

**Gametophyte-specific degradation of Centromeric  
Histone3 (CENH3) to investigate the mechanism of  
uniparental genome elimination in *Arabidopsis thaliana***

**Dissertation  
zur Erlangung des  
Doktorgrades der Naturwissenschaften (Dr. rer. nat.)**

der

Naturwissenschaftlichen Fakultät III  
Agrar- und Ernährungswissenschaften,  
Geowissenschaften und Informatik

der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von

Herrn Saravanakumar Somasundaram

Gutachter:

1. Prof. Dr. Andreas Houben
2. Prof. Dr. Nico De Storme

Tag der öffentlichen Verteidigung: 20. Januar 2025, Halle (Saale)



## **Acknowledgement**

I would like to express my deepest gratitude to my supervisor, Prof. Dr. Andreas Houben, for his guidance throughout the course of this research. I am sincerely thankful for the valuable opportunity he provided, as well as his generous contributions of time, suggestions, and assistance. His mentorship has been instrumental in shaping my professional career path.

I extend my heartfelt thanks to Dr. Joerg Fuchs for his unwavering support throughout this research project. Our discussions along with Dr. Yi-Tzu Kuo, which went beyond scientific matters, were invaluable in helping me maintain my mental well-being. I am deeply thankful to Oda Weiss for her constant support in conducting this project. Without her help, I would not have been able to complete it on time.

I owe special thanks to Dr. Andriy Kochevenko for providing key support during the early phase of this project. I also greatly appreciate the technical assistance provided by Katrin Kumke and Sylvia Swetik. I am very thankful to Dr. Jonathan Lamb (Bayer Crop Sciences, USA) for his interest and support that he provided for this project.

I would like to express my gratitude to Dr. Maria Cuacos and Dr. Stefan Heckmann for providing the essential plant materials for this research project. I am also thankful to all former and present members of the CSF group for fostering a friendly and motivating work atmosphere.

I would like to thank the support of the greenhouse and gardening staff at IPK. I am profoundly grateful to Bianka Jacobi and Dr. Britt Leps for ensuring that all necessary paperwork and procedures were correctly handled, which significantly contributed to the smooth progress of this research project.

I sincerely thank my former mentors, Dr. Gopala Krishnan, Dr. Anandhan, Dr. Ravi, and Dr. Sateesh, for their invaluable guidance and support throughout my career journey. I am equally grateful to Dr. Soumia and Dr. Ram Krishna for their encouragement and unwavering goodwill.

My heartfelt thanks go to my wife, family, and friends for their support, patience, and understanding throughout this journey. Special thanks to Suriya and Goutham for making my time in Gatersleben highly memorable.

I would like to acknowledge the financial support from DAAD, without which this research would not have been possible. Their contribution is greatly appreciated. Special thanks also to all the information-sharing systems and various computer programs used for data analysis.

# Contents

List of Figures.....	1
Abbreviations.....	3
<b>1. Introduction.....</b>	<b>5</b>
1.1. Uniparental genome elimination.....	5
1.2. Classification of haploid inducers.....	6
1.3. Centromere-based haploid induction systems.....	7
1.3.1. CENH3 manipulation in Arabidopsis.....	8
1.3.2. Influence of temperature on centromere-based UGE.....	9
1.3.3. CENH3 manipulation in crop plants.....	9
1.4. Non-centromere-based haploid induction systems.....	11
1.5. Comparison of both haploidization systems.....	15
1.6. Mechanism of centromere-based haploidization.....	16
1.7. Depletion of CENH3 protein in the gametophyte.....	20
1.8. Targeted protein degradation.....	20
1.9. Targeted CENH3 degradation for haploid induction in Arabidopsis.....	23
<b>2. Aims.....</b>	<b>24</b>
<b>3. Materials and methods.....</b>	<b>25</b>
3.1. Plant Materials.....	25
3.1.1. Description of Arabidopsis mutant alleles.....	25
3.1.2. Description of the transgenic construct from previous studies.....	25
3.2. Nomenclature of genotypes.....	25
3.3. Plant growth conditions.....	26
3.4. Preparation of transgenic constructs.....	26
3.5. Generation of transgenic Arabidopsis plants.....	26
3.6. Transient expression in Nicotiana.....	27
3.7. Plant genomic DNA isolation.....	27
3.8. PCR genotyping.....	27
3.9. SNP genotyping.....	27
3.10. Crossing and haploid screening.....	28
3.11. Purification of AtCENH3-Ntail antigen.....	28
3.12. Phage display for affinity selection.....	29
3.13. Immunostaining.....	30
3.14. Preparation of early embryos for imaging.....	30
3.15. Preparation of ovule, pollen, and anthers for imaging.....	31

3.16. Microscopy .....	31
3.17. Flow cytometry.....	31
3.18. Statistical analysis.....	32
<b>4. Results</b> .....	<b>33</b>
4.1. Different levels of EYFP-CENH3 in male and female gametes of <i>cenh3</i> null mutant complemented with EYFP-gCENH3 .....	33
4.2. Extreme asymmetry between parental chromosomes for EYFP-CENH3 at early embryogenesis is associated with uniparental genome elimination.....	34
4.3. Egg cell-specific degradation of the EYFP-CENH3 protein .....	38
4.3.1. Egg cell-specific degradation of the EYFP-CENH3 protein results in shriveled seeds .....	39
4.3.2. Egg cell-specific degradation of the EYFP-CENH3 protein leads to higher frequency of haploids .....	40
4.4. Screening for nanobody against native Arabidopsis CENH3 .....	42
4.4.1. Preparation of a suitable CENH3 antigen .....	42
4.4.2. Phage display using synthetic nanobody libraries.....	43
4.4.3. Phage display using immune nanobody library .....	44
4.5. Replacement of the EYFP-tag with an ALFA-tag .....	46
4.5.1. ALFA-CENH3 complements the Arabidopsis <i>cenh3-1</i> null mutation better than the EYFP-CENH3 .....	47
4.5.2. ALFA-CENH3 is functionally similar to CENH3 .....	48
4.5.3. ALFA-CENH3 induces haploid Arabidopsis plants at varying frequencies.....	49
4.5.4. Egg cell-specific degradation of the ALFA-CENH3 with SPOP induces higher frequency of haploids than Nslmb.....	50
4.6. Sperm-specific degradation of EYFP-CENH3 .....	51
4.6.1. Selection of candidate sperm-specific promoter .....	52
4.6.2. VHHGFP4-SPOP results in the degradation of EYFP-CENH3 in sperm nuclei .....	53
4.6.3. Sperm-specific degradation of EYFP-CENH3 produces female-like diploids	54
4.6.4. Female-like diploids from the male sterile mothers excludes possibility of contamination from self-pollination .....	56
4.6.5. Female-like diploids are genomically similar to the maternal parent.....	59
4.7. Replacement of non-plant E3-ligase with plant-derived E3-ligases .....	60
4.7.1. Transiently expressed EL5 outperforms other E3-ligase for degradation of EYFP-CENH3.....	61
4.7.2. Both EL5 AND BPM1 E3-ligases fused to anti-ALFA nanobody can induce paternal haploids.....	63
<b>5. Discussion</b> .....	<b>65</b>
5.1. A critical level of CENH3 asymmetry is necessary for haploid induction.....	65

5.2. Seed death may be caused by other reasons in addition to endosperm-failure...	67
5.3. The transgenic supply of CENH3 may also contribute to haploid induction.....	67
5.4. How to target CENH3?.....	68
5.5. Comparison of different E3-ligases.....	69
5.6. Comparison of haploidization frequencies.....	71
5.7. Maternal versus paternal haploidization.....	71
<b>6. Outlook.....</b>	<b>73</b>
<b>7. Summary.....</b>	<b>75</b>
<b>8. Zusammenfassung.....</b>	<b>77</b>
<b>9. References.....</b>	<b>79</b>
<b>10. Curriculum vitae.....</b>	<b>92</b>
<b>11. Eidesstattliche Erklärung / Declaration under Oath.....</b>	<b>94</b>
<b>12. Appendix.....</b>	<b>95</b>





## List of Figures

Figure 1: Schematic model of uniparental genome elimination (UGE) in wheat x maize hybrids .....	6
Figure 2: Different strategies of manipulation of CENH3 protein or other kinetochore proteins for haploid induction.....	16
Figure 3: Centromere size model of haploid induction.....	17
Figure 4: CENH3 dilution in female versus male gametes. ....	18
Figure 5: Proposed mechanism of GFP-tailswap mediated genome elimination by Marimuthu et al. (2021). ....	19
Figure 6: Schemata of the deGradFP system for the targeted degradation of GFP-tagged proteins of interest. ....	22
Figure 7: Haploid screening using glabrous leaf marker.....	29
Figure 8: Gametophytic asymmetry for EYFP-CENH3 in [ <i>cenh3(-/-)</i> , <i>EYFP-gCENH3(+/+)</i> ] .	34
Figure 9: EYFP-CENH3 signals from 1-cell embryos derived from crossing different CENH3 genotypes .....	35
Figure 10: EYFP-CENH3 signals from 1-cell embryos derived from inter-ploidy crosses. ....	37
Figure 11: Extreme asymmetry in CENH3 levels between parental genomes is associated with genome elimination.....	38
Figure 12: Egg cell-specific degradation of EYFP-CENH3 results in shriveled seeds.....	40
Figure 13: Egg cell-specific degradation of EYFP-CENH3 resulted in increased haploidization frequency.....	41
Figure 14: Preparation of the truncated CENH3 antigen .....	43
Figure 15: Phage display using synthetic nanobody libraries. ....	44
Figure 16: Phage display using immune nanobody libraries.....	45
Figure 17: Complementation of Arabidopsis with ALFA-CENH3. ....	47
Figure 18: Functional characterization of <i>ALFA-gCENH3</i> T1 plants.....	49
Figure 19: SPOP induces a higher frequency of haploids than Nslmb in combination with the ALFA-tag. ....	51
Figure 20: DAZ3 promoter is an ideal candidate sperm-specific promoter. ....	53
Figure 21: Expression of VHHGFP4-SPOP degrades EYFP-CENH3 in sperm cells without affecting pollen viability. ....	55
Figure 22: The frequency of EYFP-CENH3 lacking pollen grains does not correlate with the female-like diploid frequencies.....	57
Figure 23: Female-like diploids were genomically similar to the maternal parent. ....	60
Figure 24: Evaluation of efficiency of candidate E3-ligases for EYFP-CENH3 degradation..	62

Figure 25: Both EL5 and BPM1 E3-ligases can induce paternal haploids with similar frequency..... 64

Figure 26: Schemata representing gametophytic degradation of tagged-CENH3 for haploid induction..... 66

## Abbreviations

AMS	Abortive microspores
BSA	Bovine serum albumin
BAM	Binary alignment map
CENH3	Centromeric histone H3
Col	Columbia-0 ecotype
DAPI	4',6-diamidino-2-phenylindole
dCAPS	Derived cleaved amplified polymorphic sequences
DAZ3	DUO1-activated Zinc Finger 3
DMP	DUF679 Domain membrane protein
DUF	Domain of unknown function
EC1.1	Egg cell 1.1
EYFP	Enhanced yellow fluorescent protein
GFP	Green fluorescent protein
GL1	Glabrous 1
H2B	Histone 2B
HFD	Histone fold domain
IgG	Immunoglobulin G
KNL2	Kinetochores null 2
Ler	Landsberg erecta ecotype
MoClo	Modular cloning
MTL	Matrilineal
Nb	Nanobody
NLD	Not like Dad
PBS	Phosphate buffered Saline
PI	Propidium iodide
PLA1	Phospholipase A1
POI	Protein of interest
pro	Promoter
RFS	Restored Frameshift
ROS	Reactive oxygen species
SAM	Sequence alignment map

SNP	Single nucleotide polymorphism
SR2200	SCRI Renaissance 2200
TPD	Targeted protein degradation
UGE	Uniparental genome elimination

# 1. Introduction

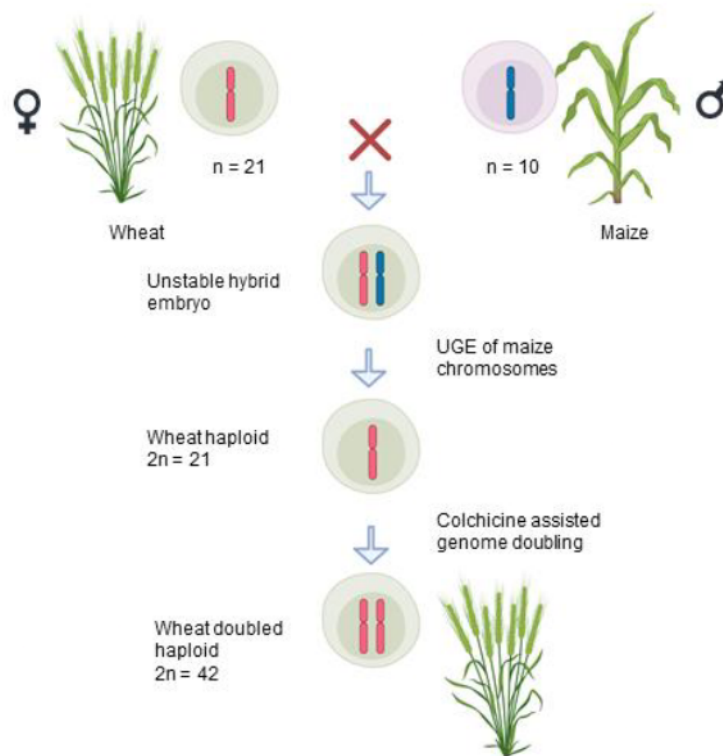
Haploid induction and subsequent chromosome doubling play a pivotal role in expediting the plant breeding process. The resulting doubled haploids are true-breeding genotypes and can be generated much more rapidly compared to recombinant inbred lines, which necessitate continuous self-pollination for a minimum of six to seven generations. While *in-vitro* cultivation of anthers (microspores) or ovules can generate haploids, this method is labor-intensive, species-specific, and could be hindered by genotype-related bottleneck issues (Kalinowska et al. 2019). *In-vivo* haploidization methods, such as wide crosses and pollen treatments, still require *in-vitro* rescue of haploid embryos (Jacquier et al. 2020).

An alternative approach involves the production of *in-planta* haploids through intra-specific hybridization with specialized genotypes known as haploid inducers. In this case, haploid induction occurs by eliminating the genome of the inducer genotypes during early embryogenesis. The primary advantage of this method is that haploids are directly obtained through seeds (Widiez 2021). The elimination of the inducer's genome facilitates genetic marker-based haploid screening. Beyond doubled haploid production, haploid inducers offer versatility in applications such as cytoplasm swapping, screening for recessive mutations, reverse breeding, and one-step haploid induction combined with genome editing (Ravi et al. 2014; Wijnker et al. 2012; Kelliher et al. 2019; Budhagatapalli et al. 2020). This broadens the scope of their application in plant breeding by providing efficient and diverse tools for genetic improvement.

## 1.1. Uniparental genome elimination

Biparental inheritance, where the offspring's nuclear genome originates from both parents, is the predominant mode among sexually reproducing organisms. An exception to this norm is uniparental genome elimination (UGE), wherein the genetic material from one of the parents is selectively excluded after fertilization, resulting in a fraction of haploid progenies. In plants, UGE after hybridization was first reported in *Nicotiana* sp., where a cross between *N. tabacum* and *N. sylvestris* resulted in *N. tabacum*-type haploids (Clausen and Mann 1924). Subsequently, post-fertilization UGE in a wide-cross has been documented across a diverse range of plant species

(as listed in (Ishii et al. 2016)). A plant breeding relevant example of an inter-specific cross leading to post-fertilization UGE is the combination of *Hordeum vulgare* x *H. bulbosum* producing *vulgare*-type haploids (Kasha and Kao 1970). Another well-known example is the elimination of maize chromosomes in the embryos of wheat x maize inter-generic cross, yielding wheat haploids (Laurie and Bennett 1988) (Fig. 1).



**Figure 1: Schematic model of uniparental genome elimination (UGE) in wheat x maize hybrids.** In unstable wheat x maize embryos, all maize chromosomes undergo elimination during early embryogenesis. Resulting embryos can be rescued *in-vitro* to develop into haploid wheat plants. The genome of the haploid plants was then doubled using colchicine treatment. Figure was prepared using Biorender.

## 1.2. Classification of haploid inducers

The approach using haploid inducers can be broadly divided into two categories: i) centromere-based systems, and ii) non-centromere-based systems. Centromere-based systems employ diverse strategies to alter the centromere-specific histone H3 (CENH3) at DNA, RNA, and protein levels. On the other hand, non-centromere-based systems involve mutants of genes that influence various aspects of the plant reproductive process, leading to haploidization.

### 1.3. Centromere-based haploid induction systems

Centromeres are the chromosomal locations of kinetochore assembly and are necessary for faithful chromosome segregation. The centromeres are inherited epigenetically and determined by a centromere-specific histone 3 variant (CENH3) in most plants and animal species (Gong et al. 2012; Mendiburo et al. 2011). CENH3 replaces the canonical histone H3 in centromeric nucleosomes. The centromeric loading of CENH3 is sufficient to initiate and perpetuate the centromeres through a positive epigenetic feedback loop, regardless of the underlying sequences (Black and Cleveland 2011; Mendiburo et al. 2011; Voullaire et al. 1993). The centromere size determined by the number of CENH3-containing nucleosomes is correlated with the chromosome segregation accuracy (Drpic et al. 2018). A critical level of CENH3 is necessary to maintain the centromere functionality (Bodor et al. 2014).

Talbert et al. (2002) identified CENH3 (also known as HTR12) in plants first. The authors confirmed the centromeric localization of CENH3 in *Arabidopsis thaliana* and showed that CENH3 is rapidly evolving, especially in its N-tail region. A CENH3 knockout mutation leads to embryo lethality; therefore, a true null mutant of CENH3 ceases to survive unless complemented by a functional CENH3 variant (Ravi et al. 2010). Reduction in endogenous CENH3 levels leads to both vegetative defects and meiotic abnormalities, whereas the modification of the N-tail region causes only meiotic abnormalities, sterility, and reduced seed setting (Lermontova et al. 2011a; Ravi et al. 2011). These observations suggest differential regulation of CENH3 loading during mitosis and meiosis; and probable involvement of the hypervariable N-terminal region in the regulation of meiotic CENH3 behavior.

Two independent studies underscored the significance of CENH3 in the uniparental genome elimination process. In the first study, manipulation of CENH3 in *A. thaliana* was found to induce uniparental genome elimination during early embryogenesis, yielding a mix of haploid, aneuploid, and diploid plants. The study utilized a chimeric protein called "GFP-tailswap", combining the N-tail of histone H3.3 and the Histone Fold Domain (HFD) of CENH3 fused to the reporter protein GFP, to complement the embryo-lethal CENH3 knockout mutant (*cenh3-1*). Crossbreeding plants complemented with GFP-tailswap as females with wild-type plants resulted in approximately 25 - 45% wild-type haploid progenies, while using the GFP-tailswap as

a male reduced haploid progeny to 5% (Ravi and Chan 2010). In the second study, centromere inactivity was found to trigger uniparental genome elimination in *Hordeum vulgare* x *H. bulbosum* inter-specific barley hybrid embryos. Loss of CENH3 from the centromeres of *H. bulbosum* led to the elimination of the corresponding chromosomes (Sanei et al. 2011).

### 1.3.1. CENH3 manipulation in Arabidopsis

After Ravi and Chan (2010) demonstrated that manipulating CENH3 can induce *in-planta* haploids, further studies showed that a broad list of CENH3 modifications can induce haploids. It was shown that a point (missense) mutation in a conserved HFD residue (L130F) of CENH3 impaired its centromeric loading in an *H. vulgare* TILLING mutant and induced haploids in the case of *A. thaliana* (Karimi-Ashtiyani et al. 2015). In parallel, a variety of point (missense) mutations in the HFD region of CENH3 was proven to result in UGE in *A. thaliana*. Haploid frequency was found ranging from 0.61 % to 12.2% (Kuppu et al. 2015). The above two studies explained a likely non-transgenic approach via TILLING to engineer haploid induction in plants. In an attempt to rescue the lethal *A. thaliana* CENH3 null mutant, [*cenh3(-/-)*], using heterologous CENH3 from *Brassica rapa* and *Lepidium oleracem*, different levels of seed death and haploid progenies were observed (Maheshwari et al. 2015). Also, combinations of point mutations, in-frame deletions in the CENH3 HFD region and complete deletion of the  $\alpha$ N helix were shown to exhibit varying levels of haploid induction ability (Kuppu et al. 2020). In addition to CENH3 manipulation, the knockout of other kinetochore proteins, such as Kinetochore Null 2 (KNL2) has been demonstrated to induce haploids in *A. thaliana* (Ahmadli et al. 2023). KNL2 is a centromere loading factor of CENH3 and the absence of KNL2 leads to a reduction of CENH3 levels in the centromere (Lermontova et al. 2013). Thus, KNL2 mutation can be considered to contribute indirectly to the haploid induction through reducing the effective CENH3 chromatin.



### **1.3.2. Influence of temperature on centromere-based UGE**

The study conducted by Ahmadli et al. (2023) demonstrated a ten-fold increase in the haploid induction ability of the *A. thaliana* KNL2 knockout mutant under short-term heat stress. Similarly, Jin et al. (2023) observed a significant rise in haploid induction, reaching approximately 95%, following post-fertilization heat stress when GFP-tailswap was used as a female parent. Notably, this condition was also associated with a high incidence of seed abortion. While the aforementioned reports highlighted the capacity of temperature manipulation to increase paternal haploids, Wang et al. (2023c) demonstrated that the frequency of maternal haploids can also be increased using temperature stress. The restoration of male fertility in GFP-tailswap at a temperature of 18°C rendered them suitable as pollen donors. The paternal haploid frequency increased the haploid induction rates up to 25% when the pollinated plants were grown at 25°C compared to 7.0% for those plants grown at 22°C. In summary, all three studies consistently reported that elevated post-fertilization temperatures were associated with heightened haploid induction rates.

However, it was underscored that temperature stress alone was inadequate for haploid induction; instead, it necessitated a combination with a parental genotype carrying a CENH3 modification (Ahmadli et al. 2023). Similar observations were made in *H. vulgare* x *H. bulbosum*, where lower temperatures (< 18°C) during embryo formation were found to suppress chromosome elimination, favoring the formation of true hybrids (Gernand et al. 2006; Pickering 1984). Conversely, higher temperatures around 21.5°C during early hybrid embryogenesis significantly boosted haploid frequency, reaching up to 99% (Pickering 1984). Despite these findings, the molecular mechanisms underlying the observed increase in haploid frequency following post-fertilization temperature elevation remain yet to be analyzed.

### **1.3.3. CENH3 manipulation in crop plants**

The manipulation of CENH3 for haploid induction in crop plants provided more insights into the mechanism of CENH3-mediated genome elimination (Table 1). In maize, Kelliher et al. (2016) evaluated transgene complemented knockout lines for CENH3 and RNAi for haploid induction rates through outcrossing. Surprisingly, only CENH3 knockout [cenh3(-/-)] lines with a GFP-tailswap transgene in hemizygous condition

exhibited significantly higher maternal haploid production compared to background levels, while no aneuploid embryos were observed. The observations mentioned above faced criticism due to the recovery of mutant alleles in homozygous condition, contradicting the embryo lethality of a true knockout CENH3 allele, as pointed out by Wang et al. (2021). Moreover, Wang et al. (2021) noted that a plant heterozygous for a null mutant (CENH3(+/-)) induced approximately 5% haploids and a low number of aneuploids when used as a female in crosses with wild-type plants. While this aligns with observations in *A. thaliana* that the inducer is more effective as a female (Ravi and Chan 2010), it contradicts the findings of Kelliher et al. (2016), who reported a higher haploid frequency when the inducer was used as a male parent.

In wheat, CRISPR/Cas9-based genome engineering led to the development of novel alleles of CENH3 termed as Restored Frameshift (RFS) alleles with modifications only in the N-tail and an intact HFD (Lv et al. 2020). Unlike in maize, knockout CENH3 alleles in heterozygous state failed to produce haploids in wheat. Interestingly, RFS CENH3 alleles exhibited a higher frequency of haploid induction in heterozygous state compared to homozygous condition.

Gao et al. (2020) demonstrated that RNAi and *in-vitro* inhibition of CENH3 resulted in pseudo-male gametes, leading to haploids and aneuploids upon selfing in cotton. When used as a pollen donor, RNAi CENH3 transformants produced around 8% of haploid progenies. However, the study hasn't done an extensive analysis of different primary transformants for their haploid induction ability.

Interesting observations were made from RNAi-mediated downregulation of CENH3 in onion (Manape et al. 2024). The study revealed a distinct correlation between the reduction in CENH3 transcript and protein levels and a consequential decrease in seed setting upon self-pollination. Homozygosity of the RNAi cassette was not obtained, possibly due to lethality caused by higher levels of CENH3 downregulation. Intriguingly, when employed as male parents in crosses with wild-type plants, the RNAi CENH3 lines induced around 5% haploids. The similarity between the RNAi strategy in both cotton and onion is that they behave as a more efficient haploid inducer when used as a pollen donor.

A range of mutants carrying amino acid substitutions for CENH3 were generated using base editing in carrot (Meyer et al. 2023). From the progenies of CENH3 mutants

crossed with wild-type male, two candidate plants were found to be potential products of UGE. Dosage analysis based on whole-genome sequencing revealed a complete loss of chromosome 7 and a partial loss of chromosome 9. Another candidate was found to be a tetraploid, and marker analysis revealed that all four copies of the chromosome were of wild-type parent origin. The study demonstrated that, despite their low efficiency, the CENH3 missense mutants could induce chromosome or genome-level elimination.

A suite of mutations carrying missense mutation, amino acid deletion and knockouts were generated for the functional copy of CENH3 in soybean (Wang et al. 2023b). While the null heterozygote did not yield haploids, an intriguing outcome was the production of one aneuploid ( $2n-2$ ) among 18 F1 plants resulting from crosses with wild-type plants. Marker analysis revealed that the two lost chromosomes were inherited from the null heterozygote. Combining the results from carrot and soybean, it is evident that the CENH3 manipulation has the potential to induce UGE, given the availability of a highly efficient methodology.

#### **1.4. Non-centromere-based haploid induction systems**

This class of haploid induction systems is represented by mutants of genes other than CENH3 or kinetochore/centromere-related functions that possess haploid induction ability. These genes are mainly involved in the plant sexual reproduction process. Mutants of these genes were shown to induce haploids in a broad range of plant species (Table 2). Two important genes that controlled haploid induction ability have been identified from maize, namely, MATRILINEAL (MTL) and a DUF679 domain membrane protein (DMP) (Zhong et al. 2019; Kelliher et al. 2017).

MATRILINEAL (MTL), also known as Phospholipase A1 (PLA1) and NOT LIKE DAD (NLD) was shown to be the gene underlying the major QTL responsible for haploid induction in Stock 6 inducer in maize (Gilles et al. 2017; Kelliher et al. 2017; Liu et al. 2017). The gene codes for a sperm-specific phospholipase, and a high frequency of DNA fragmentation was found in the sperm nuclei in the pollen of mutant genotypes (Li et al. 2017). The mutant of this gene was found to be inducing haploids in monocot species like rice, hexaploid wheat, foxtail millet and barley. However, MTL does not appear to have any orthologs in plants beyond monocots, making it impossible to utilize this strategy in dicots. Jiang et al. (2022) showed that the reactive oxygen

species (ROS) burst leads to sperm DNA fragmentation and also demonstrated that treating pollen with ROS-inducing reagents could induce haploids. The pollen treatment with methimazole (peroxidase inhibitor) or phosphatidylcholine (ROS inducer) described by Jiang et al. (2022) could still be utilized in dicots and yet remains to be analyzed (Ruban and Houben 2022). Additionally, Jiang et al. (2022) identified a pollen-specific peroxidase, POD65, gene mutant as another factor capable of inducing haploids. Another pollen-specific phospholipase, namely phospholipase D3 (ZmPLD3) was reported to trigger maternal maize haploids up to 1.2% in the knockout state (Li et al. 2021).

**Table1:** Different centromere-based systems utilized in plant species to induce haploids. The maximum reported haploid induction frequency is given for each type of modifications in different species. \*Paternal or maternal refers to the parental origin of the haploids.

Plant species	Strategy	Haploid frequency (%)		References
		Paternal*	Maternal*	
<i>Arabidopsis thaliana</i>	GFP-tailswap	34	5	(Ravi and Chan 2010)
	Point mutations (missense)	23.81	0	(Kuppu et al. 2015; Kuppu et al. 2020; Karimi-Ashtiyani et al. 2015)
	In frame deletions	25.7	0	(Kuppu et al. 2020)
	Orthologous CENH3 complementation	11	0	(Maheshwari et al. 2015)
	KNL 2 mutation	1	0	(Ahmadli et al. 2023)
Maize ( <i>Zea mays</i> )	Heterozygous null mutant	5	0.5	(Wang et al. 2021)
Wheat ( <i>Triticum aestivum</i> )	Restored frame shift (RFS) CENH3 allele	7	0	(Lv et al. 2020)
Cotton ( <i>Gossypium hirsutum</i> )	RNAi	NA	8	(Long et al. 2024)
Onion ( <i>Allium cepa</i> )	RNAi	2.68	4.63	(Manape et al. 2024)
Carrot ( <i>Daucus carota</i> )	Point mutations (missense)	0	0	(Meyer et al. 2023; Dunemann et al. 2022)
Soybean ( <i>Glycine max</i> )	Heterozygous null mutant	0	0	(Wang et al. 2023b)
Broccoli ( <i>Brassica oleracea</i> var. <i>italica</i> )	RFS alleles	1.29	0	(Gao et al. 2020)

**Table 2:** List of non-centromere-based haploid induction systems in different plant species. The maximum reported haploid induction frequency is given for the mutants in different species. \*Paternal or maternal refers to the parental origin of the haploids.

Gene	Species	Haploid frequency (%)		References
		Paternal*	Maternal*	
MTL/PLA1/ NLD	Maize ( <i>Zea mays</i> )	0	6.7	(Gilles et al. 2017; Kelliher et al. 2017; Liu et al. 2017)
	Rice ( <i>Oryza sativa</i> )	0	6.3	(Jang et al. 2023a; Liu et al. 2024b; Yao et al. 2018)
	Wheat ( <i>Triticum aestivum</i> )	0	15.6	(Liu et al. 2020)
	Foxtail millet ( <i>Setaria italica</i> )	0	2.8	(Cheng et al. 2021)
	Barley ( <i>Hordeum vulgare</i> )	0	16.4	(Tang et al. 2023)
DMP8/9	Maize ( <i>Zea mays</i> )	0	0.3	(Zhong et al. 2019)
	<i>Arabidopsis thaliana</i>	0	2.5	(Zhong et al. 2020)
	Rapeseed ( <i>Brassica napus</i> )	0	4.44	(Zhong et al. 2022b; Li et al. 2022)
	Tobacco ( <i>Nicotiana tabacum</i> )	0	1.75	(Zhang et al. 2022b; Zhong et al. 2022b)
	Tomato	0	3.68	(Zhong et al. 2022a)
	Potato ( <i>Solanum tuberosum</i> )	0	0.01	(Zhang et al. 2022a)
	<i>Medicago truncatula</i>	0	0.82	(Wang et al. 2022a)
	Cotton ( <i>Gossypium hirsutum</i> )	0	1.06	(Long et al. 2024)
	<i>Brassica oleracea</i>	0	2.35	(Zhao et al. 2022)
	Watermelon	0	1.12	(Chen et al. 2023)
PLD3	Maize ( <i>Zea mays</i> )	0	1.2	(Li et al. 2021)
POD65	Maize ( <i>Zea mays</i> )	0	7.7	(Jiang et al. 2022)
Kokopelli	<i>Arabidopsis thaliana</i>	0	0.34	(Jacquier et al. 2023)
ECS1/2	<i>Arabidopsis thaliana</i>	0	1.1	(Mao et al. 2023; Zhang et al. 2023a)
	Rice ( <i>Oryza sativa</i> )	0	3.5	(Zhang et al. 2023a)
pPLA1ly	<i>Arabidopsis thaliana</i>	0	1.12	(Jang et al. 2023b)

Haploid induction in a non-Stock6 background has been attributed to the mutation in another gene, DMP (Zhong et al. 2019). This gene is identified as a pollen-specific plasma membrane-localized protein. Unlike MTL, DMP exhibits multiple orthologues in dicots. Notably, the knockout of two pollen-expressed DMP genes, namely AtDMP8 and AtDMP9, was demonstrated to synergistically induce haploids at a rate of approximately 2.5% in *A. thaliana* (Zhong et al. 2020). The functional role of these membrane proteins was unveiled, indicating their involvement in promoting the fusion of egg cells and sperm. This fusion is facilitated by sperm adhesion mediated by the egg cell-secreted protein, EGG CELL 1 (EC1) (Cyprys et al. 2019; Wang et al. 2022b). Mutation of DMP orthologs was found to be inducing haploids in many crop species (Table 2). However, it is not evident whether the mutation of DMP induces spontaneous development of egg cell into haploid embryos or causes a post-fertilization genome elimination.

Jacquier et al. (2023) searched for genes exclusively expressed in male gametophytes and mutation, resulting in single-fertilization events in Arabidopsis. They identified the gene, KOKOPELLI (AT5G63720) as a potential candidate and a mutant of this gene was found to trigger maternal haploids at a rate of 0.34%, which is comparatively lower than the rate caused by DMP mutants. A genome-wide association study conducted in maize revealed KOKOPELLI to be one of the candidate genes involved in haploid induction (Trentin et al. 2023).

Haploid induction in Arabidopsis and rice has been demonstrated through the mutation of egg cell-specific aspartic endopeptidases (ECSs) (Mao et al. 2023; Zhang et al. 2023a). Additionally, the knockdown of gynoecium-specific phospholipase, pPLAIIy, has been found to trigger the formation of maternal haploids in approximately 1% of the progenies (Jang et al. 2023b). Notably, both of these mutants exclusively induce haploids of the mutant parent, rather than those of the wild-type crossing partner. A similar limitation is observed in transgene-induced parthenogenesis achieved by the egg cell-specific expression of embryogenesis factors such as Baby Boom (Conner et al. 2015; Chen et al. 2022). As a consequence, these haploid induction methods have limited significance in crop breeding programs, given their inability to generate haploids from wild-type crossing partners.

### 1.5. Comparison of both haploidization systems

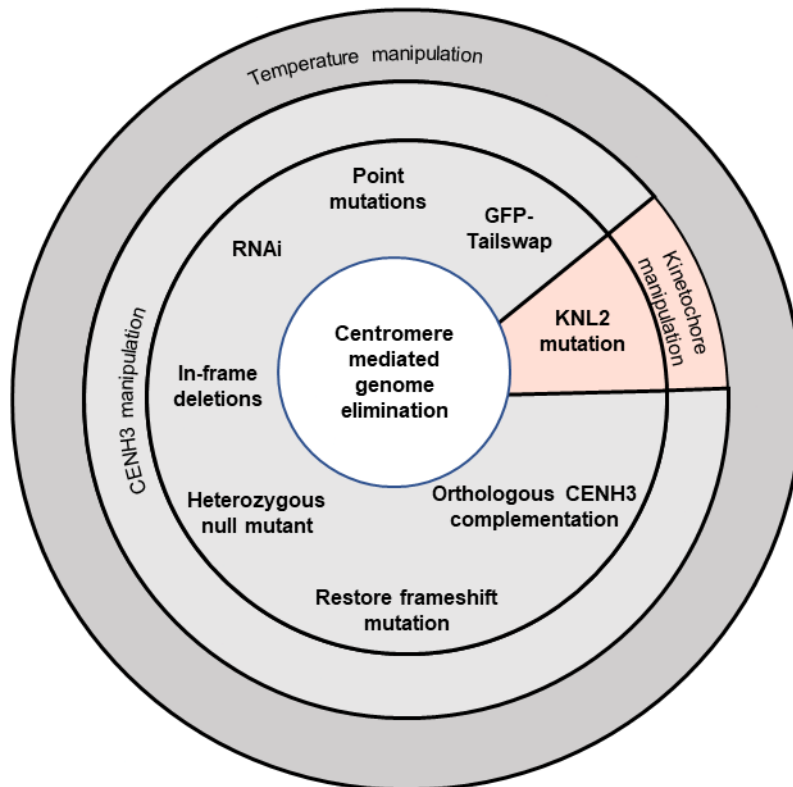
Comparing both systems, centromere-based approaches outweigh the other approaches in terms of the type of haploids induced and their applicability. All known non-centromere-based inducers are capable of inducing haploid plants of maternal origin (Table 2). However, centromere manipulation can induce maternal and paternal haploids (Table 1). In case of applicability, non-centromere-based approaches can be utilized only in a limited class of crop plants. For example, MTL has no equivalent orthologs in dicots and, therefore, is limited to monocots (Shen et al. 2023; Zhong et al. 2020). Whereas DMP mutants were found to be effective in various dicots, so far, no reports have been available from monocots except maize (Zhong et al. 2019). A recent study found that only MTL but not DMP orthologues were able to trigger haploid formation in rice (Liu et al. 2024b). On the other hand, centromere manipulations were shown to be effective in dicots like *Arabidopsis*, cotton and equally capable in monocots like wheat, maize and onion. Hence, centromere-based systems can be considered as having the properties of broad applicability.

Concerning haploid induction rates, MTL mutants in various crop species exhibit induction rates ranging from 2.8% (in foxtail millet) to 16% in barley, while achieving a 5% rate with DMP mutants is never reported (Table 2). Interestingly, CENH3 manipulation has demonstrated HIR of at least 5% in maize and more than 5% in wheat and cotton (Table 1). However, these rates are notably lower than those reported in *Arabidopsis*, particularly with GFP-tailswap line, which have shown haploid induction rates of up to 96% through temperature manipulation (Jin et al. 2023).

Notably, a single strategy of CENH3 manipulation is not universally applicable across different crops. For instance, GFP-tailswap modification of CENH3 failed to complement the null mutant in maize (Wang et al. 2021). Heterozygous null mutants could induce haploids in maize (Wang et al. 2021), with a much lower rate in *Arabidopsis* (Marimuthu et al. 2021), and not at all in wheat (Lv et al. 2020). Additionally, RNAi-mediated downregulation of CENH3 could induce haploids in cotton and onion, but this approach did not prove effective in *Arabidopsis* (Ahmadli et al. 2023) and maize (Kelliher et al. 2016).

Given the myriad of strategies available (Fig. 2), researchers may find it challenging to evaluate all possibilities of centromere manipulation in a new crop species.

Therefore, there is a need for a strategy that is universally effective with high efficiency across all crop species. To develop such a strategy, it is crucial to understand the mechanistic reason behind centromere-mediated UGE as best as possible.



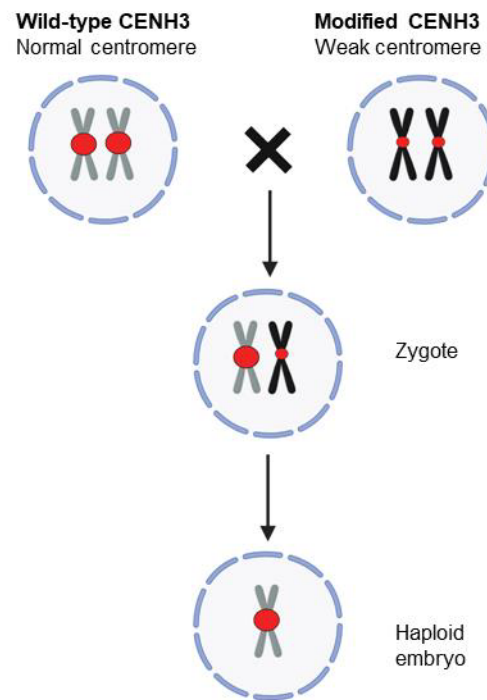
**Figure 2: Different strategies of manipulation of CENH3 protein or other kinetochore proteins for haploid induction.** An additional level of optimization is required to find the suitable temperature for efficient haploid induction.

### 1.6. Mechanism of centromere-based haploidization

The centromere size model (Wang and Dawe 2018) provides a possible mechanistic reason behind centromere-mediated UGE. It was postulated based on the observation that in most of the wide crosses causing UGE, only the chromosomes with smaller centromeres get eliminated. According to the model, genetic alterations to CENH3 impair its centromere loading capacity and reduce the CENH3 chromatin, leading to a smaller centromere size. When a plant carrying chromosomes with such small or 'weak' centromere is crossed with a plant with normal (wild-type level) centromeres, the chromosomes from plants with weak centromeres get eliminated, resulting in haploids (Fig. 3). The elimination primarily results from an inability to restore the wild-



type centromere size, as the small centromeres fail to compete effectively for the limiting CENH3 reloading factors (Marimuthu et al. 2021).

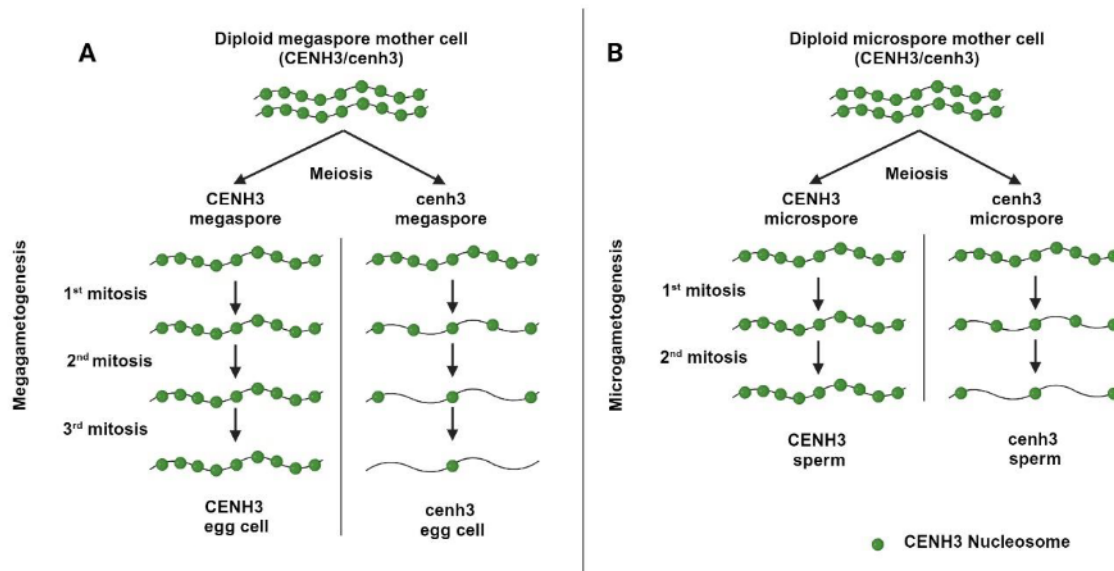


**Figure 3: Centromere size model of haploid induction.** Certain modifications to CENH3 reduce the average centromere size and functionality of the centromere. When plants with small centromeres are fertilized with wild-type gametes with larger centromeres, zygotes show centromere dimorphism in terms of size; often, the chromosomes having small centromeres are eliminated to form haploid plants. Figure was adapted from Wang and Dawe (2018) and prepared using Biorender.

The hypothesis only holds true if the parental CENH3 levels are maintained during early embryogenesis. When all the chromosomes of the haploid inducer fail to reach wild-type centromeric CENH3 levels, they get eliminated and produce haploids, whereas restoration of the level of CENH3 containing nucleosomes in some or all chromosomes during early embryogenesis results in aneuploid and diploid progenies, respectively. A study conducted in mice proposes similar strong and weak centromere states as determined by the CENP-A nucleosome (CENH3 in mammals) levels being the reason behind the female meiotic chromosome drive (Iwata-Otsubo et al. 2017).

As an experimental proof for the centromeric size model, a CENH3 dilution approach was utilized in which maize plants heterozygous for a *cenh3* null mutant induced haploid offspring (Wang et al. 2021). The authors suggested that gametes carrying

null allele undergo dilution of wild-type CENH3 during gametogenic mitosis, resulting in reduced centromere size with fewer CENH3-containing nucleosomes. The female gametophyte undergoes three rounds of mitosis, whereas the male gametophyte undergoes only two rounds (Fig. 4).

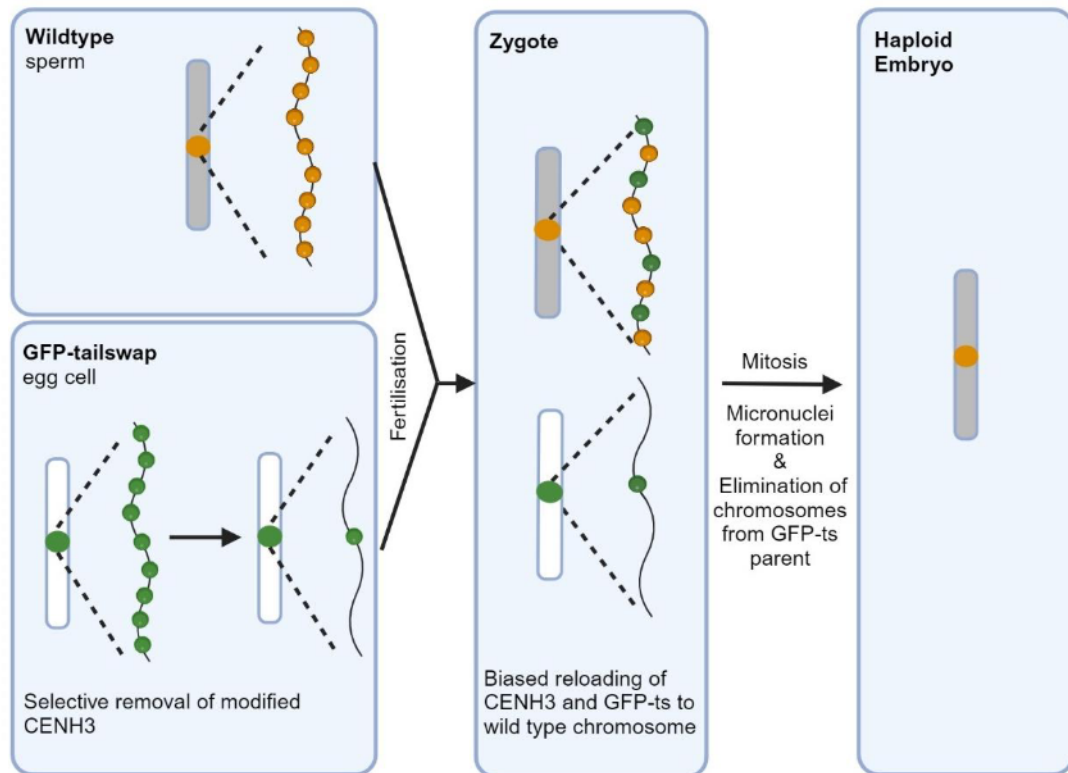


**Figure 4: CENH3 dilution in female (A) versus male (B) gametes.** The egg cell would have comparatively less CENH3 nucleosomes (in green) compared to sperm due to an additional mitotic division. Figure was prepared using Biorender.

Therefore, dilution would occur more in the female gametophytes carrying *cenh3* allele compared to their male counterparts. The egg cells may have CENH3 below the critical level to form a functional centromere compared to CENH3 of sperm cells, resulting in higher haploid frequency when a heterozygous plant is used as a female parent. However, the model does not consider the potential hindrance of gametophytic CENH3 dilution by maternally transmitted CENH3 transcripts and proteins from the sporophyte.

In the case of Arabidopsis, it was suggested that the chimeric GFP-tailswap protein was selectively removed from the egg cells (Marimuthu et al. 2021). When the GFP-tailswap was crossed as a female to a wild-type plant, CENH3 and GFP-tailswap proteins were found to be biasedly re-loading onto the chromosomes from the wild-type parent. The chromosomes from the GFP-tailswap partner failed to reassemble CENH3 nucleosomes and were lost during embryogenesis, leading to haploidization

(Fig. 5) (Marimuthu et al. 2021). Yet, it is not clear whether the GFP-tailswap protein is selectively eliminated from the egg and central cell nuclei, or if it is extensively diluted during gametophytic mitosis.



**Figure 5: Proposed mechanism of GFP-tailswap mediated genome elimination by Marimuthu et al. (2021).** The GFP-tailswap is selectively evicted from the centromeres of the matured egg cell. After fertilization with a wild-type sperm, preferential loading of CENH3 and GFP-tailswap takes place onto the wild-type centromeres. In the following mitosis, chromosomes with 'weak' centromeres are eliminated and only those from the wild-type parent is retained, resulting in a wild-type haploid embryo. Centromere loading with the GFP-tailswap protein stops after the female chromosome set encoding the GFP-tailswap construct is eliminated.

An additional study described another mutant of CENH3 in Arabidopsis, namely *cenh3-4* (Capitao et al. 2021). In this mutant, splicing of CENH3 transcripts is altered, resulting in a strong reduction of centromeric CENH3 in sporophytic tissues. Despite this, the mutant exhibits very low haploid induction ability (Capitao et al. 2021). Consequently, the authors concluded that centromere size may not be the key determinant of haploid induction efficiency. It is noteworthy, however, that the study did not assess CENH3 levels in gametophytes, rendering the conclusion incomplete.

However, independent analysis of CENH3-based haploid inducers in maize and Arabidopsis underscores the essential role of depleting CENH3 in gametes for successful haploid induction. Therefore, to validate the centromere size model, a molecular tool that enables the targeted depletion of CENH3, specifically in the gametophyte, is needed. This approach would be helpful in formulating a universal strategy to manipulate centromere for haploid induction.

### **1.7. Depletion of CENH3 protein in the gametophyte**

Protein depletion is necessary to understand the function of a specific protein in a cellular context, which is not possible using *in-vitro* analysis (Roth et al. 2019). Protein depletion can be achieved either at DNA, RNA or protein levels. Null mutations of CENH3 are lethal in homozygous state. In maize, the depletion strategy involved utilizing a heterozygous null mutant, assuming it dilutes the CENH3 in gametes carrying the null allele to a higher extent in egg cells (Wang et al. 2021). Nonetheless, this model fails to consider the contribution of CENH3 transcript or protein from the maternal pool.

Plants constitutively expressing RNAi constructs for CENH3 were shown to induce haploids in onion and cotton (Manape et al. 2024; Gao et al. 2020). However, the constitutive RNAi approach affected the plant vigour. While tissue-specific downregulation through RNAi can be accomplished using tissue-specific or cell-specific promoters, the turnover rate of the protein of interest (POI) stands as a limiting factor. For a highly stable protein such as CENH3 with very low turnover (Lermontova et al. 2011b), gametophytic RNAi would not be helpful. Therefore, gametophytic depletion of CENH3 can only be achieved through targeted protein degradation. Such a method was pioneered for *Drosophila melanogaster*. Here, engineered, sperm-specific, targeted protein degradation of CENH3 (Cid in *D. melanogaster*) resulted in gynogenetic haploid embryos (Raychaudhuri et al. 2012).

### **1.8. Targeted protein degradation**

Targeted protein degradation (TPD) approaches are highly helpful in depleting proteins where RNAi is incapable due to low turnover rates (Natsume and Kanemaki

2017; Roth et al. 2019). TPD approaches rely on the endogenous ubiquitin-proteasome system for proteolysis of the target and require the recruitment of the POI to an E3 ligase for ubiquitination. The requirement can be fulfilled using either (i) synthetic small molecules that has a bivalent affinity towards the POI and E3 ligase (Bondeson et al. 2015). (ii) tagging the POI with a degron sequence (Banaszynski et al. 2006; Nishimura et al. 2009) or, (iii) engineering E3 ligases with affinity to the POI (Caussin et al. 2011; Fulcher et al. 2017; Fulcher et al. 2016).

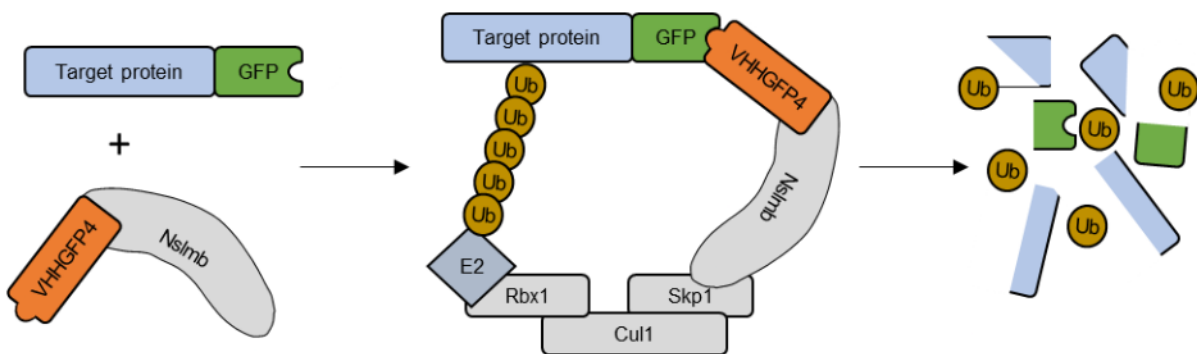
The usage of bivalent molecules poses significant challenges in terms of both the identification and delivery of these molecules into plant cells (Leon and Bassham 2024). In the case of tagging the POI with a degron sequence, auxin-based degron systems exhibit highly efficient protein degradation (Nishimura et al. 2009); however, their applicability in plant cells is limited. Daum et al. (2014) utilized the N-degron-based conditional protein degradation to degrade the Wuschel protein using the TEV protease. A primary drawback of adding a degron to the POI is the potential alteration of protein behaviour due to adding the degron tag.

The third approach, where E3 ligases were engineered for affinity against the POI, was shown to be effective in plant systems (Baudisch et al. 2018; Ma et al. 2019; Sorge et al. 2021). The E3 ligase, also known as an E3 ubiquitin ligase, is a protein that catalyzes the transfer of ubiquitin from the E2 enzyme to the specific protein substrate. Therefore, E3 ligases provide the substrate specificity to the ubiquitin-proteasome degradation. Plants encode a remarkably higher number of E3 ligases compared to animals (Gibbs et al. 2014). For example, the *A. thaliana* genome encodes for more than 1500 E3 ligases (Hua and Vierstra 2011). The E3 ligase can be customized to a target POI by substituting its natural substrate binding domain with domains capable of binding to the POI.

The antigen-binding properties of antibodies can be used for affinity engineering of E3 ligases to the POI. The coding sequence of the antibody molecule is absolutely necessary for this purpose. Phage display screening (Smith 1985) of an antibody library has been used for the selection of high-affinity antibodies (McCafferty et al. 1990). Of the commonly used antibodies, nanobodies represent the smallest (13 - 15 kDa) antibody-binding fragment (Muyldermans 2001). The term nanobodies (VHH) refers to the single domain antibodies, the variable domain of the heavy chain-only

antibodies of *Camilidae* species. The nanobodies are highly soluble, remarkably stable, and characterized by better penetration ability than conventional antibodies (Ingram et al. 2018; de Beer and Giepmans 2020).

The initial breakthrough in tailoring specificity for the 'SKP1-CUL1-F-box' (SCF) E3 ligase of *Drosophila* involved substituting the substrate recognition sequences of a F-box protein (Nslmb) with a GFP-specific nanobody (VHHGFP4), resulting in the creation of the fusion protein termed Nslmb-VHHGFP4 (Caussinus et al. 2011). This engineered F-box protein was first employed in various degradation experiments of GFP-tagged proteins (deGradFP) in *Drosophila* and human cells (Fig. 6) (Caussinus et al. 2013; Raychaudhuri et al. 2012; Daniel et al. 2018). Despite its success, the deGradFP method encountered limitations in degrading histone H2B-GFP, a nuclear protein. Subsequent efforts to devise novel synthetic ligases revealed that a modified SPOP protein, a human derived E3 ligase, demonstrated greater efficiency in the degradation of nuclear proteins compared to Nslmb-VHHGFP4 (Shin et al. 2015).



**Figure 6: Schemata of the deGradFP system for the targeted degradation of GFP-tagged proteins of interest.** When the GFP nanobody-E3 ligase fusion protein (Nslmb-VHHGFP4) is co-expressed with GFP-tagged targeted protein, it leads to polyubiquitination of the target protein mediated by the SCF complex. Next, the polyubiquitinated target protein is degraded by the proteasome. E2 – ubiquitin-conjugating enzyme, Rbx1 – Ring box protein, Cul1 – cullin 1, Skp1 –Adapter protein between Cul1 and Nslmb, Ub – Ubiquitin.

As a first attempt in plants, Baudisch et al. (2018) showed that the cytosolic GFP protein can be degraded using Nslmb-VHHGFP4 without affecting the RNA expression levels in stably transformed *Nicotiana tabacum* plants. Ma et al. (2019) utilized Nslmb-VHHGFP4 to degrade the GFP-tagged WUS proteins in the stem cells using an ethanol-inducible system in *A. thaliana*. The modified SPOP-based

degradation system was utilized for the degradation of nuclear-localized GFP-fusion proteins in *N. tabacum* and *N. benthamiana* (Sorge et al. 2021). Hence, animal-derived E3 ligases were proven to be functional in plants.

### **1.9. Targeted CENH3 degradation for haploid induction in Arabidopsis**

Sorge et al. (2021) successfully demonstrated *in-planta* degradation of EYFP-CENH3 through nanobody – guided proteosomal degradation. Demidov et al. (2022) showed that the targeted proteosomal degradation of EYFP-CENH3 during early embryogenesis can induce paternal haploids in Arabidopsis. However, the haploid induction rates (~3%) were lower compared to that of the GFP-tailswap strategy. Also, the study involved genetic modification of both crossing partners and, therefore was less significant in terms of breeding application as haploids were transgenic. A similar degradation system, spatio-temporally restricted to gametes, would be necessary to deplete CENH3 in the gametes and has the potential to be used as a universal strategy of CENH3 manipulation for haploid induction (Marimuthu et al. 2021).

## 2. Aims

Haploid induction accelerates the plant breeding process by reducing the number of breeding cycles required to achieve homozygosity. Homozygous lines exhibit consistent expression of traits, allowing for a more accurate selection of desirable characteristics. However, for breeding programs to be cost-effective, an efficient *in-planta* haploid induction method is essential.

It's been more than a decade since it was shown that the centromere manipulation can induce a higher frequency of haploids in *A. thaliana* (Ravi and Chan 2010). Despite a few successes, the frequency of haploids produced from centromere-manipulation remains low in crop plants. Additionally, different centromere manipulation strategies are required for each crop species, posing a major challenge to centromere-based haploid induction approaches.

Therefore, I attempted to develop an *in-planta* haploidization strategy based on the cellular mechanism behind the centromere-mediated uniparental genome elimination. The centromere size model provides a strong theoretical explanation of the mechanism. This study aimed to validate the centromere size model and gain additional insights using a haploid-inducing Arabidopsis genotype different from the one investigated by Marimuthu et al. (2021). Based on the conclusion drawn, a universal strategy grounded in engineering principles can be formulated.

Hence, the aims of this study are,

- To experimentally validate the centromere size model using a fluorescence-tagged CENH3-based haploid induction Arabidopsis genotype.
- To develop a gametophyte-specific protein degradation system targeting CENH3 protein in Arabidopsis and evaluate its haploid induction ability.



### 3. Materials and methods

#### 3.1. Plant Materials

##### 3.1.1. Description of Arabidopsis mutant alleles

***cenh3-1***: The *CENH3* null mutant described in the Columbia (Col-0) ecotype background of *Arabidopsis thaliana* (Ravi et al. 2010). Due to the lethality of the homozygous null [*cenh3(-/-)*], the mutation was maintained as heterozygotes [*CENH3(+/-)*]. Heterozygous lines were used for transformation with complementation constructs to recover the homozygous null mutant in the T1 generation.

***gl1-1***: A recessive glabrous (trichome-less) mutant of the gene *GL1* in the genetic background of Ler-ecotype (Oppenheimer et al. 1991; Koornneef et al. 1982)

***ams***: Recessive sporophytic male sterility due to a T-DNA insertion (SALK\_152147) in the *Abortive Micro Spores* gene in Columbia background (Xu et al. 2010). This mutation is maintained in heterozygotic state [*AMS(+/-)*] by using the kanamycin selection marker in T-DNA.

##### 3.1.2. Description of the transgenic construct from previous studies

***EYFP-gCENH3***: A transgenic construct carrying the genomic version of *AtCENH3* fused with *EYFP* coding sequence, regulated by the native *CENH3* promoter and terminator, as described by Le Goff et al. (2020).

#### 3.2. Nomenclature of genotypes

The genotype of a transgenic/mutant line is given within square brackets. The allelic status of each gene or transgene were given within curved brackets. For example, [*cenh3(-/-)*, *EYFP-gCENH3 (+/+)*] refers to the *CENH3* homozygous null mutant complemented with the *EYFP-gCENH3* transgene.

### **3.3. Plant growth conditions**

Arabidopsis plants were grown under long-day conditions in a cultivation room (16 h light/8 h dark) at 21 °C. Plants that were used for crossings were moved to a Percival growth chamber (maintained at 21 °C, 16 h light/8 h dark) at two weeks prior to emasculation to ensure a constant temperature until the complete seed set. *Nicotiana benthamiana* plants were grown at 22°C in a greenhouse under 16h light/ 8h dark conditions.

### **3.4. Preparation of transgenic constructs**

Except for the egg cell marker, all transgenic constructs were prepared using the Golden-Gate-based modular cloning (MoClo) toolkit (Engler et al. 2014). The MoClo toolkit was a gift from Sylvestre Marillonnet (IPB, Halle, Addgene kit # 1000000044). Three levels of cloning, namely level-0, level-1 and level-2 from the MoClo kit were used in this study. If not otherwise mentioned, PICSL4723 vector was used as a level-2 destination binary vector in the case of Modular cloning system. PICSL4723 was a gift from Mark Youles (The Sainsbury laboratory). The Golden-Gate reaction was set up using the protocol described by Grützner and Marillonnet (2020).

The egg cell marker, EC1.1pro:H2B-tdTomato, was prepared using multisite Gateway cloning as instructed by the manufacturer (LR Clonase™ II Plus enzyme, Thermo Fischer). pGWB501 was used as a destination vector for multisite gateway cloning.

The level-0 and entry vectors, their corresponding primer sequences and DNA templates are listed in the Appendix 1. The modules used to construct level 1 and level 2 plasmids are given in the Appendix 2 and Appendix 3, respectively. The final level-2 constructs were introduced into *Agrobacterium tumefaciens* strain GV3101 through electroporation, for stable transformation of *A. thaliana* and transient expression in *N. benthamiana*.

### **3.5. Generation of transgenic Arabidopsis plants**

The Arabidopsis plants were stably transformed using the floral dip method as described by Clough and Bent (1998). Transgenic T1 seeds were selected on ½ MS

(Murashige and Skoog 1962) plates supplemented with cefotaxime (100 µg/ml) and either kanamycin (50 µg/ml), hygromycin (15 µg/ml) or phosphinothricin (20 µg/ml) under long-day conditions (16 h light/8 h dark) at 21 °C.

### **3.6. Transient expression in *Nicotiana***

Three-week-old *N. benthamiana* plants were transiently transformed with *Agrobacterium tumefaciens* carrying appropriate constructs, diluted to an OD of 0.8 with infiltration buffer (10 mM MgCl<sub>2</sub>, 10 mM MES). The infiltrated plants were incubated for 48 hours under greenhouse conditions and used for either confocal microscopy or flow cytometry analysis.

### **3.7. Plant genomic DNA isolation**

DNA was isolated from the Arabidopsis plants in 8-well strip tubes following the method described by Kasajima et al. (2012) with slight modifications. After grinding the leaf material in the extraction buffer (200 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS), the DNA was precipitated using an equal volume of ice-cold isopropanol. The precipitated DNA was washed once using 70% ethanol and air-dried. The DNA pellet was then dissolved in 50 µl of PCR-grade water. DNA for short-read sequencing was isolated using the Nucleospin Plant II kit (Machery Nagel).

### **3.8. PCR genotyping**

PCR genotyping of transgenic plants was performed using transgene-specific primers listed in Appendix 4. The list of sequences of all the primers used in this study are given in Appendix 5. The *cenh3-1* mutation was genotyped using dCAPS marker based on *EcoRV* digestion of PCR products (Ravi et al. 2010).

### **3.9. SNP genotyping**

Genomic DNA was sequenced using short-read sequencing technology at BGI laboratory in Poland. Adapter-trimmed, high-quality reads were aligned to the TAIR10

assembly using BWA-MEM2 (Vasimuddin et al. 2019). The obtained alignments in SAM format were converted to BAM format, followed by sorting, deduplication, and quality filtering (Q value  $\geq 20$ ) using SAMtools (Danecek et al. 2021). SNP calling was performed using BCFtools (Danecek et al. 2021). Biallelic SNPs with a quality score greater than 40 and at least 5x coverage in all samples were selected, and their alternate allele frequencies were used to make genotype calls.

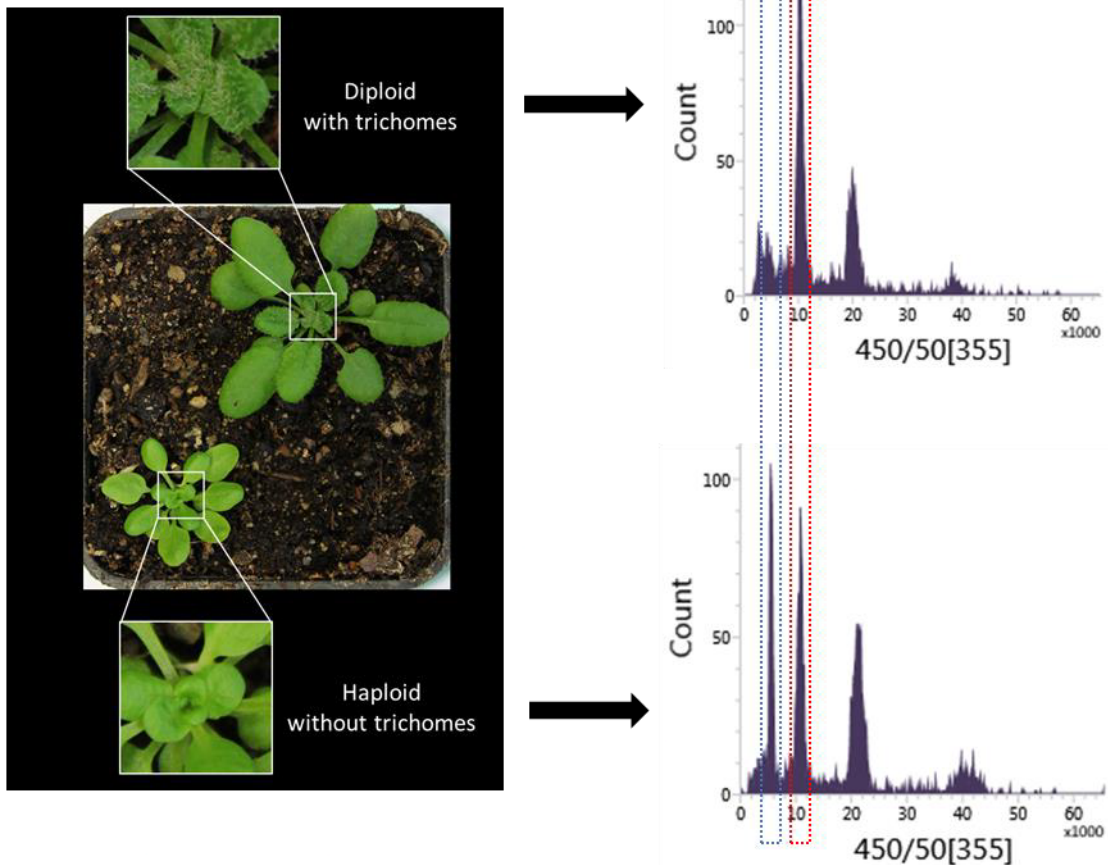
A homozygous variant genotype call was made if the alternate allele frequency was  $>95\%$ . A homozygous reference genotype call was made if the alternate allele frequency was  $<5\%$ . Heterozygous calls were made if the alternate allele frequency was between 40-60%. Graphical genotypes using the genomic SNPs polymorphic among the parents were generated using the "ggplot2" R package.

### **3.10. Crossing and haploid screening**

The crossing and haploid screening were performed as described by Kuppu et al. (2015). Briefly, the plants for which haploid induction was to be assessed were crossed with [*gl1(-/-)*] plants. All seeds from these crosses were sown in individual pots, and seedlings were screened for glabrous plants after 3 weeks of sowing. A few randomly selected plants were then analysed by flow cytometry (Fig. 7).

### **3.11. Purification of AtCENH3-Ntail antigen**

The nucleotide sequence corresponding to the amino acids of the N-terminus tail of *AtCENH3* was amplified and cloned into the pET-22b plasmid using *NotI* and *NcoI* sites. The final expression construct was transformed into Rosetta™ (DE3) strain of *E. coli* (Novagen). The recombinant protein was expressed and purified under native conditions following the manufacturer's instruction described in the Ni-NTA Fast Start Kit (Qiagen). Desalting of the elution fractions was performed using Amicon Ultra-3K spin columns (Merck Biosciences) and contaminating proteins larger than the recombinant CENH3 protein were removed using Amicon Ultra-30K spin columns (Merck Biosciences). The purified protein was verified using western blot using rabbit anti-CENH3 antibody.



**Figure 7: Haploid screening using glabrous leaf marker.** On the left, representative hairy (with trichomes) and glabrous (trichome-less) plants are shown, with zoomed-in views to clearly indicate the presence or absence of trichomes. Corresponding flow cytometry histograms are shown on the right. The red dotted boxes indicate the diploid 2C peak, while the blue dotted boxes refer to the haploid 1C peak. The haploid peak is only seen in the trichome-less plants, confirming their haploid status.

### 3.12. Phage display for affinity selection

The synthetic nanobody libraries, previously described by Gahrtz and Conrad (2009), were screened for affinity against AtCENH3 antigen using the protocol described by Gahrtz and Conrad (2009). The immune libraries were constructed and screened for affinity against AtCENH3 following the standard operating procedures at Nanotag Biotechnologies (Göttingen, Germany).

### **3.13. Immunostaining**

Young leaves were fixed with 4% paraformaldehyde in Tris-buffer (10 mM Tris, 100 mM NaCl, 10 mM EDTA, 0.1% Triton X-100, pH7.5) on ice under vacuum for 5 minutes, followed by 25 minutes without vacuum. The material was then washed twice with Tris-buffer without paraformaldehyde. The fixed leaves were finely chopped in 450  $\mu$ L of LB01 buffer (15 mM Tris, 2 mM Na<sub>2</sub>EDTA, 0.5 mM spermine, 80 mM KCl, 20 mM NaCl, 15 mM  $\beta$ -mercaptoethanol, and 0.1% Triton X-100, pH7.0) (Doležel et al. 1989). The nuclei suspension was filtered through a 50  $\mu$ m mesh. The filtered suspension was diluted 1:10 by volume and 200  $\mu$ L of the homogenate was applied onto each glass slide using a Cytocentrifuge (Cytospin3, Shandon). The slides were washed twice with PBS and blocked with 4% bovine serum albumin (BSA) for 1 hour. Either rabbit anti-CENH3 antibody or anti-ALFA minibody (anti-ALFA nanobody fused to rabbit IgG1 Fc domain) was used as the primary antibody. Anti-rabbit secondary antibodies fused with rhodamine fluorophore were used for fluorescent labelling.

### **3.14. Preparation of early embryos for imaging**

The flowers were emasculated one day prior to anthesis and subsequently pollinated two days later to synchronize ovule development. After 36 hours of post-pollination, siliques were harvested and fixed using the BVO fixative (2 mM EGTA, 1% formaldehyde, 10% DMSO, 1x PBS, 0.1% Tween 20, pH 7.5) (She et al. 2018). At this stage, the seeds predominantly contained one-cell embryos. Siliques were gently opened using an insulin needle and immersed in the fixative solution. Following 5 minutes of vacuum treatment, the siliques were fixed for 25 minutes at room temperature. They were then rinsed with 1x PBST (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, 0.1% Tween-20). and stored overnight at 4°C. The following day, the young embryos were carefully popped out of the seeds before being embedded in a 15% acrylamide film on polylysine-coated slides (Garcia-Aguilar and Autran 2018). The embryos were stained with DAPI (10  $\mu$ g/mL) and SCRI Renaissance (SR2200) (0.1%), followed by three washes with 0.1x PBS. Finally, the embryos were mounted with a 50% glycerol solution and used for microscopic imaging.

### **3.15. Preparation of ovule, pollen, and anthers for imaging**

The ovules were mounted on slides using a medium of 50% glycerol and 0.1x PBS (13.7 mM NaCl, 0.27 mM KCl, 1 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.18 mM KH<sub>2</sub>PO<sub>4</sub>). The DAPI staining of mature pollen was performed following the method described before (Park et al. 1998). For quantification of pollen with and without EYFP-CENH3 signals, the pollen was mounted with a 0.3 M mannitol solution. Alexander staining of the anthers was performed according to the procedure described in Hedhly et al. (2018).

### **3.16. Microscopy**

Ovules, DAPI-stained pollen and early embryos of *Arabidopsis* and the transiently transformed *N. benthamiana* leaf nuclei were analysed for EYFP-CENH3 signals using a Zeiss confocal laser scanning microscope (LSM780). Images were captured as Z-stacks, and maximum intensity projections were generated using Fiji (ImageJ). The egg cell in the ovule was identified using H2B-tdTomato marker driven by the EC1.1 promoter. Early embryos were identified using their cell wall structure labelled with SCR1 renaissance (SR2200). H2B-mCherry marker served as nuclei marker to identify the transformed *N. benthamiana* leaf nuclei.

Immunostained *Arabidopsis* leaf nuclei and pollen samples mounted in 0.3 M mannitol were imaged using an Olympus BX61 epifluorescence microscope equipped with a Hamamatsu Orca ER CCD camera. *Arabidopsis* seeds and anthers after Alexander staining were imaged using Zeiss Axio zoom Stereo microscope.

### **3.17. Flow cytometry**

All the flow cytometry work in this study was performed using a BD Influx cell sorter. For the ploidy analysis of *Arabidopsis* seedlings, young leaves were finely chopped with a razor blade in CyStain UV Ploidy buffer (Sysmex). The homogenate was filtered through a 50 µm nylon filter and analyzed by flow cytometry. Haploids were confirmed based on their 1C nuclei peak (Fig. 7).

For the EYFP-CENH3 quantification in *Arabidopsis* pollen nuclei, the pollen was disrupted by vortexing for 2 minutes with metallic beads balls (4mm diameter) in LB01

buffer. The filtered pollen-nuclei suspension (using 50 µm mesh) was stained with propidium iodide (PI) (50 µg/ml) supplemented with DNase-free RNase (50 µg/ml) and used for flow-cytometric analysis. Sperm nuclei were differentiated from vegetative nuclei based on DNA/RNA content (PI fluorescence intensity) versus particle size (Forward Scatter).

To analyse the efficiency of plant-derived E3-ligases, transiently transformed *N. benthamiana* leaves were chopped in Galbraith buffer (45 mM MgCl<sub>2</sub>, 20 mM MOPS, 30 mM sodium citrate, 0.1 % (v/v) Triton X-100. pH 7.0) (Galbraith et al. 1983). The filtered homogenate was stained with DAPI (1.5 µg/ml) and analyzed by flow cytometry. The transformed nuclei were identified using H2B-mCherry marker. The EYFP-CENH3 intensities of different samples were then compared using the R package, CytoExploreR.

### **3.18. Statistical analysis**

Statistical comparison of haploid induction frequencies, shrivelled seed frequencies and frequency of phosphinothricin resistant seedlings were performed using t-test. The final p values provided on the bars represent the error-corrected p values for multiple comparisons using the Benjamini–Hochberg method.



## 4. Results

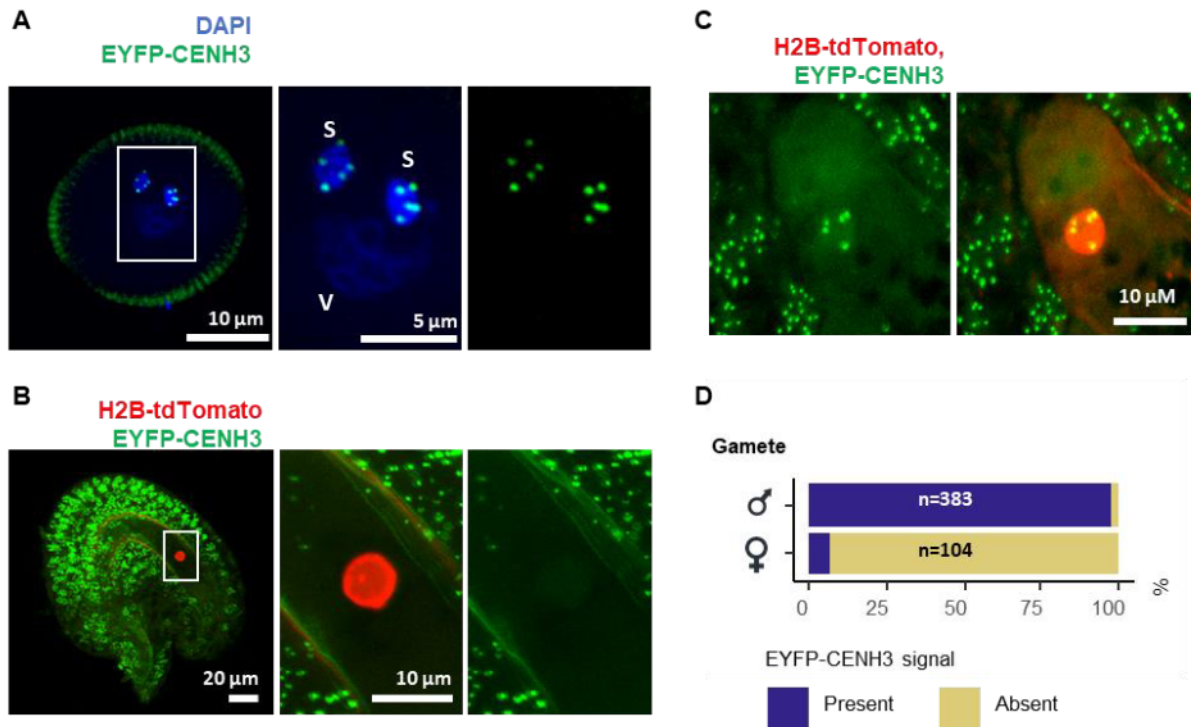
A previous study has identified epigenetic differences in centromeric CENH3 levels between the crossing partners as a contributing factor towards uniparental genome elimination (UGE) (Marimuthu et al. 2021). However, this investigation predominantly emphasized one direction of the UGE cross, specifically utilizing the GFP-tailswap mutant as the female inducer. The explanation for the reduced frequency of haploids, from 34% to 5% when the GFP-tailswap mutant was employed as the male partner (Ravi and Chan 2010) was not studied in detail. To gain additional insights into the mechanistic reason behind the biased UGE, the *cenh3* null mutant, [*cenh3(-/-)*] complemented with *EYFP-gCENH3* transgene was utilised (Le Goff et al. 2020). From previous experiments performed at our laboratory, it was found that the [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] genotype induces wild-type haploids when used as a female in a cross with wild-type plants but not when used as a male parent. This property of the haploid inducer was leveraged to investigate the mechanism behind UGE.

### 4.1. Different levels of EYFP-CENH3 in male and female gametes of *cenh3* null mutant complemented with EYFP-gCENH3

As an initial step, male and female gametes of [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] plants were analyzed to determine the presence or absence of EYFP-CENH3 signals. A maximum of 5 EYFP-CENH3 foci was expected in a haploid Arabidopsis gamete. When microscopically examined, around 97% of the mature pollen displayed 5 EYFP-CENH3 signals in both the sperm nuclei (Fig. 8A). Consistent with previous studies (Merai et al. 2014; Schoft et al. 2009), the vegetative nuclei chromatin lacked EYFP-CENH3 signals.

To facilitate the identification of egg cell nuclei, a transgenic fusion protein, histone *H2B-tdTomato*, specifically expressed in the egg cell using the *EC1.1* promoter (Sprunck et al. 2012) in the background of [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] was utilised. Two days after emasculation, unfertilized ovules were examined via confocal microscopy. It was found that 93.27% of egg-cells had no detectable EYFP-CENH3 signals (Fig. 8B). Only a small percentage of egg cell nuclei (6.73%) exhibited the presence of 4-5 EYFP-CENH3 signals (Fig. 8C). quantitative analysis of EYFP-

CENH3 foci demonstrates that male and female gametes exhibit different levels of EYFP-CENH3 (Fig. 8D).

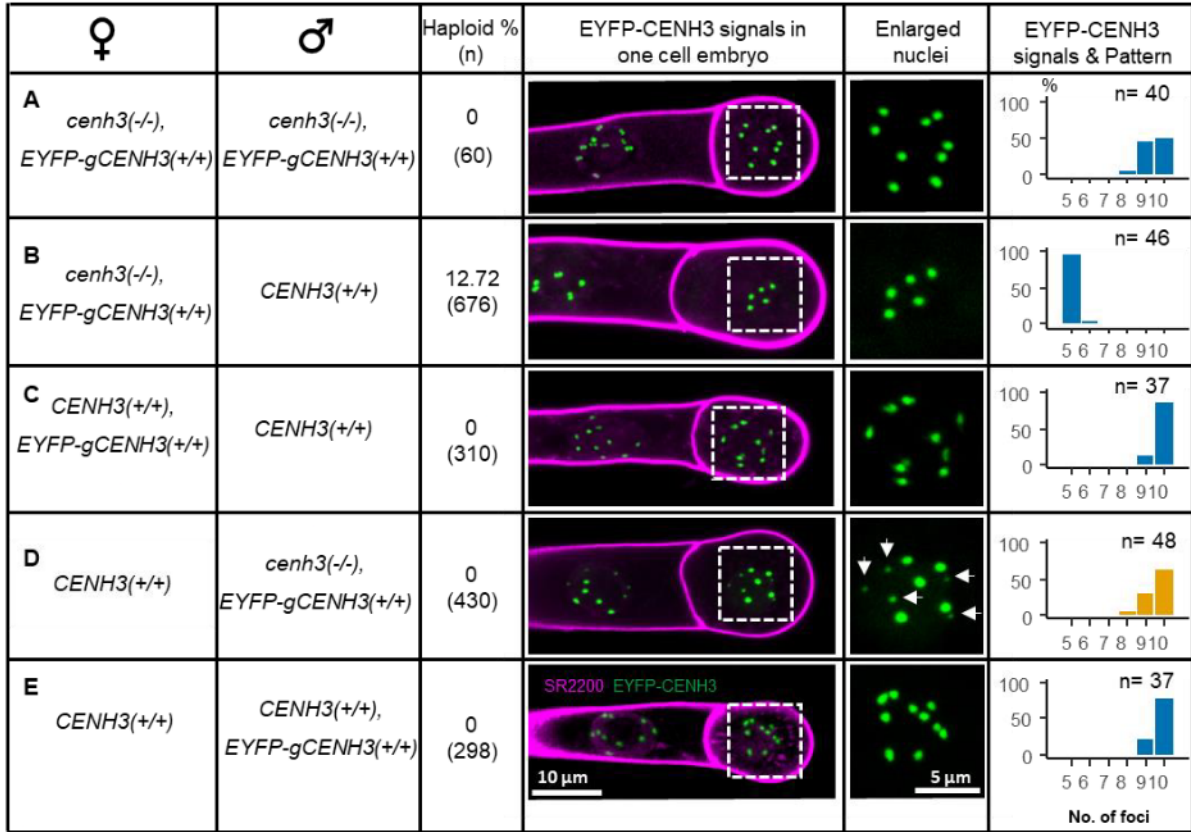


**Figure 8: Gametophytic asymmetry for EYFP-CENH3 in [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*].** **A)** EYFP-CENH3 (green) is present in sperm nuclei but absent in the vegetative nucleus. Boxed region is further enlarged. Nuclei were stained with DAPI (Blue). S - sperm nuclei, V - vegetative nucleus. **B)** An egg cell labeled with histone *H2B-tdTomato* (red) which shows no detectable EYFP-CENH3 signals. Boxed region is further enlarged. **C)** A representative picture of rare egg cells displaying 5 EYFP-CENH3 foci. **D)** Difference between male and female gametes in terms of detectable EYFP-CENH3 signals. n indicates the number of pollen/ovules analysed.

#### 4.2. Extreme asymmetry between parental chromosomes for EYFP-CENH3 at early embryogenesis is associated with uniparental genome elimination

To check if these EYFP-CENH3 differences between gametes influences the uniparental genome elimination process, the EYFP-CENH3 signals were analyzed from one to two cell embryos of different parental genotype combinations. Three different parental *CENH3* genotypes, namely [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*], [*CENH3 (+/+)*, *EYFP-gCENH3(+/+)*] and wild-type [*CENH3 (+/+)*] were utilized for the experiment. The one to two cell embryos but not the zygote was used to visualize the differences between the two parental chromosome sets observable after the zygotic

resetting of CENH3 (Ingouff et al. 2010). The EYFP-CENH3 signal patterns from different crosses were compared with their haploid induction ability to find any association between them.



**Figure 9: EYFP-CENH3 signals from 1-cell embryos derived from crossing different CENH3 genotypes. A-E)** Each row represents different cross combinations for which the parents are mentioned in first two columns. The corresponding haploid percent and the sample size (n) was mentioned in the next column. For each combination, a representative embryo labelled with SR2200 (Magenta) and EYFP-CENH3 (green) signals of the embryo-proper nuclei is shown. Boxed region is further enlarged. The bar plot represents the number of EYFP-CENH3 foci observed for the analysed 'n' number of embryos. The colour of bars in the graph indicates pattern of EYFP-CENH3 signals. Blue – monomorphic, Orange - Bimorphic. In **D**) 5 smaller sized CENH3 signals are arrowed. Haploid frequency of (A) is determined by flow cytometry of seeds, whereas in case of (**B-E**), [*gl1(-/-)*] plants were used as a proxy for [*CENH3(+/+)*] and haploid frequency were determined by glabrous plants.

Upon examination of embryos from self-fertilized [*cenh3(-/-), EYFP-gCENH3(+/+)*] plants, it was observed that EYFP-CENH3 had a modal value of 10, equally sized foci per nuclei (Fig. 9A). No haploids were found after self-pollination of this genotype. On the other hand, when [*cenh3(-/-), EYFP-gCENH3(+/+)*] plants were crossed with wild-type plants, 95.7% of analyzed embryos showed a maximum of only 5 EYFP-CENH3 foci (Fig. 9B). The remaining 4.3% had more than 5 EYFP-CENH3 foci but none of

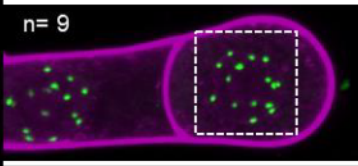
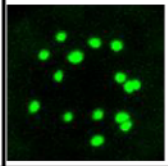
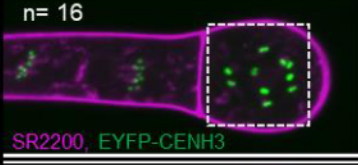
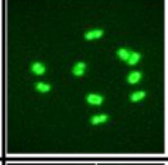
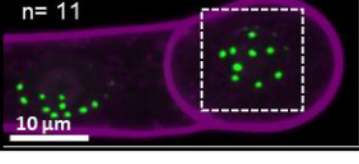
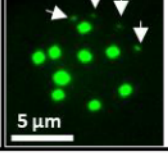
them had 10 foci. Interestingly, only this specific cross-combination resulted in haploidization frequency of 13%. When the genotype [*CENH3*(+/+), *EYFP-gCENH3*(+/+)] was crossed with wild-type plants, most of the embryos that were produced showed a total of 10 equally sized EYFP-CENH3 foci (Fig. 9C). This observation confirms the cross loading of maternally derived EYFP-CENH3 protein onto both maternal and paternal chromosomes

Contrastingly, when [*cenh3*(-/-), *EYFP-gCENH3*(+/+)] is used as a pollen donor to crosspollinate wild-type plants, 10 EYFP-CENH3 foci were detectable in most of the embryos. But a bimorphism for signal intensity was observed in all the embryos, where 5 of the foci had bright signals, and 5 had faint and smaller-sized signals (Fig. 9D). From the observation that a set of 5 foci is different from another 5, it is evident that it reflects the asymmetry contributed by the parental gametes. However, no haploids were obtained despite the contrasting size of EYFP-CENH3 signals. When [*CENH3*(+/+), *EYFP-gCENH3*(+/+)] crossed with wild-type plants, 10 monomorphic EYFP-CENH3 foci were observed, and no haploids were found among the progenies (Fig. 9E).

When [*cenh3*(-/-), *EYFP-gCENH3*(+/+)] was crossed with wild-type, it either produced embryos with 5 monomorphic EYFP-CENH3 signals or 10 bimorphic signals when used as a female or male, respectively. However, it was not clear which parental chromosome set has undetectable or low levels of EYFP-CENH3. Therefore, interploidy crossing approach described by Marimuthu et al. (2021) was exploited for the further analysis. When a tetraploid wild-type plant [*CENH3*(+/+/+/+)] (4x) was pollinated with diploid [*CENH3*(+/+), *EYFP-gCENH3*(+/+)] (2x), the obtained embryos displayed around 15 monomorphic signals (Fig. 10A).

However, when diploid [*cenh3*(-/-), *EYFP-gCENH3*(+/+)] plants were used as a mother, the embryos revealed a maximum of 10 monomorphic signals (Fig. 10B). When used as a pollen donor, 15 bimorphic signals were observed (Fig. 10C). Out of the all EYFP-CENH3 foci, a maximum of 10 foci corresponding to wild-type centromere had bright signals and 4-5 had faint small-size signals that correspond to chromosomes derived from [*cenh3*(-/-), *EYFP-gCENH3*(+/+)]. However, all the 15 signals were not seen due to either the 3D organization of the chromosomes or clustering of two or more centromeres together. Therefore, it can be concluded that

the chromosomes derived from the [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] parent have undetectable or low levels of EYFP-CENH3 and reloading of EYFP-CENH3 is biased towards the chromosomes from wild-type [*CENH3(+/+)*] plants.

♀	♂	EYFP-CENH3 signals in one cell embryo	Enlarged nuclei	EYFP-CENH3 signals & Pattern
<b>A</b> <i>CENH3(+/+/+/+)</i> 4x	<i>CENH3(+/+)</i> , <i>EYFP-gCENH3(+/+)</i> 2x	n= 9 		Monomorphic 15 foci
<b>B</b> <i>cenh3(-/-)</i> , <i>EYFP-gCENH3(+/+)</i> 2x	<i>CENH3(+/+/+/+)</i> 4x	n= 16 		Monomorphic 10 foci
<b>C</b> <i>CENH3(+/+/+/+)</i> 4x	<i>cenh3(-/-)</i> , <i>EYFP-gCENH3(+/+)</i> 2x	n= 11 		Bimorphic 10 bright + 5 faint

**Figure 10: EYFP-CENH3 signals from 1-cell embryos derived from inter-ploidy crosses. A-E)** Each row represents different cross combinations for which the parents and its ploidy are mentioned in first two columns. 4x – tetraploid, 2x – diploid. For each combination, a representative embryo and EYFP-CENH3 signals of the embryo-proper nuclei is shown. Boxed region is further enlarged. In **(C)** 4 smaller sized CENH3 signals are arrowed. The last column refers to the EYFP-CENH3 signal pattern that is characteristic of the corresponding cross combination.

With the results from the analysis of gametes and early embryos, it becomes obvious that the epigenetic strength of the centromere (i.e.,) CENH3 chromatin levels is reduced in gametes of [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] in comparison to that of wild-type plants, but the reduction is higher in egg cells compared to the sperms (Fig. 11). Haploids were only produced when the [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] genotype is used as a female crossing partner. This not only reaffirms the centromere size model but also highlights that there is a need for a critical level of asymmetry between the parental chromosomes for CENH3 levels to induce uniparental genome elimination. Therefore, it was hypothesized that by targeted degradation of the EYFP-CENH3 protein, either female or male gametes that are devoid or strongly reduced of EYFP-CENH3 can be generated, which, on crossing with wild-type plants, could generate paternal or maternal haploids, respectively.

Genotype	<i>CENH3(+/+)</i>		<i>cenh3(-/-), EYFP-gCENH3</i>	
Gamete	♀ or ♂		♂	♀
CENH3 chromatin levels				
Haploid induction				

**Figure 11: Extreme asymmetry in CENH3 levels between parental genomes is associated with genome elimination and consequently with haploid induction.** This model summarizes the findings from the analysis of EYFP-CENH3 signals in early embryos resulting from crosses between different genotypes and their corresponding haploid progenies. Based on the observations, the wild-type gametes were assumed to exhibit higher levels of CENH3 chromatin. Whereas the male gametes of [*cenh3(-/-), EYFP-gCENH3*] has a reduced CENH3 chromatin levels compared to the wild-type, and they are reduced further in the female gametes, falling below the limits of microscopic detection. The difference in CENH3 chromatin levels between wild-type and male [*cenh3(-/-), EYFP-gCENH3(+/+)*] gametes does not result in haploids. In contrast, the difference between wild-type and female [*cenh3(-/-), EYFP-gCENH3*] gametes result in haploids, indicating that a critical threshold of difference in parental chromosome sets is required for genome elimination.

#### 4.3. Egg cell-specific degradation of the EYFP-CENH3 protein

To test whether the gamete specific degradation of the EYFP-CENH3 protein increases the frequency of haploids after outcrossing with wild-type plants, the GFP-nanobody (VHHGFP4) (Rothbauer et al. 2006) fused to E3-ligases such as SPOP and Nslmb were used. These fusion proteins would lead to proteasome-mediated degradation of EYFP-tagged CENH3 (Sorge et al. 2021). The gametophytic degradation can be achieved by spatially restricted expression of these fusion proteins using a gamete-specific promoter. Initially, the female gamete was examined, because of the availability of well characterized promoters specific for the egg cell. The *EC1.1* promoter of Arabidopsis (Sprunck et al. 2012) was selected as a candidate promoter for our initial experiments.

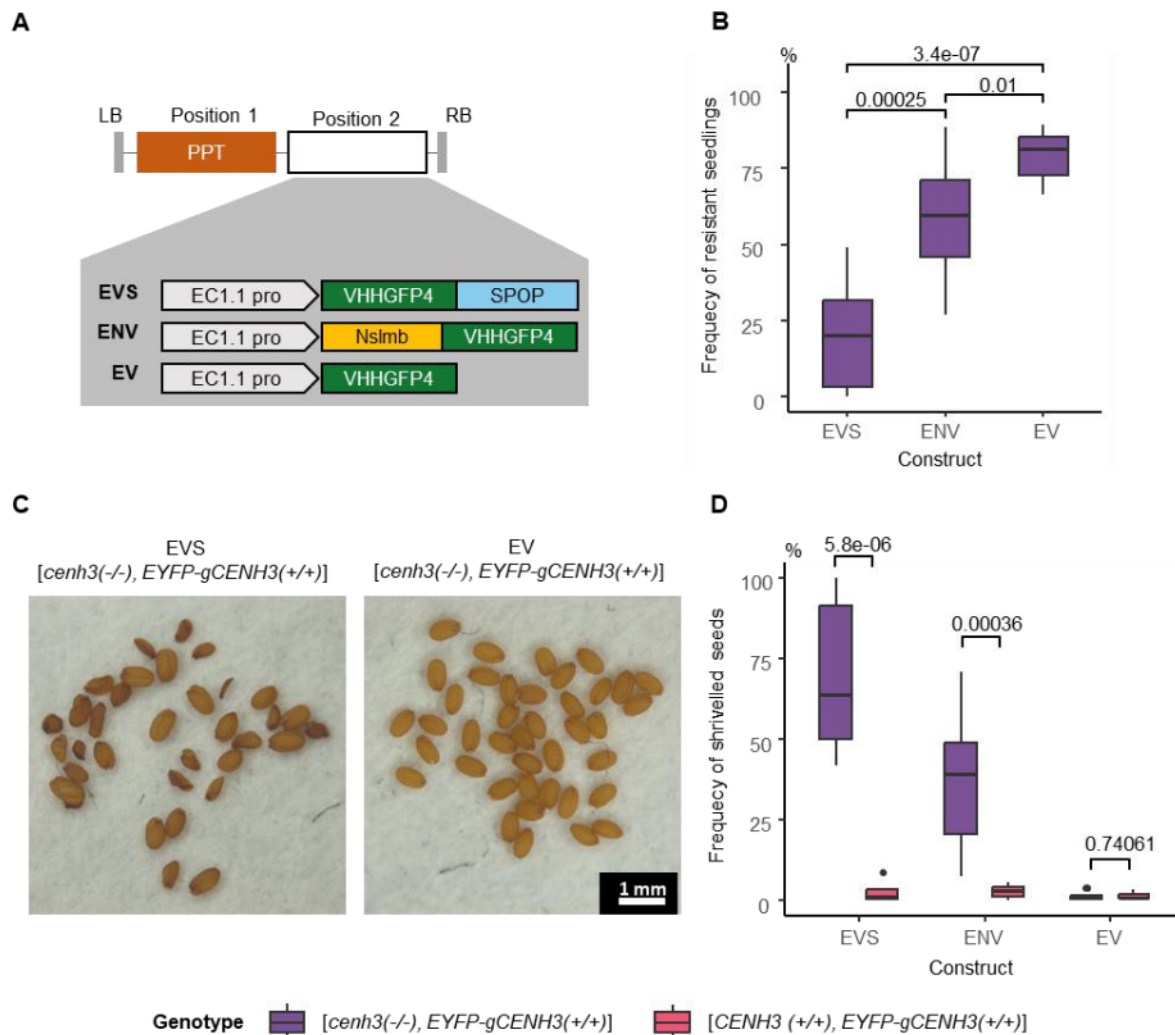
Two different E3-ligase fusion constructs namely, *EC1.1pro:VHHGFP4-SPOP* (EVS) and *EC1.1pro:Nslmb-VHHGFP4* (ENV) were designed. In addition a construct carrying only nanobody, *EC1.1pro:VHHGFP4* (EV) was prepared (Fig. 12A) and all

three variants of the constructs were used for transformation of [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] plants. After the transformation of *A. thaliana*, 10 independent T1 plants each per construct were grown, and the obtained T2 seeds were used for antibiotic selection of seedlings using phosphinothricin.

#### **4.3.1. Egg cell-specific degradation of the EYFP-CENH3 protein results in shriveled seeds**

There was a significant difference between the construct for the frequency of phosphinothricin-resistant seedlings (Fig. 12B). A single-copy insertion line should have resistant seedling of around 75% (3:1 segregation). In the case of EVS, none of the T1 family had resistant seedlings above 50%. The resistant seedlings among ENV ranged between 27.18% to 88.89%. EV recorded a minimum frequency of 66.67%, and more than half of the analyzed T1 family gave rise to resistant seedlings above 80%. From the visual evaluation of the seeds, a lot of shriveled T2 seeds was found especially among those from EVS and ENV but not from EV T1 plants (Fig. 12C). The prevalence of shriveled seeds among T2 seeds from plants containing the EVS and ENV constructs hindered the isolation of homozygous insertion lines through segregation analysis of antibiotic resistance.

The reason for the occurrence of shriveled seeds could be either due to i) the expression of E3 ligases *per-se* in the gametes or ii) genomic instability after self-pollination, caused by the degradation of EYFP-CENH3 in the egg cell. To rule out the first possibility, [*CENH3(+/+)*, *EYFP-gCENH3(+/+)*] plants were transformed with the same set of constructs. T2 seeds from 10 independent self-pollinated T1 plants per construct from both genotypes were examined for the prevalence of shriveled seeds (Fig. 12D). The difference in the shriveled seed frequency between the two genotypes ([*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] vs [*CENH3 (+/+)*, *EYFP-gCENH3(+/+)*]) was significant in the case of EVS & ENV, whereas no such difference was observed in the case of EV. A high correlation between the prevalence of shriveled *A. thaliana* seeds and the frequency of haploids was reported in a previous study (Kuppu et al. 2020). Therefore, it was concluded that the reason for shriveled seeds was likely the genomic instability during early embryogenesis caused by egg-cell specific degradation of EYFP-CENH3.



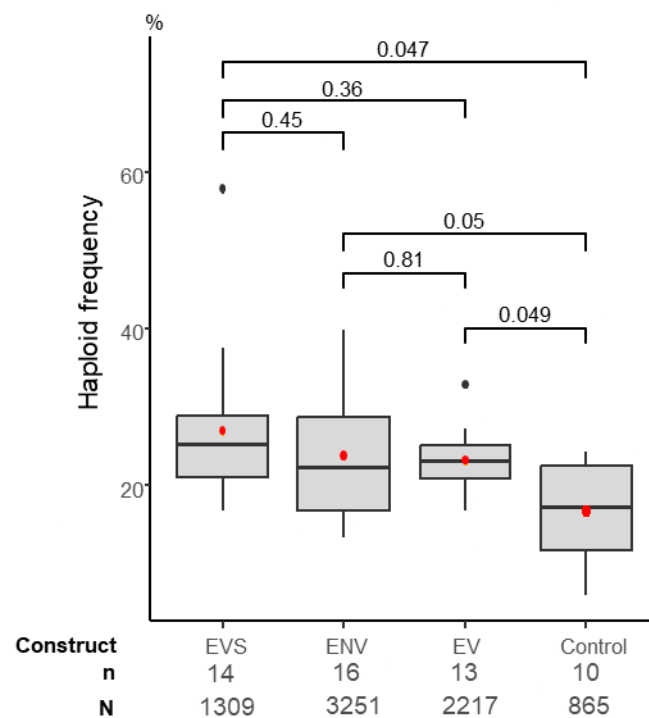
**Figure 12: Egg cell-specific degradation of EYFP-CENH3 results in shriveled seeds. A)** Schematic representation of different constructs (EVS, ENV, EV) used for the egg cell-specific degradation of EYFP-CENH3. PPT- Phosphinothricin **B)** Frequency of phosphinothricin resistant seedlings obtained from T2 seeds of individual T1 [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] plants for different egg-cell specific degradation constructs (EVS, ENV, EV). **C)** Seeds from a single silique of a representative EVS and EV T1 plants in genetic background of [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*]. Shriveled seeds can be observed among the seeds from EVS. **D)** Frequency of shriveled seeds after self-pollination of different genotypes ([*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] and [*CENH3(+/+)*, *EYFP-gCENH3(+/+)*]) transformed with different egg-cell specific degradation constructs (EVS, ENV, EV).

#### 4.3.2. Egg cell-specific degradation of the EYFP-CENH3 protein leads to higher frequency of haploids

The isolation of homozygous insertion lines through segregation analysis of antibiotic resistance was hindered due to the prevalence of shriveled seeds among T2 seeds from plants containing the EVS and ENV constructs. Therefore, the haploidization



frequency was assessed using the [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] T1 plants hemizygous for different egg-cell specific degradation constructs (EVS, ENV and EV). A fresh set of T1 plants for each construct was pollinated with pollen from [*gl1(-/-)*] plants, and haploid progenies were identified by their glabrous phenotype. The [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] plants without any of the above constructs were used as control. When comparing the mean frequency of haploidization, it was found that all three constructs resulted in significantly higher haploid frequencies than the control (Fig. 13). However, there were no significant differences in haploidization rates among the T1 plants carrying any of the three constructs. Randomly selected glabrous plants from different crosses were analysed by flow cytometry to confirm the reliability on glabrous phenotype as a tool for haploid screening. Out of 133 glabrous plants analysed, 111 plants (83.46%) were found to be haploids and remaining 22 plants (16.54%) were diploids.



**Figure 13: Egg cell-specific degradation of EYFP-CENH3 resulted in increased haploidization frequency.** Haploidization frequency of independent [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] T1 plants hemizygous for different egg cell-specific degradation constructs (EVS, ENV, EV) after pollination with pollen from [*gl1(-/-)*] plants. The [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] plants without any of the constructs were used as control. Red dots indicates the mean haploidization frequency. 'n' indicates the individual T1 plants evaluated for each construct and individual plants used in case of control. 'N' indicates the total number of plants screened for the glabrous phenotype from all the T1 plants evaluated. p values were added on top of the comparison bars.

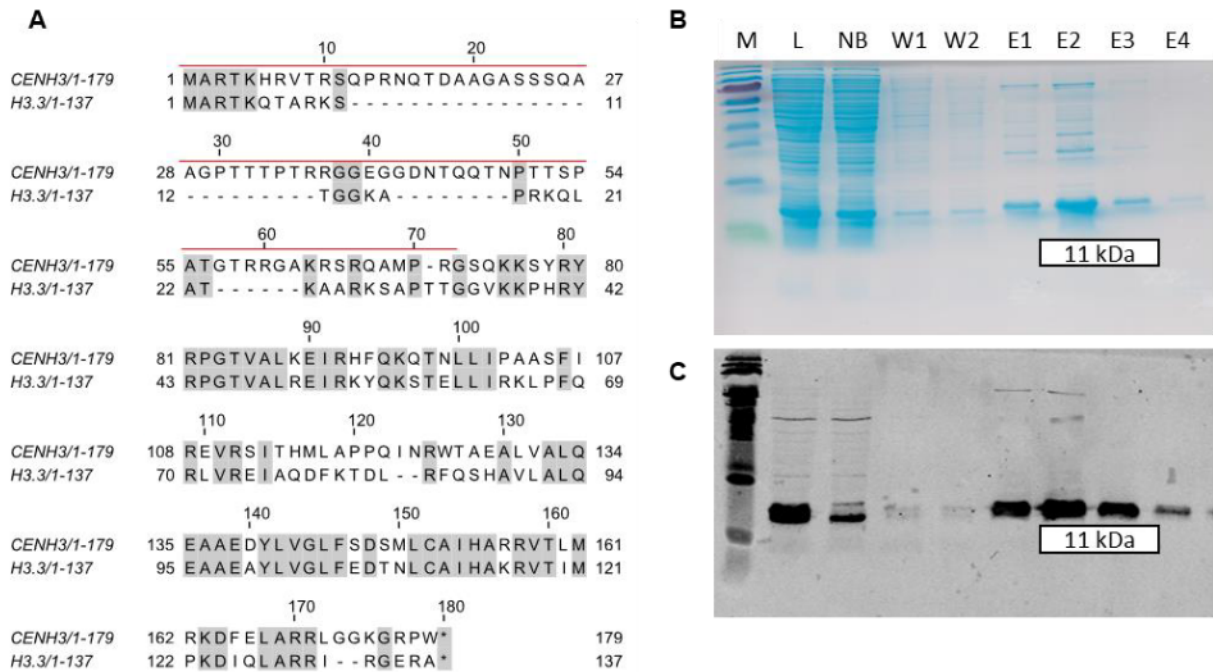
The haploid frequency among the control replicates without an egg cell-specific degradation construct was highly variable, ranging from 5.93% to 24.26%. This occurrence and variability in haploid frequency caused by the genetic background of the control makes it difficult to distinguish the influence of E3 ligases from that of the nanobody on haploidization. To overcome the issues caused by these background haploid levels, it was necessary to either identify a nanobody that facilitates in combination with an E3 ligase the degradation of native Arabidopsis CENH3 or to isolate a tagged-CENH3 complemented line with no or a very low level of haploid induction. As a preliminary step, I endeavored to isolate a nanobody targeting native CENH3 of Arabidopsis, aiming to obviate the need for complementation with a modified CENH3.

#### **4.4. Screening for nanobody against native Arabidopsis CENH3**

##### **4.4.1. Preparation of a suitable CENH3 antigen**

Nanobodies against an antigen can be isolated from either highly diverse synthetic libraries or immune libraries by affinity selection using the phage display (Muyldermans 2021). However, the full-length CENH3 protein cannot be used as an antigen owing to its similarity to histone H3, especially in the C-terminal histone fold domain. Therefore, by pairwise alignment of the amino acid sequence of both proteins, the amino acid sequence in the N-terminal part unique to CENH3 was identified (Fig. 14A).

To produce the recombinant protein of the N-terminal part of CENH3, the corresponding nucleotide sequence of *A. thaliana* was cloned into the vector pET-22b in frame with a 6x-His-tag, and the protein was affinity purified using Ni-NTA agarose. The purified protein was confirmed by Western blot using an available polyclonal antibody against AtCENH3 (Fig 14B, C).



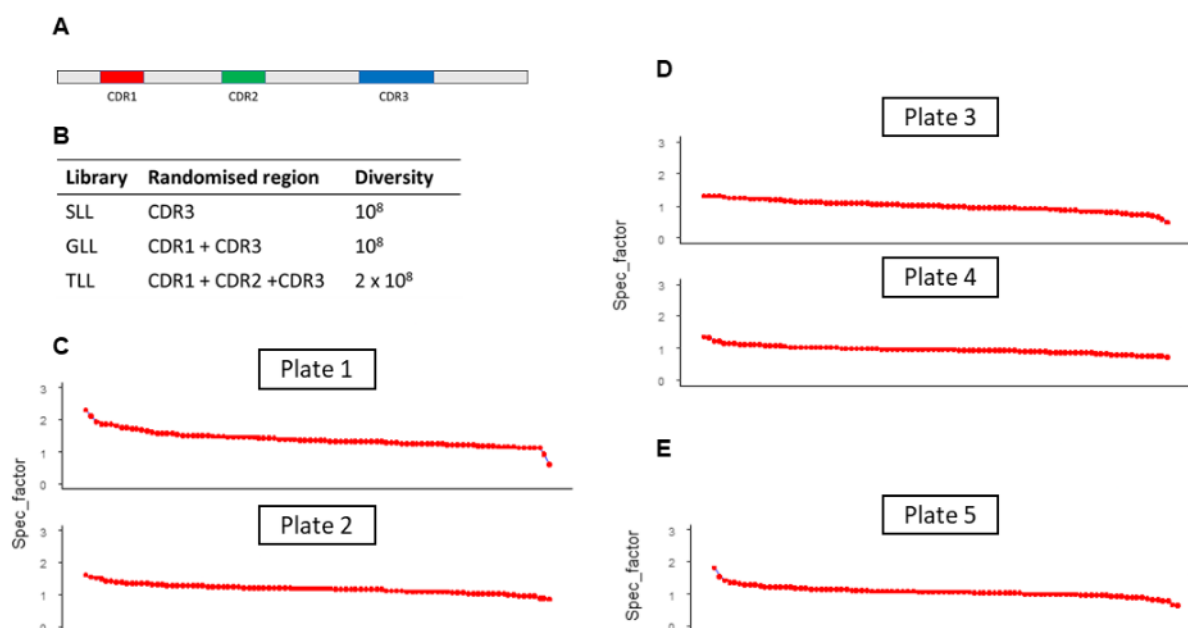
**Figure 14: Preparation of the truncated CENH3 antigen. A)** Similarity between AtCENH3 and histone H3 proteins of *A. thaliana*. The amino acid sequence highlighted with a red line on the top refers to the region used as antigen for phage display experiments. **B)** SDS-PAGE after Coomassie staining **C)** Western blot using anti-CENH3 polyclonal antibody. The band shows the expected size of 11 kDa. M – prestained protein ladder, L- crude lysate, NB – nonbinding fraction, W1 & W2 – wash fractions, E1 to E4 – elution fractions.

#### 4.4.2. Phage display using synthetic nanobody libraries

Three distinct synthetic nanobody libraries, created by randomizing the complementarity determining regions (CDRs) of the anti-hTNFalpha nanobody (Fig. 15A, B), were utilized in this study. The selected nanobody scaffold, previously reported to be highly expressible in plants (Conrad et al. 2011), formed the basis of these libraries. The combined phage libraries, with an estimated diversity of  $10^9$  (Sorge, 2022) were screened using the phage display technique. After three rounds of affinity selection, individual clones were analyzed via soluble ELISA. Three independent phage selections were performed using the same libraries. In the first two phage selection experiments, two plates of 92 clones each were analyzed, and in the third experiment, a single plate was analyzed. In total, 460 nanobody clones (5 plates) were analyzed using soluble ELISA.

The specificity factor of a clone was determined by ELISA. The specificity factor refers to the numeric value obtained by comparing the optical density (OD) value in the

presence of the antigen with the OD value in the absence of the antigen (control). An ideal clone would have a specificity factor of 3 or more. However, none of the 460 analyzed clones achieved a specificity factor of at least 3 (Fig. 15C-E). These results indicated that the synthetic library pool did not contain candidate nanobodies against the AtCENH3 antigen used for affinity selection. Consequently, an immunized library against the AtCENH3 antigen was developed for use in affinity selection.

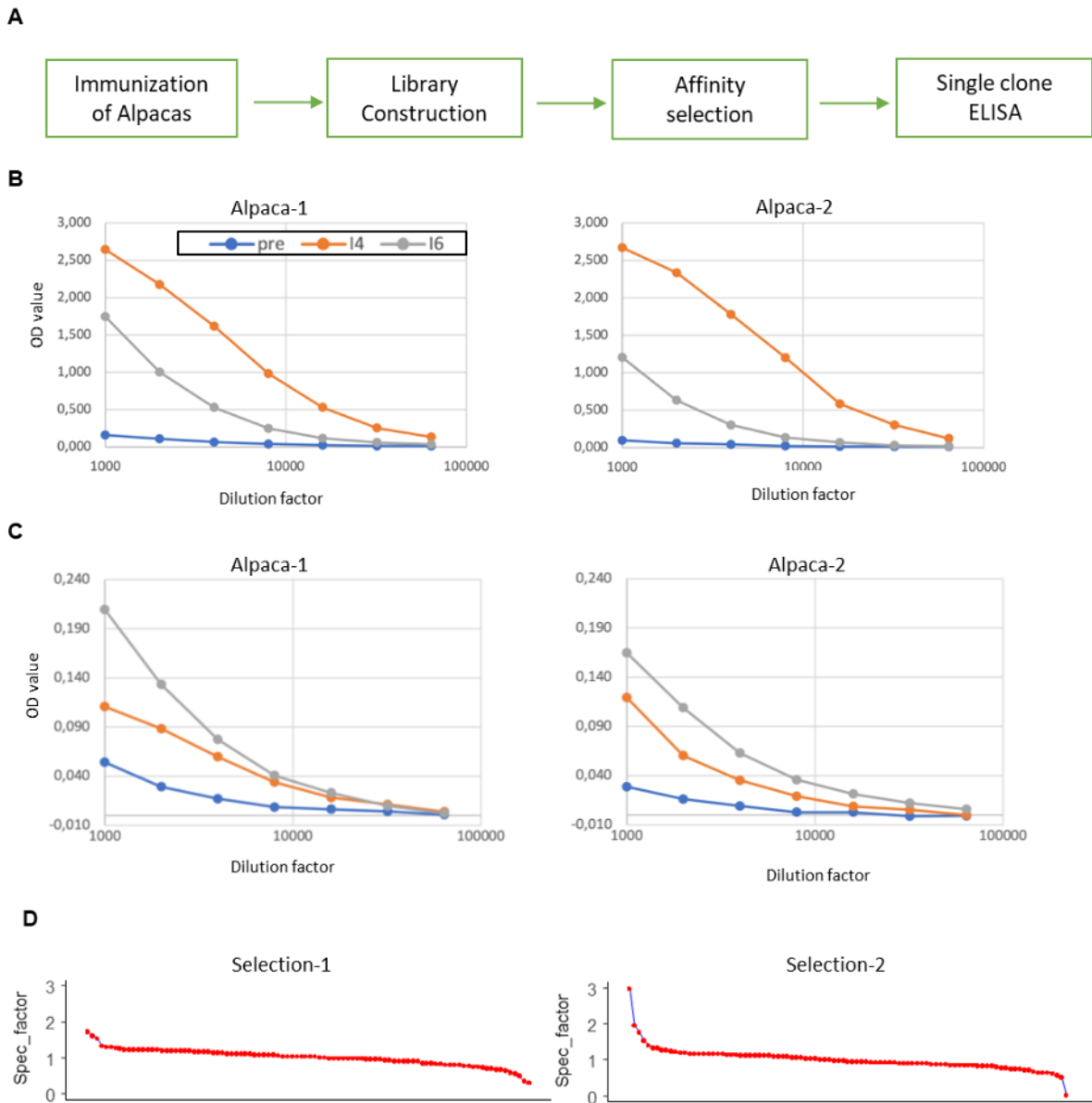


**Figure 15: Phage display using synthetic nanobody libraries.** **A)** – Diagrammatic representation of the nanobody framework. CDR - complementarity-determining regions **B)** – Characteristics of the different libraries used in the current study. **C-E)** Soluble ELISA results of single clones. Each plot represents single clone from three different phage selection experiments. Each plot represents 92 clones sorted according to the specificity factor.

#### 4.4.3. Phage display using immune nanobody library

The immune nanobody library against AtCENH3 was constructed and screened against AtCENH3 by Nanotag biotechnology (Göttingen). The immunized library was constructed using the isolated lymphocytes from the alpacas that had been immunized with the truncated Arabidopsis CENH3 antigen (Fig. 16A). The affinity selection was performed similarly to that of the synthetic library. Two alpacas were immunized six times using a KLH-conjugated AtCENH3 antigen. The immune response was evaluated five days after the fourth and final immunizations using ELISA with a serial dilution of the serum against the CENH3 N-tail antigen. Two different primary

antibodies, anti-alpaca IgG and anti-alpaca IgG2/3, were used to assess the antibody response.



**Figure 16: Phage display using immune nanobody libraries. A)** – Brief schemata of work done at Nanotag Biotechnologies, Göttingen, in preparation and screening of immune nanobody libraries. **B)** Serial titration of serum against CENH3 N-tail antigen with anti-alpaca IgG antibody **C)** Serial titration of serum against CENH3 N-tail antigen with anti-alpaca IgG2/3 antibody **D)** Soluble ELISA results of single clones from two different phage selection experiments. Each plot represents 96 clones sorted according to the specificity factor. pre - pre-immune sera, I4 – sera after fourth immunisation, I6 - sera after sixth immunisation.

The ELISA with anti-alpaca IgG provides an overview of the alpacas' overall immune response to the antigen. The immune response was high after the fourth immunization but decreased after the sixth immunization in both animals (Fig. 16B). Anti-alpaca

IgG2/3 is specific to nanobody precursors, and ELISA with this antibody provides information about the presence of nanobodies against the target antigen in the serum of the immunized animals. In contrast to the total immune response, the ELISA with the anti-alpaca IgG2/3 antibody revealed that nanobody precursors against CENH3 N-tail antigen were more enriched after the sixth immunization compared to the fourth immunization (Fig. 16C).

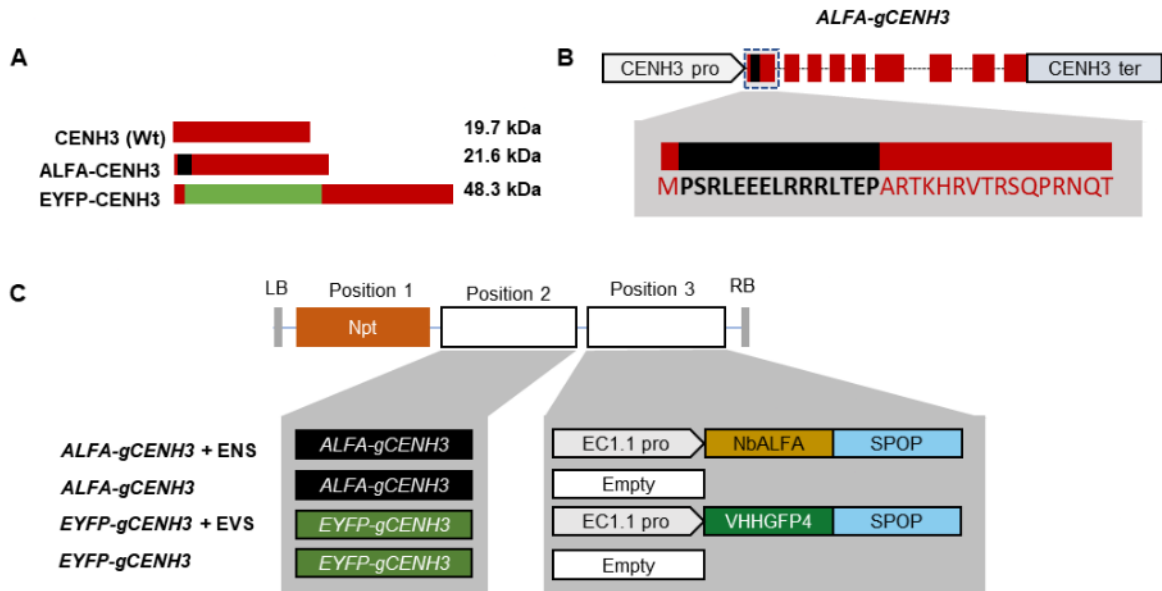
The nanobody library was prepared from lymphocytes isolated from the pooled sera of both alpacas. Three rounds of phage display screening were performed on streptavidin-coated beads using biotinylated AtCENH3 as the target. Two independent phage screening experiments were conducted, and 96 clones were analyzed by single-clone ELISA at the end of each experiment. Out of the 192 clones analyzed, none had a specificity factor of at least 3 (Fig. 16D). One clone, C2, from the second selection experiment, had a specificity factor of 2.96. However, the OD value for this clone was 0.04. A nanobody with weak affinity would have an OD of at least 0.1. Therefore, none of the clones, including clone C2, were suitable candidate nanobodies against CENH3 for further CENH3 degradation experiments.

#### **4.5. Replacement of the EYFP-tag with an ALFA-tag**

After failing to obtain a nanobody specific for native Arabidopsis CENH3, it was aimed to isolate an Arabidopsis line complemented with a tagged CENH3 that exhibits no or negligible levels of haploid induction. Adding the EYFP tag increased the size of CENH3 protein from 19.7 kDa to 48.3 kDa (Fig. 17A). Initially, it was assumed that the bulkiness of the EYFP-tag interfered with the reloading of EYFP-CENH3 during gametogenesis and triggered the formation of haploids after outcrossing with wild-type plants. Therefore, it was decided to replace the large EYFP-tag with a smaller alternative.

The candidate amino acid tag needed to meet the following criteria: i) it should be small and monomeric, ii) it should not affect the function of the tagged protein, and iii) binding protein scaffolds, such as nanobodies, should be available for the tag. The ALFA-tag, composed of 13 amino acids only (Gotzke et al. 2019), emerged as a suitable candidate. The dissociation constant of the nanobody refers to the strength of nanobody in binding the antigen. The dissociation constant of the nanobody against

the ALFA-tag (NbALFA) is 0.026 nM (Gotzke et al. 2019), which is 10-fold higher than that of the GFP nanobody (0.23 nM) (Rothbauer et al. 2006). Moreover, ALFA tagged CENH3 (ALFA-CENH3) protein is just 1.9 kDa larger than the wild-type CENH3 (Fig. 17A). Thus, ALFA-tag was selected as a perfect candidate to replace the EYFP-tag.



**Figure 17: Complementation of Arabidopsis with ALFA-CENH3.** **A)** Comparison of sizes of wild-type CENH3 (Wt), ALFA-CENH3 and EYFP-CENH3 proteins. ALFA-tag and EYFP-tag are shown in black and green, respectively. **B)** Schematic representation of ALFA-gCENH3 construct showing the site of insertion of ALFA tag. The first exon is zoomed in, and the amino-acid highlighted in black corresponds to the ALFA-tag. **C)** Schematic representation of different constructs used to assess the complementation efficiency of the ALFA-tag compared to the EYFP-tag. Npt - kanamycin resistant gene.

#### 4.5.1. ALFA-CENH3 complements the Arabidopsis *cenh3-1* null mutation better than the EYFP-CENH3

Transgenic constructs were designed to introduce the ALFA-tag, flanked by proline residues on both sides, immediately after the start codon in the genomic version of Arabidopsis *CENH3* (Fig. 17B). The ALFA-CENH3 was expressed under the control of the native promoter and terminator sequences of Arabidopsis *CENH3*. This construct was designated as *ALFA-gCENH3*. The ability of ALFA-CENH3 to complement the *cenh3-1* null mutation was compared to that of EYFP-CENH3.

Two versions of level-2 constructs were prepared for ALFA-CENH3: one containing the *ALFA-gCENH3* module alone, and another including an additional module,

*EC1.1pro:NbALFA-SPOP* (ENS) (Fig. 17C). Similarly, two versions of constructs were prepared for EYFP-CENH3: one with only the *EYFP-gCENH3* module, and another including the additional module *EC1.1pro:VHHGFP4-SPOP* (EVS) (Fig. 17C). These constructs were used to transform heterozygous [*CENH3*(+/-)] plants. The transformed T1 plants were screened for the *cenh3-1* mutation using the CENH3-specific dCAPS marker.

Complete complementation would result in a homozygous mutant recovery rate of 25%. The homozygous mutant recovery rates for the EYFP-CENH3 constructs with and without EVS were 9.09% and 8.82%, respectively, which significantly deviated from the expected *CENH3* genotypic values for the complete complementation (Table 3). In contrast, the ALFA-CENH3 constructs with and without ENS resulted in homozygous mutant recovery rates of 23.81% and 19.57%, respectively, with no significant deviation from the expected genotypic values (Table 3). These observations indicate that ALFA-CENH3 is more efficient in complementing the null mutant compared to EYFP-CENH3.

**Table3:** CENH3 genotypes of individual T1 plants for different constructs (Fig. 17c). The counts of plants were given for each *CENH3* genotype.

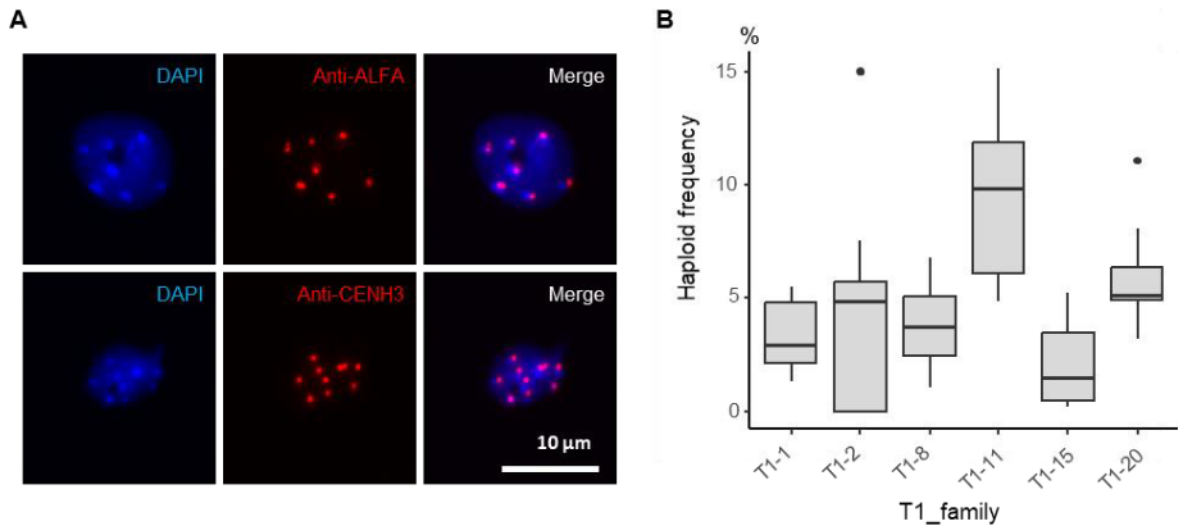
Constructs	<i>CENH3/cenh3-1</i>	<i>CENH3/CENH3</i>	<i>cenh3-1/cenh3-1</i>	Homozygous Mutant recovery (%)	Chi-square P-value
<i>ALFA-gCENH3</i> + ENS	21	11	10	23.81	0.976
<i>ALFA-gCENH3</i>	25	12	9	19.57	0.691
<i>EYFP-gCENH3</i> + EVS	30	10	4	9.09	0.024
<i>EYFP-gCENH3</i>	35	27	6	8.82	0.001

#### 4.5.2. ALFA-CENH3 is functionally similar to CENH3

A few [*cenh3*(-/-)] plants complemented with the *ALFA-gCENH3* construct were analyzed by immunostaining of centromeres using two different primary antibodies: anti-ALFA minibody and anti-AtCENH3 antibody. The anti-ALFA minibody produced immunofoci similar to those generated by the anti-CENH3 antibody, indicating the proper in-frame fusion of the ALFA tag with the CENH3 protein (Fig. 18A). The



immunoseals from both primary antibodies co-localized with the chromocenters, confirming the proper functionality of ALFA-CENH3.



**Figure 18: Functional characterization of *ALFA-gCENH3* T1 plants.** **A)** Immunolabelling of leaf nuclei from *[cenh3(-/-)]* complemented with *ALFA-gCENH3* construct with anti-ALFA and anti-CENH3. Immunoseals are shown in red. Nuclei were counterstained with DAPI (blue). **B)** Haploid induction frequencies of T2 replicates from different *[cenh3(-/-), ALFA-gCENH3(+/+)]* T1 families after outcrossing with male *[gl1(-/-)]* plants.

#### 4.5.3. ALFA-CENH3 induces haploid Arabidopsis plants at varying frequencies

To analyze whether *ALFA-gCENH3* complemented *[cenh3(-/-)]* plants also induce haploids after outcrossing with wild-type, T2 plants derived from six different T1 plants were grown. Ten randomly selected T2 replicates from each of the six T1 families were crossed as females with *[gl1(-/-)]* plants, and the resulting progenies were analyzed for glabrous plants to quantify the number of haploids. The haploid frequencies were found to be highly variable within and between the T1 families. Overall, the haploid frequency ranged from 0% to 15.15%. Despite the small size of the ALFA-tag, *[cenh3(-/-), ALFA-gCENH3(+/+)]* plants triggered haploidization after outcrossing with Arabidopsis wild-type plants.

However, this experiment was performed to isolate an *ALFA-gCENH3* line that does not induce any haploids as a control to precisely determine the effect of targeted CENH3 degradation. Therefore, it was observed that a T1 family, T1-2, in which 4 out of 10 T2 replicates had no haploids among the evaluated progenies (Fig. 18B). Consequently, T2 plant (designated T1-2-23) was selected and T3 seeds were

harvested from the plant. Ten individual T3 plants from T1-2-23 were crossed as females with [*gl1(-/-)*] plants, and the F1 progenies were analyzed for haploid offspring. None of the 3009 progenies from the 10 different T3 mother plants produced glabrous plants. Therefore, T3 plants derived from a T2 plant (T1-2-23) was used for further transformations with different constructs designed for egg cell-specific degradation of ALFA-CENH3.

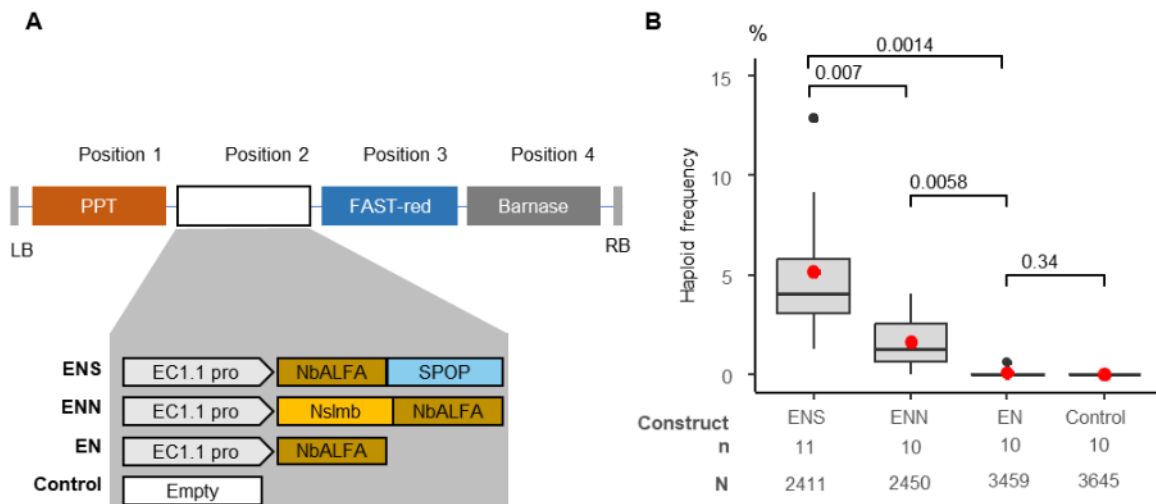
#### **4.5.4. Egg cell-specific degradation of the ALFA-CENH3 with SPOP induces higher frequency of haploids than Nslmb**

In previous experiments with EYFP-CENH3 complemented lines, the fold change from the nanobody only control (EVS & ENV vs EN) for haploid frequencies, and the difference between two E3 ligases (EVS vs ENV) were unclear. Therefore, a similar experiment was conducted using *ALFA-gCENH3* complemented lines. Four distinct transgenic constructs were prepared by replacing the anti-GFP nanobody with the anti-ALFA nanobody (NbALFA) (Fig. 19A). The constructs included either *EC1.1pro:NbALFA-SPOP* (ENS), *EC1.1pro:Nslmb-NbALFA* (ENN), *EC1.1pro:NbALFA* (EN), or an empty module (control) in the second position of the construct. To ease the laborious crossing work, a module expressing the intronized *Barnase* gene, controlled by *AtA9* (tapetum-specific) promoter, was added to all constructs. *Barnase* is a bacterial ribonuclease protein toxic to any type of cells and when combined with a tapetum-specific promoter lead to male sterility (Paul et al. 1992). *FAST* seed marker (Shimada et al. 2010) was added to all the constructs to aid with the haploid selection in the future experiments. Previous experiments shown *FAST* marker as useful marker for haploid screening in *Arabidopsis* and tomato (Zhong et al. 2020; Zhong et al. 2022a).

At least 10 individual T1 [*cenh3(-/-)*, *ALFA-gCENH3*] plants carrying different constructs (ENS, ENN, EN and control) were pollinated with [*gl1(-/-)*] pollen, and the progenies were screened for glabrous plants (Fig. 19B). When different control T1 [*cenh3(-/-)*, *ALFA-gCENH3*] plants were outcrossed with [*gl1(-/-)*] plants, none of them produced haploid plants. Most of the T1 replicates from the nanobody-only control, EN, did not produce haploids, except for one, which resulted in a haploid frequency of 0.64%. There was no significant difference between the control and EN in terms of

haploid induction frequency. However, T1 plants carrying either of the E3-ligase bearing constructs, ENS and ENN, differed significantly from EN.

ENS T1 plants resulted in haploid frequencies ranging from 1.30% to 12.90%, with a mean of 5.14%. In contrast, ENN had a mean frequency of 1.52%, with individual frequencies ranging from 0% to 4.07%. There was a 3.3-fold increase in haploid frequency in ENS compared to ENN. This indicates that the SPOP protein is more effective than Nslmb in inducing haploid plants. A randomly selected set of 20 glabrous plants from different ENS and ENN T1 female parents were analysed by flow-cytometry for ploidy levels. It was found that out of 20 glabrous plants that were analysed, 19 (95%) were confirmed to be haploid and one diploid (5%).



**Figure 19: SPOP induces a higher frequency of haploids than Nslmb in combination with the ALFA-tag. A)** Schematic representation of different constructs (ENS, ENN, EN and control) used for the egg cell-specific degradation of ALFA-CENH3. **B)** Haploidization frequency of independent hemizygous T1 [*cenh3*(-/-), *ALFA-gCENH3*] plants carrying different constructs (ENS, ENN, EN and control) after outcrossing with male [*gl1*(-/-)] plants. Red dots indicates the mean haploidization frequency. ‘n’ indicates the number of individual T1 plants evaluated for each construct. ‘N’ indicates the total number of plants screened for glabrous phenotype, from all T1 plants evaluated for a particular construct. p values were added on top of the comparison bars

#### 4.6. Sperm-specific degradation of EYFP-CENH3

It was confirmed that the egg cell-specific degradation of tagged-CENH3 protein resulted in paternal haploids. However, it remained unclear whether the gametophytic

degradation approach would lead to the formation of maternal haploids. So far, only GFP-tailswap modification of CENH3 has been shown to induce maternal haploid plants up to 5% in Arabidopsis (Ravi and Chan 2010). Therefore, a new construct was designed with the objective of sperm-specific degradation of tagged-CENH3.

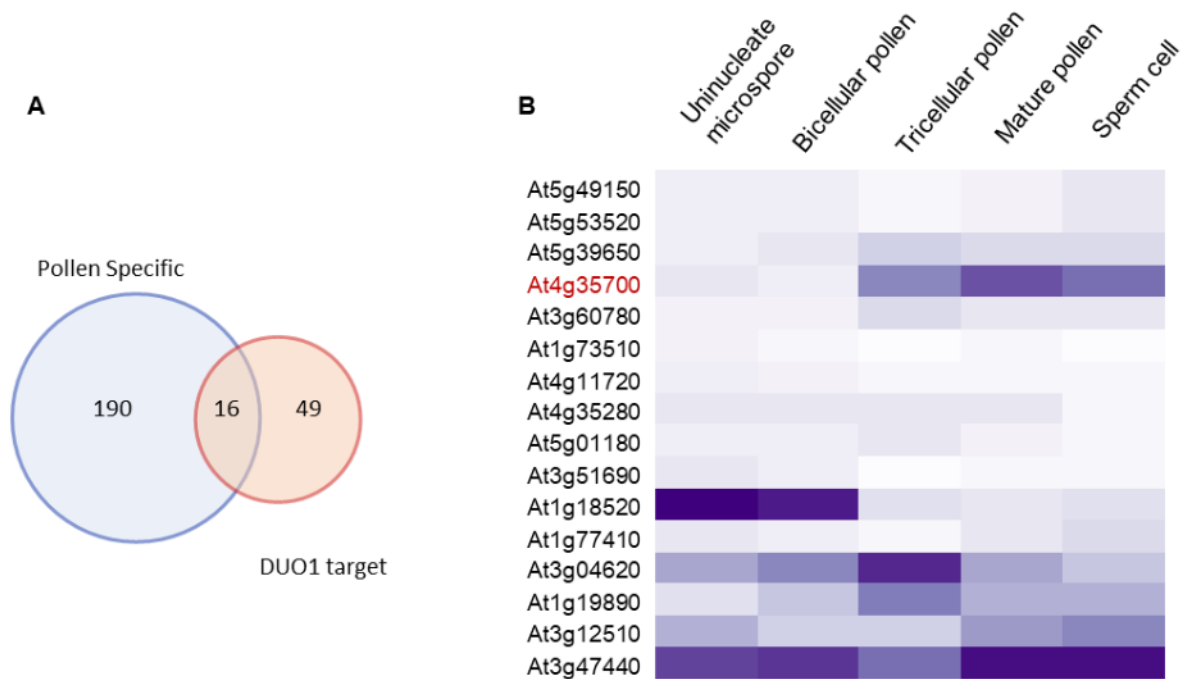
#### **4.6.1. Selection of candidate sperm-specific promoter**

In contrast to egg cells, no thoroughly researched promoter that is highly specifically expressed in sperm cells and suitable for targeted protein degradation has been described for Arabidopsis. To avoid any unnecessary consequences on plant phenotype or fertility due to CENH3 degradation outside sperm nuclei, I aimed to identify a candidate promoter that is exclusively expressed in sperm cells of mature tri-cellular pollen. Therefore, two public datasets were utilized for our analysis. One dataset provided a list of 206 pollen-specific genes, that were not expressed in the vegetative parts of Arabidopsis (Borges et al. 2008). The other dataset provided a list of 65 genes activated by the male germline-specific transcription factor, DUO1 in Arabidopsis (Borg et al. 2011).

Comparative analysis of both datasets was performed to identify the overlapping set of genes and a subset of 16 genes were identified (Fig. 20A). The candidate promoter was expected to be exclusively active in sperm cells. Thus, the expression profiles of these 16 genes using the dataset from Honys and Twell (2004) was used for comparison (Fig. 20B). An ideal candidate would have the lowest level of expression in the uni-nucleate and bicellular pollen stages but the highest in the tricellular and mature pollen stages. Three genes, namely At5g39650, At4g35700, and At3g60780, matched these selection criteria. Among these, At4g35700 (DAZ3) had the highest level of expression in sperm cells.

Previous studies reported that this promoter is active in sperm nuclei but not in vegetative nuclei (Borg et al. 2011). The datasets from Zhao et al. (2019) was also used to examine the transcript expression profile of this gene in the egg cell and zygote. For comparison, the expression profiles of an egg cell-specific gene, EC1.1, and a constitutively expressed gene, UBQ1 were utilised (Table 4). The expression of DAZ3 (At4g35700) was entirely absent in egg cells and showed negligible levels of expression in zygotes and one-cell embryos. Thus, DAZ3 fulfills all the selection

criteria, making it an ideal candidate for the subsequent CENH3 degradation experiments.



**Figure 20: DAZ3 promoter is an ideal candidate sperm-specific promoter. A)** Comparing two datasets to identify genes that are pollen-specific (Borges et al. 2008) and DUO1 targets (Borg et al. 2011). **B)** Expression levels of 16 candidates in different stages of pollen development and the sperm cells (Honys and Twell 2004). DAZ3 (At4g35700) is specifically active in tricellular, mature pollen and sperm cells.

**Table 4:** Expression profile of DAZ3 in egg cell, zygote (14 HAP), zygote (24 HAP) and one-cell embryo compared with EC1.1 and ubiquitin (UBQ1). The numbers represent normalised FPKM values from Zhao et al. (2019). HAP: hours after pollination.

Gene ID	Gene name	Expression	Egg cell	Zygote 14 HAP	Zygote 24 HAP	One-cell embryo
AT4G35700	<i>DAZ3</i>	Sperm-specific	0.00	135.02	1.30	1.13
AT1G76750	<i>EC1.1</i>	Egg cell specific	25958.07	5473.62	0.00	3.51
AT3G52590	<i>UBQ1</i>	Constitutive	948.86	717.63	1411.42	1512.40

#### 4.6.2. VHHGFP4-SPOP results in the degradation of EYFP-CENH3 in sperm nuclei

For the sperm-specific CENH3 degradation experiments, [*cenh3(-/-)*, *EYFP-gCENH3*] was used instead of [*cenh3(-/-)*, *ALFA-gCENH3*] for two reasons: (i) the sperm carries

detectable levels of EYFP-CENH3 (Fig. 8A), which facilitates to evaluate the degradation of centromeric EYFP-CENH3, and (ii) these lines do not induce haploids when used as pollen donors (Fig. 9D), making them an ideal control. Since, the SPOP E3-ligase resulted in higher frequency of haploids in previous experiments, a construct harboring a module expressing VHHGFP4-SPOP regulated by the DAZ3 promoter was prepared (Fig. 21A). This construct, named “DVS”, was transformed into [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] plants.

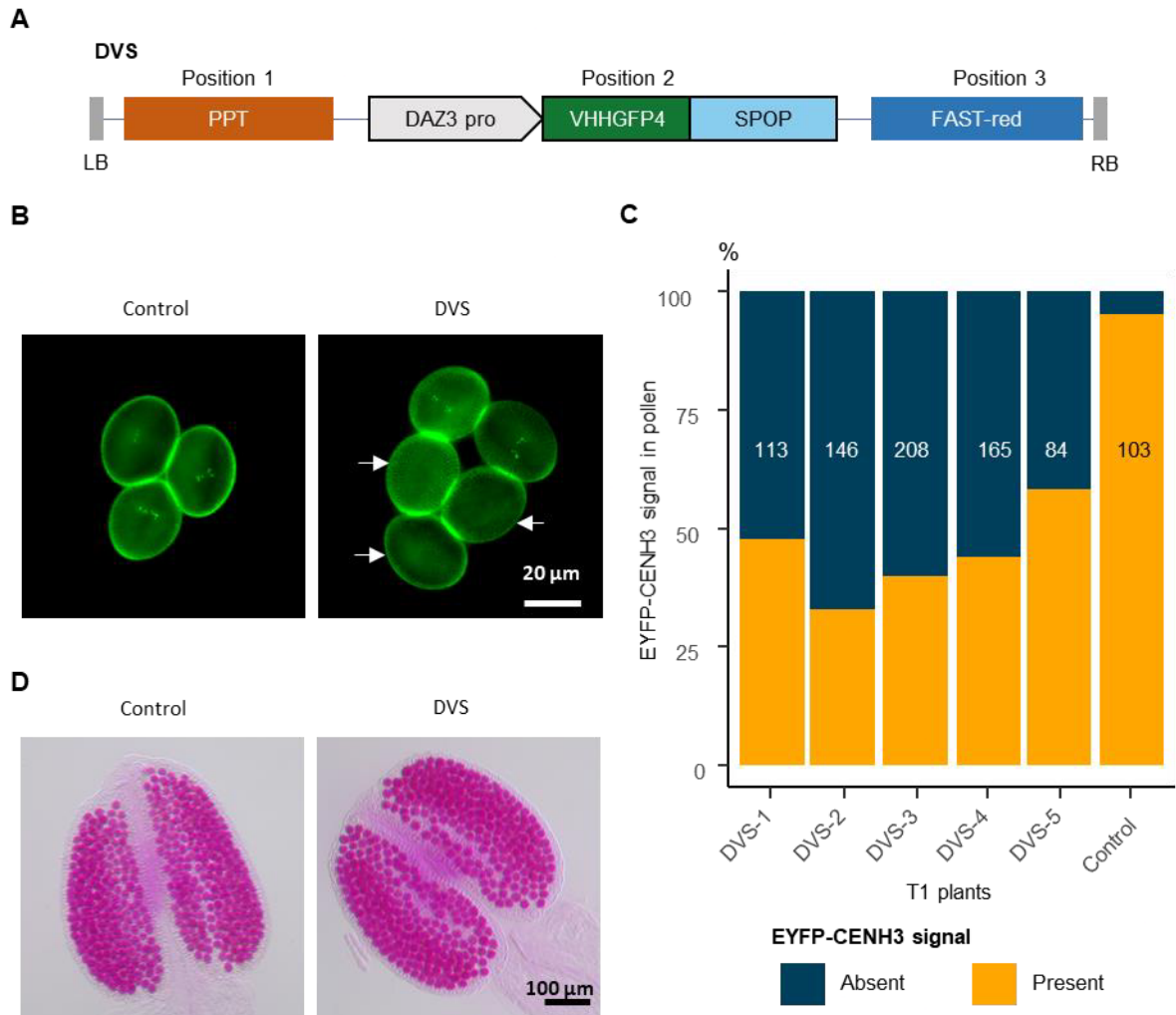
Pollen grains from a set of five DVS T1 plants, along with [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] serving as control, were evaluated using microscopy to assess EYFP-CENH3 signals (Fig. 21B). Pollen grains were assessed for the presence or absence of distinct sperm nuclei-specific EYFP-CENH3 foci. In the control, 95.15% of the pollen displayed sperm-specific EYFP-CENH3 signals. In contrast, the T1 plants exhibited a higher reduction in the number of pollen grains with EYFP-CENH3 signals (Fig. 21C). The frequency of pollen without EYFP-CENH3 signals among the T1 plants ranged from 41.67% to 67.12%. However, this reduction of pollen with EYFP-CENH3 signals has no noticeable effect on pollen viability (Fig. 21D).

#### **4.6.3. Sperm-specific degradation of EYFP-CENH3 produces female-like diploids**

The DVS T1 plants were crossed as pollen donors with [*gl1(-/-)*] mothers to investigate if sperm nuclei with degraded EYFP-CENH3 would produce maternal haploids (Table 5). No glabrous plants were observed among the F1 population from the [*gl1(-/-)*] x [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*], control cross. However, among the progeny derived from DVS T1 male parents, a varying frequency of glabrous plants was observed, ranging from 10.78% to 33.23%. Based on the data from egg cell-specific degradation experiments, it was assumed that most of the glabrous plants would be haploids.

To confirm this, a few glabrous plants were randomly selected for ploidy analysis by flow cytometry. Surprisingly, all the analysed plants were diploid. Consequently, all the remaining plants were analyzed for their ploidy by pooling. Each pool consists of six glabrous plants and the last pool for each cross consists of remaining number of glabrous plants. Out of 160 glabrous plants from a single family of crosses, only 6 pools contained haploid plants. All the remaining pools of glabrous plants from other

F1 families were found to be diploids. These diploid glabrous plants closely resembled the female parent phenotypically and I termed them as ‘female-like diploids’



**Figure 21: Expression of VHHGFP4-SPOP degrades EYFP-CENH3 in sperm cells without affecting pollen viability.** **A)** Schematic representation of the construct DVS used for the sperm-specific degradation of EYFP-CENH3. **B)** Representative pictures of pollen samples from [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] (control) and DVS T1 plants. Arrows indicate pollen without EYFP-CENH3 signals. **C)** Comparison of different DVS T1 plants and control for the number of pollen grains carrying detectable EYFP-CENH3 signals. Numbers within the bars indicates the total number of pollen grains screened by microscopy for the corresponding line. **D)** Alexander staining of anthers from the [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] control and a DVS T1 plant reveals the absence of any obvious defects.

**Table 5:** Frequency of female-like diploids from different DVS T1 fathers. Control refers to [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] as father. n refers to number of plants screened for each cross combination. Both glabrous plants% and haploid% are calculated based on the n as reference. Haploid% is calculated based on assuming one haploid plant per pool and remaining to be diploid. Six plants were pooled into single sample for flow cytometry analysis. The count of plants are given in the brackets next to the frequencies.

Male parent	Female parent	n	Glabrous plants	Haploid
DVS-1	<i>gl1(-/-)</i>	232	10.78% (25)	0%
DVS-2	<i>gl1(-/-)</i>	316	33.23% (105)	0%
DVS-3	<i>gl1(-/-)</i>	925	17.30% (160)	0.65% (6)
DVS-4	<i>gl1(-/-)</i>	652	16.10% (105)	0%
DVS-5	<i>gl1(-/-)</i>	648	11.57% (75)	0%
Control	<i>gl1(-/-)</i>	617	0%	0%

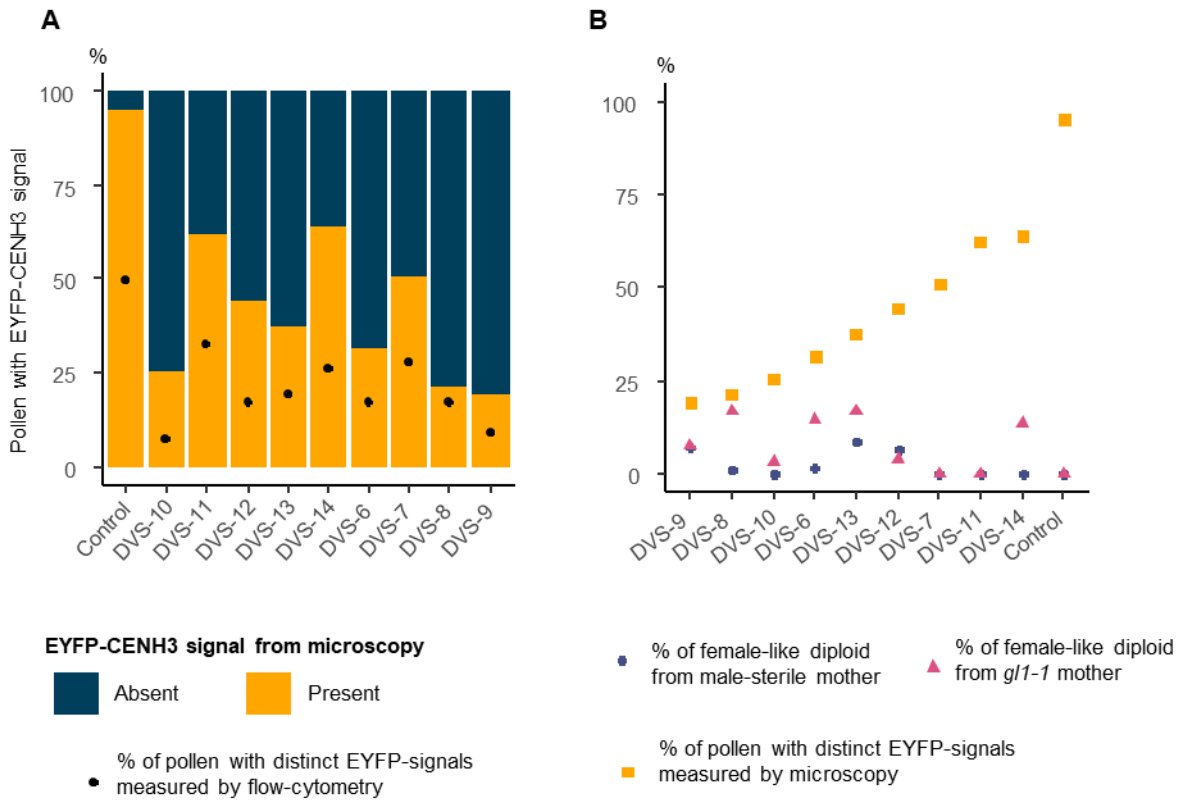
#### 4.6.4. Female-like diploids from the male sterile mothers excludes possibility of contamination from self-pollination

Two possible scenarios exist for the observation of glabrous diploids: 1) such plants might result from contamination due to self-pollination, or 2) they are genuine products of the elimination of the paternal genome. To answer this question, Attempts were made to cross the DVS transgenic lines with a male sterile mother. A T-DNA insertional mutant (SALK\_152147) for the gene *AMS* (AT2G16910) was used as a source of male-sterility (Xu et al. 2010). To facilitate haploid screening, male sterility was combined with the glabrous leaf marker within a single tester genotype. The homozygous male sterile mutant genotype [*ams(-/-)*] were crossed with [*gl1(-/-)*] plants, F1 plants of these crosses were raised to collect the F2 seeds. The F2 seeds were selected for kanamycin resistance, followed by selection for glabrous plants. The glabrous plants were grown until flowering and the sterile plants were selected based on absence of silique elongation and absence of fertile anthers. The selected plants correspond to the genotype [*gl1(-/-)*, *ams(-/-)*] and were used to cross as a female with the DVS T1 fathers.

A set of selected nine DVS T1 plants along with a control for crossing experiments with a male sterile, [*gl1(-/-)*, *ams(-/-)*] tester. Prior to crossing, the number of pollen grains carrying EYFP-CENH3 signals from the DVS T1 plants were quantified using microscopy and flow cytometry (Fig. 22A). Although the two methods yielded different



absolute values, all DVS lines exhibited a marked decrease in EYFP-CENH3 positive pollen compared to the control. Furthermore, the trend of EYFP-CENH3 reduction among the DVS lines was consistent across both methods.



**Figure 22: The frequency of EYFP-CENH3 lacking pollen grains does not correlate with the female-like diploid frequencies. A)** Microscopic and flow cytometry-based evaluation of pollen from different DVS T1 plants and control for the reduction of number of pollen grains carrying detectable EYFP-CENH3 signals. Black dots indicate the frequency measured by flow cytometry. Orange columns indicate the microscopy-based quantification. **B)** Plot showing frequencies of pollen with microscopically detectable EYFP-CENH3 signals (yellow square) (samples arranged in ascending order accordingly) and their corresponding female-like diploid frequencies from male sterile tester (blue dot) and *gl1-1* mother (pink triangle).

The DVS T1 plants and the control were used as a pollen donors for simultaneously pollinating both male sterile tester, [*gl1(-/-)*, *ams(-/-)*] and male fertile tester, [*gl1(-/-)*] female parents. The offsprings were screened for glabrous plants and those which are glabrous were analyzed for their ploidy status using flow cytometry (Table 6). Varying frequencies of glabrous plants were obtained from F1 families of both female parents for a corresponding DVS T1 plants as a male parent. Flow cytometry revealed that

most of these plants were diploid and only few being haploids (Table 6). This finding suggests that self-pollination contamination to be an unlikely cause for the occurrence of female-like diploids.

However, there was no correlation between the frequency of female-like diploids derived from the [*gl1(-/-)*, *ams(-/-)*] when compare to those from the [*gl1(-/-)*] mothers for the corresponding DVS T1 fathers (Fig. 22B). Additionally, there was no correlation between the degree of reduction in pollen numbers with EYFP-CENH3 signals and the frequency of female-like diploids for a given DVS T1 father (Fig. 22B). It appears that the sperm-specific degradation was responsible for the induction of female-like diploids but the frequency rather being stochastic.

**Table 6:** Frequency of female-like diploids from male sterile tester, [*gl1(-/-)*, *ams(-/-)*] and male fertile tester, [*gl1(-/-)*] as female parents for the corresponding DVS T1 fathers. Control refers to [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] as father. n refers to the number of plants screened for each cross combination. G - glabrous plants, H -Haploid plants, D - Diploid plants. The haploids and diploids were identified based on the flow cytometry of single plants. The G, H and D frequencies were calculated based on the n as reference. The count of plants are given in the brackets next to the frequencies.

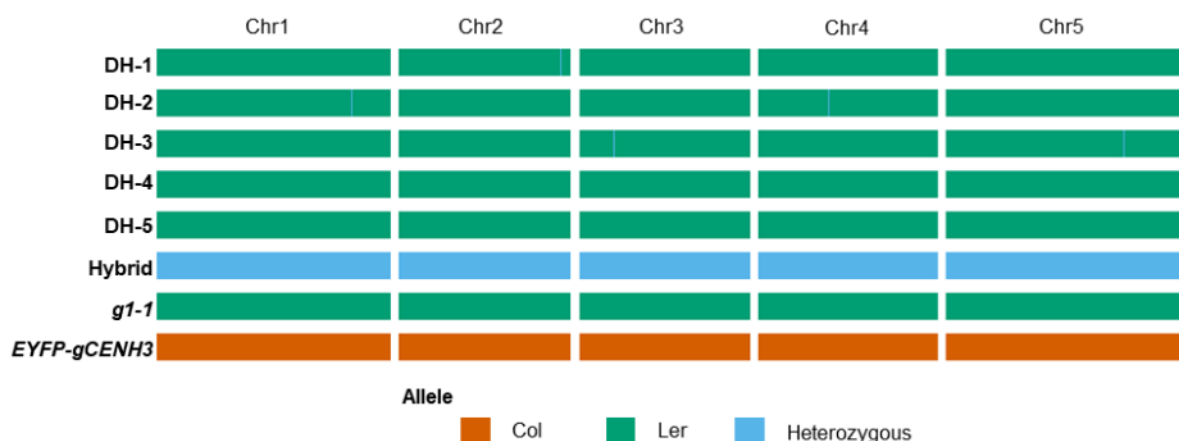
Male parents	Female parents							
	Male sterile tester [ <i>gl1(-/-)</i> , <i>ams(-/-)</i> ]				Male fertile tester [ <i>gl1(-/-)</i> ]			
	n	G	H	D	n	G	H	D
DVS-6	137	1.46% (2)	0%	1.46% (2)	14	14.29% (2)	0%	14.29% (2)
DVS-7	421	0%	0%	0%	24	0%	0%	0%
DVS-8	315	0.95% (3)	0%	0.95% (3)	18	16.67% (3)	0%	16.67% (3)
DVS-9	436	7.11% (31)	0%	7.11% (31)	366	7.65% (28)	0.27% (1)	7.38% (27)
DVS-10	187	1.07% (2)	1.07% (2)	0%	315	3.17% (10)	0%	3.17% (10)
DVS-11	300	0.33% (1)	0.33% (1)	0%	230	0.43% (1)	0.43% (1)	0%
DVS-12	257	6.61% (17)	0%	6.61% (17)	170	3.53% (6)	0	3.53% (6)
DVS-13	310	8.71% (27)	0.32% (1)	8.39% (26)	278	16.55% (46)	0%	16.55% (46)
DVS-14	345	0%	0%	0%	15	13.33% (2)	0%	13.33% (2)
Control	523	0%	0%	0%	25	0%	0%	0%

#### 4.6.5. Female-like diploids are genomically similar to the maternal parent

Given that the female-like diploids are unlikely to be the result of self-pollination, the genomic constitution of the female-like diploids were determined. The [*gl1(-/-)*] is from the genetic background of Ler ecotype and [*cenh3(-/-), EYFP-gCENH3(+/+)*] belongs to Col ecotype. Therefore, a SNP allelic profiling was performed to characterize the genomic constitution of female-like diploids. Five candidate female-like diploid plants arising from the [*gl1(-/-)*] mother and one true hybrid plant (identified by the presence of trichomes) were selected, and their genomic DNA was subjected to short-read sequencing. The [*gl1(-/-)*] and [*cenh3(-/-), EYFP-gCENH3(+/+)*] plants were also sequenced to identify the polymorphic SNPs.

SNPs that were polymorphic between these plants and perfectly heterozygous in the hybrid were identified. The allelic profiles for these SNPs among the parents, the hybrid, and the candidate female-like diploid plants were plotted as graphical genotypes (Fig. 23). It was found that all five candidates exhibited high genomic similarity to the [*gl1(-/-)*] parent, with no evidence of genomic blocks from the paternal parent. This finding confirms the complete absence of the paternal genome in the female-like diploids.

This raises the intriguing possibility that these female-like diploids might be doubled haploids. Likely, spontaneous genome doubling occurred immediately after the elimination of the paternal chromosome set. To investigate this in detail, there is a need to cross the DVS plant with a female parent heterozygous for several genetic markers and assess whether the progeny achieves complete homozygosity. However, due to time constraints, I was unable to perform these experiments within the scope of this thesis.



**Figure 23: Female-like diploids were genomically similar to the maternal parent.** Chromosome-wide SNP allelic profiles of 5 different female-like diploids (DH-1 to DH-5), a true hybrid, *gl1-1* (maternal parent), and *EYFP-gCENH3* (paternal parent). *EYFP-gCENH3* refers to [*cenh3*(-/-), *EYFP-gCENH3*(+/+)]. [*gl1-1*/*gl1-1*] contributes Ler allele and *EYFP-gCENH3* contributes the Col allele.

#### 4.7. Replacement of non-plant E3-ligase with plant-derived E3-ligases

Paternal haploids, with frequencies of up to 12.90%, can be induced through egg-cell-specific degradation of the CENH3 protein via the SPOP protein. Theoretically, a plant hemizygous for an efficient E3-ligase would produce 50% haploids. Based on this hypothesis, it was posited that the frequency of paternal haploids could be increased by employing a highly efficient E3-ligase for the degradation of target proteins. Consequently, I aimed to identify a highly efficient plant-derived E3-ligase for *in-planta* degradation of the target protein. From a previously analyzed list of E3-ligases evaluated by *in-vivo* protein degradation experiments conducted at Bayer Crop Sciences, USA (J. Lamb, personal communication), the top four candidates (Table 7) were selected for the subsequent experiments. The previously reported ubiquitination domains of the selected E3-ligase were used in this study.

**Table 7:** List of plant-originating E3-ligases utilised in this study. Type refers to the characteristics of the E3-ligase domain. Amino acid co-ordinates corresponds to the ubiquitination domain of the E3-ligases and were used in this study.

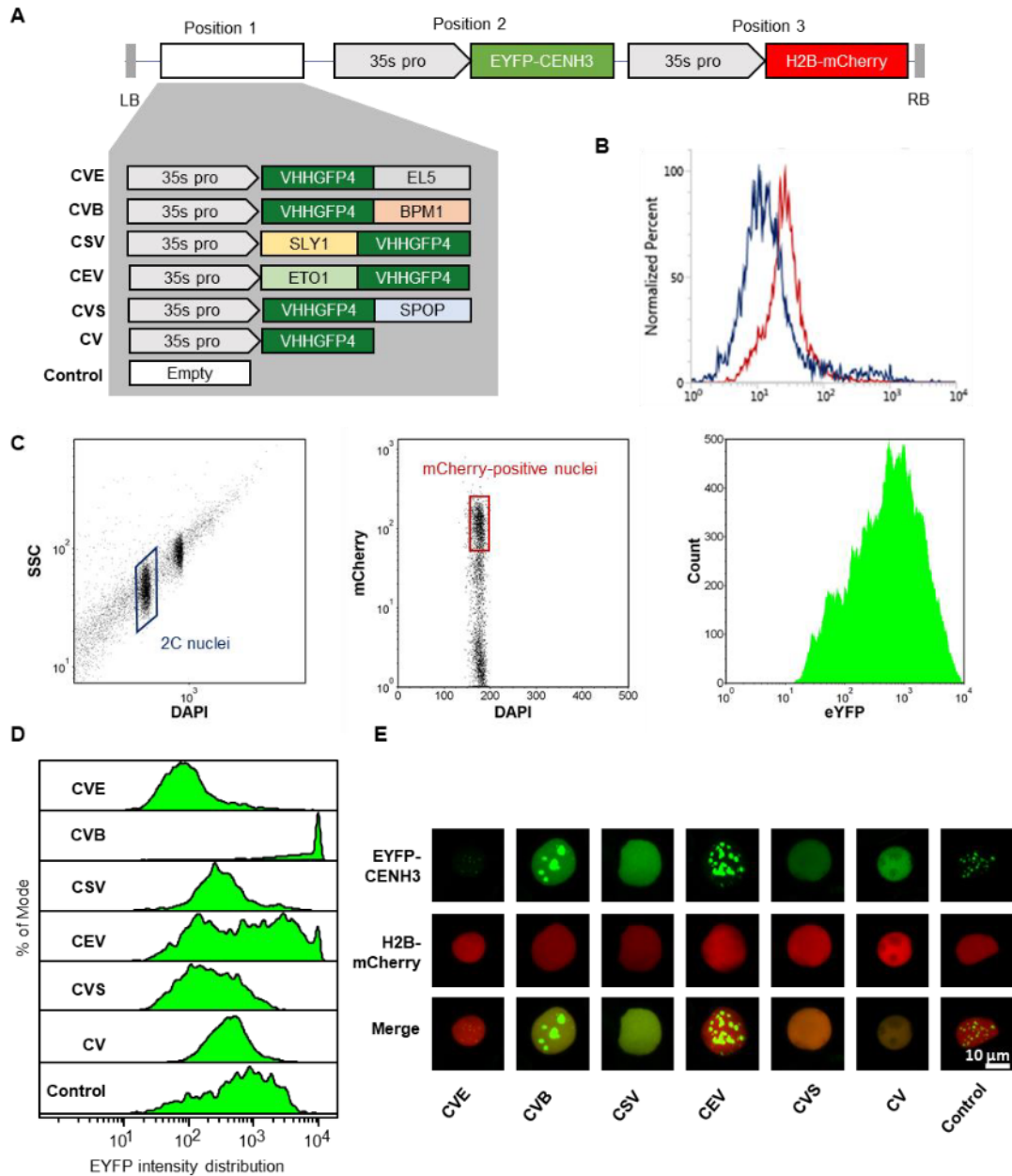
Name	Type	Species of origin	Uniprot ID	Amino acid co-ordinates	Reference
EL5	RING	<i>O. sativa</i>	Q9LRB7	93-181	Katoh et al. (2005)
BPM1	BTB	<i>A. thaliana</i>	Q8L765	170-407	Weber et al. (2005)
SLY1	F-box	<i>A. thaliana</i>	Q9STX3	1-151	Ariizumi et al. (2011)
ETO1	BTB	<i>A. thaliana</i>	O65020	85-418	Yoshida et al. (2005)

#### **4.7.1. Transiently expressed EL5 outperforms other E3-ligase for degradation of EYFP-CENH3**

To conduct *in-planta* degradation experiments, the transgenic constructs were prepared for the transient transformation of *Nicotiana benthamiana* (Fig. 24A). All constructs contain modules for the constitutive expression of EYFP-CENH3 and histone H2B-mCherry proteins. EYFP-CENH3 serves as the target protein for degradation, while H2B-mCherry, labelling the nuclei, acts as a transformation marker. Five different constructs were prepared, each expressing a different E3-ligase fused with the anti-GFP nanobody, VHHGFP4. The E3-ligases included EL5 (CVE), BPM1 (CVB), SLY1 (CSV), ETO1 (CEV), and SPOP (CVS). Additionally, two control constructs were prepared: one expressing only the nanobody (CV) and another containing an empty module (control). The VHHGFP4 is fused to a E3-ligase either of its C-terminus or N-terminus depending on its native substrate interacting domain.

Initially, the suitability of flow cytometry in measuring EYFP degradation by analyzing EYFP signal intensity was investigated. SPOP had previously been reported as efficient in degrading EYFP-CENH3. The EYFP intensity in plants transformed with the CVS construct was compared to that in plants transformed with control constructs using flow cytometry with isolated nuclei. A reduction in EYFP intensity was successfully observed in CVS-transformed plants compared to control-transformed plants (Fig. 24B). Therefore, the flow cytometry was used to compare the degradation efficiency of other E3 ligases. In parallel, the EYFP-CENH3 pattern in leaf nuclei was examined using confocal microscopy.

To measure the intensity of EYFP-CENH3, H2B-mCherry positive 2C nuclei were selected, and the events from different replicates for each construct were pooled for the EYFP-CENH3 intensity comparison (Fig.24C). It was found that CVE-transformed plants exhibited the lowest EYFP intensity, significantly reduced compared to CVS (Fig.24D). CSV had similar EYFP intensity to CVS, while CEV showed a broader distribution and higher intensity than CVS (Fig.24D). The EYFP intensity in CVB-transformed nuclei was very high, reaching the upper detection limit of the cytometer (Fig.24D).



**Figure 24: Evaluation of efficiency of candidate E3-ligases for EYFP-CENH3 degradation.** **A)** Schematic representation of different constructs used for transient transformation of *N. benthamiana*. **B)** Control flow cytometry experiment to detect the degradation of EYFP-CENH3 caused by CVS (blue histogram) compared to the control (red histogram). **C)** Gating schemata to evaluate EYFP-CENH3 intensities of 2C nuclei transformed with different constructs. **D)** Flow-cytometric quantification of EYFP intensities in nuclei expressing different constructs resulting from at least 5 replicates per construct. **E)** Representative confocal images of nuclei showing H2B-mCherry (red) and EYFP-CENH3 (green) for each construct tested.

Leaf nuclei transformed with these constructs were analyzed by confocal microscopy under similar settings (Fig. 24E). In control plants, nuclei exhibited clear centromere-like speckles of EYFP-CENH3 signals. In the nanobody-only control, EYFP-CENH3

was dispersed, but the intensity was much higher compared to CVS. CVE showed the lowest trace of EYFP-CENH3 signals. Aggregates of EYFP-CENH3 were observed in CEV and CVB, with the latter showing very large aggregates corresponding to the high signal intensity measured by the flow cytometry.

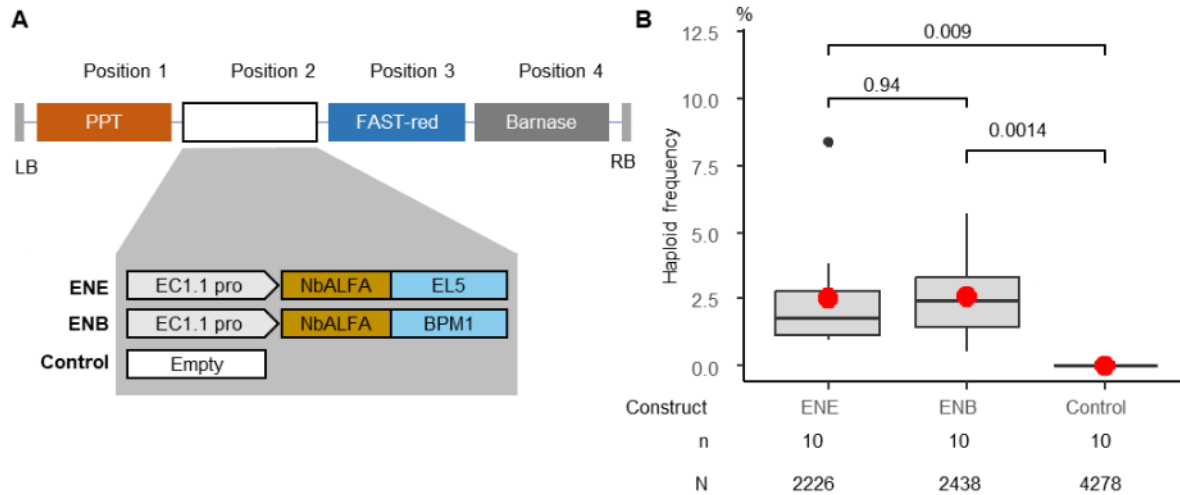
These findings indicate that EL5 (CVE) is the most effective candidate among the evaluated E3 ligases for degrading EYFP-CENH3 signals in transiently transformed *N. benthamiana*. Although BPM1 (CVB) did not result in a comparable reduction of EYFP-CENH3 intensity, it leads to higher levels of EYFP-CENH3 aggregation consistently in all replicates. BPM1 contains a nuclear localization signal at the C-terminus, causing predominant nucleolar agglomeration (Leljak Levanic et al. 2012). In this transient assay for degradation, the C-terminal part of BPM1 was used. Consequently, it was concluded that the EYFP-CENH3 aggregates were mis-localized to the nucleolus due to interaction with BPM1. Therefore, it was decided to express BPM1 in the egg cell to investigate whether the re-localization of tagged-CENH3 alone would be sufficient to induce haploids.

#### **4.7.2. Both EL5 AND BPM1 E3-ligases fused to anti-ALFA nanobody can induce paternal haploids**

Next, both EL5 and BPM1 E3-ligases were evaluated for their effectiveness in inducing paternal haploids. Transgenic constructs were prepared by fusing these E3 ligases to Nb-ALFA (Fig. 25A). The constructs contained either *EC1.1pro:NbALFA-EL5* (ENE), *EC1.1pro:NbALFA-BPM1* (ENB), or an empty module (control) in the second position. These constructs were transformed into a genetic background of [*cenh3(-/-)*, *ALFA-gCENH3(+/+)*].

Ten individual T1 plants were crossed with [*gl1(-/-)*] pollen donors, and their progenies were screened for glabrous haploid plants (Fig. 25B). No glabrous plants were found among the progenies from the control mothers. However, mothers carrying either of the engineered E3 ligases induced haploid (glabrous) plants. The haploid induction frequency of individual ENE T1 plants ranged from 0.93% to 8.37%, while for ENB it ranged from 0.52% to 5.73%. Although the median haploid induction frequency of different ENB T1 plants (2.45%) was higher than that of ENE (1.76%), the difference between their means (2.60% for ENB and 2.54% for ENE) was not statistically

significant. The results indicate that the re-localization of tagged CENH3 is equally capable of inducing haploids compared to their degradation.



**Figure 25: Both EL5 and BPM1 E3-ligases can induce paternal haploids with similar frequency.** **A)** Schematic representation of different constructs used for the egg cell-specific degradation of ALFA-CENH3 using plant-derived E3-ligase **B)** Haploidization frequency of independent hemizygous T1 plants carrying different constructs. Red dots indicates the mean haploidization frequency. 'n' indicates the number of individual T1 plants evaluated for each construct. 'N' indicates the total number of plants screened for glabrous phenotype, from all T1 plants evaluated for a particular construct. p values were added on top of the comparison bars.



## 5. Discussion

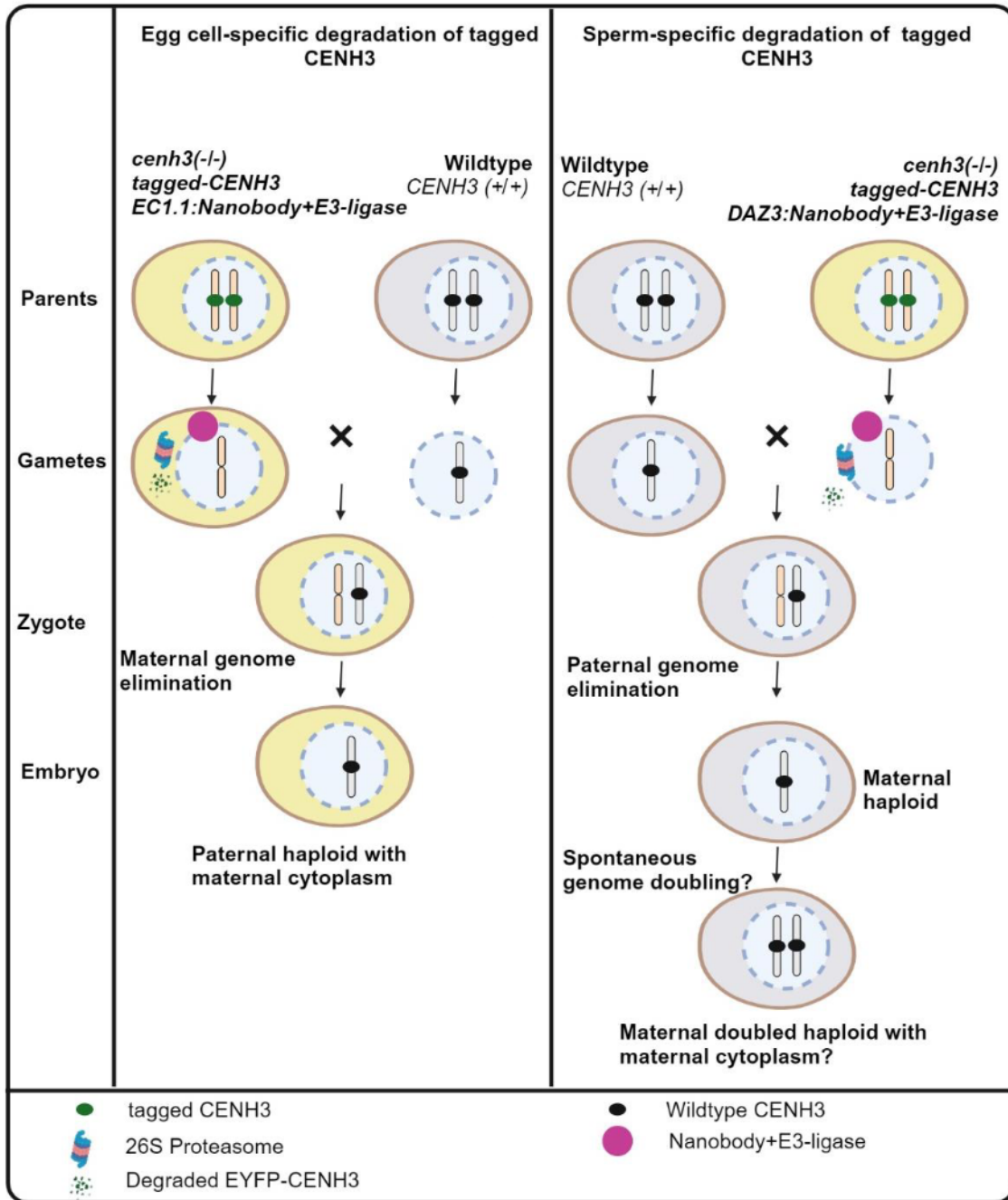
In this study, the targeted degradation of tagged versions of CENH3 in the gametic nucleus is reported as a novel method to induce haploid plants. This method was formulated based on observing the pre- and post-fertilization status of EYFP-CENH3 signals in parental chromosomes during a haploid induction cross in *A. thaliana*. The findings of this study demonstrate that degrading tagged CENH3 in the egg cell nucleus of the female parent can induce paternal haploids (Fig. 26). Additionally, it was found that degrading tagged CENH3 in the sperm nuclei of the male parent can induce female-like diploids (Fig. 26). However, it is uncertain whether the formation of these female-like diploids is due to early spontaneous haploid genome doubling.

### 5.1. A critical level of CENH3 asymmetry is necessary for haploid induction

The analysis of EYFP-CENH3 signals in early embryos indicates the existence of a threshold level of CENH3 asymmetry between parental chromosomes necessary for UGE. In embryos from [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] x wild-type crosses, the chromosome set with undetectable levels of CENH3, from the female parent, is eliminated. In contrast, differences in CENH3 levels between parental chromosomes in early embryos derived from wild-type x [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] crosses do not induce elimination of paternal chromosomes. Though, the paternal chromosomes carry comparatively less CENH3 than the maternal wild-type chromosomes, it is not eliminated. Our results align with those of Marimuthu et al. (2021) regardless of the type of CENH3 modification (GFP-tailswap vs EYFP-CENH3). Therefore, the centromere size hypothesis remains a valid model for explaining centromere-mediated UGE, with the additional requirement of a critical level of CENH3 difference between parental chromosomes.

Marimuthu et al. (2021) suggested the selective removal of modified CENH3 from the egg cell. However, I propose an alternative hypothesis: modified CENH3 is diluted in both gametes of the haploid inducer, with greater dilution occurring in the egg cell than in the sperm. Recent studies indicate that the mature egg cell of Arabidopsis is in the G2 phase of the cell cycle, while the sperm nuclei are arrested in the G1 phase (Simonini et al. 2024; Liu et al. 2021). Assuming there is no reloading of modified

CENH3, its dilution in the egg cell nucleus would be twice as much as in the sperm nuclei. This increased dilution results from an additional mitosis in the female gametophyte and DNA replication-mediated dilution in the egg cell nucleus.



**Figure 26: Schemata representing gametophytic degradation of tagged-CENH3 for haploid induction.** Left panel indicates the outcome of egg cell-specific degradation of tagged CENH3 and the right one indicates that of sperm-specific degradation of tagged-CENH3.

## **5.2. Seed death may be caused by other reasons in addition to endosperm-failure**

Shriveled seeds were common among the progeny of crosses with a high haploid frequency, likely due to endosperm failure caused by mis-segregation (Ravi et al. 2014; Kuppu et al. 2020). Observations from a previous study support this assumption. Demidov et al. (2022) found that haploid seeds were always accompanied by triploid endosperm. Therefore, the possibility of seed death due to impaired development of the endosperm cannot be excluded. However, shriveled seeds were also observed among self-pollinated T2 seeds of T1 plants carrying constructs for egg cell-specific degradation of EYFP-CENH3 using E3-ligases (Fig. 12). Additionally, shriveled seeds were noted among F1 seeds when these T1 plants were crossed with wild-type plants (data not shown).

The transgenic constructs were designed to affect only the egg cell, without impacting chromosome segregation in the endosperm. By comparing the expression pattern of the EC1.1 gene (At1G31450) in egg cells and central cells using the dataset from (Song et al. 2020), there was no evidence of EC1.1 promoter leakiness into the central cell. Therefore, I conclude that other factors contribute to seed lethality in addition to endosperm failure. I speculate that genomic imprinting (Simonini et al. 2021) and embryo-endosperm communication signals (Zhang et al. 2023b) may play a role in causing shriveled seeds beyond genome imbalance in the endosperm.

## **5.3. The transgenic supply of CENH3 may also contribute to haploid induction**

Initially, it was hypothesized that the bulkiness of the EYFP-tag might contribute to the high haploid frequency in [*cenh3(-/-)*, EYFP-gCENH3(+/+)] lines after outcrossing with wild-type. Adding the EYFP-tag increased the size of native CENH3 from 19.7 kDa to 48.3 kDa, whereas the ALFA-tagged CENH3 is just 21.6 kDa. Despite the small size of the tag, haploid induction rates of up to 15% were observed among a few ALFA-CENH3 complemented lines after outcrossing with wild-type, disproving the hypothesis. Additionally, there was high variability between T2 families for the haploid frequency despite the uniform growth conditions. Based on these observations, I argue that transgene copy number and positional effects may indirectly contribute to high haploid induction rates by altering CENH3 expression levels.

A number of substitution point mutations of CENH3 have been evaluated for their effects on haploid induction using modified CENH3 supplied as a transgene in *Arabidopsis* (Karimi-Ashtiyani et al. 2015; Kuppu et al. 2015; Kuppu et al. 2020). For example, the L130F substitution mutation has been shown to induce 0 to 4.8% haploids, depending on the transgene copy number (Karimi-Ashtiyani et al. 2015). However, base editing of the endogenous CENH3 with the L130F mutation has been shown to induce 2.9% haploids (Wang and Ouyang 2023). Therefore, transgenic supply of CENH3 point mutations may have confounding effects on haploid induction. It would be useful to confirm these findings using a TILLING mutant or by isolating a base-edited substitution mutant of the endogenous CENH3 to identify the effect of each point mutation on the haploid frequency. Alternatively, using more replicates could provide further clarity.

Nevertheless, the possible influence of the tag added to the CENH3 protein cannot be excluded. Recently it was demonstrated that the tagged CENP-A (CENH3 in mammals) was functionally inequivalent to the native CENP-A protein (Bui et al. 2024). Therefore, the tag in addition to the positional effects of the transgene insertion can also influence the haploid induction frequency.

#### **5.4. How to target CENH3?**

This study demonstrates that the targeted degradation of tagged CENH3 can induce haploids regardless of the tag used (EYFP or ALFA). Although structural modifications of CENH3 can induce haploids, a specific successful modification in one plant species may not be suitable for others (Wang et al. 2021). RNAi-mediated CENH3 dilution could help induce parental asymmetry for CENH3. However, the high dilution necessary for haploid induction might negatively impact the fitness of the potential haploid inducer (Manape et al. 2024). Also, the RNAi approach was not found to be successful in maize and *Arabidopsis* (Kelliher et al. 2016; Ahmadli et al. 2023). There exists a need for a universal strategy without involving any species-specific optimizations. Therefore, depleting CENH3, specifically in gametes, as described in this study, has significant potential to become a generic *in-planta* haploid induction method for many crop species.

The simplest approach is to isolate a native CENH3 binder, such as a nanobody, to degrade wild-type CENH3 in combination with a nanobody-E3 ligase. Theoretically, a nanobody binding to any antigen of interest can be isolated (Kourelis et al. 2023). However, despite the costly and laborious screening of synthetic and immune libraries, a nanobody against the AtCENH3 protein N-tail was not isolated. The N-terminal tail region used for immunization corresponds to the hyper-variable and unstructured region of the CENP-A protein (Tachiwana et al. 2012). A potential reason for the failure to isolate a nanobody could be the absence of any secondary structures of the applied recombinant CENH3 N-tail protein, which may favor strong interactions between the nanobody and the antigen. In addition, it is likely that candidate nanobodies against the CENH3 N-tail were lost during affinity selection due to their weak affinity.

Given this risk, it is worth identifying native CENH3-binding endogenous proteins of Arabidopsis and using their binding motifs to engineer E3 ligases in the future. So far, NASP (Sim3) is the only known candidate to interact directly with CENH3, but it does not interact with CENH3 loaded onto the centromeric chromatin (Le Goff et al. 2020). Therefore, further attempts are necessary to identify candidate proteins that can interact with free and CENH3 protein loaded onto the centromeric chromatin.

Currently, two different options are available for attempting gametophytic CENH3 degradation in crops. The first approach involves three steps: isolating a CENH3 null mutant, complementing the null mutant with a tagged-CENH3 variant, and then transforming the complemented line with a construct for gamete-specific degradation. Alternatively, the targeted insertion of a tag into the endogenous CENH3 can simplify the first two steps of the previous approach. Given its smaller size and recent advancements in CRISPR-mediated knock-in technology (Schreiber et al. 2024; Zong et al. 2022), the targeted insertion of the ALFA-tag at the N-terminus of endogenous CENH3 is now feasible in both monocots and dicots. The CRISPR-mediated homology-directed ALFA insertion in Arabidopsis is currently under investigation.

## **5.5. Comparison of different E3-ligases**

Demidov et al. (2022) reported that the E3 ligase Nslmb causes higher haploid rates than SPOP through zygotic degradation of EYFP-CENH3. In this study using EYFP-CENH3 complemented lines, no statistical differences were found between SPOP

(EVS) and Nslmb (ENV) in terms of haploid induction ability, likely due to variability from the genetic background and different E3 ligases used. However, in the ALFA-CENH3 complemented background, constructs with SPOP (ENS) resulted in a higher mean haploid induction frequency than Nslmb (ENN), contrasting with the previous study. There is no explanation for this discrepancy except for the difference in the tagged CENH3 variant used for complementation. Nevertheless, the frequency of phosphinothricin-resistant T2 seedlings and the frequency of T2 shriveled seeds from the EVS and ENV T1 plants in the [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] background provide additional indication that SPOP is more effective than Nslmb in causing uniparental genome instability.

As a further attempt to optimize gametophytic degradation components, some plant-derived E3 ligases were tested. Given that SPOP and Nslmb were derived from non-plant species, it was anticipated that using plant-derived E3 ligases would increase haploid induction frequencies. Top four candidates out of a list of seven plant-derived E3-ligases provided by Bayer Crop Sciences (J. Lamb, personal communication) were analyzed using a transient assay. Based on an *in-planta* EYFP-CENH3 degradation assay, two promising candidates, BPM1 and EL5 derived from *O. sativa* and *A. thaliana*, respectively, were selected for evaluation of their haploid induction frequencies through gametophytic CENH3 degradation. To my knowledge, this study is the first to evaluate plant-derived E3 ligases for targeted protein degradation in a plant system, whereas previous studies used those derived from only non-plant organisms (Faden et al. 2016; Huang and Rojas-Pierce 2024; Baudisch et al. 2018; Demidov et al. 2022; Sorge et al. 2021; Ma et al. 2019).

EL5 was highly efficient in degrading EYFP-CENH3 in transiently transformed *N. benthamiana* leaves. Nevertheless, SPOP had the highest median haploid induction frequency among analyzed E3 ligases when comparing T1 plants. The order of E3 ligases for median haploid induction frequency is SPOP > BPM1 > EL5 > Nslmb. The results from BPM1 confirm that CENH3 relocalization is effective in inducing haploids in addition to its degradation. The variation among E3 ligases could be due to limitations caused by the downstream adapter proteins involved in the ubiquitination process of the target protein. Possible steric hindrances that prevent conjugation to the substrate, as well as the availability of lysine residues for ubiquitin conjugation,

may also influence the efficiency of degradation. These factors must be considered for necessary optimizations in future studies.

### **5.6. Comparison of haploidization frequencies**

Although the possibility of generating haploids was successfully demonstrated through gametophytic degradation of CENH3, the maximum frequency achieved was 12.90% from a DVS T1 plant, which is significantly lower than the 34 to 45% achieved with a GFP-tailswap modification (Ravi and Chan 2010). However, given that the effects of the constructs were gametophytic and the T1 plants were hemizygous, a haploid frequency of homozygous lines is expected to be more than twice that of the corresponding hemizygous lines. This is because up to 50% of the gametes in hemizygous lines do not contain transgenic constructs, and the haploid frequencies were calculated based on germinated seedlings.

The effective haploid frequency should be calculated in terms of seeds from transgene-bearing ovules, suggesting that the frequency could be increased further by obtaining homozygous plants. Different strategies to isolate homozygous T1 plants need to be explored, as transgene segregation analysis is hampered by the presence of shriveled seeds upon self-pollination. Alternatively, the recently demonstrated toxin–antidote transgene-drive strategy (Oberhofer et al. 2024; Liu et al. 2024a) should also be investigated.

### **5.7. Maternal versus paternal haploidization**

Paternal haploid induction is crucial for cytoplasm swapping to transfer male sterility. Two recent studies demonstrated the use of centromere-based haploid induction for the rapid transfer of cytoplasmic male sterility to inbreds in broccoli and maize (Han et al. 2024; Bortiri et al. 2024). Nevertheless, maternal haploids are more relevant for breeding programs as their cytoplasm is derived from the non-inducer parent. Therefore, sperm-specific degradation of EYFP-CENH3 was attempted, and it was found that female-like diploids were generated upon outcrossing. In addition to female-like diploids, haploids were also observed among the screened progenies but comparatively less frequent than the diploids. Since there was no obvious correlation between the level of pollen without EYFP-CENH3 signals and the frequency of female-

like diploids, these plants were considered to be the result of stochastic events. However, it is uncertain if the diploidization occurred pre-fertilization due to unreduced gametes or post-fertilization by early haploid genome doubling.

Early haploid genome doubling has been previously described but is rare and has been only systematically studied in maize (Wu et al. 2014). A high frequency of spontaneously doubled maternal haploids has been reported among the progeny of interploidy crosses between *Brassica napus* and *Brassica* allo-octaploids (Fu et al. 2018). Paternal chromosome elimination coupled with maternal genome doubling was shown to be the reason behind the formation of maternal doubled haploids (Zhao et al. 2023). This phenomenon has also been described in interploidy crosses of maize (Elkonin et al. 2023). The formation of doubled haploids helps to avoid the need to treat haploid plants with hazardous chemicals like colchicine for genome doubling. However, it is necessary to confirm if the female-like diploids are indeed spontaneously doubled haploids.



## 6. Outlook

Future studies should address the following research questions:

1. The ALFA-CENH3 has been provided as a transgene in this study to induce haploids. Verifying if the results are repeatable when the ALFA-tag is added to endogenous CENH3 using CRISPR-mediated homology-directed insertion is important.
2. Homozygous transgene insertion lines for the E3 ligase construct need to be recovered using approaches other than herbicide resistance segregation. Segregating population obtained from a plant heterozygous for the ALFA-CENH3 and the E3-ligase bearing transgene can be used for isolation of such homozygous lines. These homozygous lines will be crucial for realizing the actual haploid induction potential of the gametophytic CENH3 degradation approach.
3. The effect of genotype and environment (e.g. temperature, light intensity) on the haploid induction frequencies of the gametophytic CENH3 degradation approach has to be analyzed.
4. Additional plant derived E3 ligases have to be evaluated, and the impact of adding lysine-rich flexible linkers between ALFA-tag and CENH3 on haploid induction frequencies has to be investigated.
5. Different gamete-specific promoters for the expression of E3 ligase should be assessed for their effects on haploid induction.
6. It is noteworthy that shriveled seeds were formed even when there was no effect of CENH3-degradation on the central cell (or endosperm). Therefore, the fate of endosperm derived from the haploid induction cross has to be examined for any developmental defects.
7. The suitability of the fluorescent seed marker (FAST) (Shimada et al. 2010) and additional dominant transgenic markers such as RUBY (Wang et al. 2023a) for efficient haploid screening has to be investigated.
8. To identify the origin of female-like diploids after the application of a male haploid inducer, segregation analysis of heterozygous genetic markers should

be performed. If these diploids are confirmed to be the result of early spontaneous genome doubling of haploid cells, the underlying mechanism behind this phenomenon needs to be elucidated.

9. One of the main areas of focus for the future is to validate the gametophytic CENH3 degradation approach for haploid induction using a crop plant. An economically important crop species that lacks an *in-planta* haploid induction system would be an ideal candidate for such a study.

## 7. Summary

This study presents mechanisms behind centromere-mediated uniparental genome elimination and proposes a potential universal strategy for haploid induction in plants using *Arabidopsis* as a model.

1. Analysis of the EYFP-CENH3 pattern in gametes from a haploid inducer genotype and early embryos from various crosses reveals that extreme asymmetry between parental chromosomes for the amount of centromeric CENH3 is crucial for eliminating chromosomes with 'weak' centromeres during early embryogenesis.
2. Constructs designed for egg cell-specific degradation of EYFP-CENH3 resulted in the formation of shriveled seeds, impacting transgene segregation analysis using herbicide resistance. While T1 plants for these constructs induce haploids more frequently compared to controls, statistical differences between the constructs were not observed.
3. Both synthetic and immune nanobody libraries were found to be not useful for the isolation of a nanobody against the CENH3 N-tail antigen. The lack of any secondary structure in the N-tail of recombinantly prepared CENH3 protein could be a possible reason for the failure.
4. This study is the first report on the utilization of the ALFA-tag in a plant system. ALFA-tag in combination with its nanobody, was used for gametophytic degradation of the CENH3 protein.
5. ALFA-CENH3 was found to be efficient in complementation of the CENH3 null mutation [*cenh3(-/-)*], when compared to the EYFP-CENH3. Some of the ALFA-CENH3 complemented lines were found to induce paternal haploids up to 15%.
6. Further experiments involved the transformation of [*cenh3(-/-)*, *ALFA-gCENH3(+/+)*], with constructs targeting egg cell-specific degradation of ALFA-CENH3, where the E3 ligase SPOP was identified as superior to Nslmb in inducing paternal haploids.
7. Identification of a sperm-specific promoter, *DAZ3*, was performed using publicly available *Arabidopsis* transcriptome datasets. *DAZ3* promoter was used for the

expression of E3 ligase, SPOP fused with GFP-nanobody in [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] genetic background. Microscopy and flow cytometry confirmed the degradation of EYFP-CENH3 in sperm nuclei.

8. Interestingly, sperm-specific degradation of EYFP-CENH3 led to the formation of female-like diploids when used as pollen donors with wild-type CENH3-carrying parents, necessitating further investigation into the origin of these diploids.
9. Evaluation of four plant-derived E3 ligases in *N. benthamiana* demonstrated that EL5 from *O. sativa* efficiently degrades EYFP-CENH3, while BPM1 from *Arabidopsis* relocates most of the EYFP-CENH3 protein to the nucleolus. Fusion of EL5 and BPM1 with ALFA-specific nanobody in egg cell-specific degradation or relocation of ALFA-CENH3 showed equal efficacy in inducing paternal haploids. Notably, SPOP protein derived from the mammalian system induced a higher frequency of paternal haploids compared to other plant- or animal-derived E3 ligases.

## 8. Zusammenfassung

In dieser Studie werden die Mechanismen aufgezeigt, die der durch das Zentromer vermittelten uniparentalen Genomeliminierung zugrunde liegen. Desweiteren wird eine potenzielle universelle Strategie zur universellen Haploideninduktion in Pflanzen vorgeschlagen, wobei Arabidopsis als Modell diente.

1. Die Analyse der EYFP-CENH3 Verteilungsmuster in Gameten eines haploiden Induktionsgenotyps sowie in frühen Embryonen aus verschiedenen Kreuzungen zeigt, dass eine extreme Asymmetrie zwischen den elterlichen Chromosomen in Bezug auf die Menge an zentromerischem CENH3 entscheidend für die Eliminierung von Chromosomen mit "schwachen" Zentromeren während der frühen Embryogenese ist.
2. Konstrukte, die für den Eizell-spezifischen Abbau von EYFP-CENH3 entwickelt wurden, führten zur Bildung von schrumpelig Samen, was sich auf die Segregationsanalyse von Transgenen mit Hilfe der Herbizidresistenz auswirkte. Während die T1-Pflanzen für diese Konstrukte im Vergleich zu den Kontrollen häufiger Haploide ausbilden, wurden zwischen den Konstrukten keine statistischen Unterschiede festgestellt.
3. Sowohl synthetische als auch Immun-Nanokörper Bibliotheken erwiesen sich als nicht nützlich für die Isolierung eines Nanokörpers gegen das CENH3 N-Tail-Antigen. Das Fehlen jeglicher Sekundärstruktur im N-Terminusende des rekombinant-hergestellten CENH3 Proteins könnte ein möglicher Grund für den Misserfolg sein.
4. Die vorliegende Studie ist der erste Bericht über die Verwendung von ALFA-tag in einem Pflanzensystem. ALFA-tag in Kombination mit seinem Nanokörper wurde für den gametophytischen Abbau des CENH3-Proteins verwendet.
5. ALFA-CENH3 erwies sich bei der Komplementierung der CENH3-Nullmutation [*cenh3(-/-)*] im Vergleich zu EYFP-CENH3 als effizienter. Bei einigen der ALFA-CENH3-komplementierten Linien wurde festgestellt, dass sie bis zu 15 % väterliche Haploide induzieren.
6. Weitere Experimente umfassten die Transformation von [*cenh3(-/-)*, *ALFA-gCENH3(+/+)*] Pflanzen mit Konstrukten, die auf den Eizell-spezifischen Abbau

von ALFA-CENH3 abzielen, wobei sich die E3-Ligase SPOP bei der Induktion von väterlichen Haploiden als besser erwies im Vergleich zu Nslmb.

7. Unter Zuhilfenahme öffentlich zugänglicher Arabidopsis Transkriptom-Datensätze konnte der Spermienkern-spezifische Promotor DAZ3 identifiziert werden. Der DAZ3-Promotor wurde für die Expression der E3-Ligase SPOP, fusioniert mit dem GFP-Nanobody im genetischen Hintergrund [*cenh3(-/-)*, *EYFP-gCENH3(+/+)*] verwendet. Mikroskopisch und durchflusszytometrische Analysen bestätigten den Abbau von EYFP-CENH3 in den Spermienkernen.
8. Interessanterweise führte die Spermienkern-spezifische Expression von EYFP-CENH3 zur Bildung von mutterähnlichen Diploiden, wenn sie als Pollenspender mit Wildtyp-CENH3-tragenden Eltern verwendet wurden. Weitere Untersuchungen zum Ursprung dieser Diploiden sind erforderlich.
9. Die Auswertung von vier pflanzlichen E3-Ligasen in *N. benthamiana* zeigte, dass EL5 aus *O. sativa* EYFP-CENH3 effizient abbaut, während BPM1 aus Arabidopsis den Großteil des EYFP-CENH3-Proteins in den Nukleolus verlagert. Die Fusion von EL5 und BPM1 mit ALFA-spezifischen Nanokörpern zum Eizell-spezifischen Abbau oder die Verlagerung von ALFA-CENH3 zeigte die gleiche Wirksamkeit bei der Erzeugung von väterlichen Haploiden. Insbesondere das aus dem Säugetiersystem stammende SPOP-Protein induzierte im Vergleich zu anderen aus Pflanzen oder Tieren stammenden E3-Ligasen eine höhere Häufigkeit von väterlichen Haploiden.

## 9. References

- Ahmadli U, Kalidass M, Khaitova LC, Fuchs J, Cuacos M, Demidov D, Zuo S, Pecinkova J, Mascher M, Ingouff M, Heckmann S, Houben A, Riha K, Lermontova I (2023) High temperature increases centromere-mediated genome elimination frequency and enhances haploid induction in Arabidopsis. *Plant Commun* 4 (3):100507. doi:10.1016/j.xplc.2022.100507
- Ariizumi T, Lawrence PK, Steber CM (2011) The role of two f-box proteins, SLEEPY1 and SNEEZY, in Arabidopsis gibberellin signaling. *Plant Physiol* 155 (2):765-775. doi:10.1104/pp.110.166272
- Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ (2006) A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* 126 (5):995-1004. doi:10.1016/j.cell.2006.07.025
- Baudisch B, Pfort I, Sorge E, Conrad U (2018) Nanobody-Directed Specific Degradation of Proteins by the 26S-Proteasome in Plants. *Front Plant Sci* 9:130. doi:10.3389/fpls.2018.00130
- Black BE, Cleveland DW (2011) Epigenetic centromere propagation and the nature of CENP-a nucleosomes. *Cell* 144 (4):471-479. doi:10.1016/j.cell.2011.02.002
- Bodor DL, Mata JF, Sergeev M, David AF, Salimian KJ, Panchenko T, Cleveland DW, Black BE, Shah JV, Jansen LE (2014) The quantitative architecture of centromeric chromatin. *Elife* 3:e02137. doi:10.7554/eLife.02137
- Bondeson DP, Mares A, Smith IE, Ko E, Campos S, Miah AH, Mulholland KE, Routly N, Buckley DL, Gustafson JL, Zinn N, Grandi P, Shimamura S, Bergamini G, Faeltsh-Savitski M, Bantscheff M, Cox C, Gordon DA, Willard RR, Flanagan JJ, Casillas LN, Votta BJ, den Besten W, Famm K, Kruidenier L, Carter PS, Harling JD, Churcher I, Crews CM (2015) Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat Chem Biol* 11 (8):611-617. doi:10.1038/nchembio.1858
- Borg M, Brownfield L, Khatab H, Sidorova A, Lingaya M, Twell D (2011) The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in Arabidopsis. *Plant Cell* 23 (2):534-549. doi:10.1105/tpc.110.081059
- Borges F, Gomes G, Gardner R, Moreno N, McCormick S, Feijo JA, Becker JD (2008) Comparative transcriptomics of Arabidopsis sperm cells. *Plant Physiol* 148 (2):1168-1181. doi:10.1104/pp.108.125229
- Bortiri E, Selby R, Egger R, Tolhurst L, Dong S, Beam K, Meier K, Fabish J, Delaney D, Dunn M, McNamara D, Setliff K, Castro Miranda Lunny R, Gergen S, Dawe RK, Kelliher T (2024) Cyto-swapping in maize by haploid induction with a cenH3 mutant. *Nat Plants* 10 (4):567-571. doi:10.1038/s41477-024-01630-1
- Budhagatapalli N, Halbach T, Hiekel S, Buchner H, Muller AE, Kumlehn J (2020) Site-directed mutagenesis in bread and durum wheat via pollination by cas9/guide RNA-transgenic maize used as haploidy inducer. *Plant Biotechnol J* 18 (12):2376-2378. doi:10.1111/pbi.13415
- Bui M, Baek S, Bentahar RS, Melters DP, Dalal Y (2024) Native and tagged CENP-A histones are functionally inequivalent. *Epigenetics Chromatin* 17 (1):19. doi:10.1186/s13072-024-00543-9

- Capitao C, Tanasa S, Fulnecek J, Raxwal VK, Akimcheva S, Bulankova P, Mikulkova P, Crhak Khaitova L, Kalidass M, Lermontova I, Mittelsten Scheid O, Riha K (2021) A CENH3 mutation promotes meiotic exit and restores fertility in SMG7-deficient *Arabidopsis*. *PLoS Genet* 17 (9):e1009779. doi:10.1371/journal.pgen.1009779
- Caussinus E, Kanca O, Affolter M (2011) Fluorescent fusion protein knockout mediated by anti-GFP nanobody. *Nat Struct Mol Biol* 19 (1):117-121. doi:10.1038/nsmb.2180
- Caussinus E, Kanca O, Affolter M (2013) Protein knockouts in living eukaryotes using deGradFP and green fluorescent protein fusion targets. *Curr Protoc Protein Sci* 73 (1):30 32 31-30 32 13. doi:10.1002/0471140864.ps3002s73
- Chen B, Maas L, Figueiredo D, Zhong Y, Reis R, Li M, Horstman A, Riksen T, Weemen M, Liu H, Siemons C, Chen S, Angenent GC, Boutilier K (2022) BABY BOOM regulates early embryo and endosperm development. *Proc Natl Acad Sci U S A* 119 (25):e2201761119. doi:10.1073/pnas.2201761119
- Chen X, Li Y, Ai G, Chen J, Guo D, Zhu Z, Zhu X, Tian S, Wang J, Liu M, Yuan L (2023) Creation of a watermelon haploid inducer line via CIDMP3-mediated single fertilization of the central cell. *Hortic Res* 10 (6):uhad081. doi:10.1093/hr/uhad081
- Cheng Z, Sun Y, Yang S, Zhi H, Yin T, Ma X, Zhang H, Diao X, Guo Y, Li X, Wu C, Sui Y (2021) Establishing in planta haploid inducer line by edited SiMTL in foxtail millet (*Setaria italica*). *Plant Biotechnol J* 19 (6):1089-1091. doi:10.1111/pbi.13584
- Clausen RE, Mann MC (1924) Inheritance in *Nicotiana Tabacum*: V. The Occurrence of Haploid Plants in Interspecific Progenies. *Proc Natl Acad Sci U S A* 10 (4):121-124. doi:10.1073/pnas.10.4.121
- Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16 (6):735-743. doi:10.1046/j.1365-313x.1998.00343.x
- Conner JA, Mookkan M, Huo H, Chae K, Ozias-Akins P (2015) A parthenogenesis gene of apomict origin elicits embryo formation from unfertilized eggs in a sexual plant. *Proc Natl Acad Sci USA* 112 (36):11205-11210. doi:10.1073/pnas.1505856112
- Conrad U, Plagmann I, Malchow S, Sack M, Floss DM, Kruglov AA, Nedospasov SA, Rose-John S, Scheller J (2011) ELPylated anti-human TNF therapeutic single-domain antibodies for prevention of lethal septic shock. *Plant Biotechnol J* 9 (1):22-31. doi:10.1111/j.1467-7652.2010.00523.x
- Cyprys P, Lindemeier M, Sprunck S (2019) Gamete fusion is facilitated by two sperm cell-expressed DUF679 membrane proteins. *Nat Plants* 5 (3):253-257. doi:10.1038/s41477-019-0382-3
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. *Gigascience* 10 (2). doi:10.1093/gigascience/giab008
- Daniel K, Icha J, Horenburg C, Muller D, Norden C, Mansfeld J (2018) Conditional control of fluorescent protein degradation by an auxin-dependent nanobody. *Nat Commun* 9 (1):3297. doi:10.1038/s41467-018-05855-5



- Daum G, Medzihradzky A, Suzuki T, Lohmann JU (2014) A mechanistic framework for noncell autonomous stem cell induction in Arabidopsis. *Proc Natl Acad Sci U S A* 111 (40):14619-14624. doi:10.1073/pnas.1406446111
- de Beer MA, Giepmans BNG (2020) Nanobody-Based Probes for Subcellular Protein Identification and Visualization. *Front Cell Neurosci* 14:573278. doi:10.3389/fncel.2020.573278
- Demidov D, Lermontova I, Moebes M, Kochevenko A, Fuchs J, Weiss O, Rutten T, Sorge E, Zuljan E, Giehl RFH, Mascher M, Somasundaram S, Conrad U, Houben A (2022) Haploid induction by nanobody-targeted ubiquitin-proteasome-based degradation of EYFP-tagged CENH3 in Arabidopsis thaliana. *J Exp Bot* 73 (22):7243-7254. doi:10.1093/jxb/erac359
- Doležel J, Binarová P, Lcretti S (1989) Analysis of Nuclear DNA content in plant cells by Flow cytometry. *Biologia Plantarum* 31 (2):113-120. doi:10.1007/bf02907241
- Drpic D, Almeida AC, Aguiar P, Renda F, Damas J, Lewin HA, Larkin DM, Khodjakov A, Maiato H (2018) Chromosome Segregation Is Biased by Kinetochore Size. *Curr Biol* 28 (9):1344-1356 e1345. doi:10.1016/j.cub.2018.03.023
- Dunemann F, Krüger A, Maier K, Struckmeyer S (2022) Insights from CRISPR/Cas9-mediated gene editing of centromeric histone H3 (CENH3) in carrot. *bioRxiv*. doi:10.1101/2022.09.19.508489
- Elkonin LA, Mavlyutova LI, Kolesova AY, Panin VM, Tsvetova MI (2023) Elimination of Chromosomes as a Mechanism for the Formation of Diploid Plants in Diploid–Tetraploid Crosses in Maize (*Zea mays* L.). *Russian Journal of Genetics* 59 (9):888-899. doi:10.1134/s102279542309003x
- Engler C, Youles M, Gruetzner R, Ehnert TM, Werner S, Jones JD, Patron NJ, Marillonnet S (2014) A golden gate modular cloning toolbox for plants. *ACS Synth Biol* 3 (11):839-843. doi:10.1021/sb4001504
- Faden F, Ramezani T, Mielke S, Almudi I, Nairz K, Froehlich MS, Hockendorff J, Brandt W, Hoehenwarter W, Dohmen RJ, Schnittger A, Dissmeyer N (2016) Phenotypes on demand via switchable target protein degradation in multicellular organisms. *Nat Commun* 7 (1):12202. doi:10.1038/ncomms12202
- Fu S, Yin L, Xu M, Li Y, Wang M, Yang J, Fu T, Wang J, Shen J, Ali A, Zou Q, Yi B, Wen J, Tao L, Kang Z, Tang R (2018) Maternal doubled haploid production in interploidy hybridization between Brassica napus and Brassica allooctaploids. *Planta* 247 (1):113-125. doi:10.1007/s00425-017-2772-y
- Fulcher LJ, Hutchinson LD, Macartney TJ, Turnbull C, Sapkota GP (2017) Targeting endogenous proteins for degradation through the affinity-directed protein missile system. *Open Biol* 7 (5). doi:10.1098/rsob.170066
- Fulcher LJ, Macartney T, Bozatz P, Hornberger A, Rojas-Fernandez A, Sapkota GP (2016) An affinity-directed protein missile system for targeted proteolysis. *Open Biol* 6 (10). doi:10.1098/rsob.160255
- Gahrtz M, Conrad U (2009) Immunomodulation of plant function by in vitro selected single-chain Fv intrabodies. *Methods Mol Biol* 483:289-312. doi:10.1007/978-1-59745-407-0\_17
- Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E (1983) Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220 (4601):1049-1051. doi:10.1126/science.220.4601.1049

- Gao X, Guo H, Wu J, Fan Y, Zhang L, Guo H, Lian X, Fan Y, Gou Z, Zhang C, Li T, Chen C, Zeng F (2020) Haploid Bio-Induction in Plant through Mock Sexual Reproduction. *iScience* 23 (7):101279. doi:10.1016/j.isci.2020.101279
- Garcia-Aguilar M, Autran D (2018) Localization of Chromatin Marks in Arabidopsis Early Embryos. *Methods Mol Biol* 1675:419-441. doi:10.1007/978-1-4939-7318-7\_24
- Gernand D, Rutten T, Pickering R, Houben A (2006) Elimination of chromosomes in *Hordeum vulgare* x *H. bulbosum* crosses at mitosis and interphase involves micronucleus formation and progressive heterochromatinization. *Cytogenet Genome Res* 114 (2):169-174. doi:10.1159/000093334
- Gibbs DJ, Bacardit J, Bachmair A, Holdsworth MJ (2014) The eukaryotic N-end rule pathway: conserved mechanisms and diverse functions. *Trends Cell Biol* 24 (10):603-611. doi:10.1016/j.tcb.2014.05.001
- Gilles LM, Khaled A, Laffaire JB, Chaignon S, Gendrot G, Laplaige J, Berges H, Beydon G, Bayle V, Barret P, Comadran J, Martinant JP, Rogowsky PM, Widiez T (2017) Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. *EMBO J* 36 (6):707-717. doi:10.15252/embj.201796603
- Gong Z, Wu Y, Koblizkova A, Torres GA, Wang K, Iovene M, Neumann P, Zhang W, Novak P, Buell CR, Macas J, Jiang J (2012) Repeatless and repeat-based centromeres in potato: implications for centromere evolution. *Plant Cell* 24 (9):3559-3574. doi:10.1105/tpc.112.100511
- Gotzke H, Kilisch M, Martinez-Carranza M, Sograte-Idrissi S, Rajavel A, Schlichthaerle T, Engels N, Jungmann R, Stenmark P, Opazo F, Frey S (2019) The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. *Nat Commun* 10 (1):4403. doi:10.1038/s41467-019-12301-7
- Grützner R, Marillonnet S (2020) Generation of MoClo Standard Parts Using Golden Gate Cloning. In: Chandran S, George KW (eds) *DNA Cloning and Assembly: Methods and Protocols*. *Methods in Molecular Biology*. Springer US, New York, NY, pp 107-123
- Han F, Zhang X, Liu Y, Liu Y, Zhao H, Li Z (2024) One-step creation of CMS lines using a BoCENH3-based haploid induction system in Brassica crop. *Nat Plants* 10 (4):581-586. doi:10.1038/s41477-024-01643-w
- Hedhly A, Vogler H, Eichenberger C, Grossniklaus U (2018) Whole-mount Clearing and Staining of Arabidopsis Flower Organs and Siliques. *J Vis Exp* (134). doi:10.3791/56441
- Honys D, Twell D (2004) Transcriptome analysis of haploid male gametophyte development in Arabidopsis. *Genome Biol* 5 (11):R85. doi:10.1186/gb-2004-5-11-r85
- Hua Z, Vierstra RD (2011) The cullin-RING ubiquitin-protein ligases. *Annu Rev Plant Biol* 62 (1):299-334. doi:10.1146/annurev-arplant-042809-112256
- Huang L, Rojas-Pierce M (2024) Rapid depletion of target proteins in plants by an inducible protein degradation system. *Plant Cell*. doi:10.1093/plcell/koae072
- Ingouff M, Rademacher S, Holec S, Soljic L, Xin N, Readshaw A, Foo SH, Lahouze B, Sprunck S, Berger F (2010) Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in Arabidopsis. *Curr Biol* 20 (23):2137-2143. doi:10.1016/j.cub.2010.11.012

- Ingram JR, Schmidt FI, Ploegh HL (2018) Exploiting Nanobodies' Singular Traits. *Annu Rev Immunol* 36 (1):695-715. doi:10.1146/annurev-immunol-042617-053327
- Ishii T, Karimi-Ashtiyani R, Houben A (2016) Haploidization via Chromosome Elimination: Means and Mechanisms. *Annu Rev Plant Biol* 67 (1):421-438. doi:10.1146/annurev-arplant-043014-114714
- Iwata-Otsubo A, Dawicki-McKenna JM, Akera T, Falk SJ, Chmatal L, Yang K, Sullivan BA, Schultz RM, Lampson MA, Black BE (2017) Expanded Satellite Repeats Amplify a Discrete CENP-A Nucleosome Assembly Site on Chromosomes that Drive in Female Meiosis. *Curr Biol* 27 (15):2365-2373 e2368. doi:10.1016/j.cub.2017.06.069
- Jacquier NMA, Calhau ARM, Fierlej Y, Martinant JP, Rogowsky PM, Gilles LM, Widiez T (2023) In planta haploid induction by kokopelli mutants. *Plant Physiol* 193 (1):182-185. doi:10.1093/plphys/kiad328
- Jacquier NMA, Gilles LM, Pyott DE, Martinant JP, Rogowsky PM, Widiez T (2020) Puzzling out plant reproduction by haploid induction for innovations in plant breeding. *Nat Plants* 6 (6):610-619. doi:10.1038/s41477-020-0664-9
- Jang JH, Noh G, Seo HS, Jung KH, Kim YJ, Lee OR (2023a) Loss of function of pollen-expressed phospholipase OsMATL2 triggers haploid induction in japonica rice. *Plant Physiol* 193 (3):1749-1752. doi:10.1093/plphys/kiad422
- Jang JH, Seo HS, Widiez T, Lee OR (2023b) Loss-of-function of gynoeceium-expressed phospholipase pPLAIIgamma triggers maternal haploid induction in Arabidopsis. *New Phytol* 238 (5):1813-1824. doi:10.1111/nph.18898
- Jiang C, Sun J, Li R, Yan S, Chen W, Guo L, Qin G, Wang P, Luo C, Huang W, Zhang Q, Fernie AR, Jackson D, Li X, Yan J (2022) A reactive oxygen species burst causes haploid induction in maize. *Mol Plant* 15 (6):943-955. doi:10.1016/j.molp.2022.04.001
- Jin C, Sun L, Trinh HK, Danny G (2023) Heat stress promotes haploid formation during CENH3-mediated genome elimination in Arabidopsis. *Plant Reprod* 36 (2):147-155. doi:10.1007/s00497-023-00457-8
- Kalinowska K, Chamas S, Unkel K, Demidov D, Lermontova I, Dresselhaus T, Kumlehn J, Dunemann F, Houben A (2019) State-of-the-art and novel developments of in vivo haploid technologies. *Theor Appl Genet* 132 (3):593-605. doi:10.1007/s00122-018-3261-9
- Karimi-Ashtiyani R, Ishii T, Niessen M, Stein N, Heckmann S, Gurushidze M, Banaei-Moghaddam AM, Fuchs J, Schubert V, Koch K, Weiss O, Demidov D, Schmidt K, Kumlehn J, Houben A (2015) Point mutation impairs centromeric CENH3 loading and induces haploid plants. *Proc Natl Acad Sci U S A* 112 (36):11211-11216. doi:10.1073/pnas.1504333112
- Kasajima I, Ide Y, Ohkama-Ohtsu N, Hayashi H, Yoneyama T, Fujiwara T (2012) A protocol for rapid DNA extraction from Arabidopsis thaliana for PCR analysis. *Plant Molecular Biology Reporter* 22 (1):49-52. doi:10.1007/bf02773348
- Kasha KJ, Kao KN (1970) High frequency haploid production in barley (*Hordeum vulgare* L.). *Nature* 225 (5235):874-876. doi:10.1038/225874a0
- Katoh S, Tsunoda Y, Murata K, Minami E, Katoh E (2005) Active site residues and amino acid specificity of the ubiquitin carrier protein-binding RING-H2 finger domain. *J Biol Chem* 280 (49):41015-41024. doi:10.1074/jbc.M411127200

- Kelliher T, Starr D, Richbourg L, Chintamanani S, Delzer B, Nuccio ML, Green J, Chen Z, McCuiston J, Wang W, Liebler T, Bullock P, Martin B (2017) MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* 542 (7639):105-109. doi:10.1038/nature20827
- Kelliher T, Starr D, Su X, Tang G, Chen Z, Carter J, Wittich PE, Dong S, Green J, Burch E, McCuiston J, Gu W, Sun Y, Strebe T, Roberts J, Bate NJ, Que Q (2019) One-step genome editing of elite crop germplasm during haploid induction. *Nat Biotechnol* 37 (3):287-292. doi:10.1038/s41587-019-0038-x
- Kelliher T, Starr D, Wang W, McCuiston J, Zhong H, Nuccio ML, Martin B (2016) Maternal Haploids Are Preferentially Induced by CENH3-tailswap Transgenic Complementation in Maize. *Front Plant Sci* 7:414. doi:10.3389/fpls.2016.00414
- Koornneef M, Dellaert LW, van der Veen JH (1982) EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mutat Res* 93 (1):109-123. doi:10.1016/0027-5107(82)90129-4
- Kourelis J, Marchal C, Posbeyikian A, Harant A, Kamoun S (2023) NLR immune receptor-nanobody fusions confer plant disease resistance. *Science* 379 (6635):934-939. doi:10.1126/science.abn4116
- Kuppu S, Ron M, Marimuthu MPA, Li G, Huddleson A, Siddeek MH, Terry J, Buchner R, Shabek N, Comai L, Britt AB (2020) A variety of changes, including CRISPR/Cas9-mediated deletions, in CENH3 lead to haploid induction on outcrossing. *Plant Biotechnol J* 18 (10):2068-2080. doi:10.1111/pbi.13365
- Kuppu S, Tan EH, Nguyen H, Rodgers A, Comai L, Chan SW, Britt AB (2015) Point Mutations in Centromeric Histone Induce Post-zygotic Incompatibility and Uniparental Inheritance. *PLoS Genet* 11 (9):e1005494. doi:10.1371/journal.pgen.1005494
- Laurie DA, Bennett MD (1988) The production of haploid wheat plants from wheat x maize crosses. *Theor Appl Genet* 76 (3):393-397. doi:10.1007/BF00265339
- Le Goff S, Keceli BN, Jerabkova H, Heckmann S, Rutten T, Cotterell S, Schubert V, Roitinger E, Mechtler K, Franklin FCH, Tatout C, Houben A, Geelen D, Probst AV, Lermontova I (2020) The H3 histone chaperone NASP(SIM3) escorts CenH3 in *Arabidopsis*. *Plant J* 101 (1):71-86. doi:10.1111/tpj.14518
- Leljak Levanic D, Horvat T, Martincic J, Bauer N (2012) A novel bipartite nuclear localization signal guides BPM1 protein to nucleolus suggesting its Cullin3 independent function. *PLoS One* 7 (12):e51184. doi:10.1371/journal.pone.0051184
- Leon RG, Bassham DC (2024) PROTAC for agriculture: learning from human medicine to generate new biotechnological weed control solutions. *Pest Manag Sci* 80 (2):262-266. doi:10.1002/ps.7741
- Lermontova I, Koroleva O, Rutten T, Fuchs J, Schubert V, Moraes I, Koszegi D, Schubert I (2011a) Knockdown of CENH3 in *Arabidopsis* reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. *Plant J* 68 (1):40-50. doi:10.1111/j.1365-313X.2011.04664.x
- Lermontova I, Kuhlmann M, Friedel S, Rutten T, Heckmann S, Sandmann M, Demidov D, Schubert V, Schubert I (2013) *Arabidopsis* kinetochore null2 is an upstream component for centromeric histone H3 variant cenH3 deposition at centromeres. *Plant Cell* 25 (9):3389-3404. doi:10.1105/tpc.113.114736

- Lermontova I, Rutten T, Schubert I (2011b) Deposition, turnover, and release of CENH3 at Arabidopsis centromeres. *Chromosoma* 120 (6):633-640. doi:10.1007/s00412-011-0338-5
- Li X, Meng D, Chen S, Luo H, Zhang Q, Jin W, Yan J (2017) Single nucleus sequencing reveals spermatid chromosome fragmentation as a possible cause of maize haploid induction. *Nat Commun* 8 (1):991. doi:10.1038/s41467-017-00969-8
- Li Y, Li D, Xiao Q, Wang H, Wen J, Tu J, Shen J, Fu T, Yi B (2022) An in planta haploid induction system in Brassica napus. *J Integr Plant Biol* 64 (6):1140-1144. doi:10.1111/jipb.13270
- Li Y, Lin Z, Yue Y, Zhao H, Fei X, E L, Liu C, Chen S, Lai J, Song W (2021) Loss-of-function alleles of ZmPLD3 cause haploid induction in maize. *Nat Plants* 7 (12):1579-1588. doi:10.1038/s41477-021-01037-2
- Liu C, Li X, Meng D, Zhong Y, Chen C, Dong X, Xu X, Chen B, Li W, Li L, Tian X, Zhao H, Song W, Luo H, Zhang Q, Lai J, Jin W, Yan J, Chen S (2017) A 4-bp Insertion at ZmPLA1 Encoding a Putative Phospholipase A Generates Haploid Induction in Maize. *Mol Plant* 10 (3):520-522. doi:10.1016/j.molp.2017.01.011
- Liu C, Zhong Y, Qi X, Chen M, Liu Z, Chen C, Tian X, Li J, Jiao Y, Wang D, Wang Y, Li M, Xin M, Liu W, Jin W, Chen S (2020) Extension of the in vivo haploid induction system from diploid maize to hexaploid wheat. *Plant Biotechnol J* 18 (2):316-318. doi:10.1111/pbi.13218
- Liu XQ, Shi JJ, Fan H, Jiao J, Gao L, Tan L, Nagawa S, Wang DY (2021) Nuclear DNA replicates during zygote development in Arabidopsis and Torenia fournieri. *Plant Physiol* 185 (1):137-145. doi:10.1093/plphys/kiab014
- Liu Y, Jiao B, Champer J, Qian W (2024a) Overriding Mendelian inheritance in Arabidopsis with a CRISPR toxin-antidote gene drive that impairs pollen germination. *Nat Plants* 10 (6):910-922. doi:10.1038/s41477-024-01692-1
- Liu Z, Zhong Y, Qi X, An T, Guo S, Wang D, Wang Y, Feng B, Zhu Z, Chen S, Liu C (2024b) Haploids can be induced in knockout mutants of OsPLA1, but not OsDMP3 or OsDMP6, in rice. *The Crop Journal* 12 (1):213-221. doi:10.1016/j.cj.2023.11.005
- Long L, Feng YM, Shang SZ, Zhao JR, Hu GY, Xu FC, Song CP, Jin SX, Gao W (2024) In vivo maternal haploid induction system in cotton. *Plant Physiol* 194 (3):1286-1289. doi:10.1093/plphys/kiad620
- Lv J, Yu K, Wei J, Gui H, Liu C, Liang D, Wang Y, Zhou H, Carlin R, Rich R, Lu T, Que Q, Wang WC, Zhang X, Kelliher T (2020) Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. *Nat Biotechnol* 38 (12):1397-1401. doi:10.1038/s41587-020-0728-4
- Ma Y, Miotk A, Sutikovic Z, Ermakova O, Wenzl C, Medzihradsky A, Gaillochot C, Forner J, Utan G, Brackmann K, Galvan-Ampudia CS, Vernoux T, Greb T, Lohmann JU (2019) WUSCHEL acts as an auxin response rheostat to maintain apical stem cells in Arabidopsis. *Nat Commun* 10 (1):5093. doi:10.1038/s41467-019-13074-9
- Maheshwari S, Tan EH, West A, Franklin FC, Comai L, Chan SW (2015) Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genet* 11 (1):e1004970. doi:10.1371/journal.pgen.1004970

- Manape TK, Satheesh V, Somasundaram S, Soumia PS, Khade YP, Mainkar P, Mahajan V, Singh M, Anandhan S (2024) RNAi-mediated downregulation of AcCENH3 can induce in vivo haploids in onion (*Allium cepa* L.). *Sci Rep* 14 (1):14481. doi:10.1038/s41598-024-64432-7
- Mao Y, Nakel T, Erbasol Serbes I, Joshi S, Tekleyohans DG, Baum T, Gross-Hardt R (2023) ECS1 and ECS2 suppress polyspermy and the formation of haploid plants by promoting double fertilization. *Elife* 12:e85832. doi:10.7554/eLife.85832
- Marimuthu MPA, Maruthachalam R, Bondada R, Kuppu S, Tan EH, Britt A, Chan SWL, Comai L (2021) Epigenetically mismatched parental centromeres trigger genome elimination in hybrids. *Sci Adv* 7 (47):eabk1151. doi:10.1126/sciadv.abk1151
- McCafferty J, Griffiths AD, Winter G, Chiswell DJ (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* 348 (6301):552-554. doi:10.1038/348552a0
- Mendiburo MJ, Padeken J, Fulop S, Schepers A, Heun P (2011) *Drosophila* CENH3 is sufficient for centromere formation. *Science* 334 (6056):686-690. doi:10.1126/science.1206880
- Merai Z, Chumak N, Garcia-Aguilar M, Hsieh TF, Nishimura T, Schoft VK, Bindics J, Slusarz L, Arnoux S, Opravil S, Mechtler K, Zilberman D, Fischer RL, Tamaru H (2014) The AAA-ATPase molecular chaperone Cdc48/p97 disassembles sumoylated centromeres, decondenses heterochromatin, and activates ribosomal RNA genes. *Proc Natl Acad Sci U S A* 111 (45):16166-16171. doi:10.1073/pnas.1418564111
- Meyer CM, Goldman IL, Krysan PJ (2023) Chromosome-level changes and genome elimination by manipulation of CENH3 in carrot (*Daucus carota*). *Front Plant Sci* 14:1294551. doi:10.3389/fpls.2023.1294551
- Murashige T, Skoog F (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* 15 (3):473-497. doi:10.1111/j.1399-3054.1962.tb08052.x
- Muyldermans S (2001) Single domain camel antibodies: current status. *J Biotechnol* 74 (4):277-302. doi:10.1016/s1389-0352(01)00021-6
- Muyldermans S (2021) A guide to: generation and design of nanobodies. *FEBS J* 288 (7):2084-2102. doi:10.1111/febs.15515
- Natsume T, Kanemaki MT (2017) Conditional Degrons for Controlling Protein Expression at the Protein Level. *Annu Rev Genet* 51 (1):83-102. doi:10.1146/annurev-genet-120116-024656
- Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M (2009) An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* 6 (12):917-922. doi:10.1038/nmeth.1401
- Oberhofer G, Johnson ML, Ivy T, Antoshechkin I, Hay BA (2024) Cleave and Rescue gamete killers create conditions for gene drive in plants. *Nat Plants* 10 (6):936-953. doi:10.1038/s41477-024-01701-3
- Oppenheimer DG, Herman PL, Sivakumaran S, Esch J, Marks MD (1991) A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell* 67 (3):483-493. doi:10.1016/0092-8674(91)90523-2

- Park SK, Howden R, Twell D (1998) The *Arabidopsis thaliana* gametophytic mutation *geminipollen1* disrupts microspore polarity, division asymmetry and pollen cell fate. *Development* 125 (19):3789-3799. doi:10.1242/dev.125.19.3789
- Paul W, Hodge R, Smartt S, Draper J, Scott R (1992) The isolation and characterisation of the tapetum-specific *Arabidopsis thaliana* A9 gene. *Plant Mol Biol* 19 (4):611-622. doi:10.1007/BF00026787
- Pickering RA (1984) The influence of genotype and environment on chromosome elimination in crosses between *Hordeum vulgare* L. × *Hordeum bulbosum* L. *Plant Science Letters* 34 (1-2):153-164. doi:10.1016/0304-4211(84)90138-x
- Ravi M, Chan SW (2010) Haploid plants produced by centromere-mediated genome elimination. *Nature* 464 (7288):615-618. doi:10.1038/nature08842
- Ravi M, Kwong PN, Menorca RM, Valencia JT, Ramahi JS, Stewart JL, Tran RK, Sundaresan V, Comai L, Chan SW (2010) The rapidly evolving centromere-specific histone has stringent functional requirements in *Arabidopsis thaliana*. *Genetics* 186 (2):461-471. doi:10.1534/genetics.110.120337
- Ravi M, Marimuthu MP, Tan EH, Maheshwari S, Henry IM, Marin-Rodriguez B, Urtecho G, Tan J, Thornhill K, Zhu F, Panoli A, Sundaresan V, Britt AB, Comai L, Chan SW (2014) A haploid genetics toolbox for *Arabidopsis thaliana*. *Nat Commun* 5 (1):5334. doi:10.1038/ncomms6334
- Ravi M, Shibata F, Ramahi JS, Nagaki K, Chen C, Murata M, Chan SW (2011) Meiosis-specific loading of the centromere-specific histone CENH3 in *Arabidopsis thaliana*. *PLoS Genet* 7 (6):e1002121. doi:10.1371/journal.pgen.1002121
- Raychaudhuri N, Dubrulle R, Orsi GA, Bagheri HC, Loppin B, Lehner CF (2012) Transgenerational propagation and quantitative maintenance of paternal centromeres depends on Cid/Cenp-A presence in *Drosophila* sperm. *PLoS Biol* 10 (12):e1001434. doi:10.1371/journal.pbio.1001434
- Roth S, Fulcher LJ, Sapkota GP (2019) Advances in targeted degradation of endogenous proteins. *Cell Mol Life Sci* 76 (14):2761-2777. doi:10.1007/s00018-019-03112-6
- Rothbauer U, Zolghadr K, Tillib S, Nowak D, Schermelleh L, Gahl A, Backmann N, Conrath K, Muyldermans S, Cardoso MC, Leonhardt H (2006) Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat Methods* 3 (11):887-889. doi:10.1038/nmeth953
- Ruban A, Houben A (2022) Highly reactive chemicals meet haploidization. *Mol Plant* 15 (6):937-939. doi:10.1016/j.molp.2022.05.010
- Sanei M, Pickering R, Kumke K, Nasuda S, Houben A (2011) Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. *Proc Natl Acad Sci U S A* 108 (33):E498-505. doi:10.1073/pnas.1103190108
- Schoft VK, Chumak N, Mosiolek M, Slusarz L, Komnenovic V, Brownfield L, Twell D, Kakutani T, Tamaru H (2009) Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. *EMBO Rep* 10 (9):1015-1021. doi:10.1038/embor.2009.152
- Schreiber T, Prange A, Schafer P, Iwen T, Grutzner R, Marillonnet S, Lepage A, Javelle M, Paul W, Tissier A (2024) Efficient scar-free knock-ins of several kilobases in plants by engineered CRISPR-Cas endonucleases. *Mol Plant* 17 (5):824-837. doi:10.1016/j.molp.2024.03.013

- She W, Baroux C, Grossniklaus U (2018) Cell-Type Specific Chromatin Analysis in Whole-Mount Plant Tissues by Immunostaining. *Methods Mol Biol* 1675:443-454. doi:10.1007/978-1-4939-7318-7\_25
- Shen K, Qu M, Zhao P (2023) The Roads to Haploid Embryogenesis. *Plants (Basel)* 12 (2):243. doi:10.3390/plants12020243
- 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 (3):519-528. doi:10.1111/j.1365-313X.2009.04060.x
- Shin YJ, Park SK, Jung YJ, Kim YN, Kim KS, Park OK, Kwon SH, Jeon SH, Trinh le A, Fraser SE, Kee Y, Hwang BJ (2015) Nanobody-targeted E3-ubiquitin ligase complex degrades nuclear proteins. *Sci Rep* 5 (1):14269. doi:10.1038/srep14269
- Simonini S, Bemer M, Bencivenga S, Gagliardini V, Pires ND, Desvoyes B, van der Graaff E, Gutierrez C, Grossniklaus U (2021) The Polycomb group protein MEDEA controls cell proliferation and embryonic patterning in *Arabidopsis*. *Dev Cell* 56 (13):1945-1960 e1947. doi:10.1016/j.devcel.2021.06.004
- Simonini S, Bencivenga S, Grossniklaus U (2024) A paternal signal induces endosperm proliferation upon fertilization in *Arabidopsis*. *Science* 383 (6683):646-653. doi:10.1126/science.adj4996
- Smith GP (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228 (4705):1315-1317. doi:10.1126/science.4001944
- Song Q, Ando A, Jiang N, Ikeda Y, Chen ZJ (2020) Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in *Arabidopsis* female gametophytes. *Genome Biol* 21 (1):178. doi:10.1186/s13059-020-02094-0
- Sorge E, Demidov D, Lermontova I, Houben A, Conrad U (2021) Engineered degradation of EYFP-tagged CENH3 via the 26S proteasome pathway in plants. *PLoS One* 16 (2):e0247015. doi:10.1371/journal.pone.0247015
- Sorge E (2022) Gezielter Proteinabbau in Pflanzen mittels chimärer E3-Ubiquitin-Ligasen (PhD Thesis). [opendata.uni-halle.de](https://opendata.uni-halle.de). doi: 10.25673/91833
- Sprunck S, Rademacher S, Vogler F, Gheyselinck J, Grossniklaus U, Dresselhaus T (2012) Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science* 338 (6110):1093-1097. doi:10.1126/science.1223944
- Tachiwana H, Kagawa W, Kurumizaka H (2012) Comparison between the CENP-A and histone H3 structures in nucleosomes. *Nucleus* 3 (1):6-11. doi:10.4161/nucl.18372
- Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell* 14 (5):1053-1066. doi:10.1105/tpc.010425
- Tang H, Qiu Y, Wang W, Yu M, Chang Y, Han Z, Du L, Lin Z, Wang K, Ye X (2023) Development of a haploid inducer by editing HvMTL in barley. *J Genet Genomics* 50 (5):366-369. doi:10.1016/j.jgg.2022.11.007
- Trentin HU, Krause MD, Zunjare RU, Almeida VC, Peterlini E, Rotareno V, Frei UK, Beavis WD, Lubberstedt T (2023) Genetic basis of maize maternal haploid induction beyond MATRILINEAL and ZmDMP. *Front Plant Sci* 14:1218042. doi:10.3389/fpls.2023.1218042



- Vasimuddin M, Misra S, Li H, Aluru S (2019) Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems | IEEE Conference Publication | IEEE Xplore. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). doi:10.1109/IPDPS.2019.00041
- Voullaire LE, Slater HR, Petrovic V, Choo KH (1993) A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *American Journal of Human Genetics* 52 (6):1153-1163
- Wang D, Zhong Y, Feng B, Qi X, Yan T, Liu J, Guo S, Wang Y, Liu Z, Cheng D, Zhang Y, Shi Y, Zhang S, Pan R, Liu C, Chen S (2023a) The RUBY reporter enables efficient haploid identification in maize and tomato. *Plant Biotechnol J* 21 (8):1707-1715. doi:10.1111/pbi.14071
- Wang J, Wang X-F, Yang W-C, Li H-J (2023b) Loss of function of CENH3 causes genome instability in soybean. *Seed Biology* 2 (1):0-0. doi:10.48130/SeedBio-2023-0024
- Wang N, Dawe RK (2018) Centromere Size and Its Relationship to Haploid Formation in Plants. *Mol Plant* 11 (3):398-406. doi:10.1016/j.molp.2017.12.009
- Wang N, Gent JI, Dawe RK (2021) Haploid induction by a maize cenH3 null mutant. *Sci Adv* 7 (4):eabe2299. doi:10.1126/sciadv.abe2299
- Wang N, Xia X, Jiang T, Li L, Zhang P, Niu L, Cheng H, Wang K, Lin H (2022a) In planta haploid induction by genome editing of DMP in the model legume *Medicago truncatula*. *Plant Biotechnol J* 20 (1):22-24. doi:10.1111/pbi.13740
- Wang S, Ouyang K (2023) Rapid creation of CENH3-mediated haploid induction lines using a cytosine base editor (CBE). *Plant Biol (Stuttg)* 25 (1):226-230. doi:10.1111/plb.13482
- Wang W, Xiong H, Zhao P, Peng X, Sun MX (2022b) DMP8 and 9 regulate HAP2/GCS1 trafficking for the timely acquisition of sperm fusion competence. *Proc Natl Acad Sci USA* 119 (45):e2207608119. doi:10.1073/pnas.2207608119
- Wang Z, Chen M, Yang H, Hu Z, Yu Y, Xu H, Yan S, Yi K, Li J (2023c) A simple and highly efficient strategy to induce both paternal and maternal haploids through temperature manipulation. *Nat Plants* 9 (5):699-705. doi:10.1038/s41477-023-01389-x
- Weber H, Bernhardt A, Dieterle M, Hano P, Mutlu A, Estelle M, Genschik P, Hellmann H (2005) Arabidopsis AtCUL3a and AtCUL3b form complexes with members of the BTB/POZ-MATH protein family. *Plant Physiol* 137 (1):83-93. doi:10.1104/pp.104.052654
- Widiez T (2021) Haploid Embryos: Being Like Mommy or Like Daddy? *Trends Plant Sci* 26 (5):425-427. doi:10.1016/j.tplants.2021.02.010
- Wijnker E, van Dun K, de Snoo CB, Lelivelt CL, Keurentjes JJ, Naharudin NS, Ravi M, Chan SW, de Jong H, Dirks R (2012) Reverse breeding in *Arabidopsis thaliana* generates homozygous parental lines from a heterozygous plant. *Nat Genet* 44 (4):467-470. doi:10.1038/ng.2203
- Wu P, Ren J, Li L, Chen S (2014) Early spontaneous diploidization of maternal maize haploids generated by in vivo haploid induction. *Euphytica* 200 (1):127-138. doi:10.1007/s10681-014-1166-5
- Xu J, Yang C, Yuan Z, Zhang D, Gondwe MY, Ding Z, Liang W, Zhang D, Wilson ZA (2010) The ABORTED MICROSPORES regulatory network is required for

- postmeiotic male reproductive development in *Arabidopsis thaliana*. *Plant Cell* 22 (1):91-107. doi:10.1105/tpc.109.071803
- Yao L, Zhang Y, Liu C, Liu Y, Wang Y, Liang D, Liu J, Sahoo G, Kelliher T (2018) OsMATL mutation induces haploid seed formation in indica rice. *Nat Plants* 4 (8):530-533. doi:10.1038/s41477-018-0193-y
- Yoshida H, Nagata M, Saito K, Wang KL, Ecker JR (2005) *Arabidopsis* ETO1 specifically interacts with and negatively regulates type 2 1-aminocyclopropane-1-carboxylate synthases. *BMC Plant Biol* 5 (1):14. doi:10.1186/1471-2229-5-14
- Zhang J, Yin J, Luo J, Tang D, Zhu X, Wang J, Liu Z, Wang P, Zhong Y, Liu C, Li C, Chen S, Huang S (2022a) Construction of homozygous diploid potato through maternal haploid induction. *aBIOTECH* 3 (3):163-168. doi:10.1007/s42994-022-00080-7
- Zhang X, Shi C, Li S, Zhang B, Luo P, Peng X, Zhao P, Dresselhaus T, Sun MX (2023a) A female in vivo haploid-induction system via mutagenesis of egg cell-specific peptidases. *Mol Plant* 16 (2):471-480. doi:10.1016/j.molp.2023.01.001
- Zhang X, Zhang L, Zhang J, Jia M, Cao L, Yu J, Zhao D (2022b) Haploid induction in allotetraploid tobacco using DMPs mutation. *Planta* 255 (5):98. doi:10.1007/s00425-022-03877-4
- Zhang Y, Maruyama D, Toda E, Kinoshita A, Okamoto T, Mitsuda N, Takasaki H, Ohme-Takagi M (2023b) Transcriptome analyses uncover reliance of endosperm gene expression on *Arabidopsis* embryonic development. *FEBS Lett* 597 (3):407-417. doi:10.1002/1873-3468.14570
- Zhao P, Zhou X, Shen K, Liu Z, Cheng T, Liu D, Cheng Y, Peng X, Sun MX (2019) Two-Step Maternal-to-Zygotic Transition with Two-Phase Parental Genome Contributions. *Dev Cell* 49 (6):882-893 e885. doi:10.1016/j.devcel.2019.04.016
- Zhao S, Huang L, Zhang Q, Zhou Y, Yang M, Shi H, Li Y, Yang J, Li C, Ge X, Gong W, Wang J, Zou Q, Tao L, Kang Z, Li Z, Xiao C, Hu Q, Fu S (2023) Paternal chromosome elimination of inducer triggers induction of double haploids in *Brassica napus*. *Front Plant Sci* 14:1256338. doi:10.3389/fpls.2023.1256338
- Zhao X, Yuan K, Liu Y, Zhang N, Yang L, Zhang Y, Wang Y, Ji J, Fang Z, Han F, Lv H (2022) In vivo maternal haploid induction based on genome editing of DMP in *Brassica oleracea*. *Plant Biotechnol J* 20 (12):2242-2244. doi:10.1111/pbi.13934
- Zhong Y, Chen B, Li M, Wang D, Jiao Y, Qi X, Wang M, Liu Z, Chen C, Wang Y, Chen M, Li J, Xiao Z, Cheng D, Liu W, Boutilier K, Liu C, Chen S (2020) A DMP-triggered in vivo maternal haploid induction system in the dicotyledonous *Arabidopsis*. *Nat Plants* 6 (5):466-472. doi:10.1038/s41477-020-0658-7
- Zhong Y, Chen B, Wang D, Zhu X, Li M, Zhang J, Chen M, Wang M, Riksen T, Liu J, Qi X, Wang Y, Cheng D, Liu Z, Li J, Chen C, Jiao Y, Liu W, Huang S, Liu C, Boutilier K, Chen S (2022a) In vivo maternal haploid induction in tomato. *Plant Biotechnol J* 20 (2):250-252. doi:10.1111/pbi.13755
- Zhong Y, Liu C, Qi X, Jiao Y, Wang D, Wang Y, Liu Z, Chen C, Chen B, Tian X, Li J, Chen M, Dong X, Xu X, Li L, Li W, Liu W, Jin W, Lai J, Chen S (2019) Mutation of ZmDMP enhances haploid induction in maize. *Nat Plants* 5 (6):575-580. doi:10.1038/s41477-019-0443-7
- Zhong Y, Wang Y, Chen B, Liu J, Wang D, Li M, Qi X, Liu C, Boutilier K, Chen S (2022b) Establishment of a dmp based maternal haploid induction system for

polyploid *Brassica napus* and *Nicotiana tabacum*. *J Integr Plant Biol* 64 (6):1281-1294. doi:10.1111/jipb.13244

Zong Y, Liu Y, Xue C, Li B, Li X, Wang Y, Li J, Liu G, Huang X, Cao X, Gao C (2022) An engineered prime editor with enhanced editing efficiency in plants. *Nat Biotechnol* 40 (9):1394-1402. doi:10.1038/s41587-022-01254-w

## 10. Curriculum vitae

**Saravanakumar Somasundaram**

Orcid: 0000-0002-8168-0452

### Education

- PhD**                    **Plant molecular biology and biotechnology**  
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK),  
Germany  
(August, 2020 to January 2025)  
**Thesis title:** Gametophyte-specific degradation of Centromeric  
Histone3 (CENH3) to investigate the mechanism of uniparental  
genome elimination in *Arabidopsis thaliana*.  
Advisor: Prof. Dr. Andreas Houben
- M.Sc.**                    **Genetics and Plant breeding**  
Indian Agricultural Research Institute (IARI), New Delhi, India  
(July, 2015 to July, 2017)  
**Thesis title:** Molecular characterization and evaluation of medium  
slender grain rice genotypes pyramided with bacterial blight and  
blast resistance genes.  
Advisor: Dr. Gopala Krishnan  
Final Grade: 8.86/10
- B.Sc.**                    **Agriculture**  
Tamil Nadu Agricultural University, Coimbatore, India.  
(July, 2011 to May, 2015)  
Final Grade: 9.15/10

### Work Experience

- Guest Researcher (IPK, Gatersleben)                    January 2020 to August 2020  
Project Title: CRISPR imaging to simultaneously visualize DNA repeats and  
histone modifications in onion.
- Senior Research fellow (IISER, India)                    August 2019 to January 2020  
Project Title: Targeted editing of potato genome to develop variety specific True  
Potato Seeds
- Senior Research fellow (DOGR, India)                    September 2017 to August 2019  
Project Title: Haploid induction in onion through genome elimination

## Academic Achievements

- DAAD-Research Grants for doctoral program in Germany (2020-2024)
- Jeff Schell Fellowships from Bayer foundation, Germany (2019-2020)
- ICAR fellowship for master studies, India (2015-2017)

## Publications

Manape TK, Satheesh V, **Somasundaram S**, Soumia PS, Khade YP, Mainkar P, Mahajan V, Singh M, Anandhan S (2024) RNAi-mediated downregulation of AcCENH3 can induce in vivo haploids in onion (*Allium cepa* L.). *Sci Rep* 14 (1):14481. doi:10.1038/s41598-024-64432-7.

Potlapalli BP, Ishii T, Nagaki K, **Somasundaram S**, Houben A (2023) CRISPR-FISH: A CRISPR/Cas9-Based In Situ Labeling Method. In: *Plant Cytogenetics and Cyto genomics: Methods and Protocols*. Springer, pp 315-335.

Demidov D, Lermontova I, Moebes M, Kochevenko A, Fuchs J, Weiss O, Rutten T, Sorge E, Zuljan E, Giehl RFH, Mascher M, **Somasundaram S**, Conrad U, Houben A (2022) Haploid induction by nanobody-targeted ubiquitin-proteasome-based degradation of EYFP-tagged CENH3 in *Arabidopsis thaliana*. *J Exp Bot* 73 (22):7243-7254. doi:10.1093/jxb/erac359.

**Somasundaram S**, Satheesh V, Singh M, Anandhan S (2021) A simple flow cytometry-based assay to study global methylation levels in onion, a non-model species. *Physiol Mol Biol Plants* 27 (8):1859-1865. doi:10.1007/s12298-021-01047-6.

Bondada R, **Somasundaram S**, Marimuthu MP, Badarudeen MA, Puthiyaveedu VK, Maruthachalam R (2020) Natural epialleles of *Arabidopsis* SUPERMAN display superwoman phenotypes. *Commun Biol* 3 (1):772. doi:10.1038/s42003-020-01525-9.

## Conference contributions

- Flash talk and Poster presentation, 27th International conference on sexual plant reproduction, Providence, USA (2024)
- Poster presentation, 26th International conference on sexual plant reproduction, Prague, Czech Republic (2022)
- Oral presentation, International Symposium on Edible Alliums: Challenges and Opportunities, Pune, India (2019)

## 11. Eidesstattliche Erklärung / Declaration under Oath

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

---

Datum / Date

---

Unterschrift des Antragstellers / Signature of the applicant

## 12. Appendix

**Appendix 1:** List of primers, templates and destination plasmid used for the construction of level-0 plasmids and entry vectors. '-' refers to no primers as synthetic fragments were used to build these modules

Construct	Destination	Primer-F	Primer-R	Template
pNOS-PPT	pICH41331	NOS_PROM-F	NOS_TER-R	pGWB605
pNOS-NPT	pICH41331	NOS_PROM-F	NOS_TER-R	pGWB401
pCAMV35S-pro	pICH41295	35S-gg-F	35S-gg-R	pICH86988
pDAZ3-pro	pICH41295	DAZ3-Prom-F	DAZ3-Prom-R	Arabidopsis gDNA
pEC1.1-pro	pICH41295	EC1.1-GG-F	EC1.1-GG-R	Arabidopsis gDNA
pA9-pro	pICH41295	A9-Prom-F	A9-Prom-R	Arabidopsis gDNA
pHSP-ter	pAGM9121	HSP-TER-F	HSP-TER-R	Arabidopsis gDNA
pRbcSE9-ter	pAGM9121	rbcS-TER-F	rbcS-TER-R	<i>Pisum sativum</i> gDNA
pBarnase	pAGM1287	-	-	Synthetic fragment
pH2B	pICH41258	H2B-F1	H2B-R1	Arabidopsis cDNA
pMCherry	pAGM1299	-	-	Synthetic fragment
pVHHGFP4	pAGM1287	-	-	Synthetic fragment
pVHHGFP4-BTB	pAGM1287	-	-	Synthetic fragment
pNslmb-VHHGFP4	pAGM1287	-	-	Synthetic fragment
pNbALFA-N	pICH41258	-	-	Synthetic fragment
pNbALFA-C	pAGM1299	-	-	Synthetic fragment
pNbALFA	pAGM1287	-	-	Synthetic fragment
pNslmb	pICH41258	Nslmb-F1	41258-NSLMB-R	pNslmb-VHHGFP4
pBTB	pAGM1299	BTB-F	MYC-R	pVHHGFP4-BTB
pEYFP	pICH41258	-	-	Synthetic fragment
pCENH3-cds	pAGM9121	CENH3-Ntail-F1	cenh3-gg-r	Arabidopsis cDNA
pEL5	pAGM1299	-	-	Synthetic fragment
pBPM1	pAGM1299	-	-	Synthetic fragment
PCR1	PCR fragments	CENH3-Bsa1-F1	CENH3-Bsa1-R1	Arabidopsis gDNA
PCR2	PCR fragments	CENH3-Bsa1-F2	CENH3-Bsa1-R2	Arabidopsis gDNA
PCR3	PCR fragments	CENH3-Bsa1-F1	CENH3-Bsa1-R2	EYFP-gCENH3
pVHHGFP4-N	pICH41258	VHHGFP-F1	VHHGFP4-6xGly-R	pVHHGFP4
pVHHGFP4-C	pAGM1299	gly-vhhgfp4-f	gly-vhhgfp4-R	pVHHGFP4
pSLY1	pICH41258	-	-	Synthetic fragment
pETO1	pICH41258	-	-	Synthetic fragment
p221EC1.1	pDONR221 p1-p4	-	-	Construct from UC Davis
p221H2B-tdTomato	pDONR221 p4r-p3r	-	-	Construct from UC Davis
p221RbcSE9 Ter	pDONR221 p3-p2	-	-	Construct from UC Davis

**Appendix 2:** List of level-0 modules and destination plasmid used for the construction of level-1 plasmids. '-' refers to less than four level-0 plasmids (either 3, 2 or 1) were used for the construction of corresponding level-1 plasmid.

Construct	Destination	Level-0 modules			
		1	2	3	4
pPPT	pICH47802	pNOS-PPT	-	-	-
pNPT	pICH47802	pNOS-NPT	-	-	-
pCVE	pICH47732	pCAMV35S-pro	pVHHGFP4-N	pEL5	pRbcSE9-ter
pCVB	pICH47732	pCAMV35S-pro	pVHHGFP4-N	pBPM1	pRbcSE9-ter
pCSV	pICH47732	pCAMV35S-pro	pSLY1	pVHHGFP4-C	pRbcSE9-ter
pCEV	pICH47732	pCAMV35S-pro	pETO1	pVHHGFP4-C	pRbcSE9-ter
pCVS	pICH47732	pCAMV35S-pro	pVHHGFP4-N	pBTB	pRbcSE9-ter
pCV	pICH47732	pCAMV35S-pro	pVHHGFP4	pRbcSE9-ter	-
pDVS	pICH47811	pDAZ3-pro	pVHHGFP4-BTB	pRbcSE9-ter	-
pEVS	pICH47811	pEC1.1-pro	pVHHGFP4-BTB	pRbcSE9-ter	-
pENV	pICH47811	pEC1.1-pro	pNslmb-VHHGFP4	pRbcSE9-ter	-
pEV	pICH47811	pEC1.1-pro	pVHHGFP4	pRbcSE9-ter	-
pENS	pICH47811	pEC1.1-pro	pNbALFA-N	pBTB	pRbcSE9-ter
pENN	pICH47811	pEC1.1-pro	pNslmb	pNbALFA-C	pRbcSE9-ter
pEN	pICH47811	pEC1.1-pro	pNbALFA	pRbcSE9-ter	
pENE	pICH47811	pEC1.1-pro	pNbALFA-N	pEL5	pRbcSE9-ter
pENB	pICH47811	pEC1.1-pro	pNbALFA-N	pBPM1	pRbcSE9-ter
pALFA-gCENH3	pICH47742	PCR1	PCR2		
pEYFP-gCENH3	pICH47742	PCR3	-	-	-
pEYFP-CENH3	pICH47742	pCAMV35S-pro	pEYFP	pCENH3-cds	pRbcSE9-ter
pENS2	pICH47751	pEC1.1-pro	pNbALFA-N	pBTB	pRbcSE9-ter
pEVS2	pICH47751	pEC1.1-pro	pVHHGFP4-BTB	pRbcSE9-ter	
pFAST	pICH47822	pICSL70008	-	-	-
pH2B-mCherry	pICH47751	pCAMV35S-pro	pH2B	pMCherry	pHSP-ter
pBarnase	pICH47831	pA9-pro	pBarnase	pHSP-ter	



**Appendix 3:** List of level-1 modules and destination plasmid used for the construction of level-2 plasmids. '-' refers to empty (dummy) plasmid or an unoccupied position at the end.

Construct	Destination	Position 1	Position 2	Position 3	Position 4
EVS	pICSL4723	pPPT	pEVS	-	-
ENV	pICSL4723	pPPT	pENV	-	-
EV	pICSL4723	pPPT	pEV	-	-
ALFA-gCENH3 + ENS	pICSL4723	pNPT	pALFA-gCENH3	pENS2	-
ALFA-gCENH3	pICH86966	pNPT	pALFA-gCENH3	-	-
EYFP-gCENH3 + EVS	pICSL4723	pNPT	pEYFP-gCENH3	pEVS2	-
EYFP-gCENH3	pICSL4723	pNPT	pEYFP-gCENH3	-	-
ENS	pICSL4723	pPPT	pENS	pFAST	pBarnase
ENN	pICSL4723	pPPT	pENN	pFAST	pBarnase
EN	pICSL4723	pPPT	pEN	pFAST	pBarnase
Control (only Barnase)	pICSL4723	pPPT	-	pFAST	pBarnase
DVS	pICSL4723	pPPT	pDVS	pFAST	-
CVE	pICSL4723	pCVE	pEYFP-CENH3	pH2B-mCherry	-
CVB	pICSL4723	pCVB	pEYFP-CENH3	pH2B-mCherry	-
CSV	pICSL4723	pCSV	pEYFP-CENH3	pH2B-mCherry	-
CEV	pICSL4723	pCEV	pEYFP-CENH3	pH2B-mCherry	-
CVS	pICSL4723	pCVS	pEYFP-CENH3	pH2B-mCherry	-
CV	pICSL4723	pCV	pEYFP-CENH3	pH2B-mCherry	-
Control (no E3-ligase)	pICSL4723	-	pEYFP-CENH3	pH2B-mCherry	-
ENE	pICSL4723	pPPT	pENE	pFAST	pBarnase
ENB	pICSL4723	pPPT	pENB	pFAST	pBarnase

**Appendix 4:** List of PCR primers used for the genotyping of Arabidopsis stable-transformants bearing different constructs

Transgene/mutant	Forward primer	Reverse primer
EVS	EC1.1-GG-F	BTB-R2
ENV	EC1.1-GG-F	NSLMB-R
EV	EC1.1-GG-F	VHGF-R1
ENS	EC1.1-GG-F	BTB-R2
ENN	EC1.1-GG-F	NSLMB-R
EN	EC1.1-GG-F	VHGF-R1

ENE	EC1.1-GG-F	EL5-R
ENB	EC1.1-GG-F	BPM1-R1
Control (only Barnase)	A9-Prom-F	HSP-TER-R
DVS	DAZ3-Prom-F	BTB-R2
ALFA-gCENH3 + ENS	EC1.1-GG-F	BTB-R2
	Alfa-F	CenH3-exon2-R
ALFA-gCENH3	Alfa-F	CenH3-exon2-R
EYFP-gCENH3 + EVS	EC1.1-GG-F	BTB-R2
	EYFP-F	CenH3-exon2-R
EYFP-gCENH3	EYFP-F	CenH3-exon2-R
<i>cenh3-1</i> EcoRV dCAPS for EYFP-gCENH3 lines	CenH3-dCAPS-F	CenH3-dCAPS-R
<i>cenh3-1</i> EcoRV dCAPS for ALFA-gCENH3 lines	AtCENH3-Geno-F2	CenH3-dCAPS-R

**Appendix 5:** List of all PCR primers used in this study and their corresponding nucleotide sequence.

Primer name	Sequence
NOS_PROM-F	TTGAAGACAAGGAGGCGGAGAATTAAGGGAGTCACG
NOS_TER-R	TTGAAGACAAAGCGAATTCCCGATCTAGTAACATAGATGACACCG
35S-gg-F	TTGAAGACAAGGAGGAATTCCAATCCCACAAAAATC
35S-gg-R	TTGAAGACAACATTGTATCGATAATTGTAAATGTAATTGTAATGTTGTTTG
rbcS-TER-F	TTGAAGACATCTCATTTCGATGTAATTATGGCATTGGGAAAACCTG
rbcS-TER-R	TTGAAGACAACCTCGAGCGTGTCTTACTCCTCATATTAACCTTCGGTC
HSP-TER-F	TTGAAGACATCTCATTTCGATATGAAGATGAAGATGAAATATTTG
HSP-TER-R	TTGAAGACAACCTCGAGCGCTTATCTTTAATCATATTCCATAGTC
EC1.1-GG-F	TTGAAGACAAGGAGAACGCCTATCATGAATTAGCTCTACTAAATCTAG
EC1.1-GG-R	TTGAAGACAACATTTTCTCAACAGATTGATAAGGTGCGAAAGAAGAAAGAG
DAZ3-Prom-F	TTGAAGACAAGGAGGGAAGTTGAGATGAGCATTGATG
DAZ3-Prom-R	TTGAAGACAACATTAATTTCTTCTTTGCAGAGATTTAG
A9-Prom-F	TTGAAGACAAGGAGCACCTACACTTGTGTTTCCTC
A9-Prom-R	TTGAAGACAACATTCTAATTAGATACTATATTGTTTGTACTTCTG
H2B-F1	TTGAAGACAAAATGGCACCAAGAGCCGAGAAGAAGCCCGCCGAGAAAAA ACCGCCGCTGAGAGGCCGCTG
H2B-R1	TTGAAGACAACCTCCACCACTTCCACCACCTCCAG

Nslmb-F1	TTGAAGACAAAATGATGAAAATGGAGACTGACAAAATAATG
41258-NSLMB-R	TTGAAGACAAACCTGACTCGAGGTGGCGGCCAG
BTB-F	TTGAAGACAAAGTTCTGTCAACATTTCTGGCCAG
MYC-R	TTGAAGACAACGAACCATTTCAGATCCTCTTCTGAGATG
CENH3-Ntail-F1	TTGAAGACATCTCAAGGTATGGCGAGAACCAAGCATCG
cenh3-gg-r	TTGAAGACAACCTCGCGAATCACCATGGTCTGCCTTTTC
CENH3-Bsa1-F1	TTGGTCTCATCCCGTACGGACGCATCATCAACATCTG
CENH3-Bsa1-F2	TTGGTCTCAGGAACCTCAGACGTAGGCTCACGGAACCTGCGAGAACCAAGC ATCGCG
CENH3-Bsa1-R1	TTGGTCTCATTCTCTTCCAGCCTGGACGGCATTGATTTTTACTGCTGGAG AAATC
CENH3-Bsa1-R2	TTGGTCTCACAGATTTTCGTTTAAGGACCAATGAG
gly-vhhgfp4-f	TTGAAGACAAAGGTGGAAGTGGTGGAAATGGATCAAGTCCAACCTGGTG
gly-vhhgfp4-R	TTGAAGACAACGAACTAGCTGGAGACGGTGACCTG
VHHGFP-F1	TTGAAGACAAAATGGATCAAGTCCAACCTGGTG
VHHGFP4- 6xGly-R	TTGAAGACAAACCTCCACCACTTCCACCGCTGGAGACGGTGACCTG
CenH3-dCAPS- F	GGTGCGATTTCTCCAGCAGTAAAAATC
CenH3-dCAPS- R	CTGAGAAGATGAAGCACCGGCGATAT
BTB-R2	GGTTTCCAAGACATCCGAAGCATGATAG
EYFP-F	AATGGTGAGCAAGGGCGAG
CenH3-exon2-R	CGTAGTTGGACCTGCCGCCTG
Alfa-F	TCCAGGCTGGAAGAGGAACTCAG
AtCENH3-Geno- F2	CTCCAGCAGTAAAAATCAATGGATAGA
NSLMB-R	ACCTGACTCGAGGTGGCGGCCAG
VHHGFP-R1	TCCTCATAGGAGCTACGATCACCCG
EL5-R	CGAACTACACGACGACGGTGAGGC
BPM1-R1	CGCCTAAAGGGCCAAAAAGCTGTG