG01* and *pFASTG02* were digested using the restriction enzymes AscI and PmeI (NEB) according to manufacturer's instructions. The excised and purified larger fragment of *pFASTG01* (vector backbone) and the excised and purified smaller fragment of *pFASTG02* (Bar resistance cassette) were combined in a 1:3 molar ratio and

ligated using T4 DNA ligase (NEB) according to manufacturer's instructions to create the recombinant *pFASTG01-Bar* plasmid.

The backbone of *pENTR201* containing a genomic *SWN* sequence (*gSWN*) (Turck, unpublished) was amplified using primers AF65 and AF81 (**Error! Reference source not found.**) and assembled with a *mCherry* PCR product amplified with primers AF51 and AF52 (**Error! Reference source not found.**) in a standard Gibson assembly reaction (NEB) according to manufacturer's instructions and to above mentioned specifications (3.2.2). The entry vectors *GFPgCLF* (Turck, unpublished) and *mCherrygSWN* were recombined in an LR reaction (Invitrogen) according to the manufacturer's instruction in a 10µL total reaction volume at half the indicated concentration of enzyme into *pFASTG01*-Bar binary vector (see paragraph above). These *pFASTG01-Bar* binary entry vectors containing *mCherrygSWN* and *GFPgCLF* were transformed into Agrobacterium strain GV3101 (*pSOUP*).

3.2.6 Promoter swaps

The entry clones *gCLF* and *gSWN* (Turck, unpublished) were used to amplify *CLF* and *SWN* minigene regulatory sequences and vector backbone using primers pCLFlong-fw/ pCLFlong-rw and B028/B029, respectively (Table 7). *CLF* and *SWN* CDS were amplified from Arabidopsis young inflorescence cDNA using primers B028/B029 and Z321/Z320, repectively, and combined in a Gibson assembly (NEB) with the backbone PCR products to give *CLFmini:CLF* and *SWNmini:SWN* entry clones. *CLFmini:SWN* was produced in the same manner using primers Z321/Z320 to amplify *SWN* CDS from *SWNmini:SWN*, which was then combined with the PCR product amplified from *CLFmini:CLF* using primers Z329/Z328 (Table 7). *SWNmini:CLF* was made using primers Z345-Z347 (Table 7). All entry clones were sequenced and recombined in a Gateway LR reaction (Invitrogen, details in last paragraph of 0) using the binary vector pFAST-G01-Bar (see 0), which was transformed into Agrobacterium strain GV3101 (pSOUP).

Table 7 Primers used to construct minigene constructs of CLF and SWN

Primer	ID	Sequence
1	pCLFlong-fw	CGTAGGGCGCGCCGTATATATATAATCTCCACG
2	pCLFlong-rw	CCTCGCCTCGAGGTGTCAAGAAACCAGATCGGA
3	B028-att1CLF	ATGGCGTCAGAAGCTTCGCCTT
4	B029-att2CLF	AGCAAGCTTCTTGGGTCTACCA
5	pSWN1	CGTAGGGCGCGCCAACCATCAGATATACAAATA
6	pSWN2	CCTCGCCTCGAGGTGATGACTCCTCGAGCTTTC
7	Z321-IPIPEfwd_SWNcDN	ATGGTGACGGACGATAGCAACTC
8	Z320-IPIPErevcSWN	CGAAATGATACATTTTTACTTATTGACC

9	Z321-IPIPEfwd_SWNcDN	ATGGTGACGGACGATAGCAACTC
10	z329-VpCLF:cSWNfwd	GCACCAATCTCATTGACAACAAAAGAAACAAC
11	z328-VpCLF:cSWNrev	ATCGTCCGTCACCATTGTCAAGAAACCAGATC
12	Z347-pSWNcCLFIfwd	ATGGCGTCAGAAGCTTCGCCTTC
13	Z348-pSWNcCLFIrev	CTAAGCAAGCTTCTTGGGTCTACC
14	Z346-pSWNcCLFVfwd	CAAGAAGCTTGCTTAGTGATTACTGGCTAAGA
15	Z345-pSWNcCLFVrev	AGCTTCTGACGCCATTGATGACTCCTCGAGCT

3.2.7 Domain swaps

The domain swap *CLFmini:BF* was produced by the same procedure as described in 0, using primers AF14/AF13 for backbone PCR on *CLFmini:CLF* and AF12/AF11 for insert PCR on *CLFmini:SWN*; and *CLFmini:AF* using primers AF8/AF7 for backbone PCR on *CLFmini:CLF* and AF9/AF10 for insert PCR on *CLFmini:SWN* (Table 8). *CLFmini:CLF<SWN_C5* (Zografou, unpublished), *CLFmini:BF* and *CLFmini:AF* and were used in a Gateway LR reaction (Invitrogen) into pFASTG01-Bar, which was transformed into Agrobacterium strain GV3101 (pSOUP).

Table 8 Primers used to make N-terminal domain swap

Primer	ID	Sequence
1	AF12-IrevBfus	CTAAGCAAGCTTCTTGGGTCTACCAAC
2	AF11-IfwdBfus	ACGCTTCTCAAGTCTGGAAGATTTCC
3	AF14-VrevBfus	CTTCCAGACTTGAGAAGCGTACTAGCAGCAGCATGGGGAACATC
4	AF13_VfwdBfus	CCCAAGAAGCTTGCTTAGCAACAAAAGAAACAACCATTTTTTTGTC
5	AF10-IrevAfus	TTTATAGCAGTTTGCACCACAGGTTAAA
6	AF9-IfwdAfus	ATGGCGTCAGAAGCTTCGCCTT
7	AF8-VfwdAfus	GGTGCAAACTGCTATAAAGGTGTCAGTCTGCAAGTTGAGAAGAC
8	AF7-VrevAfus	CGAAGCTTCTGACGCCATTGTCAAGAAACCAGATCGGAACCG

3.2.8 Cross species complementation

Different plant species (Table 9) were sampled for RNA extraction. Samples of Brassicaceae were gained from fresh inflorescence tissue, while tomato samples were taken from a small number of pooled, unopened flowers. Brachypodium samples were taken from above ground structures of three young seedlings 20 days after germination. Physcomitrella samples were taken from moss grown on plate and under continuous light (AG Bernd Reiss, MPIPZ). All samples were snap frozen in liquid nitrogen and ground to a fine powder using mortar and pestle in liquid nitrogen. 5g of the ground tissue powder was used to extract RNA using TRIzol reagent (Invitrogen) according to manufacturer's manual. 5 µg of the extracted RNA was

DNase1-treated using the DNAase-free kit (Ambion) according to manufacturer's manual. cDNA synthesis was performed using dT18 primer and the Superscript II reverse transcriptase enzyme (Roche). cDNA was diluted to 150 µl with water, and 2 µl of diluted cDNA was used in Phusion PCR reaction (see 3.2.1).

Table 9 List of accessions used to extract RNA used in Cross-species complementation experiment

Species and accession
Arabidopsis thaliana Col-0
Arabis alpina Pajares
Aethionema arabicum (Haudry, Platts et al., 2013)
Solanum lycopersicum 13151
Brachypodium distachyon Bd21
Physcomitrella patens ssp. *patens* isolate Gransden 1962

CDS of CLF and SWN orthologs were amplified using primers in Table 10 and BP reaction into pDONR201.BP reaction was unsuccessful for *PpCLF*. Hence, PCR product was ligated into pCR4-TOPO (Invitrogen) and the resulting clones were sequenced using T3 primer (Invitrogen). After verification, the fragment was amplified with the primers AF251-attbtopoIfwd (GTACAAAAAAGCAGGCT) and AF252-attbtopoIrev (TTAAGCAACTTTCTGTGCT) and ligated using T4 DNA ligase (NEB) in a Gibson assembly (NEB) into the pENR201 backbone, which was linearized using the primers AF249---attbtopoVfwd (AGCACAGAAAGTTGCTTAAACCCAGCTTTC TTGTACAAA) and AF250-attbtopoVrev (AGCCTGCTTTTTTGTACAAACTTGGCATTATAAAAAAGC ATTG).

Table 10 Primers used to amplify cross-species coding sequence

Primer	ID	Sequence
1	AF28-PhyscoCLFrev	AAATGGTTGTTTCTTTTGTTGTTAAGCAACTTTCTGTGCTCGTCC
2	AF27-PhyscoCLFfwd	CCGATCTGGTTTCTTGACAATGGCGTCCTCCAGCTACGC
3	AF26-SlEZ2CLFrev	AATGGTTGTTTCTTTTGTTGTTATGTATGCTTCCTAGCACGACCACTTG
4	AF25-SlEZ2CLFfwd	CCGATCTGGTTTCTTGACAATGTCGCCGGCGTCGGATA
5	AF24-SlEZ1SWNrev	AATGGTTGTTTCTTTTGTTGTTATTGGTGTTTCTTTGGTCGACCTAAAG
6	AF23-SlEZ1SWNfwd	CCGATCTGGTTTCTTGACAATGATCTCCTCCACCTCCATCTCTG
7	AF22-BradiSWNrev	ATGGTTGTTTCTTTTGTTGCTATCGAGCAACTTTGTGTGCTCGG
8	AF21-BradiSWNfwd	CCGATCTGGTTTCTTGACAATGGCGTCGTCGTCGTCCAA
9	AF20-BradiCLFrev	AATGGTTGTTTCTTTTGTTGTCAGTGGGCAACCTTCTTTGCTC
10	AF19-BradiCLFfwd	CCGATCTGGTTTCTTGACAATGACCATTCAAGAGTGTGGCATG
11	AF18-ArabisCLFrev	ATGGTTGTTTCTTTTGTTGCTAAGCAAGCTTCTTGGGTCTACCAACTG
12	AF17-ArabisCLFfwd	CCGATCTGGTTTCTTGACAATGGCGTCGGGAGCTTCGC
13	AF16-arabisSWNrev	AATGGTTGTTTCTTTTGTTGTCAATGAGATTGGTGCTTTCTGGC

14	AF15-ArabisSWNfwd	CCGATCTGGTTTCTTGACAATGGTGATGACTGATGATGATAGCGA
15	AF34-AethSWNrev	AATGGTTGTTTCTTTTGTTGCTATAAACTTTATAGTAGTACCTTGGCG
16	AF33-AethSWNfwd	CCGATCTGGTTTCTTGACAATGTTTTCTGATAGCGTTTATTGG
17	AF32-AethCLFrev	AATGGTTGTTTCTTTTGTTGCTAGACGATGAGTTTTGGTGAAAATAAA
18	AF31-AethCLFfwd	CCGATCTGGTTTCTTGACAATGGCGTCGGGAGCTTC

Entry clones contatining CDS of CLF and SWN orthologs were used in an LR reaction into CLFmini:AttB1/2, which contained a gateway cassette in place of the CLF CDS.

3.3 Material preparation for *in vitro* experiments

3.3.1 Tobacco protein expression

Agrobacterium clones carrying the recombinant vector pICH31070_CLF/SWN/MSI1/FIE/EMF2 (described in 0), pICH14011(containing an integrase) and either pICH20111 (containing the apoplastic localization signal) or pICH20155 (containing no signal peptide for cytoplasmic expression) (Icon genetics) were grown individually. Cultures were started from a 5mL preculture in Luria Broth medium at 230 RPM, 27°C over night. Main culture of 200mL total volume was grown to an optical density at 600 nm (OD600) of 0.6-0.8. The cultures were diluted to OD600 0.16 in infiltration buffer. Buffer was prepared according to Zhou, Tergemina et al. (2017b). Young leaves of three weeks old tobacco plants (*Nicotiana benthamiana*) were infiltrated with diluted and pooled Agrobacteria. Plants were incubated at 24°C in a LD climatic chamber until harvest.

3.3.2 Protein extraction from tobacco

Tobacco leaves expressing recombinant protein were removed from plant, snap frozen in liquid nitrogen and stored at -80°C. Frozen leaves were gound in a mortar with a dash of SiO_2 and cold 10 mL extraction buffer (50 mM NaPO4 pH 6.8, 10 mM EDTA, 10 mM DTT, 1% Triton-X, app. 5% PVPP, 10% glycerol, 1 tablet Roche protease inhibitor per 50 mL). Leaves were ground for 5-10 min. Crude extract was centrifuged at (11.000g, 4°C, 10 min) to remove cell debris, PVPP and SiO_2.

3.3.3 Insect cells culture conditions

Sf21 cell line was obtained from AG Classen (LMU, Munich, 2014) at a low passage number. Cells were cultured at 27°C and 120 RPM in SFM 900 II (Gibco) supplemented with 10% FBS (Gibco) and one part 100x Antibiotic/Antimycotic mix (Sigma) which gave a total of 100 units penicillin, 0.1 mg streptomycin and 0.25 µg amphotericin B. 30 mL cells culture was incubated in 125mL plastic Erlenmeyer flasks (Gibco). The cells were passaged to a density of 0.5 10^6/mL in a 48h cycle by transferring a specific volume of cell culture into fresh medium and into a clean, sterile vessel to maintain them in exponential growth phase. Typically, the cell density would have quadrupled after 48h. A new culture with a low passage number was started once the culture reached app. 30 passages.

3.3.4 Insect cell transformation

Sf21 cells were transfected in a 6-well plate in a volume of 2 mL/well and a total number of 1 10^6 cells/well. After 1h, cells were attached to bottom of the well and could be transformed using 1 µg of recombinant bacmid DNA and 8 µL of X-treme Gene transfection reagent (Roche) according to manufacturer's instructions. P0 virus was harvested after 48h by collecting supernatant (SN), centrifugation at 700xg, 4°C to remove cells and debris. Viral SN was filter-sterilized using a 0.22 µm syringe filter and stored at 4°C.

3.3.5 Baculovirus amplification in insect cells

30 µL of the P0 viral SN were used to infect 30 mL of Sf21 cells in suspension at a densitiy of 0.5 10^6/mL. The cells were passaged to a density of 0.5 10^6/mL daily for approximately 7-10 days until notable increase of average cell diameter from app. 17 µm to app. 22 µm was measurable with Scepter Cell Counter (Millipore). The 30 mL of the P1 viral SN were harvested by centrifugation at 700xg for 2 min at 4°C and stored at 4°C for up to 5 months. 5mL of P1 viral SN was used to infect 300 mL of Sf21 cells in suspension at 1 x 10^6 cells/mL and cultured. The resulting P2 viral SN was harvested by centrifugation at 700xg for 2 min at room temperature, filter-sterilized (0.22µm) and stored for up to 5 month at 4°C.

3.3.6 Plaque assay in insect cells

Plaque assay was performed to determine viral titer in P2 viral SN. Plaque assay was performed in a 6-well plate format with cells in stationary incubation at 27°C. in each well, stationary cells were mixed with viral dilutions ranging from 10^{-8} to 10^{-3} according to Aigner (2011), pages 424-426. 6-well plates were incubated in a sealed container to reduce evaporation. Low-melt Agarose (Biozym) was used for overlaying cells with serum-free, antibiotic-free Sf900 II SFM medium according to Aigner (2011), pages 424-426. After 4 days incubation, an overlay of Neutral red was applied according to Aigner (2011), pages 424-426, to increase visibility of plaques. Plaques were counted after approximately 10-14 days after infection and viral titer in P2 was calculated according to Aigner (2011), pages 424-426.

3.3.7 Protein expression in insect cells

Protein was expressed in 2-times 2L Sf21 cells in suspension in 4 L plastic Erlenmeyer falsks at a density of 1 10^6/mL. Cell culture was infected using a multiplicity of infection (MOI) MOI3 (unless otherwise indicated) of P2 virus. For protein complex expression, individual viruses were pooled, each at MOI3. After 24h of expression, Sf21 were centrifuged at 700xg for 2 min at 4°C and SN was removed by inverting to remove medium. The cell pellet was snap frozen in liquid nitrogen and stored at -80°C.

3.3.8 Protein extraction from insect cells

The cell pellet was thawed on ice together with in 50mL extraction buffer (20mM Tris-HCl pH8, 0.1% Triton-X, 150 mM NaCl, 10% glycerol, 2 mM DTT, 10 mM PMSF and one tablet Roche protease inhibitor per 50 mL buffer) per 1L cell culture. Cells were lysed by pipetting up and down repeatedly (app. 20 times) and medium-strength sonication of lysate on ice in a 50 mL conical tube using a ultrasonic finger-disrupter. The lysate was centrifuged at 8.000xg for 10 minat 4°C to remove cell debris and the crude extract was filter-sterilized (0.22µm).

3.3.9 Protein purification from insect cells

For on-bench StrepII purification, 200 µL of 50% Strep-Tactin sepharose bead slurry (IBA) was equilibrated in 10-times bead volume of extraction buffer (see 3.3.8) and were generally centrifuged at 1000xg for 2min at 4°C. Equilibrated sepharose beads were mixed with crude extract in a ratio of 0.050:1.5 (beads:crude extract) and incubated for 1.5h at 4°C while gently shaking. The sepharose beads were washed twice with 10-times bead volume of wash buffer (20mM Tris-HCl pH8, 150 mM NaCl, 10% glycerol). A third wash was done in a small chromatography column using 10-times bead volume of wash buffer. The Chromatograpy column was centrifuged at 1000xg for 2min at 4°C. Finally, protein was eluted using 2-times bead volume, 200 µL, of wash buffer supplemented with 2.5 nM desthiobiotin (IBA) and 5µl were used in western blot analysis (see 3.5.4).

FPLC chromatography was done using ÄKTA explorer (GE) in combination with a Strep-Tactin Superflow HP cartridge (IBA). Buffers were as described in the on-bench StrepII purification. crude extract was loaded to the column at a flow rate of 0.5mL/min. The column was washed at a flow rate of 1mL/min until UV absorption of the flow through returned almost to 0 for approximately 2 column volumes (CV). Elution was continued at a flow-rate of 1 mL/min. Typically the protein eluted after 1.5 CV. The elution peak was fractionized into 0.5 mL fractions, which were later on pooled (unless otherwise indicated).

The eluate was rebuffered in 10 MWCO concentrator columns (Thermofisher) into HMTase buffer (50 mM Tris-HCl pH8 adjusted at room temperature, 50 mM NaCl, 0.1% Tween20 , 1mM DTT and 1mM EDTA). Sample was aliquoted (≥ 10µL), snap frozen in liquid nitrogen and stored at -80°C.

3.3.10 Competitive pulldown

For competitive pulldown, crude extracts were prepared from Sf21 cells expression single 6xHis-TEV-CLF, single 6xHis-TEV-SWN or triple 6xHis-TEV-MSI1, 6xHis-TEV-FIE and StrepII-TEV-EMF2 (MSI1/FIE/EMF2). 1 mL of CLF and SWN crude extracts recovered from 30 mL insect culture volume (expression conditions as above) were diluted 10 fold, MSI1/FIE/EMF2 was diluted 100 fold. Diluted crude extracts of CLF and SWN were mixed in 0:10, 1:9, 5:5, 9:1 and 10:0 ratios. CLF and SWN samples

were mixed in a 1:10 ratio with the 100 fold diluted MSI1/FIE/EMF2 sample and purified using the on-bench StrepII purification (3.3.9).

3.4 Material preparation for *in vivo* experiments

3.4.1 Plant culture conditions

Plants were either grown on soil in the greenhouse (MPIPZ, Cologne) or under sterile conditions on plate. Both plate grown plants and greenhouse grown plants were stratified according to Adrian (2009). Seeds on plate were grown on 0.5x murashige skoog agar plates containing 1.5% sucrose and 0.8% agar and cultivated under 80 mmol m^{-2} sec^{-1} white light in LD conditions and at 22°C (Intellus® environmental controller, Percival). For callus induction, growth medium was supplemented with 4.5 mM 2,4-D and 0.45 mM kinetin plant hormones (Farrona, Thorpe et al., 2011).

3.4.2 Plant crossing

To produce plants with *clf-28-/- swn-7+/-* genotype, soil-grown flowering plants with *clf-28+/-swn-7-/-* were pollenated with *clf-28-/-* genotype as a pollen donor. Plants were genotyped using specific primers for *clf-28*, *swn-7* and WT alleles (Zografou, 2013) in F1generation to select *clf-28-/- swn-7+/-* genotype and selfed. In F2 generation genotyping was used to select *clf-28-/- swn-7+/-* plants for phenotypic analysis.

3.4.3 Plant transformation

Binary vectors were introduced into Agrobacterium tumefaciens GV3101 (pSOUP) (reference: Koncz and Schell, 1986). Flowering Arabidopsis plants were transformed by the floral dip method (reference: Clough & Bent 1998). Contrary to the protocol, flowering plants were arrested in development by storing them for up to 6 weeks in an illuminated chamber at 4°C under LD conditions and plants were transferred to ambient temperature two days prior to floral dip.

3.4.4 Plant selection

Transgenic lines produced with pFast-G01-Bar were selected on soil using BASTA® (Bayer) in the T1 generation and using GFP fluorescence of seeds under a fluorescence microscope (Leica MZ16 FA, filters Leica GFP3/ GFP1) in T2 and T3. Single locus insertion lines were selected in T2 generation, but no tandem repeat analysis was performed.

3.5 Sample analysis

3.5.1 Plant phenotyping

Rosette diameter at first flower opening was recorded from photographs, which were taken from a fixed angle and distance. Using the Fiji software (Schindelin, Arganda-Carreras et al., 2012) the outer rosette circumference was marked manually. Feret diameter in pixels was extracted automatically from the photographs and exported to MS excel. Pixel values were converted into cm using a

standard ruler placed on height of the plants in the photographs. Statistical test was made for eight replicates of one transgenic line using Anova Holm correction in R (Team, 2013). Between two and three significance groups were defined based on the data and averages of each transgenic line were tested for P-value<0.05 in a matrix.

3.5.2 Histochemical analysis of GUS expression

GUS staining protocol was performed according to Jessika Adrian dissertation (Adrian, 2009) and stained for app. 4h.

3.5.3 qRTPCR

10 day old seedlings were grown on plate and more than 10 seedlings from two replicate plates were harvested for cDNA production (described in 0). Expression of *CLF*, *SWN* and *PPT* was measured in a qRT PCR using 5 µL of diluted cDNA (described in 0). To check absence of genomic DNA from cDNA, primers AF360 (TCT TTG TTG GCT CTC ACA AGT) and AF361 (CCG AAC ATC AAC ATC TGG GTC) were used in a standard PCR reaction. These primers would only give a product for PP2A genomic sequence if genomic DNA contamination was present. Evagreen qRT PCR mix and program were adapted from Adrian (2009). *CLFmini:CLF, CLFmini:SWN* and a plasmid containing *PP2A* CDS (Krause, unpublished) were used to make a standard plasmid dilution ranging from 10^9-10^3 copies/µL and pRT PCR was set up with *CLF*, *SWN* and *PP2A CDS* specific primers (Table 11). DeltaCt values were converted into copy number by using an exponential regression model to the standard and the mRNA ratios *CLF/PP2A* and *SWN/PP2A* were calculated for three technical replicates of each genotype.

Table 11 QRTPCR primers used to for CLF, SWN and PP2A mRNA quantification

Primer	ID	Sequence
1	AF309 -CLFqPCR_F	GGG ATA TGG CAC CAT TGA AG
2	AF310 -CLFqPCR_R	TCA CTT GGG AAG GTC TTT GG
3	AF306 -SWNnintron15_F	GAT CAG TAC GTC CTC GAT GCT
4	AF07 -SWNnintron15_R	GCG TAG CAA TTG GGT TTA GC
5	AF358 -PP2AfwdqPCR	ACA CAA TTC GTT GCT GTC TTC T
6	AF359 -PP2ArevqPCR	TGC TTG GTG GAG CTA AGT GA

3.5.4 Western blot analysis

Proteins were separated on a 12% SDS polyacrylamide gel electrophoresis (SDS PAGE) and stacking gel (BioRad). Separation occurred at 30 mA until running front exited the gel. SDS PAGE was transferred to Immobilon membrane (GE lifescience) by wet-blotting at 4°C, 16h and 30 volts. Incubation, blocking in 3% milk and Ab probing was done according to manufacturer protocol

(Millipore, #70796) using the primary Ab at 1000-fold dilution and the corresponding secondary Ab at 5000-fold dilution (Table 12). Detection was done using 1 mL total volume of a 1:1 mixture of SuperSignal West Femto and SuperSignal West Dura (Thermo Scientific) detection reagents. Blots were typically exposed for 10 seconds during image acquisition when testing recombinant protein expression and 10-60 seconds when testing plant lysates.

Table 12 List of primary Ab used in western blot detection

Epitope	Manufacturer	Supplier ID
αHis	Merck Millipore	70796
αGFP	Abcam	ab6556
αRFP/mCherry	Abcam	ab124754
αH3K27me1	Merck Millipore	07-448
αH3K27me2	Merck Millipore	07-452
αH3K27me3	Merck Millipore	07-449
αH3	Abcam	ab12079
αEMF2	custom made, eurogentec (Turck, unpublished)	-

3.5.5 Phylogenetic analysis

Phylogenetic tree was constructed based on orthologues protein sequences of CLF and SWN identified in 3.1.2. All trimmed sequences were aligned with the Mega6 software using muscle and a maximum likelihood algorithm (unrooted, no punishments on gap openings) was used to construct 1000 independent trees. Bootstrap analysis was made for the tree construction and the tree was centered using Drosophila E(z).

3.5.6 Blue native PAGE

Blue native PAGE was performed based on the NativePAGE Novex Bis-Tris gel system (Invitrogen) in the first dimension. Proteins in native loading buffer were separated at 150V, 4°C and until running front ran out of the gel (app. 3h). The first dimension PAGE was either used for single band excision and subsequent LC-MS/MS data acquisition (see 3.5.7), or the entire lane was excised and employed in a second, denaturing dimension. After excision, the lane was equilibrated two-times in 1% SDS for 10 min, room temperature, while shaking. The excised and equilibrated lane was embedded in 0.5% low-melt agarose, which was supplemented with very little bromophenol blue, on a ZOOM gradient SDS PAGE 4-12% (Invitrogen). Proteins were separated by size applying 100V for app. 2h at room temperature. Second dimension was stained using standard silver staining and spots were excised for LC-MS/MS data acquisition (see 3.5.7)

3.5.7 LC-MS/MS data acquisition

The excised lanes and bands were trypsin-digested. Digested peptide products were dried and the peptides were dissolved in 2% acetonitrile (ACN), 0.1% TFA for analysis. Samples were analyzed using an EASY-nLC 1000 (Thermo Fisher) coupled to a Q Exactive Plus mass spectrometer (Thermo Fisher). Peptides were separated on 16 cm frit-less silica emitters (New Objective, 0.75 µm inner diameter), packed in-house with reversed-phase ReproSil-Pur C18 AQ 3 µm resin (Dr. Maisch). Peptides were loaded on the column and eluted for 50 min using a segmented linear gradient of 5% to 95% solvent B (0 min : 5%B; 0-5 min -> 5%B; 5-25 min -> 20%B; 25-35 min ->35%B; 35-40 min -> 95%B; 40-50 min ->95%B) (solvent A 0% ACN, 0.1% FA; solvent B 80% ACN, 0.1%FA) at a flow rate of 300 nL/min. Mass spectra were acquired in data-dependent acquisition mode with a TOP10 method. MS spectra were acquired in the Orbitrap analyzer with a mass range of 300–1500 m/z at a resolution of 70,000 FWHM and a target value of 3×10^6 ions. Precursors were selected with an isolation window of 1.3 m/z. HCD fragmentation was performed at a normalized collision energy of 25. MS/MS spectra were acquired with a target value of $5x10^5$ ions at a resolution of 17,500 FWHM, a maximum injection time of 85 ms and a fixed first mass of m/z 100. Peptides with a charge greater than 6, or with unassigned charge state were excluded from fragmentation for MS^2, dynamic exclusion for 20s prevented repeated selection of precursors.

3.5.8 LC-MS/MS data analysis

Raw data were processed individually using MaxQuant software (version 1.5.7.4, Weblink [10]) (Cox & Mann, 2008) iBAQ enabled (Tyanova, Temu et al., 2016) MS/MS spectra were searched by the Andromeda search engine against an insect database (Insecta (50557), reviewed database, UniProt) and the sequences of the PRC2 complex components, additionally sequences of 244 common contaminant proteins and decoy sequences were added during the search. Trypsin specificity was required and a maximum of two missed cleavages allowed. Minimal peptide length was set to seven amino acids. Carbamidomethylation of cysteine residues was set as fixed, oxidation of methionine and protein N-terminal acetylation as variable modifications. Peptide-spectrum-matches and proteins were retained if they were below a false discovery rate of 1%.

3.5.9 Activity assay

HMTase measurements were set up as three technical replicates in a 384 well plate and a total volume of 20µL. Each reaction contained 1x HMTase buffer (see 3.3.9), 20 µM SAM, 100 nM (unless otherwise indicated) enzymatic complex, 65.1µm H3.3 (reaction biology) (or the indicated concentrations of chicken oligonucleosomes/ H3 (21-44) peptides) (see Table 13). If H3 (21-44) peptides were used, the peptides were dissolved in ultrapure water, aliquoted, lyophilized and stored at -20°C. Before usage pure, lyophilized peptides were dissolved in ultrapure water and measured in a nanodrop at 280 nm wavelength using the correction for their exact molecular weight

and exact extinction coefficient. Reactions were incubated at 22°C for 2h (or the indicated times). The reaction was either detected by liquid scintillation according to Fingerman, Du et al. (2008), in which case ^{14}C-labelled SAM was used as substrate or using the EPIgeneous Methyltransferase assay kit (Cisbio), in which case 'cold' SAM was used as a substrate. Samples were stopped and detected according to manufacturer's instructions (Cisbio) together on a plate with 20 µM SAM/SAH standard (see Table 13). The 625 nm and 660 nm HTRF signal was measured at a sensitivity between 100-150 in a Synergy 4 plate reader (BioTek), which was upgraded with an HTRF filter set (BioTek).

Table 13 Materials used in HMTase assay

Name	Supplier	Supplier ID
EPIgeneous Methyltransferase assay 1,000tests	Cisbio	62SAHPEB
EZH1 complex	Reaction biology	HMT-25-115
Histone H3.3	Reaction biology	HMT-11-134
Nucleosomes (ChickenOligo)	Reaction biology	HMT-35-177
[Lys(Me2)27] - Histone H3 (21 - 44) - GK(Biotin)	Anaspec	AS-64366-1
[Lys(Me1)27] - Histone H3 (21 - 44) - GK(Biotin)	Anaspec	AS-64365-1
Histone H3 (21 - 44) - GK(Biotin)	Anaspec	AS-64440-1

4 Results

4.1 Characterisation of CLF and SWN catalytic activity

Previous studies have shown that CLF and SWN have an unequal contribution to genome-wide H3K27me3 levels *in vivo* (Zografou, unpublished). Hence one aim of this study was to highlight the underlying protein differences of CLF and SWN *in vitro*. As metazoan E(z) type proteins are inactive without a PRC2 context, I speculated that plant PRC2 behaves similarly, making it necessary to produce all PRC2 components to test catalytic activity. Prokaryotic expression systems were excluded due to the possible effect on the activity of missing/variant posttranslational modification of PRC2 proteins. Two expression systems were evaluated: (1) Magnifection ® in *Nicotiana benthamiana* (tobacco) and (2) baculovirus expression system in Sf21 insect cells (Figure 2).

4.1.1 Protein expression is a bottleneck for studying plant PRC2 biochemistry

4.1.1.1 Expression of FIE and MSI1 using magnifection®

The individual PRC2 components MSI1, FIE, EMF2, CLF and SWN were cloned into proviral vectors designed for either apoplastic or cytoplasmic expression. Western Blot analysis of a crude extract made from tobacco expressing apoplastic proteins showed that only MSI1 and FIE express to a sufficient level, while EMF2 expressed much less and CLF and SWN were undetectable (*Figure 2*, B). In cytoplasmic expression only MSI1 and FIE expressed, but no CLF, SWN or EMF2 were detectable (data not shown). In all cases tobacco leaves showed lesions leading to necrosis of infiltrated areas, most strongly pronounced in CLF and SWN expressing leaves (data not shown). As necrosis of tobacco leaves presumably negatively affected protein expression, the leaves were sampled at different time points to obviate these negative effects. Western blot indeed confirmed that an earlier harvest of tobacco leaves had a positive effect on protein quality, since after only four days of expression, the protein bands, which probably stem from degradation, were less pronounced than compared with harvest after 6 and 10 days of expression (*Figure 2*, C). However, since the PRC2 HMTases CLF and SWN did not express, the magnifection® expression system was abandoned and Sf21 cells were used alternatively.

4.1.1.2 Sf21 insect cells express all plant PRC2 components

A suspension culture of Sf21 cells was established. Cells stocks at a low passage number were made and frozen to guarantee availability of a healthy cell culture at all times. To achieve high baculoviral titer and to avoid WT baculovirus contamination, the production required optimisation (final pipeline depicted in Figure 2, A). A crucial optimisation concerned the P1 virus production (3.3.5). Here, sequential passaging of cells, which were infected with P0 virus (1:1000, P0 virus volume:culture volume), in a 24h cycle was performed until cell diameter increased and infection of the majority of Sf21 cells in a culture indicated a high viral titer (3.3.5). Starting with a high quality P1

virus, P2 virus production merely served to increase the volume of viral supernatant (Figure 2, A.) Using this pipeline, a viral titer of up to 10^8 plaque forming units/mL (PFU/mL) was measured using a plaque assay (3.3.6). Expression of individual PRC2 components MSI1, FIE, EMF2, CLF and SWN was tested using western blot with His Ab (Figure 2, E). Unlike MSI1, FIE, CLF and SWN, EMF2 carries a Strep tag and cannot be detected using His Ab. All PRC2 components accumulated to sufficient levels in Sf21 cells. Hence, co-expression of CLF-PRC2 (MSI1, FIE, EMF2, CLF) and SWN-PRC2 (MSI1, FIE, EMF2, SWN) was tested and optimised for multiplicity of infection (MOI), which indicates ratio of baculovirus to Sf21 cells, and for harvesting time (Figure 2, F). Individual viruses of MOI 3 and MOI 10 were mixed for co-expression. With MOI 3 and MOI 10, MSI1 remained to accumulate highly after 24h, 48h and 72h, while CLF and SWN protein levels dropped below detection range after 24h. MSI1 accumulated slightly higher with MOI 10, but CLF and SWN were less abundant even after only 24h (Figure 2, E). With MOI 10 and MOI 3, the FIE protein level gradually decreased comparing 24h, 48h or 72h and FIE was generally still detectable after 72h expression. In conclusion, expression of all PRC2 components was most equal at 24h after infection and using a MOI 3. Therefore, these conditions were used for subsequent experiments.

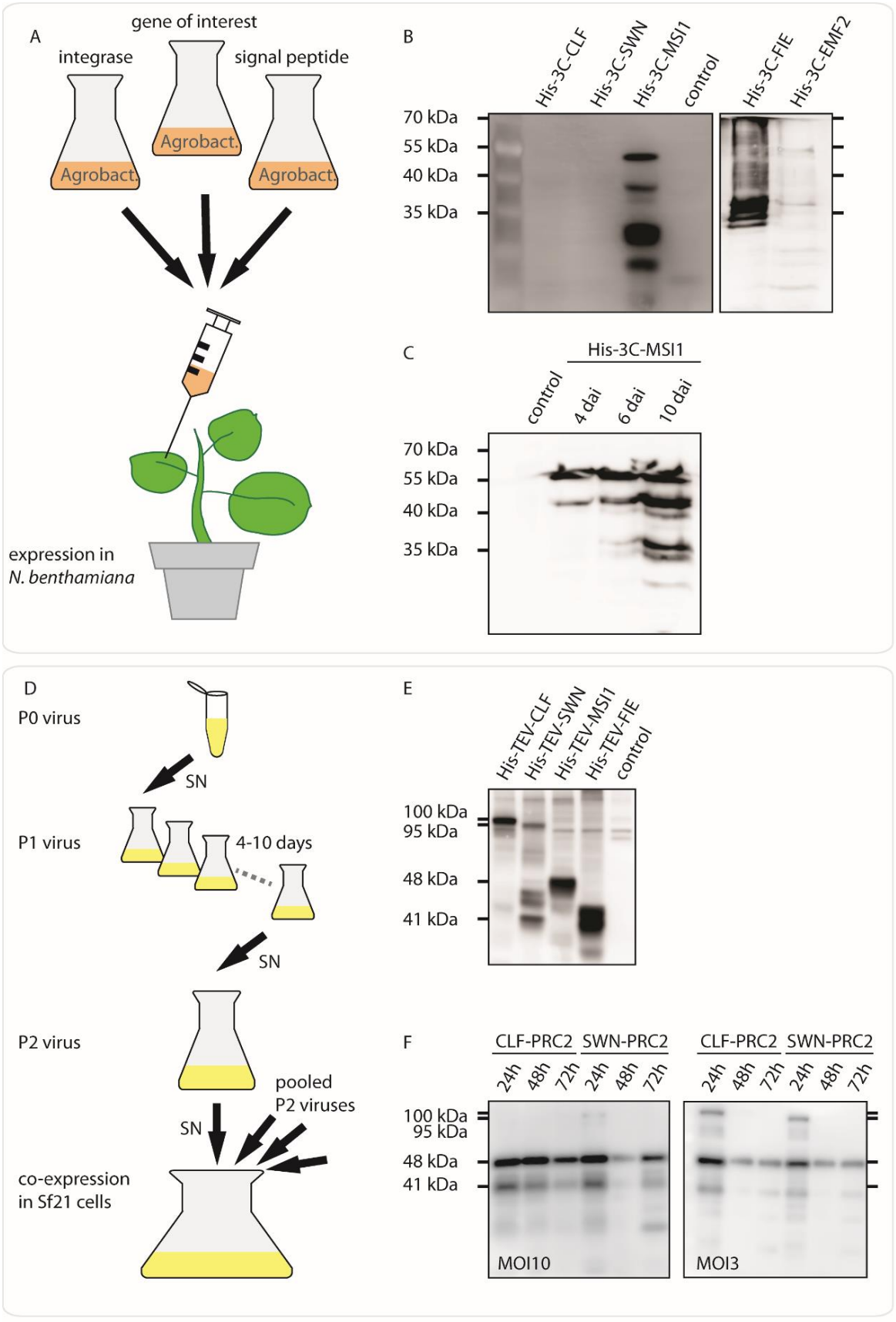


Figure 2 Superior performance of Sf21 insect cells over tobacco (*Nicotiana benthamiana*) expressing plant PRC2 components.

(A) Schematic diagram of individual PRC2 component expression in tobacco using Magnifection ®.
(B) Western blot of individual His-tagged PRC2 components in tobacco after 4 days after infection (dai).
(C) Anti-His Western blot to optimise harvesting time in tobacco comparing 4, 6 and 10 dai.
(D) Schematic diagram of PRC2 co-expression in Sf21 cells.
(E) Anti-His Western blot of individual His-tagged PRC2 components.
(F) Anti-His Western blot to optimise harvesting time and multiplicity of infection (MOI) in Sf21 cells, co-expression of PRC2. All proteins are His-tagged, except EMF2, which is Strep2-tagged (western blot not shown).
Approximate molecular weight of tagged proteins: 100 kDa CLF, 95 kDa SWN, 48 kDa MSI1, 41 kDa FIE and 72 kDa EMF2.

4.1.1.3 CLF- and SWN-PRC2 can be purified from Sf21 cells as a tetramer using EMF2-Strep2 as bait

Following the optimised pipeline, PRC2 complexes were expressed at large-scale (2-4L cultures). Purification of CLF- and SWN-PRC2 was performed using Äkta (GE) fast protein liquid chromatography (FPLC) (representative chromatograms shown in Figure 3, A and B). The chromatograms showed a distinct protein elution peak, which generally had a left-sided saddle. Although individual fractions were tested using western blot, the protein purity was comparable between fractions and peak fractions were pooled henceforth. SDS PAGE indicated a relatively high purity and band intensity of individual PRC2 components was similar (Figure 3, C). Two additional protein bands with an approximate size of 40 and 70 kDa appeared consistently (Figure 3, C). Label-free quantitative mass spectrometry (LC-MS/MS) analysis of the total sample confirmed presence of MSI1, FIE, EMF2 and CLF or SWN proteins in respective sample (Figure 3, D and E). In addition, LC-MS/MS analysis identified contaminant protein bands as insect isoforms of heatshock protein 70 (HSP70) and actin. The predicted size of 70 kDa and 40kDa matched with the observed size of HSP70 and actin, respectively (Figure 3, C, D and E). Overall, HSP70 contamination is consistent with the literature (Schmitges et al., 2011) and the relative stoichiometric abundance of HSP70 and actin with PRC2 is possibly due to a specific interaction (Justin Goodrich, personal communication). Instead of further purification of CLF and SWN-PRC2 using size-exclusion chromatography, I decided to compare their stoichiometry and overall composition by blue native gel electrophoresis. (BN PAGE).

4.1.2 CLF- and SWN-PRC2 purify from Sf21 cells with comparable composition

BN PAGE analysis showed that composition of CLF and SWN-PRC2 was highly similar (Figure 3, F). Specifically, five specific protein bands were observed, indicating five distinct protein complexes containing PRC2 components and contaminants (Figure 3, F). While three bands were running at a high molecular weight range (>146 kDa), two additional protein bands were running at a low molecular weight range (<60 dDa). To investigate the five protein bands of the BN PAGE (, F) for their composition, they were excised and subjected to LC-MS/MS analysis.

iBAQ values were extracted from the LC-MS/MS analysis to compare the abundance of individual components in both the protein bands number one to five and the total sample of three

independent CLF- and SWN- PRC2 expressions/purifications (Figure 3, G and H). IBAQ values of one band/sample were summed and relative percentages were plotted (Figure 3, G and H). Comparing quantification of PRC2 components in the total sample showed that CLF/SWN were present at similar quantity to EMF2 (Figure 3, G and H), while MSI1 exceeded EMF2 abundance roughly twofold. FIE was least present in all total samples (Figure 3, G and H).

A comparison of the iBAQ percentages of individual protein bands showed that protein band number one predominantly contained MSI1 in both the CLF- and the SWN-PRC2 purification (Figure 3, G and H). This suggested that MSI1 might disintegrate during BN PAGE separation from remaining PRC2, possibly due to its low isoelectric point (app. 2 units below other PRC2 components). Protein band number two predominantly contained EMF2 (Figure 3, G and H). The protein bands number three to five shared a similar composition between each other and between CLF- and SWN-PRC2 samples containing approximately equal percentages of CLF/SWN and EMF2. In protein bands number three to five, MSI1 contributed between 1-35% and had its highest presence in band number five. Overall, FIE was present with the lowest percentage ranging from 2-7% in bands number three to five. In sum, MSI1 was most abundant in purified PRC2 complexes, but was predominately enriched in band number one, while EMF2 either migrated as a single component in band number two or migrated together with CLF/SWN at an equimolar ratio in a dimeric sub-complex, and FIE was least abundant and migrated in complex with all components and sub-complexes.

To investigate the composition of purified CLF- and SWN-PRC2 complexes more deeply, the BN PAGE was subjected to a second dimension separation using denaturing gel electrophoresis (3.5.6). Individual spots were excised and analyzed using LC-MS/MS to acquire iBAQ percentages (complete analysis summarised in the following). In such a two-dimensional separation, single proteins generally migrate along a diagonal from top left to bottom right, while protein complexes generally migrate within a triangle left of the diagonal. Indeed, a fraction of the total FIE and MSI1 protein migrated individually and corresponded to spots 1 and 11, respectively (Figure 3, I and J). EMF2 and CLF/SWN separated along the vertical axes and were identified in spots number 12/7 and 13/8, respectively, indicating that they were separated from a single band in the first dimension, which contained both proteins in a dimeric sub-complex. In addition to spots number 12 and 7, EMF2 was also identified as the dominant component in protein band number two (Figure 3, I and J), which was likely separated from protein band number two in the BN PAGE first dimension (Figure 3, F). CLF and SWN were identified in protein bands 13, 8 and 3 in the respective sample and were probably separated from bands 5, 4 and 3 of the BN PAGE first dimension, respectively (Figure 3, F). Overall, CLF and SWN-PRC2 purification showed remarkable similarity in all sample analyses performed here. Despite the presence of sub-stoichiometric complexes, I concluded that the obtained CLF- and SWN-

samples were remarkably comparable with regard to composition and stoichiometry. I therefore proceeded to characterise kinetic parameters of the two Arabidopsis PRC2 variants.

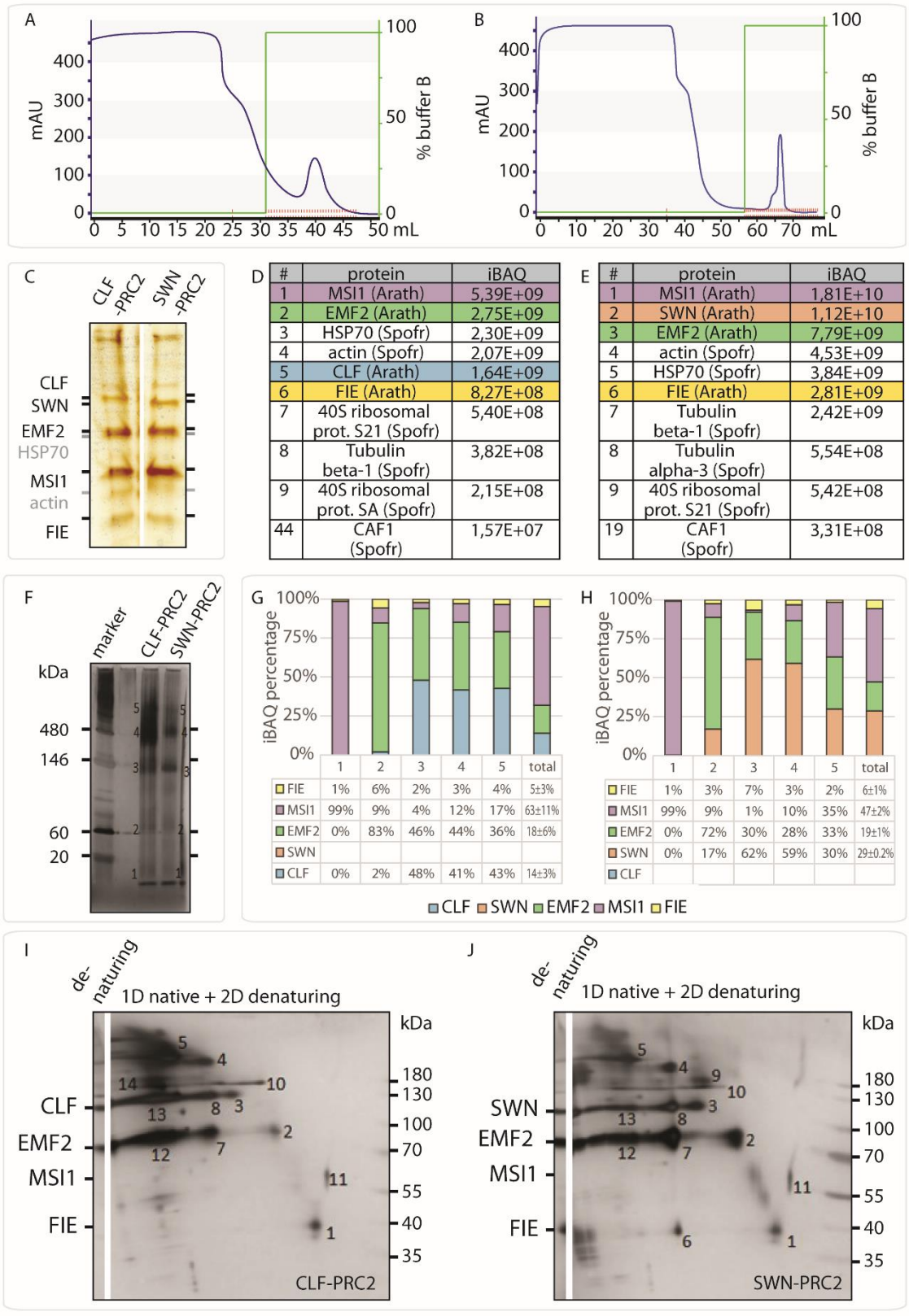


Figure 3 Comparative, representative Strep-affinity purification of Strep-tagged EMF2 co-expressed as tetrameric PRC2 (CLF/SWN, FIE, MSI1) in Sf21 cells.

(A, B) Chromatogram of *CLF- and SWN*-PRC2 purification using Äkta FPLC, respectively. 100% buffer B (green) corresponds to protein complex elution and 0.5 mL fractionation (red). Absorption by protein was measured in mAU (blue).
(C) SDS PAGE of CLF- and SWN-PRC2 (purified at pH 8) using silver staining. Equal concentration of purified complex was loaded. Contaminants HSP70 (app. 70 kDA) and actin (40 kDa) were identified in LC-MS/MS.
(D, E) LC-MS/MS analysis of total eluate from CLF- and SWN- PRC2 purifications. IBAQ values of nine most abundant proteins and CAF1.
(F) Compositional analysis of purified PRC2 using silver-stained, one-dimensional (1D) blue native PAGE.
(G, H) LC-MS/MS analysis of bands 1-5 excised from blue native PAGE in (F). Relative abundance of FIE, MSI1, EMF2, SWN and CLF measured as iBAQ percentage. iBAQ percentage for each PRC2 component was calculated as proportion from the sum of all PRC2 iBAQ values. 'total' indicates sum of three independent protein purifications, standard error indicated as ± percentage.
(I, J) Two-dimensional PAGE of CLF and SWN-PRC2 purifications. First dimension is BN PAGE and second dimension is a denaturing SDS PAGE. Reference was loaded and is shown on left side of each image ('denaturing'). Numbering refers to the individual spots that were excised for LC-MS/MS analysis.

4.1.3 CLF-and SWN-PRC2 show histone methyltransferase activity *in vitro*

4.1.3.1 *^{14}C incorporation is below detection range*

To record enzyme kinetics of CLF and SWN-PRC2, HMTase reactions were performed *in vitro* using ^{14}C isotope-labelled SAM (^{14}C SAM). Detection was carried out using liquid scintillation, phospoimager and photo screen. None of the detection methods revealed substrate conversion. Although the amount of ^{14}C SAM, which was applied in the activity assay, could be detected by liquid scintillation, a tenth of this amount could no longer be detected. Hence, I concluded that specific activity of ^{14}C SAM was not high enough in relation to the theoretical product incorporation, which expectedly would not exceed 10% (compare Figure 4, B). I decided to use an alternative method for product detection.

4.1.3.2 *Homogeneous time-resolved fluorescence to measure SAM to SAH conversion*

In an alternative attempt to record enzyme kinetics of CLF- and SWN-PRC2, HMTase reactions were performed *in vitro* and detected using homogeneous time-resolved fluorescence (HTRF). This method relies on the use of a fluorophore-labelled Ab (emission at 620nm) specific to the co-product SAH (αSAH Ab). SAH created by the enzymatic reaction competes with fluorophore-labelled SAH (emission at 665nm) in the detection mix leading to depletion of the otherwise present Förster resonance energy transfer (FRET). The HTRF signal (665/620nm) was normalised using a standard curve. Due to potential cross-reactivity of SAM with αSAH Ab, SAH and SAM were mixed in defined ratios, which amount to a total of 20 µM SAM/SAH. Owing to cooperative behaviour of the αSAH Ab to bind SAH, the standard curve had a sigmoidal shape and showed a two-sided saturation at both high and low SAH concentrations (Figure 4, A). A regression model was made based on the Hill equation describing this cooperative behavior. Using the Solver excel plugin, stochastic values were seeded in free parameters (dissociation constant, K_d; Hill coefficient, n;. maximal velocity, V_{max}). Due to the logarithmic nature of the curve and regression model, errors are possibly higher at the lower and upper saturation levels of the SAH Ab. However, the effect is probably negligible, because of the good fit of the model and because complete saturation was avoided in the experimental set-up.

Instead of 20µM SAM as co-substrate, 2µM and 40µM were also considered. The error is expectedly smaller at lower SAM concentrations, due to less saturation of the SAH Ab, while it is expectedly larger at higher SAM concentrations, due to a more pronounced saturation of the SAH Ab. Therefore, 20µM probably present a good compromise between a desirable excess of SAM co-substrate and the limitations presented by the HTRF method.

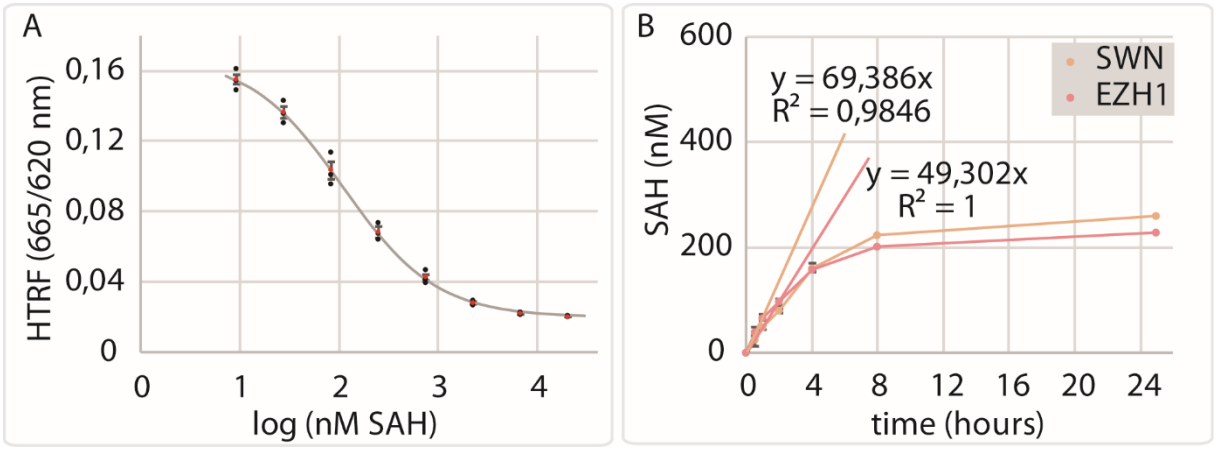


Figure 4 Optimisation of histone methyltransferases assay using homogeneous time-resolved fluorescence (HTRF).

(A) HTRF measurement of one representative SAM/SAH standard as it was included on each independent well-plate to normalise HTRF signal (665nm/620nm). Dilution ratios of SAM/SAH of 20µM total concentration. Average (red) of three technical replicates (black), error bars are standard error. Hill model ($y = V_{max} * [S]^n / (K_M^n + [S]^n) + V_{min}$) fitting to the saturation curve, free parameters V_{max}, n and K_m yielded V_{max}=0.148, n=-0.9 and K_M=110.04 nM in the computationally fitted Hill model with a least square of R^2=0.999. V_{min}=0,02 is blank sample and [S] is SAH concentration.

(B) Initial velocity measurement of SWN-PRC2 (orange) and EZH1-PRC2 (red) *in vitro* histone methyltransferase activity on H3.3 substrate in a 24 hours time course. HTRF signal normalised using standard as in (A) Linear regression model fitting for the first 2 hours of the reaction; R^2 is least square of the linear regression model. Error bars are standard error of three technical replicates.

4.1.3.3 Optimisation of HMT reaction conditions

Next, the initial velocity was measured to find the optimal incubation time over a time-course of 24h and to avoid underestimation of CLF- and SWN-PRC2 activity due to decreasing activity during incubation time. CLF- and SWN-PRC2 as well as a commercially available EZH1-PRC2 were consistently within linear phase of reaction within 2h after the beginning of the enzymatic reaction (Figure 4, B; Figure 5, A). After 8-19h of incubation, no more SAH was created by SWN or EZH1, while with CLF this occurred much earlier, between 2-8h (Figure 5, A). In a direct comparison of CLF- and SWN-PRC2, the former maximally converted 8.8% SAM into SAH even after prolonged incubation, whereas the latter only converted up to 1.7%. Apparent saturation of the reaction could either be a result of product inhibition or decreasing protein stability. Indeed, CLF-PRC2 proved more fragile than SWN-PRC2 during protein purification, when using the Strep2-tagged EMF2 as bait for

remaining PRC2 components (CLF/SWN, FIE, MSI1). Using a pH 7.5 buffer series during purification had a more pronounced negative effect on protein abundance of CLF and FIE than compared to SWN and FIE in a parallel purification (Figure 5, B). The incubation time of the enzymatic reaction was therefore reduced. Since SAH levels were below detection range at 1h, finally, an incubation time of 2h was chosen in order to guarantee linear enzyme kinetics.

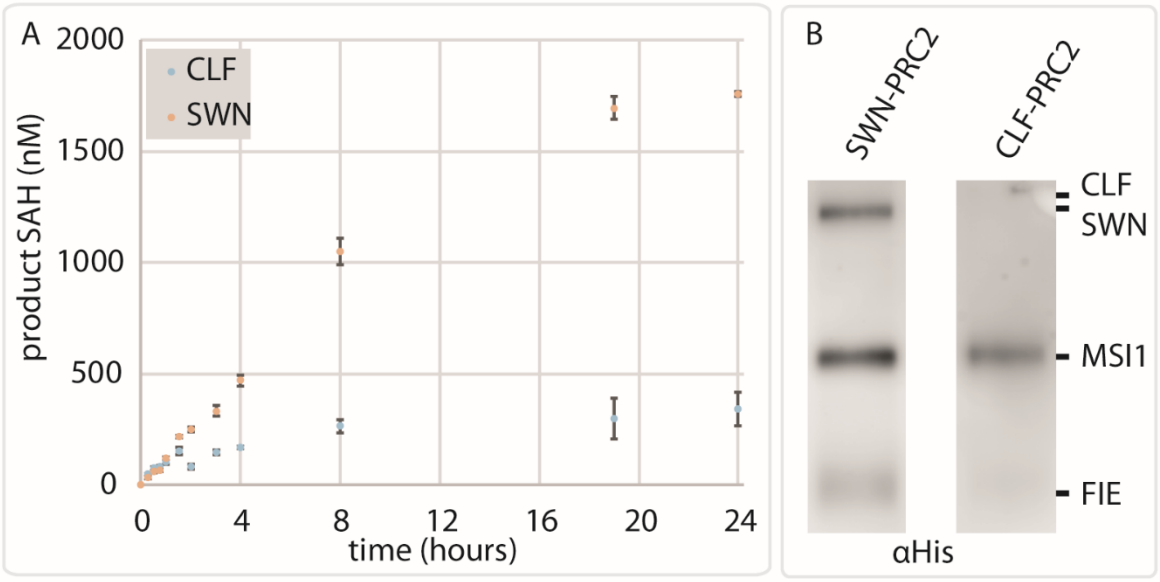


Figure 5 CLF-PRC2 activity depletes due to instability.

(A) Initial velocity measurement of CLF-PRC2 (blue) and SWN-PRC2 (orange) *in vitro* histone methyltransferase activity on H3.3 substrate in a 24 hours time course. HTRF signal normalised using standard as in (figure 4, A). Error bars are standard error of three technical replicates.

(B) Anti-His Western blot of CLF- and SWN-PRC2 complexes, which were purified at pH 7.5. Protein complexes were purified using Step2-tagged EMF2 as a bait.

4.1.4 *In vitro* methylation predominantly produces H3K27 dimethylation

To ascertain whether H3K27 methylation in the *in vitro* reaction is H3K27me1, H3K27me2 or H3K27me3 production, specific antibodies were used to analyze the reaction products (Figure 6, A). In a reaction using recombinant human H3.3 produced in *E. coli*, predominantly H3K27me2 was produced at high enzyme concentrations. H3K27me3 was undetectable and H3K27me1 was barely detectable (Figure 6, A). The H3K27me2 Ab showed unspecific binding to unmethylated H3.3, which I missed to avoid by blocking the Western blot using un-methylated H3 peptide.

In order to test whether CLF-PRC2 or SWN-PRC2 have differential activity depending on the substrate and its respective methylation state, three substrates were tested: recombinant H3.3, chicken oligonucleosomes and H3 peptide (amino acids 21-44) (Figure 6, A). Plant PRC2 was compared with human EZH1-PRC2 as a reference. While H3.3 and H3 peptide are un-methylated substrates, chicken oligonucleosomes probably contain a mix of all four possible methylation states (me0, me1, me2,

me3). Presumably, chicken oligonucleosomes may have the highest resemblance with chromatin substrate of PRC2 *in vivo*. All three PRC2s showed the highest normalised activity on recombinant H3.3, intermediate activity on chicken oligonucleosomes and the least on histone H3 peptide (21-44).

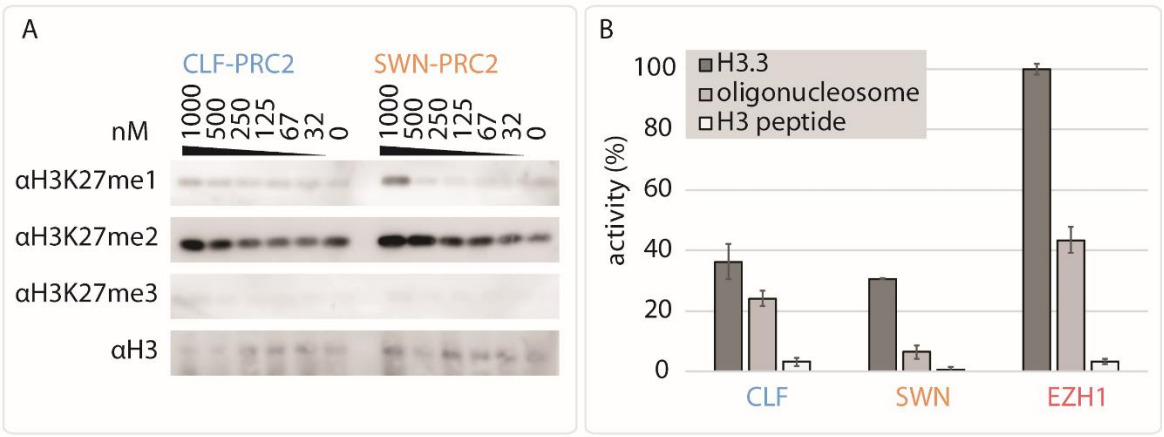


Figure 6 Plant PRC2 produces H3K27me2 *in vitro* and favours H3.3 (un-methylated) and chicken oligonucleosomes (extract) over peptides as substrate.

(A) *In vitro* histone methyltransferases reaction, 24h incubation time. Anti-H3K27me1, anti-H3K27me2 and anti-H3K27me3 Western blot at different concentrations of purified PRC2. Anti-H3 is used to detect total substrate and serves as loading control. Three technical replicates were pooled and immuno-detected on individual blots.

(B) *In vitro* HMTase reaction using 70 nM PRC2 (CLF, SWN and EZH1), 2h incubation time. Normalised activity (%) for three substrates: H3.3, chicken oligonucleosomes and H3 peptide (amino acids 21-44, unmodified). Normalisation to EZH1 and H3.3; fixed substrate concentration (312.5 nM). Error bars are standard error of three technical replicates.

4.1.5 CLF-PRC2 and SWN-PRC2 differ in specificity of H3K27 methylation

To investigate whether CLF- or SWN-PRC2 is more active in general, classic kinetic reaction series were carried out using chicken oligonucleosomes. Here, enzyme and SAM substrate concentration were kept constant and chicken oligonucleosomes were invested at different concentration (0-2500nM). As expected, increased substrate concentration led to increased SAH production per hour (nM SAH/h) in all three cases indicating relatively simple enzyme kinetics (Figure 7, A). Steepness and saturation level increased with the following order for the three enzymes: EZH1>CLF>SWN.

To extract kinetic parameters, a Hill model was fitted using Solver. Stochastic seeds were provided to the model to find solutions of free parameters: Michaelis-Menten constant (K_m) and Hill coefficient (n). Maximal velocity (V_{max}) was seeded in the model based on an educated guess of 30, 40 and 60 nM/h for CLF-, SWN- and EZH1-PRC2, respectively. Computation of K_m, n and V_{max} free parameters found the following least square solutions: *K_m(CLF-PRC2)=0.33 µM, K_m (SWN-PRC2)=1.96 µM and K_m (EZH1-PRC2)=0.34 µM; V_{max} (CLF-PRC2)=33.6 nM/h, V_{max} (SWN-PRC2)=56.7 nM/h and V_{max} (EZH1-PRC2)=61.2 nM/h; n (CLF-PRC2)=1.25, n (SWN-PRC2)=0.87* and *n (EZH1-PRC2)=0.97* (Figure 7, A). From

these solutions of the model *kcat* can be calculated: *kcat (CLF-PRC2)=0.1 h^{-1}, kcat (SWN-PRC2)=0.03 h^{-1} and kcat (EZH1-PRC2)=0.18 h^{-1}*. The specificity constant *kcat/Km* serves as an indicator of catalytic efficiency and commonly serves as comparator between different enzymes: *$kcat/K_m$ (CLF-PRC2)=0.303 μM^{-1} h^{-1}, $kcat/K_m$ (SWN-PRC2)=0.015 μM^{-1} h^{-1}* and *$kcat/K_m$ (EZH1-PRC2)=0,529 μM^{-1} h^{-1}*. In conclusion, CLF-PRC2 is more catalytically active than SWN-PRC2 on chicken nucleosomes.

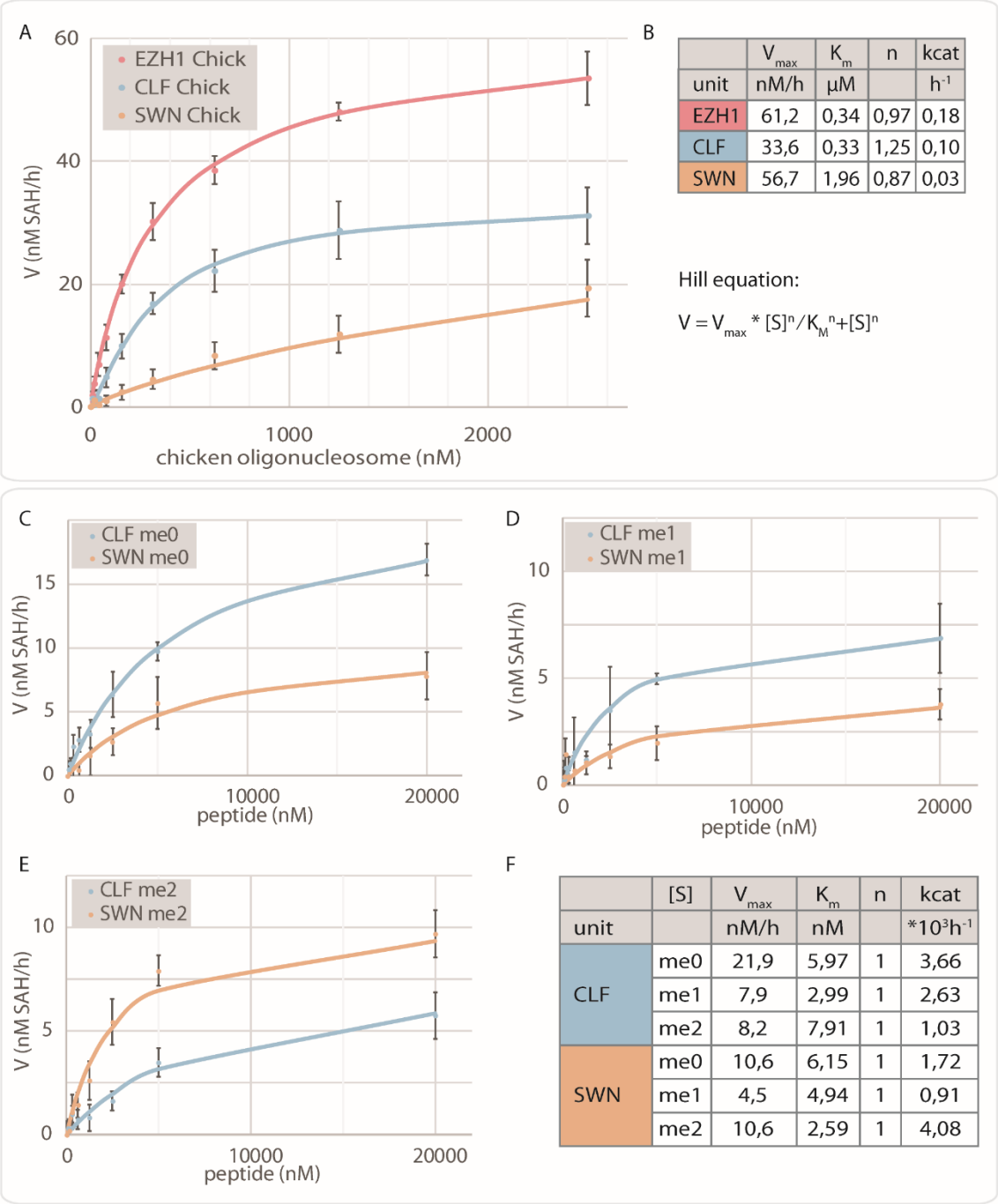
Interestingly, CLF-PRC2 and SWN-PRC2 showed opposing values for the Hill coefficient (Figure *7*, A). The Hill coefficient serves as an indicator of cooperativity of substrate binding. EZH1-PRC2 showed near non-cooperativity indicating simple Michaelis-Menten kinetics. Hill coefficient of SWN-PRC2 on the other hand was n<1, which indicates negative cooperativity, while the Hill coefficient of CLF-PRC2 was n>1, which indicates positive cooperativity. Chicken oligonucleosomes might be a complex substrate with H3K27 in variable contexts and methylation states. In order to get a better approximation using a Michaelis-Menten kinetic model, I sought out more uniform substrates.

To dissect the steps of H3K27 methylation, non- and pre-methylated peptides were used as a substrate and Michaelis-Menten kinetic behavior (n=1) was imposed (Figure *7*, B-F). As indicated previously, H3 peptide (21-44) performs poorly as a substrate compared to chicken oligonucleosomes (Figure 6, B) and the overall activity towards H3K27me0, H3K27me1 and H3K27me2 substrates was indeed much lower (Figure 7, A-E). Nevertheless, kinetic behavior of the reaction could be observed by increasing substrate investment (0-20 µM) compared to chicken oligonucleosomes (0-0.25 µM). Similar to chicken oligonucleosomes, increased substrate concentration led to increased SAH production per hour (nM SAH/h) of CLF- and SWN-PRC2. A curve regression was set as described for chicken oligonucleosomes. CLF-PRC2 outperformed SWN-PRC2 in reaching saturation with H3K27m0 and H3K27me1 peptides used as substrate (Figure 7, C, D). In contrast, SWN-PRC2 outperformed CLF-PRC2in reaching saturation with H3K27me2 peptide used as substrate (Figure 7, E).

Computation of Km and Vmax free parameters found the following least square solutions of CLF-PRC2: K_m(H3K27me0)=5.97 nM, K_m (H3K27me1)=2.99 nM and K_m (H3K27me2)=7.91 nM; V_{max} (H3K27me0)= 21.9 nM/h, V_{max}(H3K27me1)=7.9 nM/h and V_{max}(H3K27me2)=8.2 nM/h. Computation of SWN-PRC2 kinetics found the flowing solutions: *K_m(H3K27me0)=6.15 µM, K_m (H3K27me1)=4.94 µM* and *K_m (H3K27me2)=2.59 µM; V_{max} (H3K27me0)=10.6 nM/h, V_{max}(H3K27me1)=4.5 nM/h* and *V_{max}(H3K27me2)=10.6 nM/h* (Figure 7, F). From these solutions of the model *kcat* was calculated for CLF-PRC2: *kcat(H3K27me0)=3.66 10^{-3} h^{-1}, kcat(H3K27me1)= 2.63 10^{-3} h^{-1}* and *kcat(H3K27me2)= 1.03 10^{-3} h^{-1}*. SWN-PRC2 has the following *kcat* values: *kcat(H3K27me0)=1.72 10^{-3} h^{-1}, kcat(H3K27me1)= 0.91 10^{-3} h^{-1}* and *kcat(H3K27me2)= 4,08 10^{-3} h^{-1}*. The specificity constant *$kcat/K_m$* was calculated for CLF-PRC2: *$kcat/K_m$(H3K27me0)=6.1 10^{-4} nM^{-1} h^{-1}, $kcat/K_m$ (H3K27me1)= 8.8 10^{-4} nM^{-1} h^{-1}* and *$kcat/K_m$(H3K27me2)= 1.3 10^{-4} nM^{-1} h^{-1}*. SWN has the following specificity constants *$kcat/K_m$: $kcat/K_m$(H3K27me0)=2.8 10^{-4} nM^{-}*

^{1} h^{-1}, $kcat/K_m$ *(H3K27me1)= 1.8* 10^{-4} nM^{-1} h^{-1} and $kcat/K_m$*(H3K27me2)= 15.8* 10^{-4} nM^{-1} h^{-1}. Although overall activity towards H3 peptide was very low, it could be determined that CLF- and SWN-PRC2 catalyze the sequential steps of H3K27 methylation with different catalytic specificity. While CLF-PRC2 had a $kcat/K_m$ ratio of approximately 6:9:1 (me0:me1:me2), SWN has a ratio of 3:2:16. In conclusion, CLF has higher specificity towards H3K27me0 and H3K27me1 peptide substrate, while SWN has higher catalytic specificity towards H3K27me2.

Although, I characterised clear biochemical differences in this work, another goal was to better understand the interplay of CLF and SWN *in vivo*. Therefore, experiments in Arabidopsis were performed to better support the biochemical findings.



	V_{max}	K_m	n	kcat
unit	nM/h	µM		h^{-1}
EZH1	61,2	0,34	0,97	0,18
CLF	33,6	0,33	1,25	0,10
SWN	56,7	1,96	0,87	0,03

	[S]	V_{max}	K_m	n	kcat
unit		nM/h	nM		$*10^3h^{-1}$
CLF	me0	21,9	5,97	1	3,66
	me1	7,9	2,99	1	2,63
	me2	8,2	7,91	1	1,03
SWN	me0	10,6	6,15	1	1,72
	me1	4,5	4,94	1	0,91
	me2	10,6	2,59	1	4,08

Figure 7 CLF is a more active histone methyltransferase (HMTase) acting on oligonucleosomes and excels at *de novo* methylation, while SWN excels at conversion of H3K27me2 to H3K27me3.

(A) *In vitro* HMTase reaction using Chicken oligonucleosomes (extract) as substrate of CLF-PRC2 (blue), SWN-PRC2 (orange) and EZH1-PRC2 (red), SAH product detection by HTRF. Normalisation using a 20 µM SAM/SAH standard (same as figure 4, A). Chicken nucleosome (extract) serving as substrate. 2h incubation time within range of maximal initial velocity (figure 4, B; figure 5, A). Error bars are standard error of three technical replicates. Hill model applied graphical regression using Solver.

(B) Table of kinetic parameters of the *in vitro* HMTase reaction using chicken oligonucleosomes as substrate of CLF-, SWN- and EZH1-PRC2, and Hill equation (panel A, bottom right). [S], substrate concentration; V_{max}, maximal velocity; K_m, Michaelis-Menten constant, n, Hill coefficient; n>1; positive cooperative; n < 1; negative coop.; n = 1; non-coop./Michaelis-Menten kinetics and *kcat*, specificity constant. Stochastic seeding of free parameters: Michaelis-Menten

constant (K_m) and Hill coefficient (n). Maximal velocity (V_{max}) was seeded in the model based on an educated guess of 30, 40 and 60 nM/h for CLF-, SWN- and EZH1-PRC2, respectively. Kinetic parameter error not shown due to lack of biological replicates.

(C-E) *In vitro* HMTase reaction using modified histone H3 peptides (amino acids 21-44) H3K27me0 (C, me0), H3K27me2 (D, me1) and H3K27me2 (E, me2) as substrate of CLF- (blue) and SWN-PRC2 (orange). SAH product detection by HTRF. Normalisation using a 20 µM SAM/SAH standard (same as figure 4, A). 2h incubation time within range of maximal initial velocity (figure 4, B; figure 5, A). Error bars are standard error of three technical replicates. Hill regression model fitted using Solver. Stochastic seeding of the Michaelis-Menten constant (K_m) as free parameter. Maximal velocity (V_{max}) was seeded in the model based on educated guess ($V_{max, guess}$ (CLF-PRC2, me0)=20 nM SAH/h; $V_{max, guess}$ (SWN-PRC2, me0)=10 nM SAH/h; $V_{max, guess}$ (CLF-PRC2, me1)=7.5 nM SAH/h; $V_{max, guess}$ (SWN-PRC2, me1)=5 nM SAH/h; $V_{max, guess}$ (CLF-PRC2, me2)=7.5 nM SAH/h; $V_{max, guess}$ (SWN-PRC2, me2)=10 nM SAH/h). Kinetic parameter error not shown due to lack of biological replicates.

4.2 Experiments highlighting redundancy

4.2.1 Signatures of redundancy in expression, localisation and complex formation

While one study showed that CLF mRNA accumulates in floral and vegetative meristems, another study showed the same for SWN (Chanvivattana et al., 2004, Goodrich et al., 1997). To confirm an overlap of expression in a direct comparison, promoter GUS fusions of *CLF* and *SWN* were made and analyzed (Figure 8, A). A strong GUS signal was observed in apical and root meristems of seedlings. Unlike the SWN promoter, the CLF promoter was also active in meristematic regions of the leaf margin (Figure 8, A). Hence, the expression patterns here not only overlap with what previous studies have found, but also showed expression in other meristematic regions of Arabidopsis.

In light of the meristematic expression of CLF and SWN, I assumed that cell division positively affects both *CLF* and *SWN* expression. To test this hypothesis, Arabidopsis seedlings were grown on callus inducing medium. Indeed, the GUS signal expanded widely and thereby confirmed this hypothesis (Figure 8, A).

To make CLF and SWN accessible to immunoprecipitation studies, N-terminal fusion of GFP CDS to genomic constructs (*GFPgCLF*, *mCherrygSWN*) were made. In T1 generation, inflorescences of *GFPgCLF* and *mCherrygSWN* transgenic lines were sampled and expression of labelled protein was confirmed by western blot (Figure 8, B). Despite positive complementation of the *clf28* phenotype by *GFPgCLF* constructs, signal in western blot was not detectable in T3 generation of selected *mCherrygSWN* or *GFPgCLF* lines (data not shown). Hence, despite repeated trials, immunoprecipitation failed to enrich CLF and SWN protein (data not shown). Based on the high expression level in induced calli, samples were taken from *mCherrygSWN* and *GFPgCLF* lines after callus induction. However, sufficient enrichment could still not be achieved. Confocal microscopy was used as an independent experiment to detect mCherry-SWN and GFP-CLF fusion protein (Figure 8, C). Signal was only detectable in very few callus cells (Figure 8, C), but not in shoot nor root meristems of seedlings. The callus cells, which showed expression, confirmed nuclear localisation of both CLF and SWN fusion protein. Since *mCherrygSWN* and *GFPgCLF* transgenic lines proved to express the fusion protein below detection limit under most conditions, they were not used in any further experiments. Instead, I found that complementation experiments were more appropriate. In summary, the expression analysis performed here suggest near full redundancy of CLF and SWN with regard to protein nuclear localisation and gene expression in dividing cells.

Given their redundant expression and localisation, I considered whether a different affinity of CLF and SWN to PRC2 could contribute to the different phenotypes of their respective single mutants. To test this hypothesis, a competitive protein pulldown was performed using protein extracted from Sf21 cells. Here, EMF2 was used as a bait in combination with equal amounts (10:10) or in defined ratios (0:10; 1:9, 9:1 or 10:0) of CLF and SWN in presence of MSI1 and FIE proteins. EMF2, MSI1 and FIE

were applied at limitation (100-fold dilution), while CLF and SWN were applied at saturation (10-fold dilution). The experiment showed that CLF and SWN were pulled out with similar amounts indicating for roughly similar affinity to PRC2 (Figure 9, A).

In summary, my experiments showed that CLF and SWN share tissue-specific expression, subcellular localisation and can replace each other effectively in EMF2-PRC2.

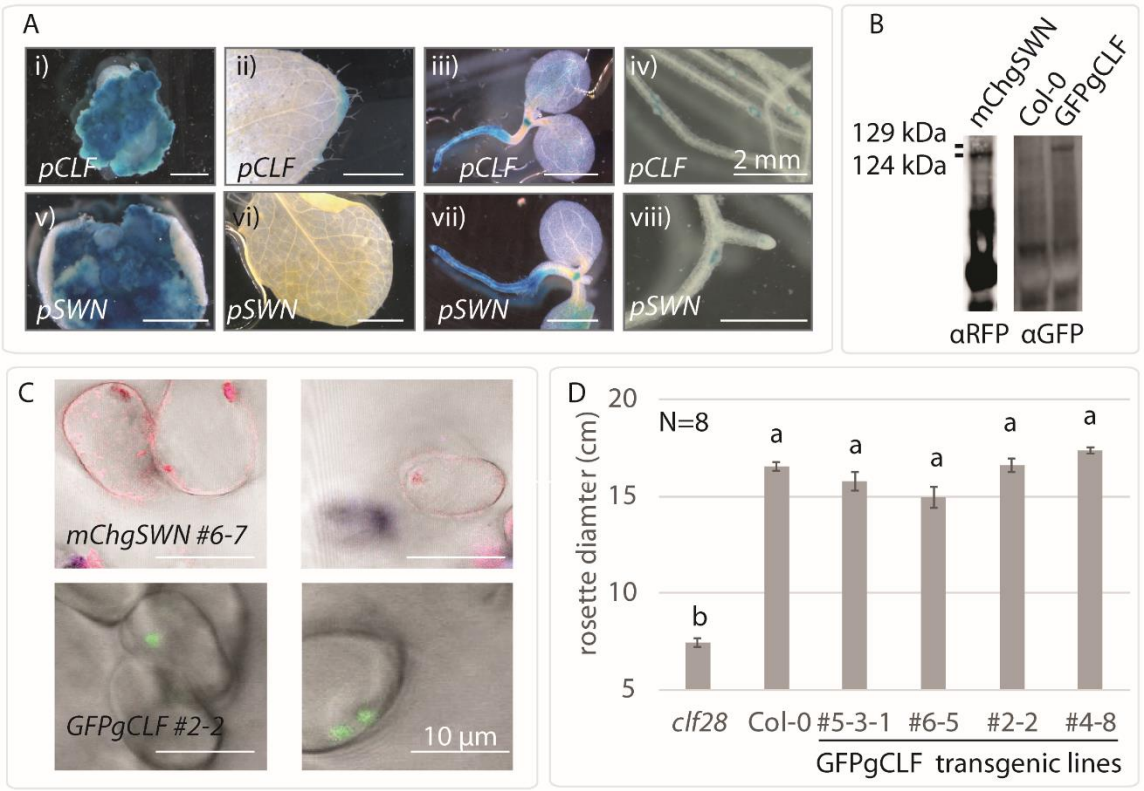


Figure 8 CLF and SWN overlap in meristematic expression and nuclear localisation.

(A) Histochemical GUS staining of *CLF* promoter (*pCLF*, i-iv) and of *SWN* promoter (*pSWN*, v-viii) plantlets. The pictures show representative distribution pattern of GUS staining in: 21-day old induced callus (i, v), leaf margin of 21-day old seedling (ii, vi), 5-day old whole seedling (iii, vii) and root tips of 21-day old seedling (iv, viii). Scale bars: 2 mm.

(B) Anti-RFP and anti-GFP Western blot on total extract from pooled T1 transgenic plants. Floral meristems of approximately 10 individual plants were pooled to prepare sample. Col-0 sample as control. Estimated size of mCherry-SWN fusion protein is 124 kDa and of GFP-CLF is 129 kDa.

(C) Fusion protein localisation of mCherry-SWN (upper) and GFP-CLF (lower) in 21-day old callus cells induced from germination. Amino (N)-terminal fusion protein expressed from the genomic constructs *mCherrygSWN* and *GFPgCLF* in mutant backgrounds *swn7* and *clf28*, respectively. Shown are two relatively high-expressing T3 lines *(mCherrygSWN* #6-7 and *GFPgCLF* #2-2), which were selected by anti-GFP/anti-RFP Western blot in T2 generation (data not shown). Scale bars: 10 µm.

(D) Complementation of GFPgCLF construct in *clf28* mutant. Complementation measured by rosette diameter at first flower opening of four independent T3 transgenic lines in comparison to *clf28* mutant and Col-0. Error bars indicate mean standard error of eight individuals per transgenic line or reference (N=8). Significance determined by two-sided ANOVA with multiple comparison correction by Holm-Sidak. Letters (a, b) indicate two significance groups ($P < 0.05$).

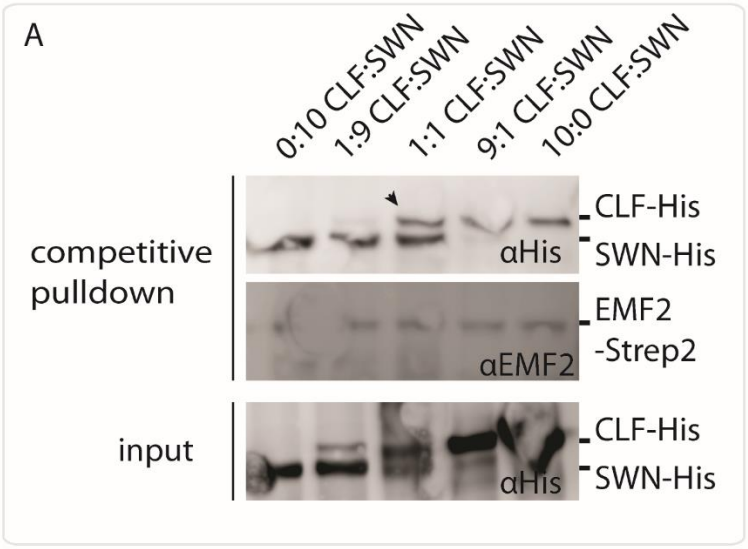


Figure 9 CLF and SWN compete for PRC2 with equal affinity.

Competitive pull down of protein expressed in Sf21 cells using EMF2-Strep as bait. PRC2 components MSI1, FIE and EMF2 were applied at limitation (100-fold dilution) and CLF and SWN at saturation (10 fold dilution) with the ratios 0:10, 1:9, 1:1, 9:1 or 10:0 (CLF:SWN). Western blot detection using anti-His (upper) and specific anti-EMF2 (middle) antibodies on eluate, and anti-His antibody on input (lower). Arrow indicates bands of CLF and SWN after purification at 1:1 ratio.

4.3 Experiments highlighting functional divergence

4.3.1 Functional divergence at coding sequence level can be observed *in vivo*

Expression analysis in Arabidopsis showed that *CLF* and *SWN* promoters are active in mostly the same tissues, while biochemical analysis highlighted some clear differences in methylation specificity. In

addition, the *clf28* mutant proved to be sensitive to decreased allelic dose of *swn* mutation resulting in enhancement of phenotypic defects of *clf* mutant (Figure 10, A, B).

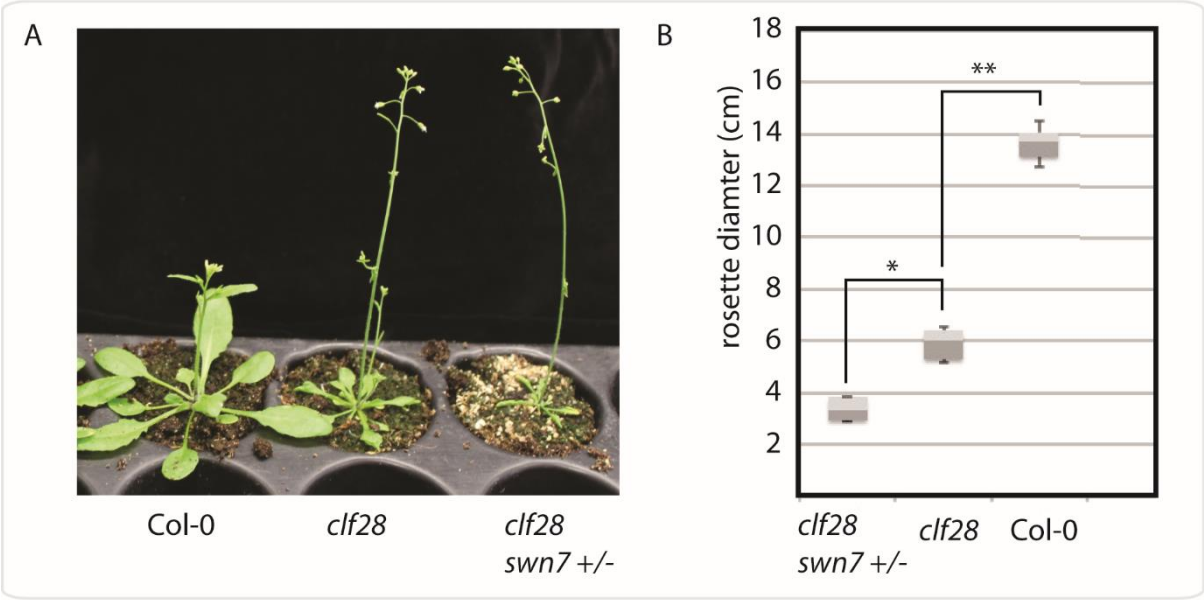


Figure 10 *SWN* allelic dose-dependency of *clf28* mutant plants.
(A) Representative plants at flowering of Col-0, *clf28* mutant and *clf28*, which carries one mutant *swn7* allele (*clf28 swn7* +/-).
(B) Quantitative analysis of rosette diameter at first flower opening of *clf28 swn7* +/-, *clf28* and Col-0 plants. Twelve individual plants per genotype (N=12). Statistical t-test with significance levels **p<0.01, *p<0.05.

Conversely, overexpression of *SWN* in *clf28* background should therefore alleviate phenotypic defects of the mutation. To test this hypothesis and to investigate redundancy of *CLF* and *SWN* regulatory sequences, promoter swap experiments were set up (Figure 11, A). C*lf28* was partially complemented in the rosette diameter phenotype in four out of eight transgenic lines using *SWN* CDS driven by *CLF* regulatory sequences (*CLFmini:SWN*) (Figure 11, B). Whereas a construct of *CLF* CDS driven by a *CLF* minigene (*CLFmini:CLF*) fully complemented in five out of five transgenic lines (Figure 11, B). The complementation was more robust in T3 than in T2 generation (data not shown) indicating that the allelic dose also played a role in this experimental setup. However, the expression of the *CLF* CDS from a *SWN* minigene construct resulted into a less robust complementation than the expression of the *CLFmini:CLF* construct and only one out of seven *SWNmini:CLF* transgenic lines could partially complement (Figure 11, B). Expression of the transgene in the transgenic lines measured by absolute qRT PCR using plasmid dilution standards roughly reflected the complementation (Figure 11, B, C, D). While the genomic constructs *GFPgCLF* and *mCherrygSWN* had a more similar expression profile to Col-0 WT, the minigene constructs varied more significantly.

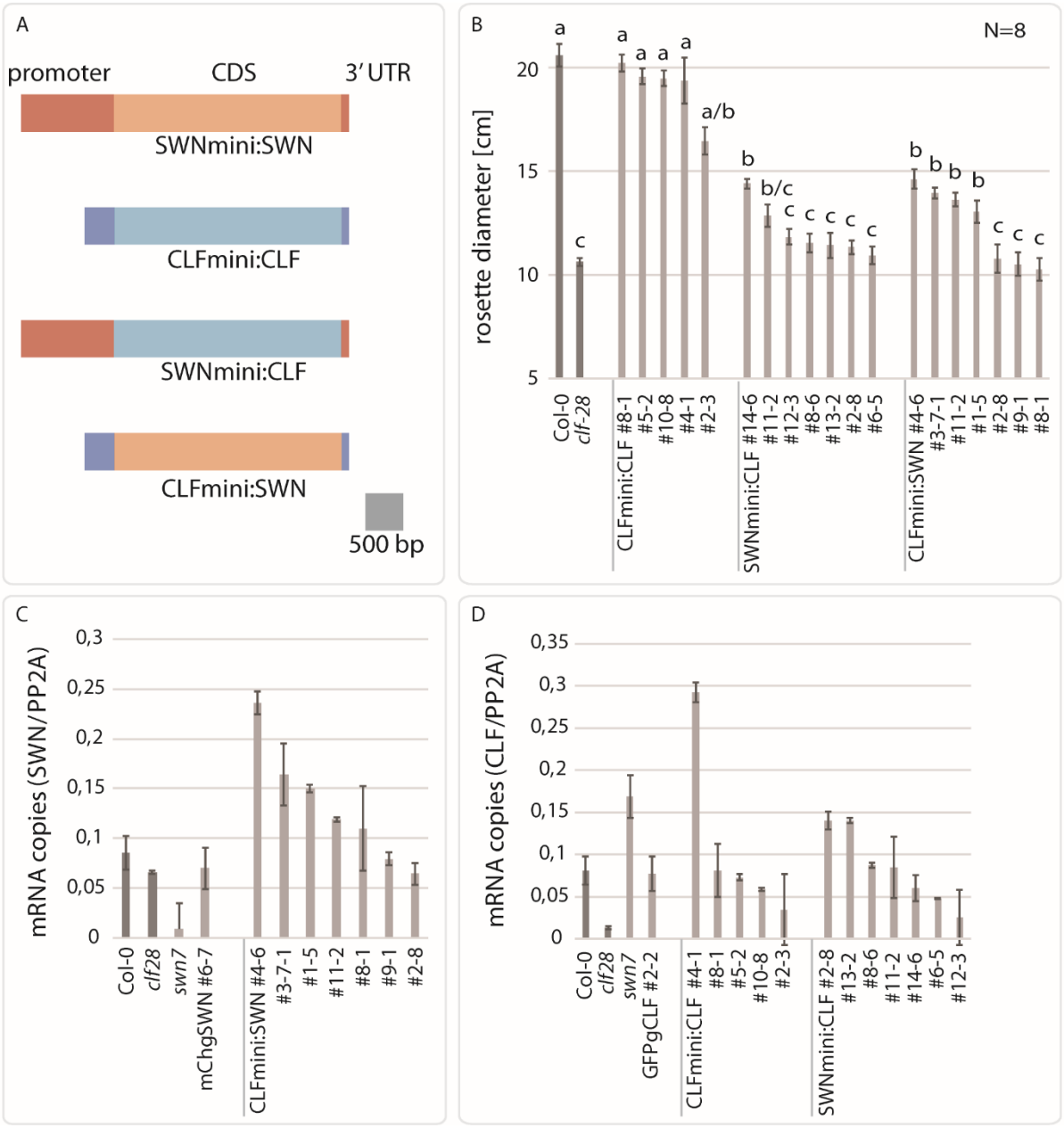


Figure 11 CLF and SWN coding sequence underlies functional divergence *in vivo*.

(A) Illustration of promoter-swap minigene constructs used in B.

(B) Complementation analysis of promoter-swap transgenic lines in *clf28* mutant background. Complementation measured by rosette diameter at first flower opening of independent T3 transgenic lines in comparison to *clf28* mutant and Col-0. Error bars indicate standard error of eight individuals per transgenic line or reference (N=8). Two-tailed ANOVA using Holm-Sidak correction. Letters (a, b, c) indicate three significance groups ($p<0.05$).

(C) Quantitative real-time PCR (qRT PCR) of references (Col-0, *clf28*, *swn7*), transgenic line *mChgSWN* #6-7 and individual *CLFmini:SWN* transgenic lines using specific primers to detect *SWN* mRNA. Absolute mRNA quantification of SWN normalised to PP2A and plasmid dilution standard of *PP2A* and *SWN* CDS (mRNA copies *SWN/PP2A*). Three technical replicates of approximately twenty 10 day-old, plate grown seedlings per transgenic line or reference. Three technical replicates of qRT PCR. Error bars are standard error.

(D) Quantitative real-time PCR (qRT PCR) of references (Col-0, *clf28*, *swn7*), transgenic line *GFPgCLF* #2-2, individual *CLFmini:CLF* and *SWNmini:CLF* transgenic lines using specific primers to detect *CLF* mRNA. Absolute mRNA quantification

of *CLF* normalised to *PP2A* and plasmid dilution standard of *PP2A* and *CLF* CDS (mRNA copies *CLF/PP2A*). Three technical replicates of approximately twenty 10 day-old, plate grown seedlings per transgenic line or reference. Three technical replicates of qRT PCR. Error bars are standard error.

4.3.2 Signatures of diversification in N-terminus

The full complementation of the *CLF* minigene constructs clearly highlighted the functional differences of *CLF* and *SWN in vivo*. To specify which part of the *CLF* and *SWN* CDS encodes the functional difference of the protein, domain swaps were made (Figure 12, A). The CDS of the C5 domain of SWN protein, which was shown to bind EMF/VRN2/FIS2 in a yeast two-hybrid interaction study (Chanvivattana et al., 2004), was swapped with the endogenous CLF protein C5 domain CDS and the altered CDS was expressed from a *CLF* minigene in the *clf28* mutant background (Figure 12, A). Most transgenic lines showed WT or near-WT phenotype overall and regarding the length of their sixth leaf after germination ((Figure 12, A). In conclusion, the SWN C5 domain probably takes redundant role in a CLF protein *in vivo*.

In another domain swap experiment, approximately one third of the total CDS was exchanged at 5' end of *CLF* and *SWN* CDS (N-terminal swap). While the construct harboring the CLF N-terminus (*CLFmini:AF*) complemented significantly to near WT-level in two out of three transgenic lines, the *SWN* N-terminal constructs (*CLFmini:BF*) did not complement (Figure 12, C). In summary, the CLF N-terminus likely comprises complementation capacity that is unmatched by SWN N-terminus.

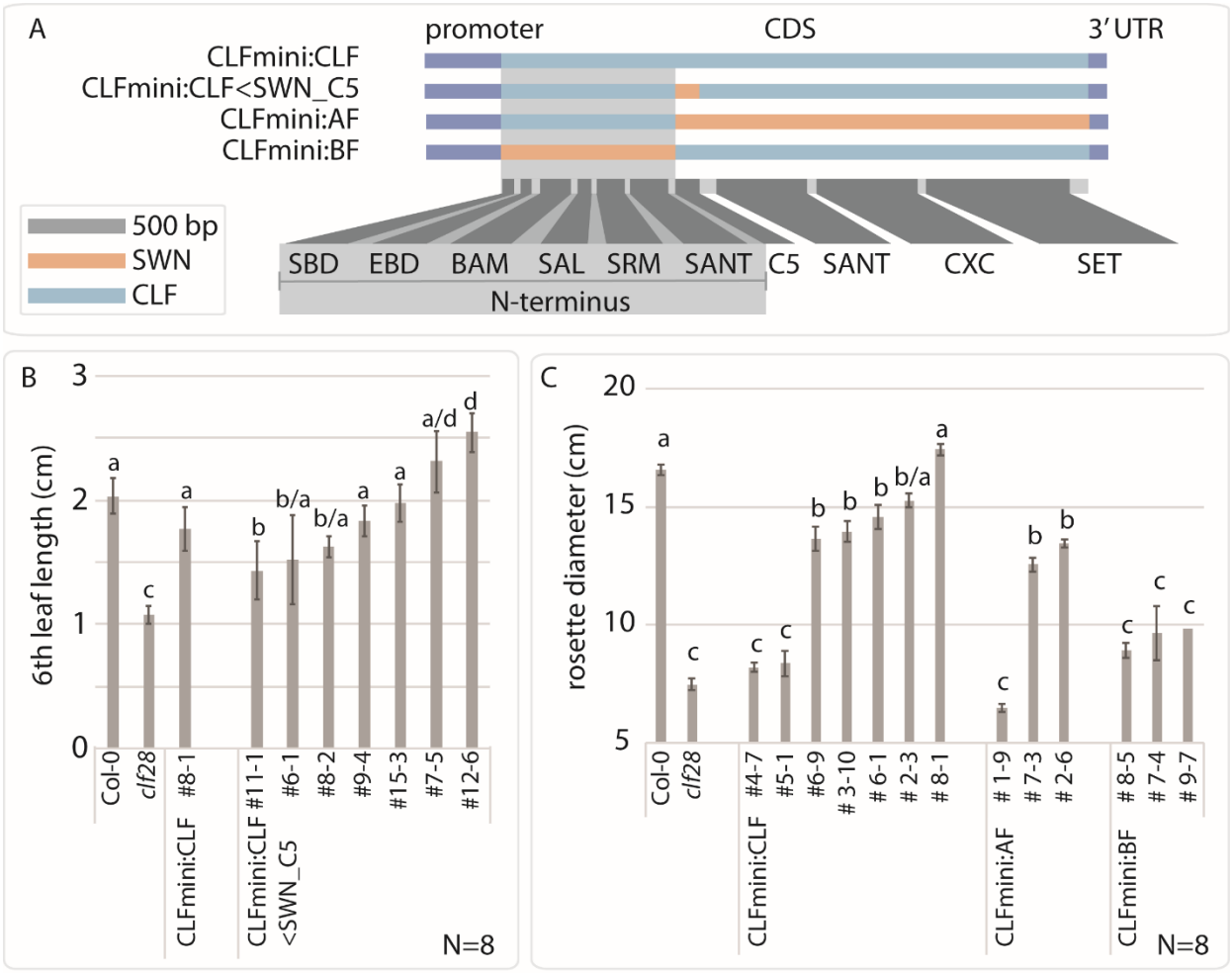


Figure 12 Amino (N)-terminus of CLF comprises CLF complementation capacity.

(A) Illustration of domain-swapped transgenic constructs used in B and C. *CLFmini* regulatory sequences (promoter and 3' UTR) were fused with *CLF* CDS (CLFmini:CLF). Based on CLFmini:CLF, *CLF* CDS with endogenous C5 domain that was replaced by SWN C5 domain (CLFmini:CLF<SWN_C5), *CLF* CDS with endogenous Carboxyl (C)-terminus that was replaced by SWN C-terminus (CLFmini:AF) and *CLF* CDS with endogenous N-terminus that was replaced by SWN N-terminus (CLFmini:BF). N-termini harbor the SANT binding domain (SBD), EED binding domain (EBD), beta-addition-motif (BAM), SET domain activation loop (SAL), stimulation response motif (SRM) and the first SANT domain. C-terminus harbors the VEFS binding domain (C5), second SANT domain, CXC domain and SET domain.

(B) Complementation analysis of individual CLFmini:CLF<SWN_C5 transgenic lines in *clf28* mutant background. Complementation measured by sixth leaf length of independent T3 transgenic lines in comparison to *clf28* mutant, Col-0 and CLFmini:CLF (*clf28* background) control. Error bars indicate standard error of eight individuals per transgenic line or reference (N=8). Two-tailed ANOVA using Holm-Sidak correction. Letters (a-d) indicate four significance groups ($p<0.05$).

(C) Complementation analysis of individual CLFmini:AF and CLFmini:BF transgenic lines in *clf28* mutant background. Complementation measured by rosette diameter at first flower opening of independent T3 transgenic lines in comparison to *clf28* mutant, Col-0 and CLFmini:CLF (*clf28* background) control. Error bars indicate standard error of eight individuals per transgenic line or reference (N=8). Two-tailed ANOVA using Holm-Sidak correction. Letters (a-c) indicate four significance groups ($p<0.05$).

To better understand the protein function of the CLF and SWN N-termini, an alignment was made with the human EZH2 N-terminus harboring domains annotated by atomic structure resolution. Due to a relatively poor conservation in the N-terminus between human EZH2 and plant EZH2 homologs, the EZH2 homolog of Physcomitrella, *PpCLF*, was used as an intermediate in the alignment of CLF and human EZH2 (data not shown). Subsequent alignment of CLF and SWN identified the following N-terminal domains: SBD, EBD, BAM, SAL, SRM and SANT. All domains of EZH2 were aligned and precisely annotated using this strategy (Figure 13). Overall, the CLF and SWN amino acid percentage identity (PID) was determined at 44.6% using ClustalOmega algorithm. In contrast, the CLF and SWN N-termini only contribute 36.5% to that PID and the C-termini contribute 52.7%. In summary, the CLF and SWN N-termini are diverged on sequence level as well as functionally, whereas the C-terminus is more conserved and functional divergence is unclear.

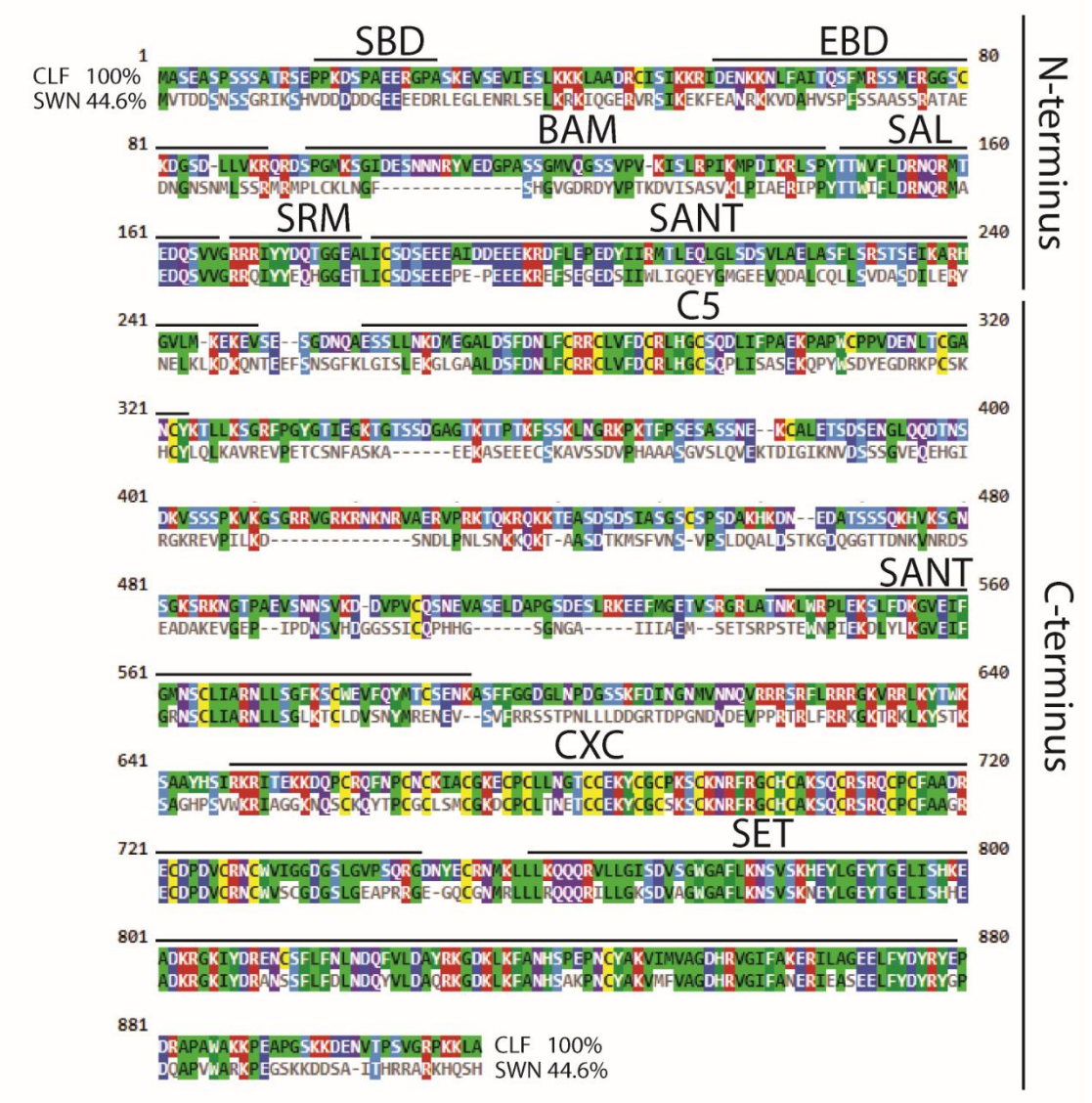


Figure 13 Alignment and domain annotation CLF and SWN protein. SANT binding domain (SBD), EED binding domain (EBD), beta-addition-motif (BAM), SET domain activation loop (SAL), stimulation response motif (SRM), first SANT domain, VEFS-binding domain (C5), second SANT domain, CXC domain and SET domain. In this study, amino (N)-terminus was defined as amino acids 1-257 of CLF protein sequence.

4.4 Evolutionary context of CLF and SWN functional divergence

4.4.1 CLF and SWN homologs are found in all Angiosperms

A phylogenic analysis was made to see how general might be the functional differences that were observed on CLF and SWN *in vivo* and *in vitro*. The SET domain was strongest conserved of all CLF and SWN domains. Therefore, 100 amino acids in the SET domain of CLF and SWN homologs were chosen to compute a maximum likelihood phylogenetic tree using over 20 different species of various clades (monocots, dicots , Brassicaceae, core Brassicaceae, mosses, lycophytes, gymnosperms, green algae) (Figure 14). The tree showed a clear distinction between CLF- and SWN-

clade. Even the most basal angiosperm species included, *Amborella trichopoda*, showed this distinction. In contrast, moss, lycophyte and gymnosperm species generally only possess one E(z) homolog and these homologs locate roughly centrally between the CLF- and SWN-clade. Similarly, green algae also possess only one E(z) homolog that did not cluster together with either of the two clades. Homologs of metazoans and green algae form a very distant outgroup, which did not branch off at mosses, which are the closest relatives in this tree. The low bootstrap value of 17 indicated for a relatively high uncertainty of this branch point. Similarly, MEA branched off from the distinct SWN clade formed by the Brassicaceae and located at the end of a relatively long branch. MEA homologs were only identified in species belonging to the core Brassicaceae, whereas the basal Brassicaceae specie *Aethionema arabicum* only possessed two homologs, which individually appeared in the distinct CLF- and the Brassicaceae SWN-clades. Overall, the phylogenetic tree reflected well the evolutionary history of land plants. Likely, the ancient duplication gave rise to the distinct CLF- and SWN-clades of the green lineage. A CLF and SWN duplication might have coincided with the divergence of angiosperms. A duplication of SWN that gave rise to MEA likely coincided with divergence of core Brassicaceae.

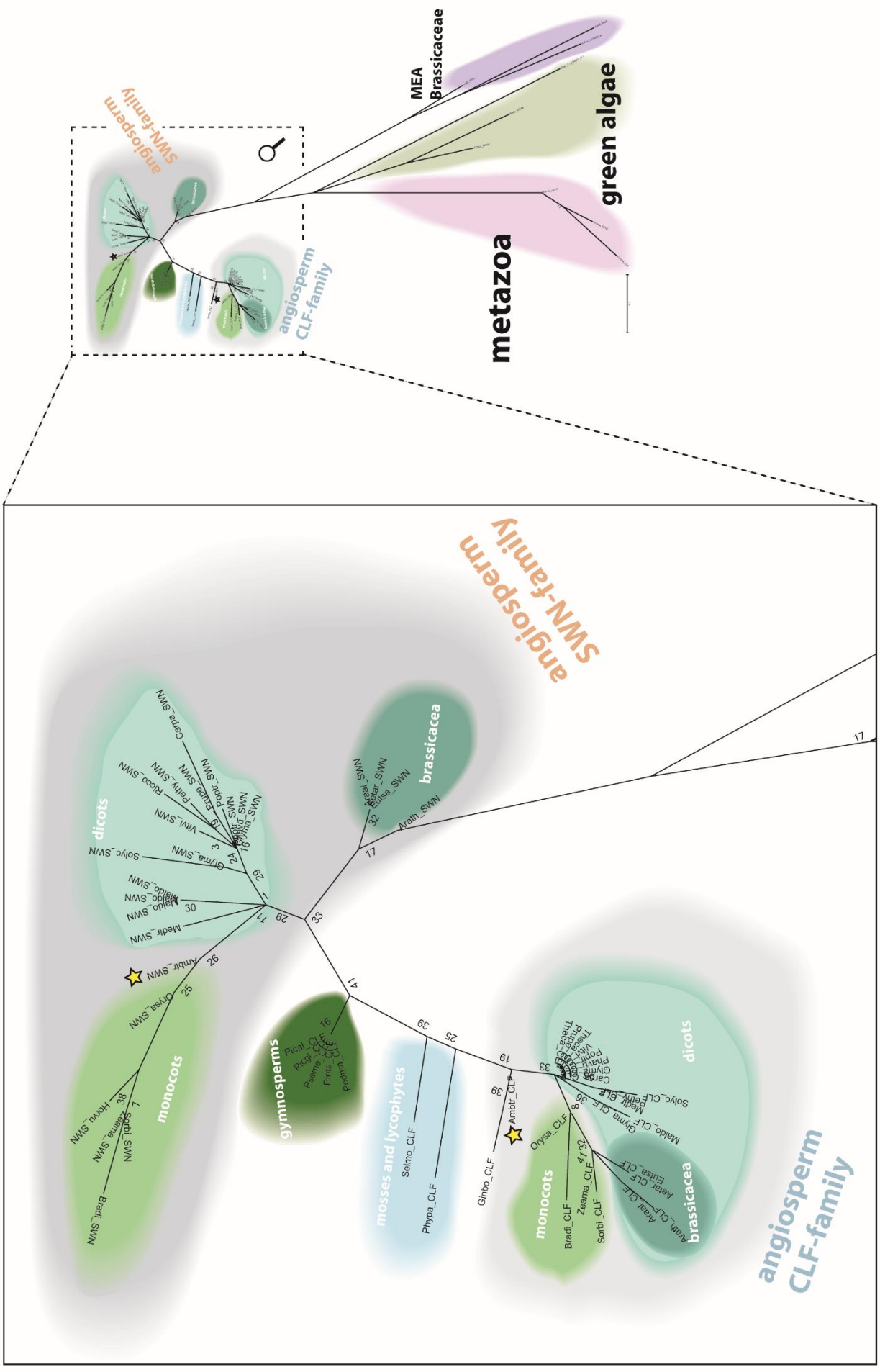

MEA
Brassicaceae
green algae
metazoa
angiosperm
SWN-family
angiosperm
CLF-family
angiosperm
SWN-family
dicots
brassicacea
monocots
gymnosperms
mosses and lycophytes
monocots
dicots
brassicacea
angiosperm
CLF-family

Figure 14 CLF duplication likely occurred at the base of Angiosperm evolution and occurred after the divergence of Gymnosperms, mosses and lycophytes.

Maximum likelihood phylogenetic tree (unrooted), which is based on highly conserved 210 amino acids in SET domain of CLF and SWN protein homologs. Included were representative plant species in diverse phyla (e.g. mosses, lycophytes, gymnosperms and angiosperms). Relatively more angiosperm phyla (e.g. monocots, dicots, Fabaceae, Solanaceae, Brassicaceae and core Brassicaceae) were included to increase resolution in this phylum. Metazoa and green algae were included as outgroup. Alignment performed using ClustalW. Bootstrap values of 1000 stochastic simulations is shown if below 50. Stars mark CLF and SWN paralogs of the basal Angiosperm *Amborella trichopoda*.

4.4.2 Cross-species complementation of *CLF* and *SWN* homologs is masked by poor expression of transgene

The proteins CLF and SWN belong to an ancient gene clade in the angiosperms and this study highlighted their functional divergence. To investigate whether these differences are conserved in the green lineage, a cross-species complementation experiment in Arabidopsis *clf28* mutant was performed. The CDS of *CLF* and *SWN* homologs of selected plant species was cloned. Included in the analysis were the moss Physcomitrella (Pp), the monocot species *Brachypodium distachyon* (Bd), the dicot species *Solanum lycopersicum* (Sl), the basal Brassicaceae species Aethionema (Ae) and the perennial core Brassicaceae species *Arabis alpina* (Aa). Their *CLF* and *SWN* CDS was expressed under the control of a *CLF* minigene construct (Figure 15, A). Despite full complementation of rosette diameter of Arabidopsis *CLF* (*AtCLF*) and the expected, mild complementation of Arabidopsis *SWN* (*AtSWN*), all other CLF and SWN homologs complemented poorly in T3 generation (Figure 15, B, C). Two out of five lines expressing the Aa *CLF* homolog (*AaCLF*) complemented significantly, but the complementation capacity still did not reach WT-level (Figure 15, B). Although, a mild increase of rosette diameter was observable in transgenic lines expressing Ae *CLF* (*AeCLF*) (2 out of 10 lines) (Figure 15, B), Bd CLF (*BdCLF*)(2 out of 11 lines) and Pp *CLF* (*PpCLF*) (5 out of 7 lines) (Figure 15, C), these were not significantly different from *clf28* mutant. In a previous T2 generation experiment, a similar trend was observably and even milder complementation effects were present. The expression of the transgenes was tested using qRT PCR, which showed that also *mRNA* levels were below detection range (data not shown). In conclusion, the expression of *CLF* and *SWN* homologs of other species expressed from a *CLF* minigene could only mildly affect the rosette diameter phenotype and significant differences could only be detected in a few lines, which expressed homologs most closely related to Arabidopsis CLF protein sequence (Figure 15, B).

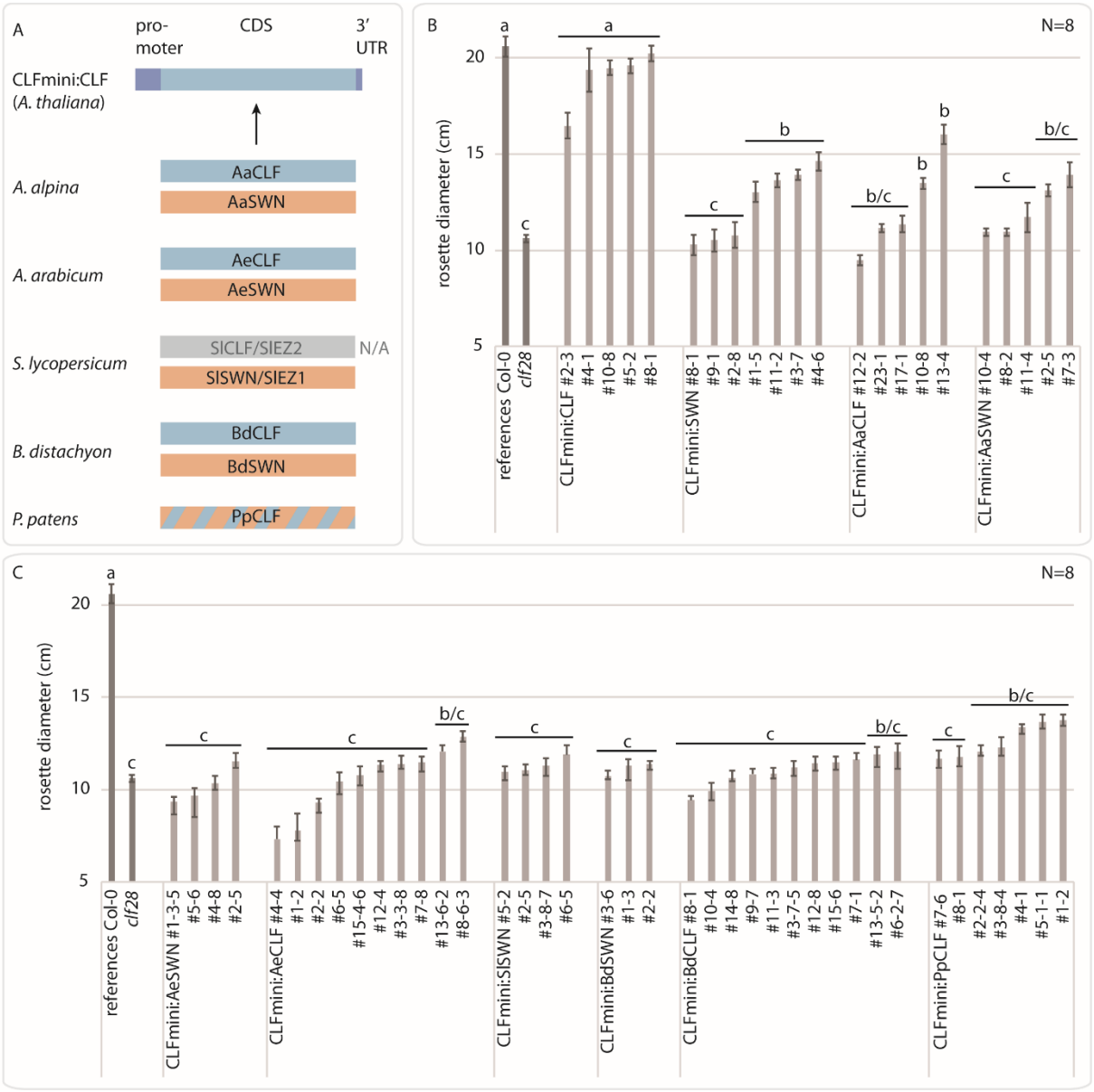


Figure 15 Cross-species complementation experiments highlights strong functional divergence of *Arabidopsis thaliana* CLF and SWN homologs from their homologs in close and distantly related plant species.

(A) Illustration of cross-species transgenic constructs used in B and C. *CLFmini* regulatory sequences (promoter and 3' UTR) were fused with *CLF* and *SWN* CDS of diverse plant species. Species included in the analysis were: moss *Physcomitrella patens* (PpCLF), grass *Brachipodium distachyon* (BdCLF, BdSWN), tomato *Solanum lycopersicum* (SlSWN/SlEZ1), basal Brassicaceae *Aethionema arabicum* (AeCLF, AeSWN)and perennial core Brassicaceae *Arabis alpina* (AaCLF, AaSWN).The second tomato homolog SlCLF/SlEZ2 was not available (N/A) because of cloning difficulties.

(B) Complementation analysis of transgenic lines expressing *A. arabicum* and *A. alpina CLF* and *SWN* homologs compared to *A. thaliana CLF* and *SWN*. Complementation measured by rosette diameter at first flower opening of independent T3 transgenic lines in comparison to *clf28* mutant and Col-0. Error bars indicate standard error of eight individuals per transgenic line or reference (N=8). Two-tailed ANOVA using Holm-Sidak correction. Letters (a-c) indicate three significance groups ($p<0.05$).

(C) Complementation analysis of transgenic lines expressing *S. lycopersicum*, *B. distachyon* and *P. patens CLF* and/or *SWN* homologs compared to *A. thaliana CLF* and *SWN*. Complementation measured by rosette diameter at first flower opening of independent T3 transgenic lines in comparison to *clf28* mutant and Col-0. Error bars indicate standard error of eight

individuals per transgenic line or reference (N=8). Two-tailed ANOVA using Holm-Sidak correction. Letters (a-c) indicate three significance groups ($p<0.05$).

5 Discussion

5.1 Redundancy of CLF and SWN *in vivo*

5.1.1 *CLF* and *SWN* expression domains overlap

The GUS reporter analysis confirmed the overlap of *CLF* and *SWN* tissue-specific expression (4.2.1). Furthermore, fluorescence-protein fusion data confirmed that both proteins localise to the nucleus (4.2.1). It is clear from my analysis, and that of others (Chanvivattana et al. (2004), that *CLF* and *SWN* predominantly differ in the coding region. This is corroborated by the clear difference in complementation of *CLF* and *SWN* CDS when expressed from the same regulatory sequence (*CLFmini*) in the *clf28* mutant background (4.3.1). Nonetheless, the promoter swap could not verify a possible redundancy of the regulatory sequences of *CLF* and *SWN*, since *SWNmini:CLF* did not complement well and expression of the transgene was random between transgenic lines (4.3.1). One explanation is that repressive elements are present in the 3' UTR of the *SWNmini* construct that are missing in the *CLFmini* construct and the *SWN* promoter GUS fusion. In addition, it is possible that subtle differences exist between the expression of *CLF* and *SWN* that either are beyond the resolution of GUS assays, e.g. cell cycle dependence, or not sufficiently analyzed here, e.g. developmental stage.

5.1.2 CLF and SWN are unequally redundant

QRT PCR in the complementation lines also showed that complementation of *clf28* mutant correlates with total *SWN* mRNA level (4.3.1). This suggests that increased *SWN* mRNA levels also result in increased protein levels, which can compensate for the absence of functional CLF protein. The inverse effect that decreased dose of *SWN* can affect the *clf* mutant can be observed in the *clf-/- swn+/-* mutant, which shows phenotypic enhancement. The single mutant phenotypes, as well as the double mutant phenotype of *clf* and *swn,* and the data presented here, are all in agreement with the definition of unequal genetic redundancy between paralogous genes (Briggs, Osmont et al., 2006). In this scenario, *CLF* contributes more activity (*sensu lato*) than *SWN* to the quantitative modulation of the Arabidopsis phenotype, e.g. rosette diameter, leaf curling. The absence of a strong *swn* mutant phenotype suggests that the observed molecular differences between CLF and SWN protein in *ex vivo* experiments are in fact less important *in vivo*. I speculate that the mild phenotypic defects of *swn* mutants would be more pronounced in a non-laboratory setting and the presence of *SWN* would therefore be adaptive. To confirm this statement, it would be necessary to measure fitness of *swn* mutant plants in field conditions.

5.1.3 CLF protein level is fixed

Despite overexpression of *CLF* in some of the *CLFmini:CLF* transgenic lines, they have a WT-like phenotype and show no developmental defects (4.3.1). This is surprising, because overexpression alleles of human EZH2 are classified as oncogenes (Yamaguchi & Hung, 2014). Instead,

complementation of the *clf* mutant by *CLF* follows the principle of 'all-or-nothing'. One possible explanation could be that the CLF protein level is tightly regulated. Indeed, CLF was shown to be regulated through binding of UPWARD CURLY LEAF 1 (UCL1), which functions as a recruiter of the proteasomal degradation pathway (Jeong, Roh et al., 2011). Another possibility is that higher than normal CLF protein levels are toxic and are selected against in my experiments. Overexpression in tobacco indeed showed that plants react with necrosis to attempted overexpression (4.1.1.1). As a next step, it would be essential to test the CLF and SWN protein levels in WT, the *clf* and *swn* single mutants, and my transgenic lines. This could be achieved either by using specific antibodies or by using quantitative MS. In summary, overexpression of *CLF* results in qualitative complementation, while both increased and decreased expression of *SWN* results in a continuous, qualitative variation of the *clf* phenotype.

5.1.4 CLF and SWN protein level low

The low accumulation of CLF and SWN fluorescence fusion protein and the strong GUS signal in callus cells appear contradictory (4.2.1). However, the low protein decay of GUS protein and faster decay of fluorescence fusion protein may pose an explanation. Zografou (2013) showed that CLF and SWN translational fusions could not express to high levels despite usage of various tags, overexpression and/or proteasome inhibitors. I therefore agree with the conclusion that the endogenous CLF and SWN protein levels are relatively low and tightly regulated.

However, this contradicts the current literature, which reports higher fluorescent protein accumulation levels of both CLF and SWN translational fusions to fluorescent proteins (Wang, Tyson et al., 2006). My CLF and SWN fusion protein lines are N-terminal and I present data showing full complementation of all *GFPgCLF* transgenic lines (4.2.1). The WT-like *CLF* expression level in the *GFPgCLF* #2-2 transgenic line (4.3.1) further indicates that the mRNA expressed from the transgene is processed efficiently and that the fusion protein is fully functional. The complementation capacity of *mCherrygSWN* construct was not analyzed here. Next steps should therefore include a complementation analysis in the *clf swn* double mutant background or complementation analysis of the known, mild phenotype of the *swn* single mutant, e.g. premature trichome formation (Xu et al., 2015, Xu et al., 2016), shorter root elongation zone (de Lucas et al., 2016). Nevertheless, it is surprising that despite full complementation, the protein is barely detectable. In contrast, published lines are C-terminal and complementation data is not available (Wang et al., 2006). Additionally, structural data of human PRC2 allows for the speculation that a large C-terminal fluorescence protein could possibly occlude the catalytic SET domain of CLF or SWN. It would be crucial to make a thorough complementation analysis and compare protein levels between the published transgenic lines and mine.

5.2 Divergence of CLF and SWN in PRC2

5.2.1 Plant PRC2 - truly a tetramer?

Based on CoIP-MS, a previous study suggested that CLF and SWN proteins are mutually exclusive in PRC2 (Liang et al., 2015). PRC2 structural data also does not support binding of two E(z) homologs. My competitive pulldown as well as the purification of the single complexes demonstrated that SWN and CLF can effectively pull down EMF2 with the same efficiency and can also pull down MSI1 and FIE (4.2.1). Liang et al. (2015) and Derkacheva et al. (2013) found SWN to be more abundant in their CoIP-MS experiments when using EMF2 or MSI, and ALP1, a known PRC2 interactor, as a bait, respectively, and a 50mM Tris pH 7.5 buffer. Similarly, I found CLF-PRC2 to be less stable in a 50mM Tris pH 7.5 buffer (4.1.3.3), while purification of CLF- and SWN-PRC2 is comparable at pH 8 (). This difference of CLF- and SWN-PRC2 under similar purification conditions not only indicates that CLF-PRC2 might generally be less stable, but also gives an explanation for the apparent higher abundance of SWN in Liang's and Derkacheva's experiments. A more thorough interrogation of the *in vivo* abundance of CLF- and SWN-PRC2 is needed and, here, buffer artifacts ought to be minimised.

Based on homology, I earlier proposed a model of Arabidopsis PRC2 according to which the EMF2:CLF/SWN dimer, referred to as the catalytic core, binds the two monomeric WD40 accessories MSI1 and FIE in opposite sites of the complex (Figure 1) (F148rderer et al., 2016). Based on the work presented here I would like to revisit this proposed model.

The iBAQ values of purified CLF- and SWN-PRC2 (4.1.1.3) can be used to approximate the relative abundance of individual components and might serve to get insight into PRC2 stoichiometry after the purification and possibly *in vivo*. IBAQ comparison showed that the WD40 accessory MSI1 exceeds a 1:1 stoichiometry with the catalytic core and rather purifies in 2:1 stoichiometry (MSI1:catalytic core) (4.1.1.3). One possibility is that MSI1 is 'sticky' and has unspecific binding to the catalytic core. Another possibility is that the binding is indeed specific and MSI1 can take the place of FIE in variant PRC2s.

The iBAQ values recorded of purified PRC2 also showed that the other WD40 accessory FIE has low abundance in the total CLF- and SWN-PRC2 sample and the bands 3-5 of the native gel separation (4.1.1.3). Based on the poor association of FIE with the catalytic core, I propose that FIE is not an obligatory component of specific PRC2 variants *in vivo*. It is however also possible that the seemingly low abundance of FIE in my PRC2 purifications is (1) due poor FIE expression in insect cells or (2) due to underestimation by the iBAQ algorithm. Although MS was not performed on the crude extract, a clearly lower expression of FIE (1) was not found on western blot (data not shown). The error of the iBAQ (2) could be reduced by using heavy-labeled peptide standards of FIE during MS. Finally, I

speculate that the WD40 protein FIE can either be absent in a variant PRC2 or even that FIE can be replaced by the WD40 protein MSI1, which purified in a 2:1 stoichiometry with the catalytic core.

Clearly, a more thorough investigation of plant PRC2 stoichiometry is needed. Subsequent to the Strep purifications presented here, size-exclusion chromatography would be necessary to separate the different sub-complexes that the blue native analysis suggested. Recent advances in cryo-electronmicroscopy technology would make it possible to analyze individual fractions recovered in the size-exclusion chromatography with the goal to differentiate between PRC2 compositions and to resolve variant PRC2 structure.

CoIP-MS using EED as a bait to pull out PRC2 from HeLa cells found that EZH2, SUZ12 and EED form a trimer *in vivo* with 1:1:1 stoichiometry, while the MSI1 homolog NURF55 purifies in in a 0.5:1 ratio with this trimer (Smits, Jansen et al., 2013). In contrast, EZH2 or SUZ12 as a bait recovered stoichiometric trimers containing EZH2, SUZ12 and EED, or SUZ12, EED and NURF55, respectively (Oliviero, Brien et al., 2016). Probably, human PRC2 is not an obligate tetramer *in vivo* and interrogation of multiple individual subunits has proven crucial in dissecting variant PRC2 stoichiometry *in vivo*. In Arabidopsis, analysis of PRC2 stoichiometry is missing. However, CoIP-MS datasets exist for EMF2, MSI1, ALP1, CLF (Derkacheva et al., 2013, Liang et al., 2015) and each component showed different strongest associations. In addition, PcG mutants in plants generally have distinct morphological and molecular phenotypes, which also suggests that a static tetrameric PRC2 is not always present *in vivo*.

5.2.2 A possible role of PRC2 during DNA replication

Biochemical evidence from animal PRC2 suggested activation of PRC2 by H3K27me3 through EED binding (Margueron et al., 2009). My comparison of the cooperativity of CLF and SWN-PRC2 reaction kinetics, which is indicated by the Hill coefficient, suggests that CLF-PRC2 is stimulated by oligonucleosomes (4.1.5). Based on the homology to human PRC2, it is possible that H3K27me3 acts as a stimulator of CLF-PRC2 activity but not of SWN-PRC2.

It is intriguing question whether CLF, SWN and LHP1 might play individual roles in the initial deposition (nucleation), spreading and maintenance during DNA replication of H3K27me3. Indeed, CLF and LHP1 are essential in spreading and nucleation of H3K27me3, whereas SWN is only present at nucleation sites (Yang et al., 2017). Moreover, *clf* and *lhp1* mutants have widely overlapping molecular phenotypes, whereas *swn* mutants have a mild molecular phenotype (Wang et al., 2016). LHP1 was also suggested to function as a mediator between SWN-PRC2 and the DNA replication machinery (Zhou et al., 2017b). CLF and LHP1 on the other hand were shown to interact with CAF-complex, which is important for H3.1 repositioning after DNA replication and defects lead to H3K27me3 depletion (Jiang 2017). Indeed, I also found insect CAF1 co-purifies with both CLF- and

SWN-PRC2 at low quantities (4.1.1.3), indicating that SWN and CLF might share CAF1 binding domains.

I propose that LHP1 has dual functions, both as a recruiter of CLF-PRC2 to the nucleation site where it is essential in spreading of H3K27me3 and as a stimulator of SWN-PRC2 at the replication fork during S-phase. In both cases, the LHP1-H3K27me3 interaction module could function as a stimulator of PRC2 activity. HMTase assays could serve to test the stimulatory effect of H3K27me3 peptides and LHP1 on SWN- and CLF-PRC2 activity.

5.2.3 CLF can hypermethylate *in vivo* and is more active on oligonucleosomes *in vitro*

My biochemical analysis is the first demonstration of SWN catalytic activity *in vitro* (4.1.3; 4.1.4; 4.1.5). It is clear from this data that CLF-PRC2 and SWN-PRC2 are both able to act as methyltransferases of H3K27me1, H3K27me2 and H3K27me3. Generally, SWN-PRC2 activity is lower compared to CLF-PRC2 when provided with oligonucleosomes as a substrate (4.1.5). This is in agreement with the morphological phenotypes of the *clf* and *swn* single mutants.

One goal of this study was to explain the molecular phenotype of *clf* and *swn* mutants that was described by Zografou for CLF-dependent genes (1.7). The increased H3K27me3 level in *swn* mutants and the decreased H3K27me3 in *clf* mutant suggests that CLF and SWN compete for these common targets genes. This could also mean that CLF and SWN compete for other PRC2 subunits FIE, MSI1 and EMF2 at these genes. However, my competitive pulldown indicated that indeed CLF and SWN have similar affinity towards MSI1, FIE and EMF2 (4.1.1.3), which means that they could essentially replace each other *in vivo*. Based on the activity of CLF-PRC2 on oligonucleosomes, a plausible explanation might be that in the absence of SWN, CLF would hypermethylate available histone tail substrates.

Nevertheless, comparing the catalytic activity of CLF and SWN-PRC2 on peptide substrate (4.1.5) showed that the dynamic relationship of CLF and SWN is more complex than previously anticipated and is not always consistent with the morphological and molecular phenotypes.

5.2.4 SWN catalyzes conversion of H3K27me2-me3 and CLF catalyzes H3K27me2 *de novo* deposition

SWN methylates H3K27me2 with 16-times higher specificity than CLF (4.1.5). Western blot analysis and genome wide ChIP showed that *swn* mutants show overall little change in H3K27me3 levels and *clf* mutants show more drastic change in H3K27me3 (Lafos et al., 2011, Zografou, 2013). Clearly, more genome-wide data sets are needed that provide annotations of H3K27me1 and H3K27me2 in *clf* and *swn* single mutants. It is possible that the catalysis of me2-me3 by CLF-PRC2 depends on additional factors such as its unique interactors VERNALIZATION 5 (VRN5) and VERNALIZATION-LIKE 1 (VEL1) (Liang 2015), while SWN-PRC2 exhibits this activity independently. In this study, only EMF-PRC2 was

examined. Alternatively, VRN2 can act in the sporophytic PRC2 (Mozgova & Hennig, 2015a). It is possible that VRN-PRC2 has different biochemical properties with regard to H3K27me0-me3 conversion and that *in vivo* both EMF-PRC2 and VRN-PRC2 have different roles depending on CLF or SWN incorporation.

In conclusion, H3K27me2-me3 conversion by SWN-PRC2 might require prior deposition of H3K27me2 by CLF-PRC2. If so, H3K27me0-me2 deposition and H3K27me2-me3 conversion would be temporally and spatially separated *in planta*.

5.2.5 A possible role of CLF and SWN in cell division

To answer the question of spatial separation of H3K27me2 deposition and H3K27me2-me3 conversion, it would be necessary to dissect the function of CLF- and SWN-PRC2 in the course of cellular differentiation. Tissue-specific sampling is a major bottleneck in Arabidopsis research. However, the distinct defects in the root differentiation and the root meristematic zone mutations of *swn* and *clf*, respectively, (de Lucas et al., 2016) might pose an opportunity to overcome that obstacle. Hand-dissection or flow cytometry would be feasible techniques to separate the root meristematic zone from the root differentiation zone (Li, Yamada et al., 2016). These samples could subsequently analyzed using quantitative mass-spectrometry or western blot to measure overall levels of H3K27me2 and H3K27me3 in *clf* and *swn* single mutants.

5.2.6 H3K27me3 spreading

In animal models, PRC2 was shown to form higher order nuclear foci and multimeric states of PRC2 were found *in vitro* and *in vivo* (Casanova, Preissner et al., 2011, Davidovich, Goodrich et al., 2014, Margueron et al., 2008, Tie, Prasad-Sinha et al., 2003). Native gel separation of my purified PRC2 also suggested that CLF- and SWN-PRC2 form three distinct bands with different molecular weight despite similar composition (4.1.1.3). In addition, similar to my findings, H3K27me2 is the dominant reaction product of animal PRC2, whereas H3K27me3 requires H3K27me2 as a substrate. A recent study proposed a model of H3K27me3 spreading (Oksuz 2018): Upon PRC2 recruitment, H3K27me2-me3 conversion takes place within a nucleation site. Stimulated by its own reaction product, PRC2 moves in to subsequently spread H3K27me2 from this nucleation site and H3K27me3 spreading trails behind H3K27me2. It is possible that CLF and SWN have distinct roles in such a process. To test this, protoplasts could be made from *clf swn* double mutants that have either CLF or SWN expressing in an inducible fashion. To test the sequence of events upon induction, H3K27me2 and H3K27me3 levels at known PcG target regions, e.g. *AG*, *FLC* or *SEP3*, could be analyzed by ChIP.

5.2.7 H3K27me2, not only PRC2

H3K27me2 is a euchromatic mark in Arabidopsis and might be correlated with gene repression (Berry et al., 2017). While H3K27me2 is exclusively deposited by PRC2, evidence suggests other

mechanisms exist for H3K27me2: (1) H3K27me2 is retained in Physcomitrella when the single *CLF* homolog *PpCLF* is knocked out (Peremann 2016, Baumbusch 2001). (2) *clf swn* double mutants overaccumulate H3K27me2 (Lafos et al., 2011). It is therefore possible that a strong *swn* phenotype is masked by redundancy with an unknown H3K27 dimethylase (HMTase X). The identification of unknown enzymes that catalyze a specific PTM is a major challenge. It could be possible to identify this HMTase X in a categorical screen of all known Arabidopsis SET domain methyltransferases. Alternatively, a phage display assay or a CoIP-MS using an H3K27me2-probe could serve to identify novel readers, writers and erasers of H3K27me2.

5.2.8 Domains underlying different function

Previous studies suggested that CLF and SWN assemble in PRC2 through the interaction of their C5 domain with the VEFS domain of VRN2 and EMF2 (Chanvivattana et al., 2004). Additionally, the VEFS domain of Su(z) homologs is minimally required to confer catalytic activity toward H3K27me3 *in vitro* in an EED- and EZH2- containing complex (Yuan 2012). Indeed, I found that the C5 domain of SWN is functionally equivalent to the CLF C5 domain (4.3.2). This further strengthens the point, that CLF and SWN share similar affinity to PRC2 or at least, that the C5 domain is not responsible for potential differences. Exchanging the divergent N-terminus of CLF and SWN showed that the CLF N-terminus is sufficient to confer complementation capacity to a SWN catalytic moiety, whereas the inverse is less true (4.3.2). Guided by structural data and alignment, further exchanges within the N-terminus could be made to achieve a finer resolution. Indeed, in an initial analysis I could find significant differences of CLF and SWN with the BAM, EBD and SANT1L domains (4.3.2, summarised in Figure 16). It is possible that the higher efficiency of SWN for H3K27me2 substrate is orchestrated by the N-terminus domains, which hypothetically could have intimate contact and embrace FIE in a belt-like structure (see Figure 16). *Vice versa*, these domains could also underlie allosteric activation of CLF, as was suggested by the Hill coefficient. Again, these hypotheses could be tested in HMTase assays using domain swapped CLF and SWN protein combined with H3K27me3 peptide as a stimulator. In addition, CLF and SWN vary in the first amino acids N-terminal of the SBD and it is possible that these strongly affect processing and stability of the protein (N-end rule) (see Figure 16). The HMTase assay would also be an appropriate method to confirm functionality of the protein prior to plant transformation. It is possible, that the N-terminus of CLF and SWN also harbours the observed differences in catalytic efficiency as N-terminal domains are in intricate contact with the catalytic moiety of the C-terminus as well as with FIE, the plant homolog of the allosteric sensor unit EED (see Figure 16).

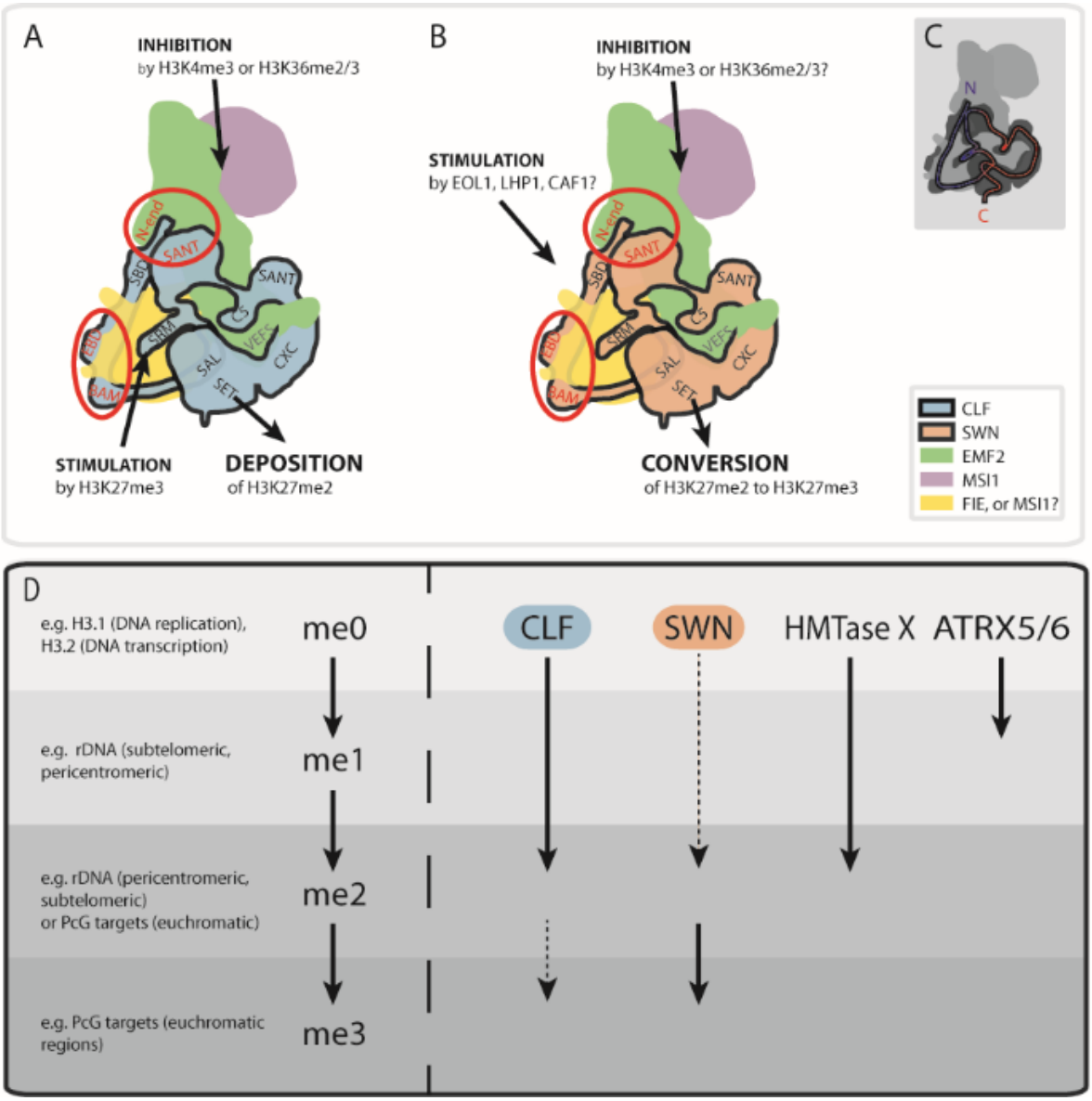


Figure 16 New working model based on findings of this study and literature.

(A) hypothetical structure of plant PRC2 containing CLF (A) or SWN (B), EMF2, FIE and MSI1 based on metazoan crystal and electron microscopy structures (Brooun et al., 2016, Ciferri et al., 2012, Jiao & Liu, 2015, Justin et al., 2016). Annotated are the domains of CLF and SWN in their proposed structural context (black) and the domains that are most divergent in primary sequence (red) in the functionally divergent N-terminus. H3K27me2 deposition by CLF is located in the SET domain, while SWN SET domain catalyzes H3K27me2 to H3K27me3 conversion. Activity of CLF and SWN might depend on allosteric activation by two main interactions of the remaining complex: allosteric inhibition by the MSI1/EMF2 module upon H3K4me3 or H3K36me2/3 binding (only shown for CLF-PRC2, (Schmitges et al., 2011)) and allosteric stimulation by the FIE/SRM module, which probably reacts to H3K27me3 (no direct evidence in case of plant PRC2). Hypothetically EOL1, LHP1 and CAF1 could have a stimulatory effect on SWN catalytic subunits, but the exact interaction points are unknown. (C) Direction of CLF and SWN amino acid chain in the model; N-terminal (blue), C-terminal (red). (D) Graphic illustration of processivity of H3K27 methylation states, me0, me1, me2 and me3 (light blue), which are mediated by CLF, SWN, ATRX5/6 and a hypothetical, unknown H3K27 dimethylase (HMTase X) in the Arabidopsis sporophyte. Me0, me1, me2 and me3 states have specific genomic accumulation and the underlying HMTases might also have specific genomic distributions.

5.3 Evolution of CLF and SWN in the green lineage

5.3.1 CLF and SWN duplication early in Angiosperm evolution

My phylogentic analysis and recently published data demonstrated the origin of the CLF and SWN gene clade in the Angiosperms (4.4.1) (Wang 2016), while Gymnosperms, lycophytes, mosses and green algae predate the duplication. Except for green algae, plant species with a single gene copy locate in between the distinct CLF and SWN gene clade (4.4.1) indicating that both CLF and SWN have subspecialised by the same extent from this ancestral state.

5.3.2 Out of context analysis of CLF and SWN homologs poses a bottleneck

The potential genetic redundancy between SWN and CLF raises the question of why a fully functional but dispensable gene copy is maintained. One possibility is that the current state is a transition towards gene copy loss. The alternative explanation is that the duplicate gene contributes to the overall activity (*sensu lato*) of the gene pair and the balance between the two is essential. However, this might not become apparent in a laboratory setting and, in this case, the mutant analyses would be misleading. I speculate that CLF and SWN in Arabidopsis could be in such a balanced relationship where CLF is generally more active, shows gene-specific recruitment, establishes the majority of H3K27me2 upon recruitment and is also able to convert H3K27me2-me3 in the course of differentiation (even if slower than SWN). SWN is present, but redundant with CLF in DNA replication-dependent H3K27me2- me3 conversion. However, given that MEA originates from a duplication in the core Brassicaceae, it is possible that SWN function in Arabidopsis is not representative for most plants in the green lineage.

To answer this question, the cross-species experiment was made (4.4.2). However, due to the great evolutionary distance, I suggest that complementation cannot be qualitative but is rather quantitative. This is consistent with the partial complementation of Arabidopsis *fie* mutant expressing *PpFIE* and the partial complementation of Physcomitrella *ppfie* mutants expressing *AtFIE* (Mosquna 2009). Indeed, the regulatory sequences used here were fully functional, as is visible from the *CLFmini:CLF* complementation (4.4.2; 4.3.1). However, the complementation is generally too mild to achieve statistical significance for more than 1-3 transgenic T3 lines with the number of tested lines here (4.4.2). To overcome the dependency of the dose of transgene expression and to get more qualitative result, overexpression driven by the constitutive 35S promoter could be used or sporophytic H3K27me3 null mutants *clf swn* could be transformed. Alternatively, analysis of many (>20) T1 plants could be made using the *CLFmini* promoter in a *clf28* background to get statistically testable results of the variation between transgenic lines.

Tomato *CLF* and *SWN* homologs express in different tissues, localise in different compartments of the nucleus when co-expressed and regulate a different set of target genes (Boureau et al., 2016,

How Kit et al., 2010). It is therefore possible, that species-specific subfunctionalisation of the proteins is underlying the poor heterologous complementation that I observed. E(z) and its metazoan homologs were reported to presumably act in different cellular processes other than H3K27 trimethylation, e.g. methylation of cytoplasmic proteins (He, Shen et al., 2012) and chromatin compaction by EZH1 (Margueron et al., 2008). Perhaps plant E(z) homologs have evolved to serve species-specific purposes and are adapted to the endogenous genome. Poor complementation in Arabidopsis could therefore be caused by the missing cellular context. However, my results presented here are not fully conclusive about whether or not the Arabidopsis CLF and SWN duplicate pair is rather an exception in the green lineage or the molecular differences described here are more general.

6 Conclusion and future perspectives

In planta, CLF and SWN display highly redundant behaviour at the resolution scale of this work: shared temporal and spatial expression, shared protein localisation, exchangeable non-coding sequences and similar affinity towards EMF2/FIE/MSI1 trimer. Although increased resolution scale might possibly bring undiscovered differences to light, the major difference most probably lies in the protein function.

This work is the first comparative biochemical analysis of CLF and SWN protein function. Further repetitions are needed of HMTase assays using methylated peptides to allow exact, statistically sound measurement of the specificity constant. The recovery of active PRC2 complex will allow further description of the biochemical properties of CLF and SWN in PRC context, e.g. structure determination and affinity measurement between CLF/SWN with other PRC2 components.

The biochemical analysis of CLF and SWN-PRC2 in this work focused on a mixture of different stoichiometric complexes. The recovered CLF and SWN complexes were scrutinised for comparability between CLF- and SWN-PRC2 purification. The PRC2 purifications in this work also showed that a tetrameric state of PRC2 might not be the only possible stoichiometry in which CLF and SWN assemble into PRC2. The existence and relevance of such variant PRC2 complexes *in vivo* will be the focus of future research.

The ability of SWN to better use H3K27me2 substrate and the fact that H3K27me2 accumulates in *clf swn* double mutants, suggest that H3K27me2 dynamics are highly relevant to achieve repression of target genes. In future, H3K27me2 should receive more attention. To initiate such research, genome-wide ChIP data using the specific Ab towards H3K27me2 will be instrumental (e.g. in *clf* and *swn* single mutants).

It remains enigmatic whether the fine-tuned balance of CLF and SWN, which is suggested by this work, is the general rule for CLF and SWN homologs of green plants. To fully answer this question, CLF and SWN homolog function would probably best be studied within the original species context.

7 Literature

Adrian J (2009) Transcriptional control of FLOWERING LOCUS T in Arabidopsis. In Universität zu Köln, Dissertation

Aigner A (2011) Genexpression in *E. coli* und Insektenzellen: Produktion und Reinigung rekombinanter Proteine. In Gentechnische Methoden: eine Sammlung von Arbeitsanleitungen für das molekularbiologische Labor, Jansohn M, Rothhämel S (eds) pp 407-427. Springer Spektrum

Alabert C, Barth TK, Reverón-Gómez N, Sidoli S, Schmidt A, Jensen ON, Imhof A, Groth A (2015) Two distinct modes for propagation of histone PTMs across the cell cycle. Genes & Development 29: 585-590

Antonysamy S, Condon B, Druzina Z, Bonanno JB, Gheyi T, Zhang F, MacEwan I, Zhang A, Ashok S, Rodgers L, Russell M, Gately Luz J (2013) Structural Context of Disease-Associated Mutations and Putative Mechanism of Autoinhibition Revealed by X-Ray Crystallographic Analysis of the EZH2-SET Domain. PLOS ONE 8: 84147

Arnaudo AM, Garcia BA (2013) Proteomic characterization of novel histone post-translational modifications. Epigenetics & Chromatin 6: 24

Baroux C, Pecinka A, Fuchs J, Schubert I, Grossniklaus U (2007) The Triploid Endosperm Genome of *Arabidopsis* Adopts a Peculiar, Parental-Dosage-Dependent Chromatin Organization. The Plant Cell 19: 1782-1794

Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, Dean C (2004) Vernalization requires epigenetic silencing of FLC by histone methylation. Nature 427: 164

Beh LY, Colwell LJ, Francis NJ (2012) A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence. P Natl Acad Sci USA 109: 1063-1071

Berke L, Snel B (2015) The plant Polycomb repressive complex 1 (PRC1) existed in the ancestor of seed plants and has a complex duplication history. BMC Evolutionary Biology 15: 44

Berry S, Dean C, Howard M (2017) Slow Chromatin Dynamics Allow Polycomb Target Genes to Filter Fluctuations in Transcription Factor Activity. Cell Systems 4: 445-457

Boureau L, How-Kit A, Teyssier E, Drevensek S, Rainieri M, Joubès J, Stammitti L, Pribat A, Bowler C, Hong Y, Gallusci P (2016) A CURLY LEAF homologue controls both vegetative and reproductive development of tomato plants. Plant Molecular Biology 90: 485-501

Bouyer D, Roudier F, Heese M, Andersen ED, Gey D, Nowack MK, Goodrich J, Renou J-P, Grini PE, Colot V (2011) Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. PLoS genetics 7: 1002014

Bratzel F, Lopez-Torrejon G, Koch M, Del Pozo JC, Calonje M (2010) Keeping cell identity in Arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. Current Biology 20: 1853-1859

Bratzel F, Yang C, Angelova A, Lopez-Torrejon G, Koch M, Carlos del Pozo J, Calonje M (2012) Regulation of the New Arabidopsis Imprinted Gene AtBMI1C Requires the Interplay of Different Epigenetic Mechanisms. Molecular Plant 5: 260-269

Briggs GC, Osmont KS, Shindo C, Sibout R, Hardtke CS (2006) Unequal genetic redundancies in Arabidopsis – a neglected phenomenon? Trends in Plant Science 11: 492-498

Brookes E, de Santiago I, Hebenstreit D, Morris Kelly J, Carroll T, Xie Sheila Q, Stock Julie K, Heidemann M, Eick D, Nozaki N, Kimura H, Ragoussis J, Teichmann Sarah A, Pombo A (2012) Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs. Cell Stem Cell 10: 157-170

Brooun A, Gajiwala KS, Deng Y-L, Liu W, Bolaños B, Bingham P, He Y-A, Diehl W, Grable N, Kung P-P, Sutton S, Maegley KA, Yu X, Stewart AE (2016) Polycomb repressive complex 2 structure with inhibitor reveals a mechanism of activation and drug resistance. Nature Communications 7: 11384

Calonje M (2014) PRC1 Marks the Difference in Plant PcG Repression. Molecular Plant 7: 459-471

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

Casanova M, Preissner T, Cerase A, Poot R, Yamada D, Li X, Appanah R, Bezstarosti K, Demmers J, Koseki H (2011) Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. Development 138: 1471-1482

Chanvivattana Y, Bishopp A, Schubert D, Stock C, Moon YH, Sung ZR, Goodrich J (2004) Interaction of polycomb-group proteins controlling flowering in Arabidopsis. Development 131: 5263-5276

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

Chen L-J, Diao Z-Y, Specht C, Sung ZR (2009) Molecular Evolution of VEF-Domain-Containing PcG Genes in Plants. Molecular Plant 2: 738-754

Ciferri C, Lander GC, Maiolica A, Herzog F, Aebersold R, Nogales E (2012) Molecular architecture of human polycomb repressive complex 2. Elife 1: 5

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology 26: 1367

Cremer T, Cremer C, Baumann H, Luedtke EK, Sperling K, Teuber V, Zorn C (1982) Rabl's model of the interphase chromosome arrangement tested in Chinise hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. Human Genetics 60: 46-56

Crevillén P, Yang H, Cui X, Greeff C, Trick M, Qiu Q, Cao X, Dean C (2014) Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. Nature 515: 587-590

Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V (2002) Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. Cell 111: 185-196

Davidovich C, Goodrich KJ, Gooding AR, Cech TR (2014) A dimeric state for PRC2. Nucleic Acids Research 42: 9236-9248

de Lucas M, Pu L, Turco G, Gaudinier A, Morao AK, Harashima H, Kim D, Ron M, Sugimoto K, Roudier F, Brady SM (2016) Transcriptional Regulation of Arabidopsis Polycomb Repressive Complex 2 Coordinates Cell-Type Proliferation and Differentiation. Plant Cell 28: 2616-2631

Deaton AM, Gómez-Rodríguez M, Mieczkowski J, Tolstorukov MY, Kundu S, Sadreyev RI, Jansen LET, Kingston RE (2016) Enhancer regions show high histone H3.3 turnover that changes during differentiation. eLife 5: 15316

Derkacheva M, Steinbach Y, Wildhaber T, Mozgova I, Mahrez W, Nanni P, Bischof S, Gruissem W, Hennig L (2013) Arabidopsis MSI1 connects LHP1 to PRC2 complexes. Embo Journal 32: 2073-2085

Dumbliauskas E, Lechner E, Jaciubek M, Berr A, Pazhouhandeh M, Alioua M, Cognat V, Brukhin V, Koncz C, Grossniklaus U, Molinier J, Genschik P (2011) The Arabidopsis CUL4-DDB1 complex interacts with MSI1 and is required to maintain MEDEA parental imprinting. Embo Journal 30: 731-743

Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, Dash C, Bazett-Jones DP, Le Grice S, McKay RDG, Buetow KH, Gingeras TR, Misteli T, Meshorer E (2008) Global Transcription in Pluripotent Embryonic Stem Cells. Cell Stem Cell 2: 437-447

Engler C, Marillonnet S (2014) Golden gate cloning. In DNA Cloning and Assembly Methods, Svein V, Rahmi L (eds) pp 119-131. Springer Spektrum

Eskeland R, Freyer E, Leeb M, Wutz A, Bickmore WA (2010) Histone acetylation and the maintenance of chromatin compaction by Polycomb repressive complexes. Cold Spring Harbor Symposia on Quantitative Biology 75: 71-78

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: 5335

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

Finch JT, Klug A (1976) Solenoidal model for superstructure in chromatin. P Natl Acad Sci USA 73: 1897-1901

Fingerman IM, Du H-N, Briggs SD (2008) *In Vitro* Histone Methyltransferase Assay. In Cold Spring Harbor Protocols, p 4939.

Forderer A, Zhou Y, Turck F (2016) The age of multiplexity: recruitment and interactions of Polycomb complexes in plants. Curr Opin Plant Biol 29: 169-178

Fransz P, de Jong H (2011) From nucleosome to chromosome: a dynamic organization of genetic information. The Plant Journal 66: 4-17

Fuchs J, Demidov D, Houben A, Schubert I (2006) Chromosomal histone modification patterns; from conservation to diversity. Trends in Plant Science 11: 199-208

Fussner E, Strauss M, Djuric U, Li R, Ahmed K, Hart M, Ellis J, Bazett-Jones DP (2012) Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. EMBO reports 13: 992-996

Fyodorov DV, Zhou B-R, Skoultchi AI, Bai Y (2017) Emerging roles of linker histones in regulating chromatin structure and function. Nature Reviews Molecular Cell Biology 19: 192-206

Gallusci P, Hodgman C, Teyssier E, Seymour GB (2016) DNA Methylation and Chromatin Regulation during Fleshy Fruit Development and Ripening. Frontiers in Plant Science 7: 807

Gan E-S, Xu Y, Ito T (2015) Dynamics of H3K27me3 methylation and demethylation in plant development. Plant Signaling & Behavior 10: 1027851

Gan E-S, Xu Y, Wong J-Y, Geraldine Goh J, Sun B, Wee W-Y, Huang J, Ito T (2014) Jumonji demethylases moderate precocious flowering at elevated temperature via regulation of FLC in Arabidopsis. Nature Communications 5: 5098

Goldberg AD, Allis CD, Bernstein E (2007) Epigenetics: A Landscape Takes Shape. Cell 128: 635-638

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

Grewal SIS, Moazed D (2003) Heterochromatin and Epigenetic Control of Gene Expression. Science 301: 798-802

Grimaud C, Nègre N, Cavalli G (2006) From genetics to epigenetics: the tale of Polycomb group and trithorax group genes. Chromosome Research 14: 363-375

Grob S, Grossniklaus U (2017) Chromosome conformation capture-based studies reveal novel features of plant nuclear architecture. Curr Opin Plant Biol 36: 149-157

Gu X, Jiang D, Yang W, Jacob Y, Michaels SD, He Y (2011) Arabidopsis Homologs of Retinoblastoma-Associated Protein 46/48 Associate with a Histone Deacetylase to Act Redundantly in Chromatin Silencing. PLoS Genetics 7: 1002366

Guitton AE, Page DR, Chambrier P, Lionnet C, Faure JE, Grossniklaus U, Berger F (2004) Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in *Arabidopsis thaliana*. Development 131: 2971-2981

He A, Shen X, Ma Q, Cao J, von Gise A, Zhou P, Wang G, Marquez VE, Orkin SH, Pu WT (2012) PRC2 directly methylates GATA4 and represses its transcriptional activity. Genes & Development 26: 37-42

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

Hohmann N, Wolf EM, Lysak MA, Koch MA (2015) A Time-Calibrated Road Map of Brassicaceae Species Radiation and Evolutionary History. The Plant Cell 27: 2770-2784

How Kit A, Boureau L, Stammitti-Bert L, Rolin D, Teyssier E, Gallusci P (2010) Functional analysis of SlEZ1 a tomato Enhancer of zeste (E(z)) gene demonstrates a role in flower development. Plant Molecular Biology 74: 201-213

Ikeuchi M, Iwase A, Rymen B, Harashima H, Shibata M, Ohnuma M, Breuer C, Morao AK, de Lucas M, De Veylder L, Goodrich J, Brady SM, Roudier F, Sugimoto K (2015) PRC2 represses dedifferentiation of mature somatic cells in Arabidopsis. Nature Plants 1: 15089

Jacob Y, Bergamin E, Donoghue MT, Mongeon V, LeBlanc C, Voigt P, Underwood CJ, Brunzelle JS, Michaels SD, Reinberg D, Couture JF, Martienssen RA (2014) Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication. Science 343: 1249-1253

Jacob Y, Feng S, LeBlanc CA, Bernatavichute YV, Stroud H, Cokus S, Johnson LM, Pellegrini M, Jacobsen SE, Michaels SD (2009) ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nature Structural & Molecular Biology 16: 763-768

Jacob Y, Stroud H, Leblanc C, Feng S, Zhuo L, Caro E, Hassel C, Gutierrez C, Michaels SD, Jacobsen SE (2010) Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. Nature 466: 987-991

Jenuwein T, Allis CD (2001) Translating the Histone Code. Science 293: 1074-1080

Jeong CW, Roh H, Dang TV, Choi YD, Fischer RL, Lee JS, Choi Y (2011) An E3 ligase complex regulates SET-domain polycomb group protein activity in Arabidopsis thaliana. P Natl Acad Sci USA 108: 8036-8041

Jiang D, Berger F (2017) Histone variants in plant transcriptional regulation. Biochimica et Biophysica Acta Gene Regulatory Mechanisms 1860: 123-130

Jiao L, Liu X (2015) Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science 350: 4383

Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, Jacobsen SE (2004) Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Research 32: 6511-6518

Joti Y, Hikima T, Nishino Y, Kamada F, Hihara S, Takata H, Ishikawa T, Maeshima K (2012) Chromosomes without a 30-nm chromatin fiber. Nucleus 3: 404-410

Jürgens G (1985) A group of genes controlling the spatial expression of the bithorax complex in Drosophila. Nature 316: 153

Justin N, Zhang Y, Tarricone C, Martin SR, Chen S, Underwood E, De Marco V, Haire LF, Walker PA, Reinberg D, Wilson JR, Gamblin SJ (2016) Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. Nature Communications 7: 11316

Kahn TG, Dorafshan E, Schultheis D, Zare A, Stenberg P, Reim I, Pirrotta V, Schwartz YB (2016) Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. Nucleic Acids Research 44: 10132-10149

Kim J-M, Kim K, Punj V, Liang G, Ulmer TS, Lu W, An W (2015) Linker histone H1.2 establishes chromatin compaction and gene silencing through recognition of H3K27me3. Scientific Reports 5: 16714

Kim SY, Lee J, Eshed-Williams L, Zilberman D, Sung ZR (2012) EMF1 and PRC2 Cooperate to Repress Key Regulators of Arabidopsis Development. PLOS Genetics 8: 1002512

Kohler C, Hennig L, Spillane C, Pien S, Gruissem W, Grossniklaus U (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. Genes & Development 17: 1540-1553

Kouzarides T (2010) Chromatin Modifications and Their Function. Cell 128: 693-705

Lafos M, Kroll P, Hohenstatt ML, Thorpe FL, Clarenz O, Schubert D (2011) Dynamic Regulation of H3K27 Trimethylation during Arabidopsis Differentiation. PLOS Genetics 7: 1002040

Lee H-G, Kahn TG, Simcox A, Schwartz YB, Pirrotta V (2015) Genome-wide activities of Polycomb complexes control pervasive transcription. Genome Research 25: 1170-1181

Li S, Yamada M, Han X, Ohler U, Benfey Philip N (2016) High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. Developmental Cell 39: 508-522

Liang SC, Hartwig B, Perera P, Mora-Garcia S, de Leau E, Thornton H, Lima de Alves F, Rapsilber J, Yang S, James GV, Schneeberger K, Finnegan EJ, Turck F, Goodrich J (2015) Kicking against the PRCs - A

Domesticated Transposase Antagonises Silencing Mediated by Polycomb Group Proteins and Is an Accessory Component of Polycomb Repressive Complex 2. Plos Genetics 11: 1005660

Lindroth AM, Shultis D, Jasencakova Z, Fuchs J, Johnson L, Schubert D, Patnaik D, Pradhan S, Goodrich J, Schubert I, Jenuwein T, Khorasanizadeh S, Jacobsen SE (2004) Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. Embo Journal 23: 4286-4296

Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, Richard McCombie W, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430: 471

Lopez-Vernaza M, Yang S, Müller R, Thorpe F, de Leau E, Goodrich J (2012) Antagonistic Roles of SEPALLATA3, FT and FLC Genes as Targets of the Polycomb Group Gene CURLY LEAF. PLOS ONE 7: 30715

Lu F, Cui X, Zhang S, Jenuwein T, Cao X (2011) Arabidopsis REF6 is a histone H3 lysine 27 demethylase. Nature Genetics 43: 715

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389: 251

Luo M (2012) Current Chemical Biology Approaches to Interrogate Protein Methyltransferases. ACS Chemical Biology 7: 443-463

Luo M, Bilodeau P, Koltunow A, Dennis ES, Peacock WJ, Chaudhury AM (1999) Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. P Natl Acad Sci USA 96: 296-301

Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ, III, Voigt P, Martin SR, Taylor WR, De Marco V, Pirrotta V, Reinberg D, Gamblin SJ (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. Nature 461: 762-767

Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, Dynlacht BD, Reinberg D (2008) Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. Molecular Cell 32: 503-518

Margueron R, Trojer P, Reinberg D (2005) The key to development: interpreting the histone code? Current Opinion in Genetics & Development 15: 163-176

Martin C, Cao R, Zhang Y (2006) Substrate Preferences of the EZH2 Histone Methyltransferase Complex. Journal of Biological Chemistry 281: 8365-8370

Mathieu O, Probst AV, Paszkowski J (2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. The EMBO Journal 24: 2783-2791

McCabe MT, Graves AP, Ganji G, Diaz E, Halsey WS, Jiang Y, Smitheman KN, Ott HM, Pappalardi MB, Allen KE, Chen SB, Della Pietra A, Dul E, Hughes AM, Gilbert SA, Thrall SH, Tummino PJ, Kruger RG, Brandt M, Schwartz B et al. (2012) Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). P Natl Acad Sci USA 109: 2989-2994

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim T-K, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: 553

Mosquna A, Katz A, Decker EL, Rensing SA, Reski R, Ohad N (2009) Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. Development 136: 2433-2444

Mozgova I, Hennig L (2015a) The Polycomb Group Protein Regulatory Network. In Annual Review of Plant Biology, Vol 66, Merchant SS (ed) pp 269-296.

Mozgova I, Hennig L (2015b) The Polycomb Group Protein Regulatory Network. Annual Review of Plant Biology 66: 269-296

Nekrasov M, Wild B, Muller J (2005) Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. Embo Reports 6: 348-353

Nishino Y, Eltsov M, Joti Y, Ito K, Takata H, Takahashi Y, Hihara S, Frangakis AS, Imamoto N, Ishikawa T, Maeshima K (2012) Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. The EMBO Journal 31: 1644-1653

Nowack MK, Shirzadi R, Dissmeyer N, Dolf A, Endl E, Grini PE, Schnittger A (2007) Bypassing genomic imprinting allows seed development. Nature 447: 312

Oh S, Park S, van Nocker S (2008) Genic and Global Functions for Paf1C in Chromatin Modification and Gene Expression in Arabidopsis. PLoS Genetics 4: 1000077

Okano Y, Aono N, Hiwatashi Y, Murata T, Nishiyama T, Ishikawa T, Kubo M, Hasebe M (2009) A polycomb repressive complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution. P Natl Acad Sci USA 106: 16321-16326

Oliviero G, Brien GL, Waston A, Streubel G, Jerman E, Andrews D, Doyle B, Munawar N, Wynne K, Crean J, Bracken AP, Cagney G (2016) Dynamic Protein Interactions of the Polycomb Repressive Complex 2 during Differentiation of Pluripotent Cells. Molecular & Cellular Proteomics 15: 3450-3460

Panchy N, Lehti-Shiu M, Shiu S-H (2016) Evolution of Gene Duplication in Plants. Plant Physiology 171: 2294-2316

Park S, Oh S, van Nocker S (2012) Genomic and Gene-Level Distribution of Histone H3 Dimethyl Lysine-27 (H3K27me2) in Arabidopsis. PLOS ONE 7: 52855

Pasini D, Malatesta M, Jung HR, Walfridsson J, Willer A, Olsson L, Skotte J, Wutz A, Porse B, Jensen ON, Helin K (2010) Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucleic Acids Research 38: 4958-4969

Pombo A, Dillon N (2015) Three-dimensional genome architecture: players and mechanisms. Nature Reviews Molecular Cell Biology 16: 245

Pontvianne F, Blevins T, Chandrasekhara C, Feng W, Stroud H, Jacobsen SE, Michaels SD, Pikaard CS (2012) Histone methyltransferases regulating rRNA gene dose and dosage control in Arabidopsis. Genes & Development 26: 945-957

Qiu Y, Liu S-L, Adams KL (2017) Concerted divergence after gene duplication in Polycomb Repressor complexes. Plant Physiology 174: 1192-1204

Ricci MA, Cosma MP, Lakadamyali M (2017) Super resolution imaging of chromatin in pluripotency, differentiation, and reprogramming. Current opinion in genetics & development 46: 186-193

Ricci Maria A, Manzo C, García-Parajo MF, Lakadamyali M, Cosma Maria P (2015) Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes *In Vivo*. Cell 160: 1145-1158

Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, Giraut L, Després B, Drevensek S, Barneche F, Dèrozier S, Brunaud V, Aubourg S, Schnittger A, Bowler C, Martin-Magniette ML et al. (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. The EMBO Journal 30: 1928-1938

Sanchez-Pulido L, Devos D, Sung ZR, Calonje M (2008) RAWUL: A new ubiquitin-like domain in PRC1 ring finger proteins that unveils putative plant and worm PRC1 orthologs. Bmc Genomics 9: 308

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B (2012) Fiji: an open-source platform for biological-image analysis. Nature methods 9: 676

Schmitges FW, Prusty AB, Faty M, Stuetzer A, Lingaraju GM, Aiwazian J, Sack R, Hess D, Li L, Zhou S, Bunker RD, Wirth U, Bouwmeester T, Bauer A, Ly-Hartig N, Zhao K, Chan H, Gu J, Gut H, Fischle W et al. (2011) Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Molecular Cell 42: 330-341

Schubert D, Clarenz O, Goodrich J (2005) Epigenetic control of plant development by Polycomb-group proteins. Curr Opin Plant Biol 8: 553-561

Sequeira-Mendes J, Gutierrez C (2015) Links between genome replication and chromatin landscapes. The Plant Journal 83: 38-51

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

Shimada TL, Shimada T, Hara-Nishimura I (2010) A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana. Plant J 61: 519-528

Simon JA, Kingston RE (2009) Mechanisms of Polycomb gene silencing: knowns and unknowns. Nature Reviews Molecular Cell Biology 10: 697

Smits AH, Jansen PWTC, Poser I, Hyman AA, Vermeulen M (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Research 41: 28

Sneeringer CJ, Scott MP, Kuntz KW, Knutson SK, Pollock RM, Richon VM, Copeland RA (2010) Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. P Natl Acad Sci USA 107: 20980

Spillane C, Schmid KJ, Laoueillé-Duprat S, Pien S, Escobar-Restrepo J-M, Baroux C, Gagliardini V, Page DR, Wolfe KH, Grossniklaus U (2007) Positive darwinian selection at the imprinted MEDEA locus in plants. Nature 448: 349

Stroud H, Hale CJ, Feng S, Caro E, Jacob Y, Michaels SD, Jacobsen SE (2012) DNA methyltransferases are required to induce heterochromatic re-replication in Arabidopsis. PLoS Genet 8: 1002808

Sung S, Amasino RM (2004) Vernalization and epigenetics: how plants remember winter. Curr Opin Plant Biol 7: 4-10

Tan J-z, Yan Y, Wang X-x, Jiang Y, Xu HE (2013) EZH2: biology, disease, and structure-based drug discovery. Acta Pharmacologica Sinica 35: 161

Tanurdzic M, Vaughn MW, Jiang H, Lee T-J, Slotkin RK, Sosinski B, Thompson WF, Doerge RW, Martienssen RA (2008) Epigenomic Consequences of Immortalized Plant Cell Suspension Culture. PLoS Biology 6: 302

Team RC (2013) R: A language and environment for statistical computing.

Tie F, Prasad-Sinha J, Birve A, Rasmuson-Lestander Å, Harte PJ (2003) A 1-megadalton ESC/E (Z) complex from Drosophila that contains polycomblike and RPD3. Molecular and cellular biology 23: 3352-3362

Tonosaki K, Kinoshita T (2015) Possible roles for polycomb repressive complex 2 in cereal endosperm. Frontiers in Plant Science 6: 144

Turck F, Roudier F, Farrona S, Martin-Magniette M-L, Guillaume E, Buisine N, Gagnot S, Martienssen RA, Coupland G, Colot V (2007) Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. Plos Genetics 3: 855-866

Tyanova S, Temu T, Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nature Protocols 11: 2301

Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K The flowering world: a tale of duplications. Trends in Plant Science 14: 680-688

Vision TJ, Brown DG, Tanksley SD (2000) The Origins of Genomic Duplications in Arabidopsis. Science 290: 2114-2117

Wang D, Tyson MD, Jackson SS, Yadegari R (2006) Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in *Arabidopsis*. P Natl Acad Sci USA 103: 13244-13249

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: 1005771

Wu H, Zeng H, Dong A, Li F, He H, Senisterra G, Seitova A, Duan S, Brown PJ, Vedadi M, Arrowsmith CH, Schapira M (2013) Structure of the Catalytic Domain of EZH2 Reveals Conformational Plasticity in Cofactor and Substrate Binding Sites and Explains Oncogenic Mutations. PLOS ONE 8: 83737

Xu M, Hu T, Smith MR, Poethig RS (2015) Epigenetic Regulation of Vegetative Phase Change in Arabidopsis. The Plant Cell: 28-41

Xu Y, Guo C, Zhou B, Li C, Wang H, Zheng B, Ding H, Zhu Z, Peragine A, Cui Y, Poethig S, Wu G (2016) Regulation of Vegetative Phase Change by SWI2/SNF2 Chromatin Remodeling ATPase BRAHMA. Plant Physiology 172: 2416-2428

Yamaguchi H, Hung M-C (2014) Regulation and Role of EZH2 in Cancer. Cancer Research and Treatment : Official Journal of Korean Cancer Association 46: 209-222

Yang H, Berry S, Olsson TSG, Hartley M, Howard M, Dean C (2017) Distinct phases of Polycomb silencing to hold epigenetic memory of cold in *Arabidopsis*. Science: 1142-1145

Zee BM, Britton L-MP, Wolle D, Haberman DM, Garcia BA (2012) Origins and Formation of Histone Methylation across the Human Cell Cycle. Molecular and Cellular Biology 32: 2503-2514

Zhang T, Cooper S, Brockdorff N (2015) The interplay of histone modifications – writers that read. EMBO reports 16: 1467-1481

Zhang X, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, Goodrich J, Jacobsen SE (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS biology 5: 129

Zhao D, Zhang X, Guan H, Xiong X, Shi X, Deng H, Li H (2016) The BAH domain of BAHD1 is a histone H3K27me3 reader. Protein & cell 7: 222–226

Zhao S, Zhang B, Yang M, Zhu J, Li H (2018) Systematic Profiling of Histone Readers in Arabidopsis thaliana. Cell Reports 22: 1090-1102

Zhou Y, Romero-Campero FJ, Gomez-Zambrano A, Turck F, Calonje M (2017a) H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and PRC2 activity. Genome Biology 18: 69

Zhou Y, Tergemina E, Cui H, Forderer A, Hartwig B, Velikkakam James G, Schneeberger K, Turck F (2017b) Ctf4-related protein recruits LHP1-PRC2 to maintain H3K27me3 levels in dividing cells in Arabidopsis thaliana. P Natl Acad Sci USA 114: 4833-4838

Zografou T (2013) Distinct impact of CURLY LEAF and SWINGER on the Arabidopsis Histone H3 Lysine 27 trimethylation pattern is linked to the underlying genetic code. In Universität zu Köln, Dissertation

Weblinks

[1] https://en.wikipedia.org/wiki/Association_fallacy

[2] http://phytozome.jgi.doe.gov/pz/portal.html

[3] http://congenie.org/gbrowse

[4] http://www.amborella.org

[5] http://medicinalplantgenomics.msu.edu/mpgr_blast.shtml

[6] http://web.expasy.org/translate/

[7] http://www.ebi.ac.uk/Tools/msa/clustalo/

[8] www.arabidopsis.org/aboutarabidopsis.html

[9] https://nebiocalculator.neb.com

[10] http://www.maxquant.org

8 Acknowledgements/Danksagung

Ich danke von Herzen:

Dr. Franziska Turck, for teaching me her approach towards science. Your way of doing science not only inspired me, but also made me want to do more. I thank you deeply for giving me the freedom to express myself as a scientist and as a person.

Prof. Dr. George Coupland, for providing an excellent research environment and enabling my research by making the right diplomatic decisions. I am also thankful for the examination of this work.

Prof. Dr. Ute Höcker and **Prof. Dr. Maria Albani**, for examining this work and for asking challenging questions (past and future).

Prof. Dr. Iris Finkemeier and **Dr. Peter Tessarz**, for being in my thesis advisory committee. The discussions always brought me forward by light years.

Team Turck, for helping me out with experiments, for making the PhD such a fun time, for having such great team spirit and for bringing out the best in me.

Everybody of the **Coupland department** for making the PhD such a joyful time.

Mass spectrometry team, especially Sara Stolze, Katharina Kramer and Anne Harzen for analyzing my samples and for discussions and suggestions.

Petra Tänzler, aka ‚Labordrachen' und ‚cloning wizzard', für die viiiiiielen Dinge die ich von dir gelernt habe. Nicht zuletzt, dass Ordnung das A und O sind (ok, rein theoretisch hab ich das gelernt). Es war eine riesen Freude und Ehre mit dir Zusammenzuarbeiten.

Christine Tölzer und **Dr. Karsten Niefind** für das Interesse an meinem Projekt. Es war extrem hilfreich für mich mit euch über meine biochmischen Experimente zu diskutieren und ohne die praktische Hilfe und Lehre von Christine wären viele dieser Experimente garnicht möglich gewesen.

Addie Kolybaba und **Dr. Anne Classen** für die Kooperation bzgl. der Insektenzellen und das freundliche Willkommen an der LMU München.

Regina in der Spülküche für die freundliche Gesellschaft im Keller, die Bereitstellung von Verbrauchsmaterialien und das makellose Reinigen der Kulturflaschen.

Dem Gärtnerteam, die mit mir dabei geholfen haben meine Pflanzen gesund zu halten, mir bei Patzern aus der Klemme geholfen haben und dabei immer ein Lachen übrig hatten.

Emmanuel Tergemina, for being my one person with whom I could share everything. You provided emotional support all the way through.

Leon Langereis for joining my research with joy and curiosity. Finally yet importantly, for taking at least some load off my shoulders in culturing the insect cells.

Meiner WG in der Dasselstraße, besonders Katina, Dao und Lola. Ihr habt mir ein Gefühl von zu Hause gegeben für so lange Zeit und habt mir immer dabei geholfen auch mal abzuschalten.

Konrad for giving me so much support during this time. I thank you from my heart for motivating me, for helping in so many ways, for emotionally grounding me and for inspiring me to be a better person. I don't know if could have done it without you as my companion.

Meiner ganzen Famile, besonders **Marie und Kathi**, für die Lebenszeichen die ihr von mir erzwingen musstet. Ich liebe euch sehr. Ich weiß, dass ihr unheimlich stolz auf mich seid.

Meinen Eltern die mich seit jeher in meinen Interessen unterstützt und gefördert haben. Ohne euch, eure großherzige Unterstützung und bedingunglose Liebe wär diese Arbeit niemals möglich gewesen.

9 Erklärung

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

Köln, den 12. Juli 2018 _____ Alexander Förderer

Es liegen keine Teilpublikationen vor.