Fall 12-16-2022

Genetic Characterization of Centromere-Mediated Uniparental Genome Elimination in Arabidopsis

Diana Spencer
University of Maine, diana.spencer@maine.edu

GENETIC CHARACTERIZATION OF CENTROMERE-MEDIATED UNIPARENTAL GENOME ELIMINATION IN *ARABIDOPSIS*

By

Diana Spencer

B.Eng. Universiti Malaysia Sarawak, Malaysia, 2004

M.Env.Sc. Universiti Malaysia Sarawak, Malaysia, 2019

A THESIS

Submitted in the Partial Fulfilment of the

Requirements for the Degree of

Master of Science

(in Botany and Plant Pathology)

The Graduate School

The University of Maine

December 2022

Advisory Committee:

Ek Han Tan, Ph.D., Associate Professor of Plant Genetics, Advisor

Yongjiang Zhang, Ph.D., Assistant Professor of Plant Physiology

Benjamin King, Ph.D., Assistant Professor of Bioinformatics

GENETIC CHARACTERIZATION OF CENTROMERE-MEDIATED UNIPARENTAL GENOME ELIMINATION IN *ARABIDOPSIS*

By Diana Spencer

Thesis Advisor: Dr. Ek Han Tan

An Abstract of the Thesis Presented
In Partial Fulfilment of the Requirements for the
Degree of Master of Science
(in Botany and Plant Pathology)
December 2022

The production of haploids through crossing was first discovered from instances of interspecies crosses in *Nicotiana spp.* in 1924. Since then, haploid induction crosses has been used to improve plant breeding programs and commercially utilized in a number of crop industries. The generation of doubled haploid instantaneously creates a pure homozygous line, therefore eliminating the need for several generations of inbreeding. There are several pathways to induce haploids in plants: of these methods, centromere-mediated genome elimination pathway engenders the highest haploid induction rate (HIR) with up to 45% in *Arabidopsis* compared to 15% through phospholipase-mediated haploid induction in maize. Centromere-mediated genome elimination operates through the manipulation of CENH3, a histone H3 variant that is associated with the formation of centromere on chromosomes. Although centromere-mediated genome elimination is highly efficient in inducing haploids, the best haploid inducer in the *Arabidopsis* system is stunted in its growth and is partially male sterile. In the work presented here, we were able to induce haploids in *Arabidopsis* using lines that were vigorous and can be crossed as either a male or female. This was achieved using four mutant allele combinations based on two recessive *cenh3* alleles: *cenh3-1*, a null allele and *cenh3-2*, a missense allele. Our

results demonstrated that we could induce haploids while balancing the trade-off between the efficacy to induce haploids and haploid inducer vigour. As *CENH3* can be found across all plant species, centromere-mediated genome elimination pathway can be employed by other plants as well. On top of that, the concept that haploids can be induced without the introduction of transgene is attractive for crop industries as it can eliminate the need to go through regulatory bodies for plant breeding programs or crop improvement efforts. Meanwhile, detailed molecular characterization of events that govern haploid induction via centromere-mediated genome elimination is still largely not known. Here, we also provide a framework and potential protocol that would eventually allow expression profiling of early *Arabidopsis* embryos undergoing centromere-mediated genome elimination.

ACKNOWLEDGEMENTS

First and foremost, heartfelt gratitude is extended to my advisor, Dr. Han Tan. Han is the epitome of intelligence mixed with tons of kindness and I am lucky to have such a person as my advisor. Indeed, no words will be enough to express how grateful I am for his constant mentoring, guidance and support. I would also like to thank my committee members, Dr. Yongjiang Zhang and Dr. Benjamin King for taking time from their busy schedule to guide and mentor me.

I am also thankful to my labmates, past and present. Special shout out to Benjamin Moore, Kat Klebon and Jake Caruson, who had helped me with my lab work. I am also thankful to Kristen Brown-Donovan and Oluwafemi Alaba for guiding and motivating me especially when things got rough at times. Special thank you is especially reserved for Margaret Perotta who had helped me with my experiment when I was away.

I am lucky to have a group of friends that always believe in me. To my best friends of 28 years, Wendy Jitos and Stephanie Stamford, thank you for always cheering me on. To Campbell Apau, Dr. Kismet Hong Ping, Dr. Danielle Levesque, Jason Teo and Pratima Pahadi, thank you for your unwavering belief and support.

Last but not least, I am very thankful for the support that my family gave me. To my daughter Aria Mason, I thank you for always being there for me. My parents, Sylvia Jaraup and Spencer Birai, and sister, Kathleen Spencer for always believing in me. I also would like to thank Charles Bernard and Dr. Cordelia Mason for the support they extended to me.

And Michael – thank you. Without you, I wouldn't be where I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

CURRENT PERSPECTIVE ON THE PRODUCTION OF HAPLOIDS IN PLANTS

1.1 Abstract

The production of haploids in plants has revolutionized plant breeding programs. Doubled haploids instantaneously create homozygous plants in a single generation, carving off years from the lengthy cycles of the traditional plant breeding programs. The existence of haploids in plants was first discovered in *Datura stramonium* in 1922. Since then, various studies had been conducted to elucidate the pathways to induce and improve haploids production. Generally, haploids can be induced *in vivo* and *in vitro.* Yet, most plants are recalcitrant to the *in vitro* methods. In the last decade, much of the research on haploid induction has been focused on the *in vivo* methods. Stock 6 is a tried and tested haploid inducer line in maize and has been utilized in the maize breeding program since 1959. However, the haploid induction rate (HIR) of Stock 6 and its derivatives have only seen increases from 3% to 15% in the last half century. In 2010, haploid induction via centromere-mediated genome elimination in *Arabidopsis* was found to induce haploids as efficiently as 45%. This breakthrough was achieved by the manipulation of CENH3, a H3 histone protein that epigenetically determines the centromeric identity of chromosomes. Besides *CENH3*, discoveries from maize identified mutations in pollen-specific phospholipases *MATRILINEAL* and *ZmPLD3,* and pollen-specific protein of unknown function, *DMP* as the causative genes for haploid induction. Efficient haploid inducers either can only be crossed as a female or a male, rely on transgenes to increase the HIR, or require high maintenance due to their lack of vigor. The ability to induce haploids through the heterozygous *cenh3* null mutant haploid inducer line has revitalized the advancement in haploid induction technology. This haploid inducer can be crossed either as a female or male and do not require transgenes. Overall, the studies of *in vivo* haploid induction

pathways in the past decade have revealed that the outlook of plant haploid induction is promising.

1.2 Haploid induction in plants

In flowering plants, the discovery of a naturally occurring haploid was first reported in *Datura stramonium* and has since captured the imagination of plant geneticists for a century (Blakeslee et al., 1922). Haploids refers to adult individuals that inherit just the gametophytic chromosome number within a single somatic cell, which is half of the parental ploidy. Therefore, a haploid derived from a diploid species ($2n = 2x$) will be monoploid ($2n = 1x$), and inherits only one set of chromosomes from one parent instead of the expected two sets of chromosomes from both parents (Forster et al., 2007; Valero et al., 1992). Developing an efficient haploid induction technology is advantageous to the advancement of science as well as economically important to the plant breeding industries (Dwivedi et al., 2015; Ravi et al., 2014). Specifically, the production of haploids is highly desirable in the crop industries and plant breeding programs because breeding lines can be inbred efficiently when haploids are doubled. A doubled haploid plant is instantaneously homozygous in a single generation, effectively bypassing several generations of traditional inbreeding (Chase, 1969; Chase, 1949; Dwivedi et al., 2015; Geiger & Gordillo, 2009).

In plants, the production of haploids by crossing was first reported in 1924 from two populations of *Nicotiana* interspecies crosses in Clausen & Mann (1924). In the hybrid between tetraploid *N. tabacum* ($2n = 4x = 48$) and diploid *N. sylvestris* ($2n = 2x = 24$), the F1 progenies were expected to be vigorous although nearly sterile replica of its *N. tabacum* parent due to triploidy. However, a progeny from each of the two populations found from the growing season of 1923 was reported to be in a smaller scale of the *N. tabacum* parent and had chromosome

numbers of $2n = 2x = 24$. These progenies were evidently *N. tabacum* haploids induced from the interspecific cross. At the time, the mechanism of haploid induction was still unknown and it was not until much later that several pathways to haploid induction were described.

Generally in plants, haploids can be induced *in vitro* via cultured gametophyte cells or *in vivo* through interspecific or intraspecific crosses as shown in Figure 1.1(A) (Laurie & Bennett, 1988; Riera-Lizarazu et al., 1996; Sanei et al., 2011; Tulecke, 1964). *In vitro* haploidization can be further categorised by its parental origin. In gynogenesis, embryo originated exclusively from the maternal genetic materials without any contribution from paternal influence, with the exception of embryogenesis stimulation by the sperm (Niazian & Shariatpanahi, 2020). In contrast, androgenesis is the development of embryo containing only the paternal genome in the cytoplasm of the maternal parent. Crops such as *Allium cepa* L. (onion), *Triticum aestivum* (wheat) and *Oryza sativa* L. (rice) are among plants that utilized *in vitro* haploid induction method to induce haploids (Bohanec et al., 1995; Chaudhary, Dhaliwal et al., 2003; Javornik et al., 1998; Mishra & Rao, 2016).

In vivo haploid induction pathways in plants involve making crosses between plants and can be further categorised as either interspecific or intraspecific crosses. Interspecific cross refers to a wide cross performed between two different plant species. The mechanism of haploid induction through interspecific cross is thought due to the centromeric dimorphism between the two species (Figure 1.1(B)) which renders one of the parental genome with a weaker centromere than the other parent's centromere (Wang & Dawe, 2018). In the case of interspecific cross between *Hordeum vulgarae* and *Hordeum bulbosum*, the depleted CENH3 in the *H. bulbosum* resulted in the uniparental genome elimination of *H. bulbosum* (Sanei et al., 2011).

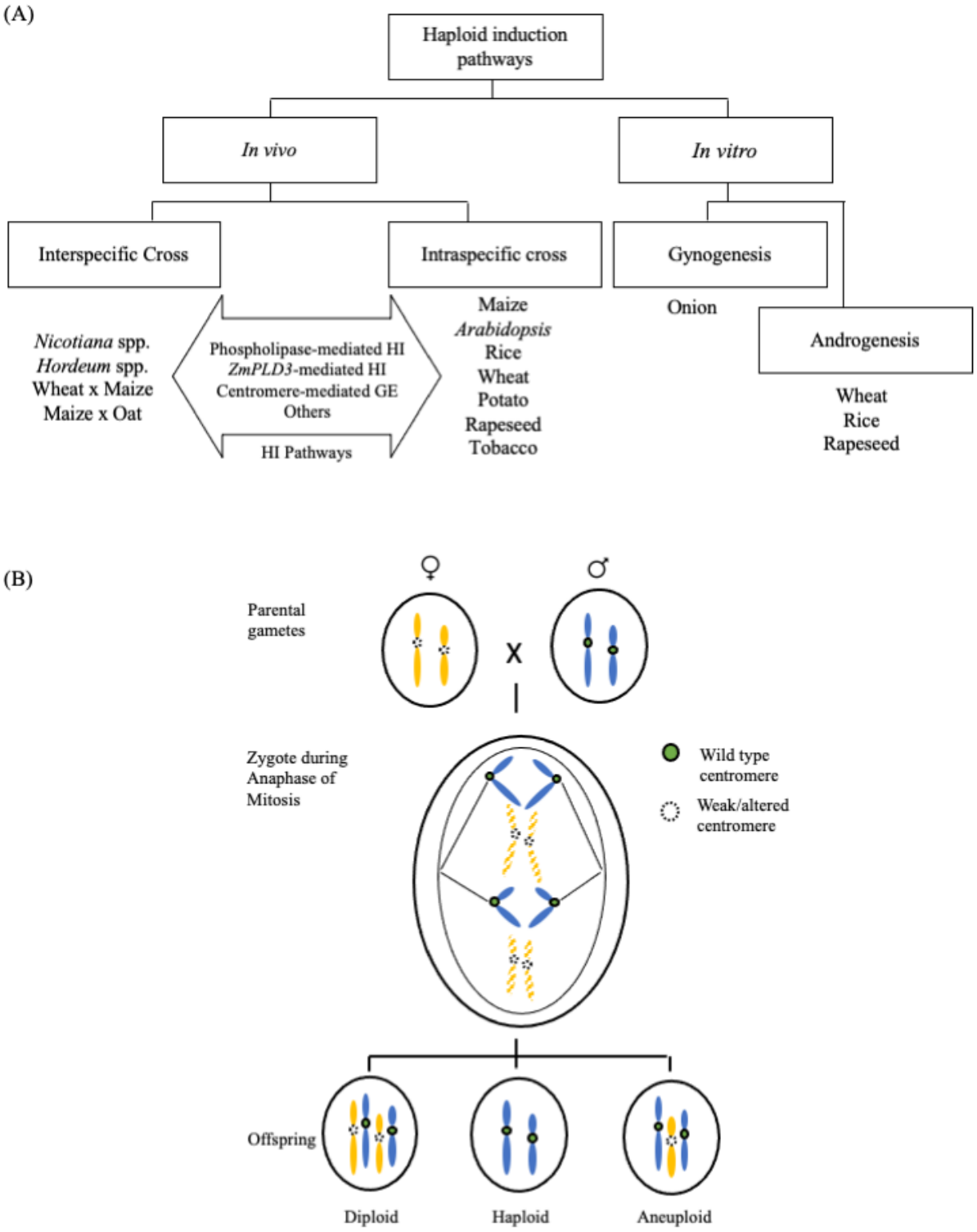


Figure 1.2 Haploid induction methods, pathways and mechanism in plants. (A) The chart is showing the methods and pathways in inducing haploids. Examples of plants induced via each method are listed underneath. (B) The cartoon is depicting the mechanism of haploid induction in the intraspecific and centromere-mediated genome elimination crosses.

As the name suggests, intraspecific cross refers to haploid induction crosses performed between the plants of the same species, often requiring a line that has been developed specifically as a "haploid inducer". There are several pathways in intraspecific haploidization, although the mechanisms that lead to haploids differ from species to species. *Zea mays* (maize) is one of the plants that produce haploids through interspecific cross. Haploids were produced by crossing to the maize haploid inducer line Stock 6. Stock 6 is an inbred line and the progenies of this line were able to produce 3% haploids when self-pollinated (Coe, 1959). *Arabidopsis* is another example of plants that utilised interspecific cross to produce haploids. Haploids in *Arabidopsis* were induced post-zygotically via centromere-mediated genome elimination and the haploid induction rate (HIR) can be as high as 45%. More detailed discussion on the postzygotic centromere-mediated genome elimination as well as about the parthenogenic phospholipase-mediated and *DOMAIN OF UNKNOWN FUNCTION 679 membrane protein (DMP)*-mediated haploid inductions are provided in the next sections (Kelliher et al., 2017; Zhong et al., 2020). Many crop plant species are resistant to gametophyte *in vitro* methods for haploid production, and this method is difficult and costly. Therefore *in vivo* haploid induction crosses remain the most widely used for haploid production purposes in crop plants (Forster et al., 2007; Kelliher et al., 2017).

1.3 Phospholipase-mediated haploid induction

Zea mays (maize) is one of the most valuable crops in the world and the industry has relied on the haploid inducer line Stock 6 and its derivatives to support the production of inbred lines for the hybrid maize seed industry (Coe, 1959). Three independent research groups identified the causative locus in *quantitative haploid induction rate 1* (*qhir1*), one of the quantitative trait loci (QTL) related to the haploid induction ability in Stock 6 as the mutation in *MATRILINEAL/Zea mays PHOSPHOLIPASE A1/NOT LIKE DAD (MTL/ZmPLA1/NLD)* gene

(Gilles et al., 2017; Kelliher et al., 2017; C. Liu et al., 2017). For simplicity, this gene will be referred to as *MATRILINEAL (MTL)*. *MTL* is a gene that can only be found in monocot plants, and is specifically expressed in pollen as a phospholipase A1 protein. The mutation that confers Stock 6 the ability to induce haploids is a 4 bp insertion that causes the alteration of 20 amino acids and ultimately, the premature transcription termination which delayed pollen germination as well as pollen tube growth (Gilles et al., 2017; Kelliher et al., 2017; Liu et al., 2017). The original Stock 6 line has a haploid induction rate (HIR) of 2-3% and workers in the field have since created improved lines with HIR up to 15% (Kelliher et al., 2017). However, the increase in the HIR also increases the kernel abortion rate and the two phenomena are thought to be linked.

The identification of *ZmMTL* as the causative gene in maize Stock 6 haploid inducer had open the door to evaluate the application of *mtl*-like gene to induce haploids in other cereal crops such as *Oryza sativa* (rice) and *Triticum aestivum* (wheat). The mutation in rice *OsMATL,* an ortholog of maize *MTL* was engineered through CRISPR-Cas9 and was able to induce haploids between 1.8% and 6% when crossed to the female inducer line (Yao et al., 2018). In wheat, the mutation in *TaMTL* was found to induce haploids at 18.9% (Liu et al., 2020).

In a further development of the phospholipase-mediated haploid induction, a mutation in maize *PHOSLIPASE D3 (ZmPLD3)* was discovered to synergistically improved the HIR induced by *zmmtl* in maize (Li et al., 2021). *ZmPLD3* encodes phospholipase D3, a subgroup of the phospholipase family and is also specifically expressed in matured pollen. When used as a single mutation, *zmmtl* was able to induce haploids at 1.2% while *zmpld3* induced haploids much at the same rate. When *zmmtl* and *zmpld3* were used as a double mutant, the HIR in maize increased almost 2-fold. The significant increase of HIR when both *zmmtl* and *zmpld3* were

used in tandem as well as the upregulation of *ZmPLD3* gene expression in the presence of *ZmMTL* suggested a synergetic interaction between the two genes during fertilization, and contributed to the doubled HIR when mutated.

1.4 *DMP*-mediated haploid induction

Recently, Zhong et al., (2019) discovered the mutation of *DOMAIN OF UNKNOWN FUNCTION 679 MEMBRANE PROTEIN* (*ZmDMP*) gene as the causative allele in the *quantitative haploid inducer rate 8 (qhir8)* of the maize CAU5 haploid inducer line. To verify *ZmDMP* as the gene that confers CAU5 the ability to induce haploids, a knockout *zmdmp-ko* mutation was engineered through clustered regularly interspaced short palindromic repeats (CRISPR-Cas9) which resulted in the frameshift mutation within *zmdmp-ko*. When maize with *zmmtl/zmmtl ZmDMP/ZmDMP* were crossed to the test line, 1% haploids were induced by this haploid induction cross. However, when the double mutant strategy of *zmmtl/zmmtl zmdmp-ko/zmdmp-ko* haploid inducers were crossed to the test line, haploids were produced at a much higher haploid induction rate (HIR) of 7%. This finding suggested that the mutation in *zmdmp* synergistically increased the haploid induction efficiency of haploid inducers with *zmmtl* mutation.

Unlike *ZmMTL*, a gene that has only been found in monocots, *DMP* gene can be found in both monocots and eudicots, paving the way for the possibility of inducing haploids in eudicot crops via *DMP*-mediated haploid induction (Zhong et al., 2020). In a eudicot such as *Arabidopsis*, two *ZmDMP*-like genes were identified as *AtDMP8* and *AtDMP9*. A translational frameshift mutation was created through CRISPR-Cas9 to engineer mutations in *AtDMP8* and *AtDMP9*. Haploid inducers with the knockout genes in the genotype configuration of single mutant *atdmp8/atdmp8*, single mutant *atdmp9/atdmp9* and double mutant *atdmp8/atdmp8*

atdmp9/atdmp9 were crossed to wild-type plants to test the genes' ability to induce haploids in *Arabidopsis*. From these crosses, haploid inducers with homozygous single mutants, *atdmp8/atdmp8* and *atdmp9/atdmp9* were found to be able to induce haploids at 0.03% and 0.39% respectively. However, when the double mutant *atdmp8/atdmp8 atdmp9/atdmp9* haploid inducers were used, they were able to induce haploids up to 2.11%.

The success in inducing haploids through *AtDMP8* and *AtDMP9* in *Arabidopsis* suggested the possibility that haploids can be induced through *DMP*-mediated strategy in other eudicots crops. In the model legume *Medicago truncatula;* mutating the orthologs of *ZmDMP, MtDMP8* and *MtDMP9* were found to be successful in inducing haploids (Wang et al., 2022). When 6 lines of legume homozygous with either *mtdmp8-1/mtdmp8-1, mtdmp8-2/mtdmp8-2, mtdmp8-3/mtdmp8-3, mtdmp9-1/mtdmp9-1, mtdmp9-2/mtdmp9-2* or *mtdmp9-3/mtdmp9-3* were self-pollinated, no haploids were produced. However, when legumes homozygous with double mutant combination of *mtdmp8/mtdmp8 mtdmp9-1/mtdmp9-1, mtdmp8/mtdmp8 mtdmp9-2/mtdmp9-2* or *mtdmp8/mtdmp8 mtdmp9-3/mtdmp9-3* were self-pollinated; haploids were induced at HIR between 0.29% and up to 0.82%. When crossed to the test lines, double mutant *mtdmp8/mtdmp8 mtdmp9-1/mtdmp9-1* had a HIR of 0.55%. These findings in *M. truncatula* and *Arabidopsis* corroborated that the simultaneous mutation of *DMP* genes were able to induce haploids in eudicot plants. In another diploid eudicot, *Solanum lycopersicon* (tomato) was also shown to be able to utilize the *DMP* gene to induce haploids. When crossed to the breeding line, tomato with ortholog *sldmp* had the collective HIR of 1.94% (Zhong et al., 2022). Meanwhile, the *stdmp* in diploid *Solanum tuberosum* group Phureja (potato) was able to induce haploids at much reduced HIR of 0.005-0.01% when crossed to female parents with different genetic background (Zhang et al., 2022).

The ability to induce haploids through *DMP*-mediated haploid induction in diploid crops as demonstrated by legume, tomato and potato was a promising start to investigate the feasibility of this pathway to induce amphihaploids in allotetraploid crops. *Brassica napus* (rapeseed) and *Nicotiana tabacum* (tobacco) are all allotetraploid and were shown to induce amphihaploids when the haploid inducer harboring *zmdmp*-like gene were crossed to the wild-type. In rapeseed, the knockout *bnadmp* was able to trigger up to 2.53% HIR when crossed to the test line while in another study, triple mutation of *bnadmp* was able to induce 2.4% ampihaploids (Li et al., 2022; Zhong et al., 2022).

The mechanism to induce haploids in the *DMP*-mediated haploid induction is still unknown although it was hypothesized that the deficiency of *dmp* gene might trigger maternal haploids (Zhang et al., 2022). Similar to *ZmMTL* and *ZmPLD3*, *DMP* was found to be abundantly expressed in mature pollen of CAU5 maize, legume, potato and *Arabidopsis* (Wang et al., 2022; Zhang et al., 2022; Zhong et al., 2020, 2019). Therefore, a shortcoming in *DMP*-mediated haploid inducer strategy is that the haploid inducers could only be crossed as males as the genes at play are pollen-specific. Nevertheless, the feasibility of engineering monocot and eudicot plants via orthologous *DMP*-like genes remains promising as it had been shown that the HIR increased from less than 1% when applied as a single mutation to about 4% and 7% when used in tandem with *zmmtl* and *zmmtl zmpld3* respectively (Li et al., 2021). These findings suggested that there are more than one pathways for haploid inductions to work. Another haploid induction pathway, mediated by the centromeric histone H3 variant (CENH3) will be discussed in the next section.

1.5 Haploid induction via centromere-mediated genome elimination

Haploids induced via uniparental centromere-mediated genome elimination (also known as *CENH3*-mediated genome elimination in other literatures) were previously thought to occur only in rare interspecific crosses (*Nicotiana* spp., *Hordeum* spp., wheat x maize and maize x oat) in the wild (Clausen & Mann, 1924; Laurie & Bennett, 1988; Rines & Dahleen, 1990; Subrahmanyam & Kasha, 1973). Previous studies had shown that the centromeric dimorphism between the two interspecific parents could result in post-zygotic genome elimination of the parent with the "smaller" or "weaker" centromere as shown in Figure 1.1(B) (Ishii et al., 2016; Wang & Dawe, 2018). It was not until 2010 when the ground-breaking study published by Ravi and Chan (2010) demonstrated that haploids could be induced in the intraspecific crosses as well, specifically in *Arabidopsis*. As with interspecific crosses, the centromere plays a prominent role in intraspecific genome elimination, specifically through the manipulation of *CENH3* gene in haploid inducers.

CENH3, also known as CENP-A, is a histone H3 variant that epigenetically determines the centromeric region in chromosomal nucleosomes (Malik & Henikoff, 2003, 2009). CENH3 protein consists of a construct that contains a conserved Histone Fold Domain (HFD) and a highly variable N-terminal tail. The centromere is critical for chromosomal segregation as CENH3-containing nucleosomes bind to kinetochore complexes which attach to the microtubules during mitosis and meiosis (Darlington, 1939; Gairdner & Darlington, 1931). The ability to attach to microtubules mediates the success of chromosomes segregation in both mitosis and meiosis (Lermontova et al., 2011). Therefore, alteration or mutation to CENH3 may hinder the centromere attachment to the kinetochore complex, thus hamper the chromosomal segregation in either mitosis or meiosis. In extreme cases, the chromosomes missegregation may cause chromothripsis, which is the breakage and rearrangement in

chromosomes or loss of a whole uniparental genome altogether (Comai & Tan, 2019; Tan et al., 2015).

Arabidopsis with variant CENH3 has been shown to be able to induce haploids up to 45% efficacy (Ravi & Chan, 2010). In that study, the embryo-lethal null mutant *cenh3-1* was complemented by *GFP-tailswap* or other transgenic *CENH3* variants. When haploid inducers with CENH3 variant were crossed to wild-type *Arabidopsis*; chromosomes from the haploid inducer parent were eliminated post-zygotically, leaving the progeny with only chromosomes from the wild-type parent (Kuppu et al., 2015; Maheshwari et al., 2015; Ravi & Chan, 2010). When the transgenic *Arabidopsis* was crossed as a female to the wild-type, haploids were produced at a higher frequency than when it was crossed as a male (Ravi & Chan, 2010). These studies also showed that self-pollinated haploid inducers did not produce haploids. This observation supports the theory that dimorphism between parentals centromeres caused unequal competition for the microtubule, resulting in the loss of chromosomes for the parent with "smaller" or "weaker" centromere. In the case of self-pollinated haploid inducer, both fusing gametes have "weak" centromeres thus no competition for microtubule occurred in the mitotic or meiotic plates.

The differences in the severity of mutation in *CENH3* seemed to be a factor in the quality of the haploid inducers. Haploid inducer with *cenh3-1/cenh3-1* complemented by *GFP-CENH3/GFP-CENH3* was able to induce haploids at just 5% compared to up to 45% haploid induction rate (HIR) by haploid inducer with *cenh3-1/cenh3-1* complemented by *GFP-tailswap/GFP-tailswap*. The variant GFP-CENH3 protein is less severe than GFP-tailswap as the former only had the GFP protein attached to the end of its hyper-variable tail while the latter had its hyper-variable tail replaced by the H3.3 histone variant on top of GFP attached to

its tail. In haploid inducer with *cenh3-1/cenh3-1* rescued by *cenh3* variant that contains L130F missense mutation, the single point amino acid exchange was able to induce haploid at HIR of 2.8% (Karimi-Ashtiyani et al., 2015). Meanwhile, *Arabidopsis* homozygous with *cenh3-2/cenh3-2* was able to induce haploids up to 3.87%. The allele, *cenh3-2* is also a missense mutation in which a single nucleotide change of C-to-T at 717 relative to ATG = +1 of *AT1G01370* gene results in the single amino acid change of alanine to valine at the HFD of the CENH3 histone protein (Kuppu et al., 2015). In another study, variant *CENH3s* were derived from progressively distant relatives of *Arabidopsis* such as *Lepidium oleraceum*, *Brassica rapa* as well as *Zea mays* (Maheshwari et al., 2015). To integrate the *CENH3s* of *Arabidopsis* distant relatives, haploid inducers with *cenh3-1/cenh3-1* null mutant were rescued by either *LoCENH3, BrCENH3, AtNTT-LoHFD* or *LoNTT-AtHFD.* In *AtNTT-LoHFD* and *LoNTT-AtHFD,* the hyperterminal tail and HFD of *Arabidopsis* and *L. oleraceum* were swapped so that AtNTT-LoHFD construct had the *Arabidopsis* tail and *L. oleraceum* HFD and vice versa in LoNTT-AtHFD. It was found that *LoCENH3* was able to induce haploids in *Arabidopsis* up to 11%.

The details on the exact molecular pathways leading to haploid induction via centromere-mediated genome elimination is still not known. Genome elimination in *Arabidopsis* is thought to occur during the pre-globular stage of the early embryogenesis during the first post-zygotic mitosis. This hypothesis was strengthened by the evident that in the first mitosis following the fusion of egg and sperm following genome elimination cross utilising *GFP-tailswap/GFP tailswap* as the haploid inducer, wild-type CENH3 and GFP-tailswap were biasedly loaded into the centromeric nucleosomes of the wild-type parent and were absent from the chromosomes of the parent with the mutant *cenh3-1/cenh3-1 GFP-tailswap/GFP-tailswap* (Marimuthu et al., 2021). However, it was shown that uniparental genome with mutant *cenh3-1/cenh3-1 GFP-*

tailswap/GFP-tailswap persisted and the loading of GFP-tailswap into its centromeric nucleosomes steadily increased between 2-6 days after pollination. Therefore, the uniparental genome elimination with "weaker" centromere had to take place before the recovery of repopulated centromere.

1.5.1 Haploid induction via centromere-mediated genome elimination in maize

The production of haploids in *Zea mays* (maize) has historically relied on the Stock 6 haploid inducer and its derivative, which we now know are based on the phospholipase-mediated pathway. However, the haploid induction rate (HIR) of phospholipase-mediated pathway in maize is much lower compared to the HIR induced through centromere-mediated genome elimination in *Arabidopsis.* Therefore, a similar study built from the classic centromere-mediated genome elimination in *Arabidopsis* was conducted in the maize system to investigate whether the manipulation of *ZmCENH3,* the orthologue of *CENH3* in maize, would be able to induce haploids and at much higher HIR than the HIR of phospholipase-mediated pathway.

In the maize study, native *ZmCENH3* was either knocked out or knocked down by RNAi and rescued by *AcGREEN-CENH3 and AcGREEN-TAILSWAP-CENH3;* constructs that were similar to GFP-CENH3 and GFP-tailswap in *Arabidopsis.* AcGREEN-CENH3 protein was derived from ZmCENH3 with AcGreen tagged to its hyper-variable tail while AcGREEN-TAILSWAP-CENH3 protein had its hyper-variable tail replaced by maize H3.3 tail and tagged with AcGreen. The complementation of transgenic *ZmCENH3* in maize haploid inducer lines yielded much lower HIR when compared to the HIR in *Arabidopsis* or maize phospholipase-mediated haploid induction (Kelliher et al., 2016). Haploid inducer with knockout *ZmCENH3* rescued by hemizygous *AcGREEN-TAILSWAP-CENH3* had the highest HIR at 0.86% when crossed as a male to the wild-type *ZmCENH3* compared to the male knockout *ZmCENH3*

haploid inducer complemented with hemizygous *AcGREEN-CENH3*, which yielded a mean of 0.31% HIR. Although the mean HIR from the two haploid inducers were quite low, the highest single event HIR via centromere-mediated genome elimination in maize was recorded at 3.6% with the utilization of *AcGREEN-TAILSWAP-CENH3* in the knockout line. This finding was in sync with the *Arabidopsis* study where the *GFP-tailswap* transgene catalysed the highest HIR. However, unlike *Arabidopsis* haploid inducers, maize haploid inducers did not suffer from the sterility issue and therefore can be crossed in both direction. Interestingly, male haploid inducers in maize outperformed female haploid inducers, which was a contradiction to the *GFP-tailswap* haploid inducer in the *Arabidopsis* system.

1.5.2 Heterozygous *cenh3* null mutant in centromere-mediated genome elimination

Another study of centromere-mediated genome elimination in *Zea mays* (maize) system recently showed that haploid can be induced via a heterozygous *cenh3* null mutant line (Wang et al., 2021). *CENH3/cenh3* line was able to induce haploids up to 5.2% when crossed as a female in maize. *cenh3-1* null mutant as the source of CENH3 variant has not been used in *Arabidopsis* centromere-mediated genome elimination experiments. Instead, as a null mutant, *cenh3-1/ cenh3-1* was utilized to replace the native *CENH3* with variant transgene. In the maize study, the authors hypothesized that in a heterozygous *CENH3/cenh3* mutant, the "weak" centromere effect can be achieved via the dilution of wild-type CENH3 in the gamete inheriting the *cenh3* null mutant allele in its genome (Figure 1.2). Following meiotic division, female gametophyte undergoes 3 rounds of mitoses while male gamete undergoes 2 rounds of mitoses. The mature gametes inheriting *cenh3* mutant allele would have to depend on the existing wild-type CENH3 carried over prior to the meiotic divisions to sustain their CENH3 supplies. Female gamete with *cenh3* mutant allele would only have 12.5% of wild-type CENH3 while male gamete with *cenh3* mutant allele would only have 25% of wild-type CENH3. Therefore,

gametes with diluted wild-type CENH3 can be expected to have weaker centromeres compared to gametes with wild-type *CENH3* allele. When the two gametes fused upon fertilization, genome with the weaker centromere would be lost, resulting in haploids inheriting uniparental genomes from the parent with wild-type *CENH3*.

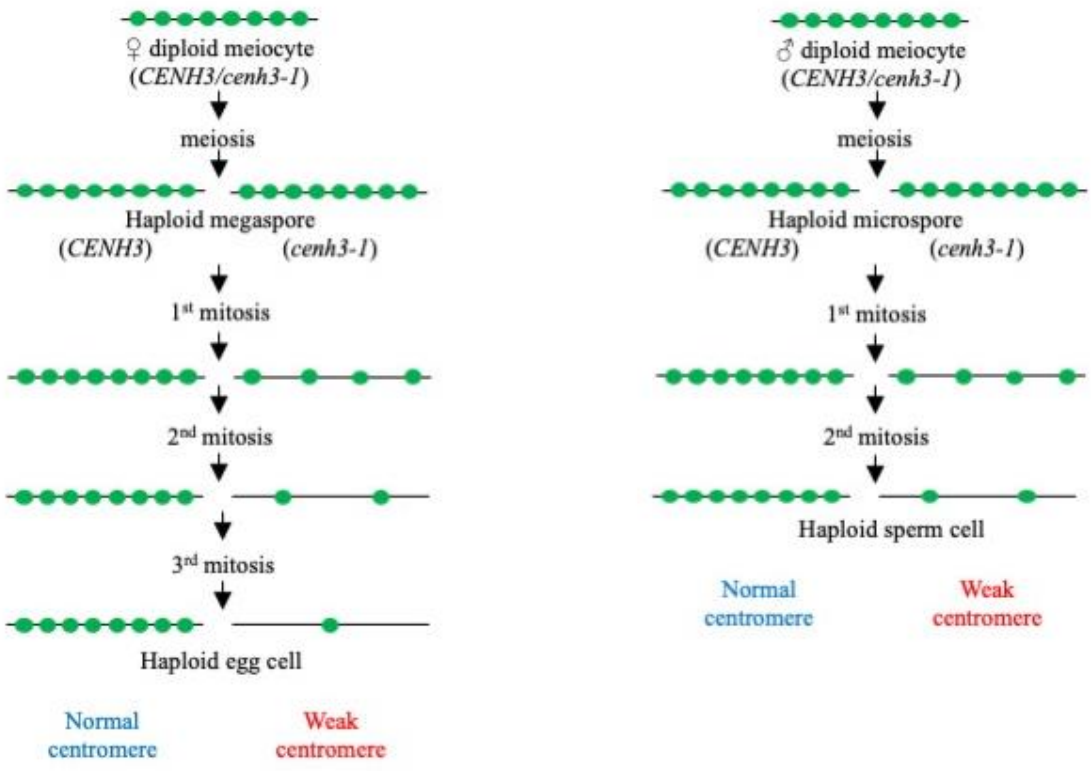


Figure 1.3 Depletion of CENH3 at each cycle of mitoses following meiosis in haploid spores inheriting the *cenh3-1* allele during gametogenesis. The green bubbles represent wild-type CENH3 and estimated retention levels in a *cenh3-1* cell. The cartoon on the left represents mitotic events that lead to the generation of a haploid egg cell and the cartoon on the right represents mitotic events that lead to the generation of a haploid sperm cell.

Although the haploid induction rate (HIR) in the heterozygous *cenh3* null mutant method was less than 1/8 of the HIR in the traditional centromere-mediated genome elimination pathway, this method offers several advantages over the traditional method. As the haploid inducer is phenotypically wild-type, the plant would have no sterility issue to hamper the direction of the crossing and it would be easy to breed. Consequently, the prospect of centromere-mediated

genome elimination whether through the quantitative reduction of the wild-type CENH3 or over the quality of variant CENH3 is promising. And as *CENH3* can be found across plant species, haploid induction through centromere-mediated genome elimination can be employed in both monocot and eudicot plants (Pal & Negi, 2019).

1.6 Conclusion

Throughout the history of haploid induction in plants, modern genetic discoveries have led to the discovery of a number of key players that are involved in *in vivo* genome elimination crosses. Major players that are well studied include *CENH3*, and a number of pollen-specific genes that were described from the maize system. Phospholipase-mediated and *DMP*-mediated haploid inducer lines can only be crossed as male due to the phospholipases encoded by *MTL* and *ZmPLD3* as well as the protein encoded by *ZmDMP* is primarily expressed in the mature pollen, while the *GFP-tailswap* haploid inducers in centromere-mediated genome elimination are mostly male sterile (Kelliher et al., 2017; Li et al., 2021; Ravi & Chan, 2010; Zhong et al., 2019). Developing vigorous, high haploid induction rate (HIR) haploid inducer lines should be prioritized and the ability to use a haploid inducer as either a male or female parent would be useful especially in cases when cytoplasmic incompatibility is an issue. Recent breakthroughs in haploid induction pathways from the last few years have shown that haploid inducers and their respective HIR could be continually improved and the work presented here provides a working model for improving the *CENH3*-based genome elimination system.

CHAPTER 2

EVALUATING CENTROMERE-MEDIATED GENOME ELIMINATION VIA LOSS OF FUNCTION *CENH3* MUTANT ALLELES IN *ARABIDOPSIS*

2.1 Abstract

The efficacy of inducing haploids is highly coveted in plants as doubled haploids created an instantaneous pure homozygous line in a single generation, eliminating the need for a complex plant breeding program. Through the modification of CENH3, a centromeric histone H3 variant, transgenic GFP-tailswap haploid inducers can induce haploids with an efficiency of up to 45% in *Arabidopsis* when crossed as a female, but not as a male. Here we show that two mutant alleles of *cenh3* in *Arabidopsis; cenh3-1*, a null allele, and *cenh3-2*, a missense allele, can induce haploids without the introduction of transgene and can be used as a male or a female. Haploid inducer with homozygous *cenh3-2* was the most efficient haploid inducer when crossed as a male with a haploid induction rate (HIR) of 2.58% while the double mutant *cenh3-1/cenh3-2* had a HIR of 2.27% when crossed as a female. All heterozygous and homozygous inducers tested were able to maintain their vigor. As *CENH3* can be found across plants systems, the same strategy to employ variant *cenh3* as either a female or a male inducer can be adopted in other plant system. The ability to induce haploids without introducing transgene to plants is advantageous to crop industries as it eliminates the need to go through regulatory bodies while the vigorous haploid inducers allow for low maintenance of the inducer lines.

2.2 Introduction

The advancement in haploid induction technology is desirable in plant breeding. Through haploid induction, plant breeders are able to create pure homozygous lines in diploid species in a single generation, effectively bypassing several generations of inbreeding (Dwivedi et al.,

2015). In plants, haploids can be induced via cultured gametophyte cells or via haploid induction crosses (Laurie & Bennett, 1988; Riera-Lizarazu et al., 1996; Sanei et al., 2011; Tulecke, 1964). Generally, the haploid induction cross method to induce haploids is preferred as most plants are resistant to the culture gametophyte methods (Forster *et al.*, 2007). In the haploid induction cross method, plants are crossed to the haploid inducer lines with the progenies inheriting the genetic materials of the non-haploid inducer parents. The first reported event of haploid induction by crossing was an interspecific cross between *Nicotinia tabacum* var. *purpurea* ($2n = 4x = 48$) and *Nicotinia sylvestris* ($2n = 2x = 24$) which yielded haploids that resembled the scaled down parental *N. tabacum* with diploid chromosome counts (Clausen & Mann, 1924). The pathway of the haploid induction was not known at the time. Molecular understanding of haploid induction by crossing was only recently described and spans the centromere-mediated genome elimination, pollen specific phospholipase-based basis for haploid and *DMP*-mediated haploidization (Kelliher et al., 2017; Marimuthu et al., 2021; Wang, 2022).

Centromere-mediated genome elimination, one of the pathways to induce haploids was pioneered by Ravi & Chan (2010) and has the ability to induce haploids up to 45% in *Arabidopsis* and 5.2% in maize (Wang et al., 2021). Central to the haploid induction via centromere-mediated genome elimination technology is the alteration of CENH3 (also known as CENP-A in the animal system), a functionally conserved centromeric H3 histone variant that defines centromeric identity in eukaryotes, and is essential to mediate faithful segregation of chromosomes during mitosis and meiosis (Malik & Henikoff, 2003, 2009). In the original study, alteration of CENH3 was achieved by the complete replacement of the native CENH3 with transgenic CENH3. This was made possible by the complementation of embryo-lethal null mutant *cenh3-1* with transgene *GFP-tailswap* which encoded CENH3 with its hyper-

variable n-terminal tail replaced by a H3.3 variant and fused with a green fluorescent protein (GFP) tag (Ravi & Chan, 2010). In *Arabidopsis*, the haploid inducer derived through *GFP-tailswap* is 30% more efficient in inducing haploids than *Zea mays* (maize) Stock 6 haploid inducer. Other variant *CENH3* genes with less extreme alteration than the *GFP-tailswap* were also able to induce haploids albeit at a lower haploid induction rate (HIR) (Karimi-Ashtiyani et al., 2015; Kuppu et al., 2015; Maheshwari et al., 2015). It is hypothesized that plant expressing variant *CENH3* has a defective centromere compared to the wild-type and the chromosomes with defective centromere will be lost at the mitotic plate in the presence of chromosomes with regular centromere. The more extreme the alteration of *CENH3*, such as that of *GFP-tailswap*, the more efficient the line is as a haploid inducer, but the plants may also have phenotypically undesirable characteristics such as dwarfism and male sterility.

Due to the promising efficacy of centromere-mediated genome elimination as a haploid inducer in *Arabidopsis*, a similar study was conducted in maize to evaluate if haploids can be induced in crop plants at a comparable HIR. The most promising HIR was when the orthologous *ZmCENH3* was knocked down and rescued by the transgenic *AcGREEN-TAILSWAP-CENH3* (Kelliher et al., 2016). This transgenic variant retained ZmCENH3's histone fold domain (HFD) and similar to the GFP-tailswap construct, the native hyper-variable tail was replaced by maize's H3.3 histone variant with fluorophore AcGFP1 (denoted as AcGreen) fused to the altered tail. However, maize expressing *AcGREEN-TAILSWAP-CENH3* can only induce haploid with average of 0.86% and up to 3.6% when crossed as a male to wild-type maize, a much lower HIR compared to *Arabidopsis* HIR when employing haploid induction method via a similar GFP-tailswap strategy.

Following this, progress to the haploid induction technology was made when Wang et al., (2021) addressed low HIR of the RNAi/AcGREEN-TAILSWAP-CENH3 construct in maize using a null mutation of *cenh3*. Using haploid inducer line that was heterozygous for *cenh3* null mutation instead of homozygous for variant *CENH3*, the HIR for the revamped method increased to 5.2% when the heterozygous haploid inducer line was crossed as a female and 0.5% when crossed as a male. Heterozygous *CENH/cenh3* is thought to enable centromere-mediated genome elimination as the gametophyte harbouring the *cenh3* allele causes weaker centromeres in a sperm cell or an egg cell. Immediately following meiotic division, female and male gametophytes undergo 3 and 2 rounds of mitoses respectively (see Chapter 1, Figure 1.2). Haploid gamete inheriting the *cenh3* null mutant allele following the segregation in meiosis only has functional CENH3 proteins carried over from sporophytic phase and these CENH3s were further diluted at each mitotic step. Therefore, the depletion of the CENH3s burdened the egg or sperm carrying variant *cenh3* allele with weaker centromeres *(*Marimuthu et al., 2021).

In this study, we wanted to test if the recessive *cenh3-1* null mutant allele or the recessive *cenh3-2* missense mutant allele can be utilized as haploid inducers for centromere-mediated genome elimination method as heterozygotes, double mutant (*cenh3-1*/*cenh3-2*) or homozygote (as in the case for *cenh3-2*). We were also interested to see how efficient these haploid inducers are when crossed as a male, which have not been tested before. The first mutant allele, *cenh3-1* harbours a G-to-A mutation at nucleotide 161 relative to ATG = +1 of *AT1G01370*. This mutation results in exon 1 being spliced incorrectly to nucleotide 140 relative to ATG = +1 in the middle of the first intron. Although the first 18 amino acids were encoded correctly, the single nucleotide change resulted in the frameshift that encoded stop codon just after 46 amino acids. Because *CENH3* is an essential gene, homozygous *cenh3-1*/*cenh3-1* is embryo lethal (Kuppu et al., 2015; Maheshwari et al., 2015; Ravi & Chan, 2010). The second

mutant allele, *cenh3-2* is a missense mutation as a result of a C-to-T nucleotide change at 717 relative to ATG = +1 of *AT1G01370* which elicits a single amino acid change from alanine to valine (A86V) in the HFD of CENH3. The *cenh3-2* loss of function allele is therefore less severe compared to the *cenh3-1* allele (Kuppu et al., 2015). Homozygous *cenh3-2/cenh3-2* plants are viable and can be self-pollinated. We use *Arabidopsis* ecotype Landsberg *erecta gl1* (L*er gl1*) that harbors wild-type *CENH3* as the tester line for our experiments. The *GL1* gene in *Arabidopsis* encodes the protein that is required for the development of trichomes on its leaves and the homozygous recessive *gl1* mutation confers smooth, trichomeless leave surfaces while the homozygous recessive *erecta* mutation confers a plant with compact rosette and stature. Therefore, crossing L*er gl1* with a haploid inducer will yield haploids of L*er gl1*, which would exhibit trichomeless leaves with compact rosette, an efficient and economical means to screen for haploid progeny.

The study of the *cenh3* null mutation in the maize system demonstrated that the haploid inducer can be crossed either as a female or a male which can induced either paternal or maternal haploids that confers some added benefits. These haploid inducers do not require transgenes, and are easy to manage and grow since they are phenotypically wild-type. Therefore, this study seeks to evaluate the efficacy of recessive *cenh3* alleles in inducing haploids in *Arabidopsis* system and the results from this study can potentially be used to inform the manner in which mutant alleles can be employed to utilize centromere-mediated genome elimination in other crops.

2.3 Materials and methods

2.3.1 Plant materials and growth conditions

Seeds were sown in PRO-MIX BX, an all-purpose soil consisting of peat moss, peat hummus, perlite and limestone. Pots containing seeds were stratified in a dark 4°C chamber for three days to elicit germination. Plants were moved into Conviron ADAPTIS growth chamber and were grown under a controlled environment at 16 hours of light/8 hours of dark at 21°C with relative humidity of 65%. Plants were crossed within the first week of bolting on the primary emerging inflorescences.

2.3.2 Genomic DNA extraction

One small *Arabidopsis* leaf (approximately 50 - 100 mg) was submerged in 125 µl quick DNA extraction buffer (200 mM Tris HCl pH 8.0, 250 mM NaCl, 25 mM EDTA pH 8.0, 0.5% SDS) to extract genomic DNA and incubated at 99°C for 10 minutes. 100 µl of isopropanol was added to the 100 µl of the extracted supernatant and incubated for 15 minutes under room temperature to precipitate the DNA. The mixture was centrifuged for 15 minutes to pellet the DNA, followed by the removal of the supernatant by aspiration. 200 µl 70% ethanol was used to wash the pellet and the pellet was left to dry at room temperature for 15 minutes. 50µl Tris-EDTA (10 mM Tris HCl pH 8.0, 0.5 mM EDTA pH 8.0) buffer was added to the pellet and incubated at 55°C to elute the DNA.

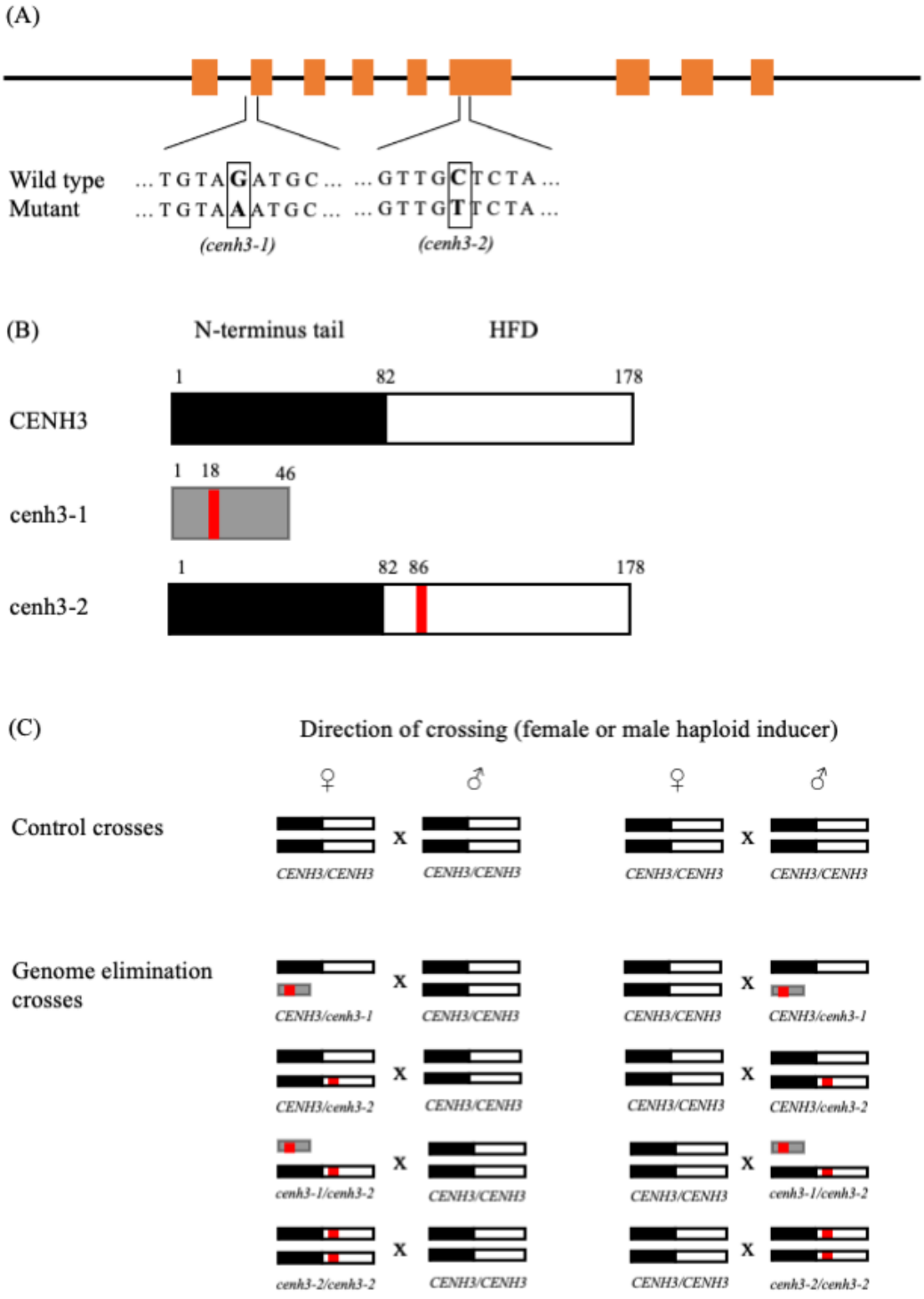


Figure 2.1 Experimental approach to evaluate the efficacy of *cenh3-1* and *cenh3-2* mutant alleles. All cartoons are not to scale. (A) Cartoon of *Arabidopsis CENH3* gene showing the nucleotide sequences of the wild-type as well as *cenh3-1* and *cenh3-2* mutant allele sequences. Orange boxes denoted exons. (B) The protein structure of CENH3 and variants as a result of the corresponding mutant alleles. Amino acid changes due to mutation are indicated by a red line and the grey box symbolized the null mutant protein structure. (C) Control and genome elimination crosses performed. The copy number and structure of CENH3 and its variants are showing the correspondent alleles of haploid inducer when reciprocally crossed as a female or a male in genome elimination crosses.

2.3.3 Genotyping for *cenh3* mutant alleles

Extracted genomic DNA was amplified using polymerase chain reaction (PCR). Each reaction was carried out in 15 µl volume, consisting of 7.5 µl Promega's 2X GoTaq® G2 Green Master Mix Catalog #M7823, 1 µl of 10 µM forward and reverse primers, 5.5 µl nuclease-free H_20 and 1 µl of DNA. The forward and reverse primers were specially designed derived Cleaved Amplification Polymorphic Sequences (dCAPS) primer sets (Table 2.1) using Geneious version 11.1.5. The *cenh3-1* dCAPS XbaI forward and reverse primer set targeted 60 bp upstream from the *cenh3-1* SNP and 148 bp downstream, while *cenh3-2* dCAPS HhaI forward and reverse primer set targeted 165 bp upstream from the *cenh3-2* SNP and 60 bp downstream. The PCR reaction was carried out in the BIO-RAD T100^{TM} thermal cycler programmed to 5 minutes of denaturation at 95°C; 40 cycles of 15 seconds of denaturation at 95°C, 15 seconds of primer annealing at 65°C and 30 seconds of primer extension at 55°C. Next, the amplified DNA was incubated overnight at 37°C for the XbaI or the HhaI digestion. The restriction enzymes recognized the wildtype allele at 5'…T^CTAGA…3' for XbaI and at 5'…GCG^C…3' for HhaI and cleaved the site denoted by the carrot sign while leaving the mutant alleles intact. The final PCR product was then resolved in a 2% agarose gel, stained in ethidium bromide, and visualized under UV illumination.

Table 2.3 Forward and reverse dCAPS primers design and digestion enzymes for *cenh3-1* and *cenh3-2* alleles. The carrot sign shows the digestion site of the enzyme**.**

Allele	Primer	Forward Sequence	Annealing Temperature	Digestion Enzyme	Enzyme Digestion Site
cenh3-1	*cenh3-1* XbaI dCAPS forward	AGAATTTTAGGTTTTTTAT TTCGATTTTGTTAACCCTA GATTTCGAATCTGAAATTT cTA	66.0°C	XbaI	5'…T^CTA GA…3' 3'…AGAT C^T…5'
	cenh3-1 XbaI dCAPS reverse	GCCTCTCCTTGTCGGGGTC TTCA			
cenh3-2	*cenh3-2* HhaI dCAPS forward	AAGAGATCCAGACAGGCT ATGCC	61.8°C	HhaI	5'…GCG^C …3' 3'…C^GCG …5'
	cenh3-2 HhaI dCAPS reverse	AAACTGGCAGCCGGAATA AGAAGGTTTGTCTGCTTCT GGAAATGGCGAATCTCTTT TAGc			

2.3.4 Plant growth assessment

The vigour of the (i) *CENH3/cenh3-1* plants, (ii) *CENH3/cenh3-2* plants, (iii) *cenh3-1/cenh3-2* plants, (iv) *cenh3-2/cenh3-2* plants and (v) wild-type Col-0 were assessed and scored based on the phenotypes of the plants at 3rd and 4th weeks. The phenotypes of the plants that were scored were their overall size, leaf morphology, flower morphology, silique length, mean total seeds per silique and viable seed per silique. For silique length, a total of n = 32 siliques from each haploid inducer line were measured for comparison to Col-0 and mean of total seeds per silique and viable seed per silique were derived from the same silique sets.

2.3.5 Haploid induction crosses

A total of 4 reciprocal genome elimination crosses to the test line L*er gl1* were carried out to assess the efficacy of *cenh3-1* and *cenh3-2* mutant alleles as a haploid inducer (Figure 2.1(C)). The reciprocal genome elimination crosses were: (i) *CENH3/cenh3*-1 plants crossed to wild-type plants, (ii) *CENH3/cenh3-2* plants crossed to wild-type plants, (iii) *cenh3-1/cenh3*-2 plants crossed to wild-type plants and (iv) *cenh3-2/cenh3*-2 plants crossed to wild-type plants. Col-0

with wild-type *CENH3* plants were reciprocally crossed to the L*er gl1* with wild-type *CENH3* plants as the control/non-genome elimination crosses.

In both the genome elimination and control crosses, unopened buds were emasculated to prevent self-pollination. The sepal, petal and stamen from the unopened bud were carefully removed under a magnifier with forceps, leaving only the exposed pistil. Pollination was carried out immediately by brushing the anthers from the designated other parent to the exposed stigma and followed by a second pollination event the next day. Between 20 to 60 pollinations were made for each cross that was carried out.

2.3.6 Seeds evaluation and haploid induction rate screening

Previous work on centromere-mediated genome elimination in *Arabidopsis* has shown that seed death frequency from haploid induction crosses provides a proxy for estimating the efficacy of haploid production (Kuppu et al., 2020). Under this assumption, we scored seed death frequencies from Col-0 and the four haploid inducers when crossed as either a female or a male. Only seeds from siliques with a length more than 1 cm were screened to eliminate screening injured siliques due to emasculation during the crossing process. In the female haploid inducers, n = 9, 23, 14, 18 and 28 siliques of control Col-0, heterozygote *CENH3/cenh3-1*, heterozygote *CENH3/cenh3-2*, double mutant *cenh3-1/cenh3-2* and homozygote *cenh3-2/cenh3-2* were used. In the male haploid inducers, n = 10, 11, 11, 17 and 13 siliques of control Col-0, heterozygote *CENH3/cenh3-1*, heterozygote *CENH3/cenh3-2*, double mutant *cenh3-1/cenh3-2* and homozygote *cenh3-2/cenh3-2* were used. Seeds were screened silique by silique under a dissecting microscope. The number of seeds per silique was recorded as well as the number of viable seeds and dead seeds. The viable seeds looked big and plump relative to the dead seeds which looked shrunken (Henry et al., 2005).

Seeds were sterilized in chlorine gas for at least 5 hrs (Lindsey et al., 2017). Then, the seeds were sown in 1/2X MS media from Caisson Labs with 1% sucrose to maximize germination efficiency and were stratified for 3 days in a dark 4°C fridge before they were moved to the growth chamber. After the true leaves emerged, the seedlings were transplanted from the media to soil. Ploidy assessment was performed on 3-week-old seedlings at the rosette stage to maximize the accuracy of identifying L*er gl1* haploids, which were products from the genome elimination crosses. Individuals that appear wild-type with trichomes were scored as diploid, while trichomeless seedlings with compacted rosette were scored as haploid. Plants that have trichomes and present the typical developmental defects associated with aneuploidy were scored as aneuploid (Ravi & Chan, 2010).

2.3.7 Statistical analysis

All statistical analyses were performed in RStudio using R version 4.2.1. Package *psych* version 2.2.5 was used to run the analysis of variance (ANOVA) and post-hoc Tukey HSD, package *stats* version 4.2.1 for chi-squared test and linear regression model, and package *ggplot2* version 3.3.6 was used to plot the data analyses. Data were analysed with one-way ANOVA followed by post-hoc Tukey HSD for mean of siliques length, mean of total seeds per silique and self-pollinated seed death frequency. Chi-squared test was used to determine if the seed death frequency of self-pollinated haploid inducers with heterozygous *CENH3/cenh3-1* and double mutant *cenh3-1/cenh3-2* was equal with the Mendelian frequency. For seed death frequencies of the reciprocal genome elimination crosses, data were analysed with two-way ANOVA with interaction followed by Tukey HSD for post-hoc analyses. Linear regression model was fitted for the seed death frequencies from genome elimination crosses and haploid induction rates (HIR) in female and male haploid inducers. For all statistical tests performed, differences between groups were considered significant at $p < 0.05$.

2.4 Results

Four haploid inducer lines were generated to evaluate the efficacy of mutant *cenh3-1* and *cenh3-2* alleles as haploid inducers in *Arabidopsis*: a *CENH3/cenh3-1* heterozygous mutant, a *CENH3/cenh3-2* heterozygous mutant, a *cenh3-1/cenh3-2* double mutant, and a *cenh3-2/cenh3-2* homozygous mutant. Haploid inducers heterozygous with *CENH3/cenh3-1* or double mutant *cenh3-1/cenh3-2* were identified by the double band of 148 bp and 208 bp through the XbaI derived Cleaved Amplification Polymorphic Sequences (dCAPS) assay while Col-0 and haploid inducers with either heterozygous *CENH3/cenh3-2* or homozygous *cenh3-2/cenh3-2* only have the cleaved product of 148 bp (Figure 2.2(A)). For the HhaI dCAPS assay, Col-0 and haploid inducers heterozygous for *CENH3/cenh3-1* only have the cleaved product of 165 bp while haploid inducers heterozygous for *CENH3/cenh3-2* and double mutant *cenh3-1/cenh3-2* have a double band of 165 bp and 225 bp (Figure 2.2(B)). Lastly, homozygous *cenh3-2/cenh3-2* mutant displayed an undigested band at 225 bp in the presence of HhaI.

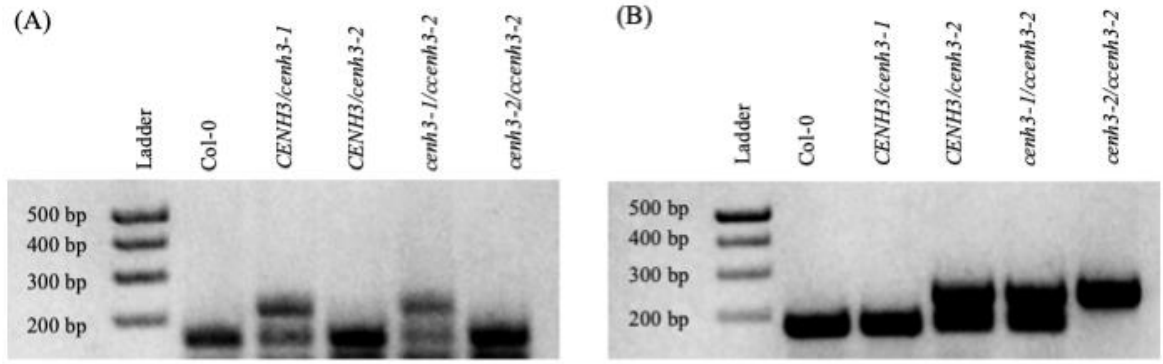


Figure 2.2 Electrophoretic gel image of haploid inducer genotypes. From left to right, genotype for control Col-0 and haploid inducers heterozygous for *CENH3/cenh3-1*, heterozygous for *CENH3/cenh3-2*, double mutant *cenh3-1/cenh3-2* and homozygous *cenh3-2/cenh3-2*. (A) dCAPS assay for the *cenh3-1* allele digested with XbaI which cleaves wild-type but not the mutant allele. (B) dCAPS assay for the *cenh3-2* allele digested with HhaI which cleaves wild-type but not the mutant allele.

2.4.1 Heterozygous and double mutant *cenh3* displayed wild-type phenotypes

Plants were phenotypically scored against Col-0 for overall plant growth, flower and leaf morphology, silique length, and general fertility by self-pollination to assess the vigour of four haploid inducers. For the overall plant growth (Figure 2.3(A)), heterozygous *CENH3/cenh3-1,* heterozygous *CENH3/cenh3-2* and double mutant *cenh3-1/cenh3-2* lines generally had the wild-type phenotypes and did not display defects presented by *GFP-tailswap* (a *cenh3-1/cenh3-1* line complemented with the *GFP-tailswap/GFP-tailswap* transgene). However, the homozygous *cenh3-2/cenh3-2* haploid inducer appeared to be shorter in stature than the other haploid inducers at 4-weeks, and would eventually grow to approximately the same height as the others. The flowers for all the haploid inducer lines were morphologically wild-type (Figure 2.3(B)). This is also true for the leaf morphology as the rosette leaves for the haploid inducers displayed the same wild-type phenotype as Col-0 (Figure 2.3(C)).

The mean of siliques length for the heterozygote *CENH3/cenh3-1,* heterozygote *CENH3/cenh3-2* and double mutant *cenh3-1/cenh3-2* were 1.6 cm on average while the mean siliques length of homozygote *cenh3-2/cenh3-2* was 1.46 cm. One-way ANOVA was conducted followed by Tukey HSD to test for a statistically significant difference between the siliques length of haploid inducers compared to wild-type *Arabidopsis*. With the exception of siliques from homozygous *cenh3-2/cenh3-2* haploid inducer, the mean of siliques length of all other haploid inducers were not significantly different from the mean siliques length of control siliques (Figure 2.4(A)).

Figure 2.3 Assessment of haploid inducers morphological phenotypes. (A) From left to right, photograph of 4-week-old Col-0, *cenh3-1/cenh3-1 GFP-tailswap/GFP-tailswap, CENH3/cenh3-1, CENH3/cenh3-2, cenh3-1/cenh3-2* and *cenh3-2/cenh3-2* whole *Arabidopsis* plants. (B) Photographs of 4-week-old flowers from Col-0 and the four haploid inducer lines: *CENH3/cenh3-1, CENH3/cenh3-2, cenh3-1/cenh3-2* and *cenh3-2/cenh3-2*. (C) Photograph of 3-week-old rosette leaves from Col-0 and the four haploid inducer lines: *CENH3/cenh3-1, CENH3/cenh3-2, cenh3-1/cenh3-2* and *cenh3-2/cenh3-2*.

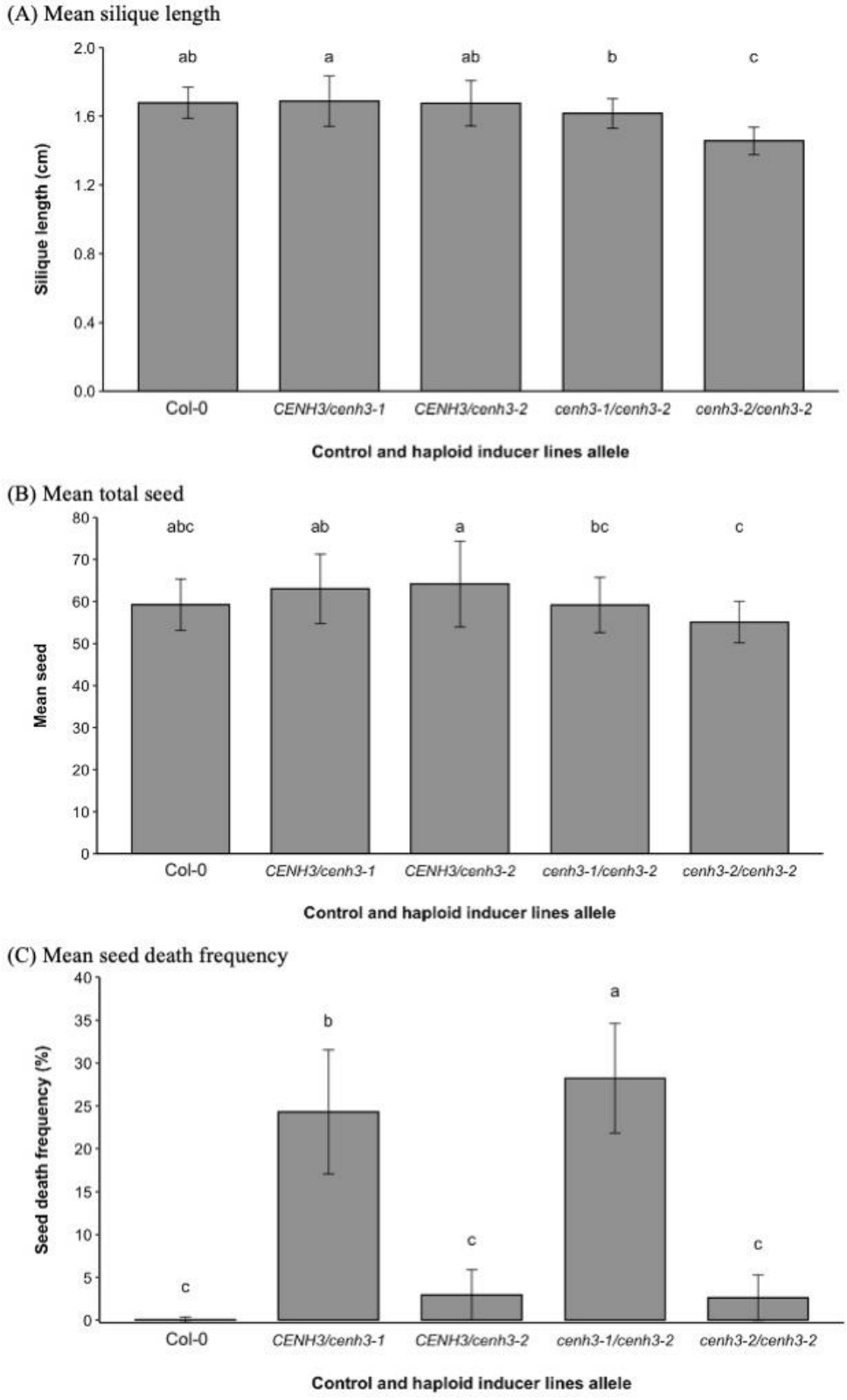


Figure 2.4 Bar plots of mean silique length (A), mean of total seeds per silique (B) and self-pollinated seed death frequencies (C) of Col-0 and haploid inducer lines: *CENH3/cenh3-1, CENH3/cenh3-2, cenh3-1/cenh3-2* and *cenh3-2/cenh3-2*. All error bars are from standard deviation. For all statistical analyses in (A), (B) and (C), one-way ANOVA was used to determine mean separation followed by post-hoc Tukey HSD. Different letters indicate significant difference with p-value < 0.05.

Mean of total seeds per silique and seed death frequency were scored to assess the fertility of the haploid inducers. We found Col-0 had 60 mean seeds per silique while haploid inducer with heterozygous *CENH3/cenh3-1* had 64 seeds per silique, haploid inducers with heterozygous *CENH3/cenh3-2* had 64 seeds per silique, haploid inducer with double mutant *cenh3-1/cenh3-2* had 59 seeds per silique and homozygote *cenh3-2/cenh3-2* had 55 seeds per silique. One-way ANOVA and post-hoc Tukey HSD analyses revealed that there were no significant differences between the mean of total seeds per silique for Col-0 and the haploid inducer lines (Figure 2.4(B)).

Next, we scored the self-pollinated seed death frequency from Col-0 and all the haploid inducer lines. Haploid inducer with double mutant *cenh3-1/cenh3-2* had the highest mean of self-pollinated seed death frequency at 28.35% followed by haploid inducer with heterozygous *CENH3/cenh3-1* at 24.07% while Col-0, *CENH3/cenh3-2* and *cenh3-2/cenh3-2 Arabidopsis* had a mean of self-pollinated seed death frequency at 0.05%, 2.96% and 2.62% respectively (Figure 2.4(C)). One-way ANOVA and post-hoc Tukey HSD revealed that means of self-pollinated seed death frequencies of haploid inducers heterozygous with *CENH3*/*cenh3-1* and double mutant *cenh3-1/cenh3-2* were significantly different than Col-0. Self-pollinated seed death frequency from haploid inducers with heterozygous *CENH3/cenh3-2* and homozygous for *cenh3-2/cenh3-2* were not significantly different than Col-0. Statistical analyses also revealed that the self-pollinated seed death frequency of heterozygous *CENH3/cenh3-1* and double mutant *cenh3-1/cenh3-2* were significantly different than each other. Therefore, we tested the seed death frequency of haploid inducers heterozygous with *CENH3/cenh3-1* and double mutant *cenh3-1/cenh3-2* using chi-squared test to determine if they were equal to the Mendelian frequency and found that the seed death frequencies did not deviate from 25%.

2.4.2 Dead seed frequency from genome elimination crosses

When crossed as a female, double mutant *cenh3-1/cenh3-2* haploid inducer showed the highest seed death frequency at 36.79% followed by *cenh3-2/cenh3-2* homozygous mutant at 15.56%, *CENH3/cenh3-2* heterozygote at 5.69% and *CENH3/cenh3-1* heterozygote at 4.32% (Table 2.2). When crossed as a male, we found that *cenh3-2/cenh3-2* homozygous haploid inducer had the highest seed death frequency at 6.07%, followed by *cenh3-1/cenh3*-2 double mutant at 3.06%, C*ENH3/cenh3-2* heterozygote at 0.48% and *CENH3/cenh3-1* heterozygote at 0.34% respectively. We performed mean separation using two-way ANOVA followed by Tukey HSD to compare the effects of haploid inducers allele and the direction of crossing on seed death frequencies. Haploid inducers with double mutant *cenh3-1/cenh3-3* and homozygous *cenh3-2/cenh3-2* were found to be statistically significant than the control Col-0 while both haploid inducers with heterozygous *CENH3/cenh3-1* and *CENH3/cenh3-2* were not (Figure 2.5(B)). The statistical analyses using two-way ANOVA followed by Tukey HSD also revealed that collectively, the seed death frequencies when the haploid inducers were crossed as females was consistently higher ($p < 0.001$) than the male haploid inducers. In the presence of the direction of crossing, the seed death frequency for the double mutant *cenh3-1/cenh3-2* haploid inducer when crossed as a female was significantly different than all of the haploid inducers regardless of the direction of crossing (Figure 2.5(B)). We also observed that haploid inducer with double mutant *cenh3-1/cenh3-2* had the highest variation in seed death frequency when crossed as a female.

(A) Viable and dead seeds

(B) Seed death frequency of *Arabidopsis* with wild type *CENH3* when reciprocally crossed to haploid inducer with variant alleles

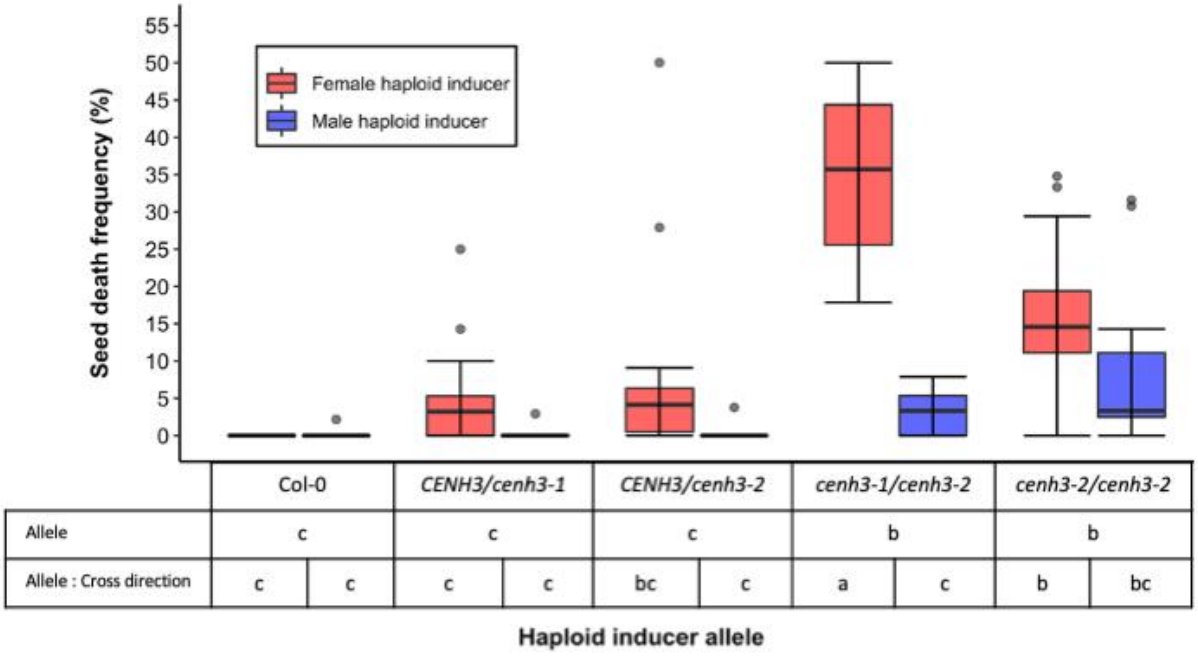


Figure 2.5 Photographs of seeds and seed death frequency from reciprocal genome elimination crosses. (A) The three panels showing viable seeds on top, and dead seeds on the bottom outlined within boxes with dotted lines. (B) Boxplot of seed death frequencies per silique of reciprocal crosses from Col-0 and haploid inducer lines crossed to L*er gl1*. Red fill indicates the line was crossed as female and blue fill denotes the line was used as male. Two-way ANOVA was used to analyze mean separation followed by post-hoc Tukey HSD. Different letters in the table below the boxplot indicate significant differences with p-value < 0.05.

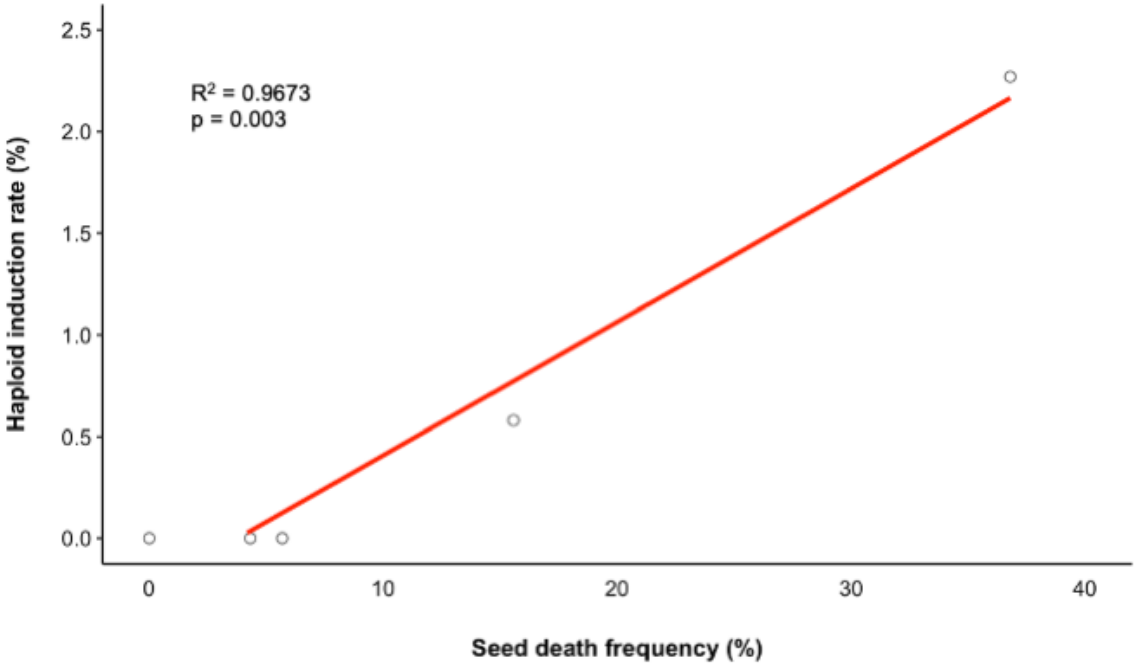
2.4.3 Haploid induction rate of haploid inducers with *cenh3-1* and *cenh3-2* mutant alleles

We evaluated the ploidy of F1 progenies derived from the reciprocal genome elimination crosses between Col-0 and the four haploid inducer lines with L*er gl1* to determine the haploid induction rate (HIR) (Table 2.2). When crossed as a female, haploid inducer with double mutant *cenh3-1/cenh3-2* had the highest HIR at 2.27% followed by homozygote *cenh3-2/cenh3-2* at 0.58%. We did not detect haploids when heterozygous *CENH3/cenh3-1* or heterozygous *CENH3/cenh3-2* were crossed as females. When crossed as a male, homozygote *cenh3-2/cenh3-2* had the highest HIR at 2.58% followed by the *CENH3/cenh3-2* heterozygote at 0.26%. We did not detect haploids when *CENH3/cenh3-1* heterozygote or *cenh3-1/cenh3-2* double mutant when crossed as males. Aneuploid progenies were observed in all progenies of reciprocal genome elimination crosses, with the highest aneuploids detected in progenies of female *CENH3/cenh3-1* haploid inducer.

Next, we fitted the HIR and seed death frequency of all haploid inducers on a linear regression model to determine if there was a correlation between our seed death frequency and the HIR. We observed that when haploid inducers were crossed as females, the seed death frequency and HIR were fitted in the linear regression model with R^2 at 0.9673 and a p-value of 0.003 (Figure 2.7(A)). When crossed as males, the haploid inducers were fitted in the linear regression model with R^2 at 0.7576 although less correlated with the p-value at borderline 0.055 (Figure 2.7(B)).

Figure 2.6 Haploid progeny from genome elimination cross of male homozygous *cenh3-2/cenh3-2* haploid inducer to the wild-type *Arabidopsis*. The three panels above display the haploid progeny on the left and the diploid parent on the right. Haploids phenocopy the L*er-gl1* diploid parent of which the leaves were without trichomes (smooth leaves surface).

(A) Linear regression model between seed death frequency and HIR of female haploid inducer



(B) Linear regression model between seed death frequency and HIR of female haploid inducer

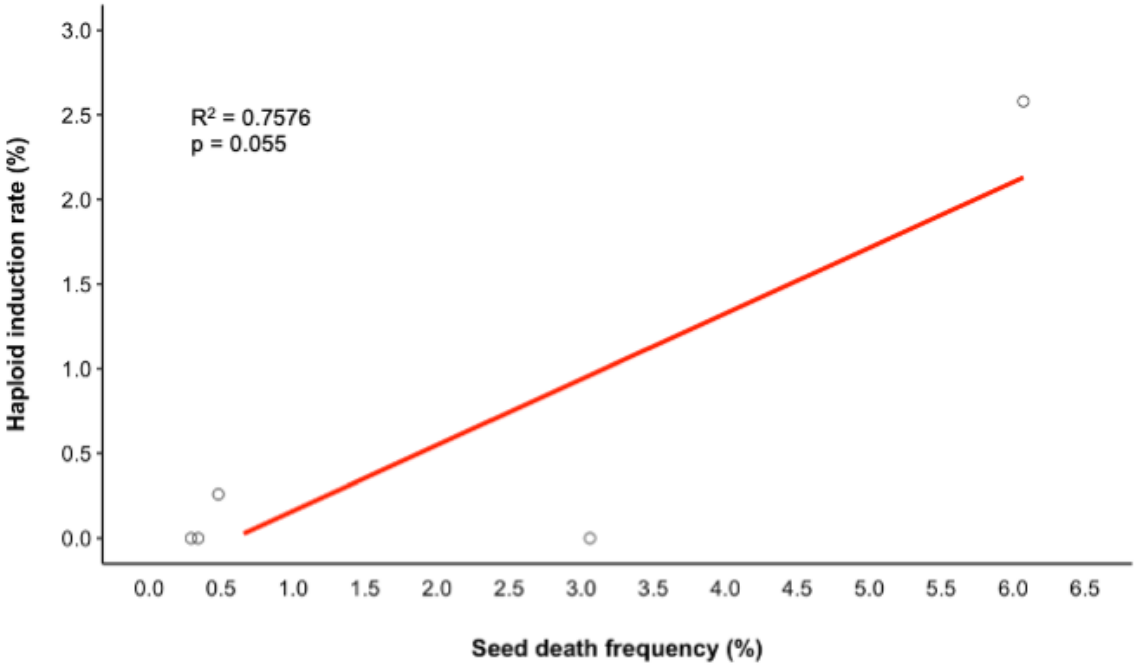


Figure 2.7 Regression between seed death frequency and haploid induction rate of haploid inducers used in this study. (A) Red line on the graph represents the regression between seed death frequency and HIR when haploid inducers were crossed as a female in genome elimination crosses and (B) when haploid inducers were crossed as a male in genome elimination crosses. R^2 signifies the goodness of fit while p-value < 0.05 indicates statistically significant correlation between seed death frequency and HIR of the haploid inducers.

Table 2.4 Seed death evaluation, ploidy evaluation and haploid induction rate from reciprocal genome elimination crosses of Col-0 and haploid inducers to L*er gl1*.

Crosses	Seed Death Evaluation		Ploidy Evaluation				HIR (%)
	No. seeds	Seed death frequency (%)	No. seedlings	Haploid	Diploid	Aneuploid	
Col-0 ♀ x L*er gl1* ♂	151	0	173	0 (0%)	173 (100%)	0 (0%)	0
CENH3/cenh3-1 ♀ x L*er gl1* ♂	509	4.32	433	0 (0%)	380 (88%)	53 (12%)	0
CENH3/cenh3-2 ♀ x L*er gl1* ♂	439	5.69	399	0 (0%)	392 (98%)	7 (2%)	0
cenh3-1/cenh3-2 ♀ x L*er gl1* ♂	443	36.79	264	6 (2%)	247 (94%)	10 (4%)	2.27
cenh3-2/cenh3-2 ♀ x L*er gl1* ♂	1189	15.56	516	3 (1%)	486 (94%)	27 (5%)	0.58
L*er gl1* ♀ x Col-0 ♂	343	0.29	208	0 (0%)	208 (100%)	0 (0%)	0
L*er gl1* ♀ x *CENH3/cenh3-1* ♂	291	0.34	274	0 (0%)	257 (94%)	17 (6%)	0
L*er gl1* ♀ x *CENH3/cenh3-2* ♂	418	0.48	378	1 (1%)	368 (97%)	9 (2%)	0.26
L*er gl1* ♀ x *cenh3*-1/*cenh3-2* ♂	720	3.06	532	0 (0%)	523 (98%)	9 (2%)	0
L*er gl1* ♀ x *cenh3-2/cenh3-2* ♂	428	6.07	233	6 (3%)	220 (94%)	7 (3%)	2.58

2.5 Discussion

CENH3 can be found across plants and therefore we were invested to determine the efficacy of *cenh3* mutant alleles in inducing haploids in centromere-mediated genome elimination crosses (Pal & Negi, 2019). Previous works had extensively investigated variant CENH3 constructs that were used to complement the homozygous null mutant allele *cenh3-1/cenh3-1* in *Arabidopsis* (Maheshwari et al., 2015; Ravi & Chan, 2010). However, haploid inducers complemented by *CENH3* variants such as *GFP-tailswap* or *CENH3* from other plant species were inundated by morphological defects such as smaller and stunted plants (Maheshwari et al., 2015; Ravi & Chan, 2010). Haploid inducer with *cenh3-1/cenh3-1* complemented by *GFP-tailswap/GFP-tailswap* also exhibit male sterility and can only be crossed as a female for the most part. More recently, Wang et al. (2021) demonstrated that haploid inducer heterozygous with null mutant *cenh3* in *Zea mays* (maize) was able to induce haploids and the heterozygous haploid inducer had wild-type phenotypes. These features allowed the heterozygous null mutant *cenh3* haploid inducer to be reciprocally crossed either as female or male. Therefore, we were interested to test the efficacy of heterozygous *cenh3* mutant alleles in *Arabidopsis* in inducing haploids, and assess their ability to be crossed as a female or a male. We generated four haploid inducer lines in this study: heterozygous *CENH3/cenh3-1*, heterozygous *CENH3/cenh3-2*, double mutant *cenh3-1/cenh3-2* and homozygous *cenh3-2/cenh3-2*.

We initially hypothesized that the haploid inducer with double mutant *cenh3-1/cenh3-2* would lack vigour because it had the most severe combination of loss of function alleles followed by haploid inducer with homozygous *cenh3-2/cenh3-2*. We also expected the heterozygous *CENH3/cenh3-1* and heterozygous *CENH3/cenh3-2* haploid inducers to have wild-type phenotypes. Contrary to our expectation, the haploid inducer with double mutant *cenh3-1/cenh3-2* had the wild-type phenotypes while the homozygous *cenh3-2/cenh3-2* appeared to

be lagging in its growth (Figure 2.3(A)). Through statistical analyses, we found that the siliques length of haploid inducers with heterozygous *CENH3/cen3-1,* heterozygous *CENH3/cenh3-2* and double mutant *cenh3-1/cenh3-2* were wild-type when compared to the Col-0 while homozygote *cenh3-2/cenh3-2* siliques length were significantly shorter than the Col-0. We also found that all the haploid inducers had wild-type phenotype for the leaf and flower phenotypes. These observations suggested that heterozygous *CENH3/cenh3-1,* heterozygous *CENH3/cenh3-2* and double mutant *cenh3-1/cenh3-2 Arabidopsis* were vigorous haploid inducers while homozygous *cenh3-2/cenh3-2* haploid inducer was generally vigorous with a slight delayed growth rate.

Haploid inducer with *cenh3-1/cenh3-1* rescued by *GFP-tailswap/GFP-tailswap* is the most efficient haploid inducer to date but is mostly male sterile (Ravi & Chan, 2010). Therefore, we wanted to establish if any of our haploid inducers were affected by sterility issues. We found the mean self-pollinated seeds per silique for haploid inducers with heterozygous *CENH3/cenh3-1,* heterozygous *CENH3/cenh3-2* and double mutant *cenh3-1/cenh-3* to be phenotypically wild-type when compared to the mean seeds per silique of the control Col-0. In the case of the *cenh3-2/cenh3-2* homozygote, we found that mean silique length as well as mean total seeds per silique appear to be significantly lower compared to Col-0. The data here suggests that the partial loss of function from the *cenh3-2* missense mutation does appear to incur a fitness cost when grown under similar condition. However, this effect was not detected from the double mutant *cenh3-1/cenh3-2* which only carries a single copy of the *cenh3-2* missense allele and should phenocopy the *cenh3-2/cenh3-2* homozygote. There are two possible explanations: one is that the *cenh3-2* allele is partially neomorphic, in which it is dosage sensitive and behaves like a dominant negative allele when present at higher copy numbers, or second: a secondary mutation on a closely linked gene in the *cenh3-2* mutant

background is implicated. Additional experiments will have to be performed to distinguish the effects of these *cenh3* mutant alleles.

The rated seed death frequency of self-pollinated haploid inducers did not suggest sterility issues from the haploid inducer lines tested. For haploid inducers with heterozygous *CENH3/cenh3-1,* heterozygous *CENH3/cenh3-2* and double mutant *cenh3-1/cenh-2*: the fusion of gametes with different *CENH3* alleles could render dimorphism between the two parental centromere. The parent with "weaker" centromere could in effect be the haploid inducer. This could lead to haploid induction or genome instability during the first post-zygotic mitosis resulting in seed death. Hence, the low level of seed death frequency in haploid inducer with homozygous *cenh3-2/cenh3-2* was also expected as female and male gametes had the same allele. Besides centromeric dimorphism due to different *CENH3* alleles in the fusing gametes, the self-pollinating haploid inducers with heterozygous *CENH3/cenh3-1* or double mutant *cenh3-1/cenh3-2* had one copy of the *cenh3-1* allele, which is homozygous embryo-lethal. Therefore, we expected 25% of seed death based on the Mendelian frequency in the self-pollination of haploid inducers with heterozygous *CENH3/cenh3-1* or double mutant *cenh3-1/cenh3-2*. This was evident as the seed death frequency of our haploid inducers with heterozygous *CENH3/cenh3-1* or double mutant *cenh3-1/cenh3-2* did not deviate from the expected 25% when we performed a chi-squared test. However, we noticed that the self-pollinated seed death frequency between heterozygous *CENH3/cenh3-1* and double mutant *cenh3-1/cenh3-2* were also significantly different than each other, suggesting some influence from the *cenh3-2* allele in the haploid inducer. Incidentally, the seed death frequencies of self-pollinated haploid inducers with heterozygous *CENH3/cenh3-1* or double mutant *cenh3-1/cenh3-2* were significantly different than the control Col-0, heterozygote *CENH3/cenh3-2* and homozygote *cenh3-2/cenh3-2*.

Seed death frequency can be used as a proxy for haploid induction rate (HIR) and therefore can be a good indicator to predict the haploid induction rate at seeds level (Kuppu et al., 2020). The development of the endosperm, one of the two products of double fertilization in angiosperms such as *Arabidopsis* is critical for the fate of a seed. Incorrect genome dosage in the endosperm due to genome elimination would be catastrophic for the endosperm and thus may not be able to sustain the growth of the embryo resulting in seed death (Johnston et al., 1980). Hence, we predicted that the double mutant *cenh3-1/cenh3-2* haploid inducer would be the most efficient haploid inducer due to the combination of its loss of function *cenh3* alleles and therefore would illicit the highest seed death frequency. We also deduced that haploid inducer homozygous with *cenh3-2/cenh-2* would be second most efficient followed by heterozygotes *CENH3/cenh3-1* and lastly haploid inducer heterozygous with *CENH3/cenh3-2*. Our prediction was proven to be true when the haploid inducers were crossed as female. However, when the haploid inducers were crossed as male, homozygote *cenh3-2/cenh3-2* induced higher seed death frequency than double mutant *cenh3-1/cenh3-2* haploid inducer.

Collectively, all female haploid inducers produced higher and significantly different seed death frequency than their male counterparts. We found that not only double mutant *cenh3-1/cenh3-2* haploid inducer had the highest mean of seed death frequency when crossed as a female but was also significantly different from all of the other female or male haploid inducers. However, double mutant *cenh3-1/cenh3-2* haploid inducer was also most inconsistent in inducing haploids as the seed death frequency for each siliques were spread out more than the others.

The seed death frequency indicated that double mutant *cenh3-1/cenh3-2* would be one of the most efficient haploid inducers when crossed either as a female or a male. The seedlings ploidy assessment revealed that this was indeed true when the double mutant *cenh3-1/cenh3-2* haploid

inducer was crossed as a female but not as a male. Surprisingly, male double mutant *cenh3-1/cenh3-2* haploid inducer failed to produce any haploids at all. In fact, male homozygous *cenh3-2/cenh3-2* haploid inducer out-performed all of the other haploid inducers whether they were crossed as a female or a male. Taken together with data analyses from mean of total seeds per silique and seed death frequency of self-pollinated haploid inducers, *cenh3-2* allele seemed to have some influence on the pollen of homozygous *cenh3-2/cenh3-2* haploid inducer which increased the capability of *cenh3-2/cenh3-2* homozygote as a haploid inducer when crossed as a male. Further experiment should be conducted to investigate the effect of *cenh3-2* allele on the pollens of *Arabidopsis*.

Seed death frequencies were able to predict the HIRs of the female haploid inducers although it was less predictive for the HIRs of the male haploid inducers. It is possible that not enough progeny were screened to report haploids, but the occurrence of aneuploids from all haploid inducers tested was indicative of centromeric dysfunction that is associated with centromere-mediated genome elimination. We noted that the HIR for the *cenh3-2/cenh3-2* homozygous mutant was lower than what was reported (Kuppu et al. (2015), which may be caused by the environmental conditions of our lab. Our male seed death frequency data through the genome elimination crosses, HIR of the male haploid inducers and the efficacy of haploid inducer homozygous with *cenh3-2/cenh3-2* in inducing haploids seem to imply that *cenh3-2* allele may affect the male gametophyte and therefore influenced the haploid induction trough male haploid inducers. Nevertheless, we could see a theme where the combination of the most severe form of *cenh3* mutant alleles would be the most efficient haploid inducer while still maintaining the haploid inducer vigour.

2.6 Conclusion

The trade-off between haploid inducer plant vigour and its efficacy to induce haploids is one of the roadblocks in developing efficient haploid induction for plant breeding technology. Our experiment demonstrated that double mutant *cenh3-1/cenh3-2* haploid inducer was as vigorous as wild-type Col-0 while able to induce haploids up to 2.27% when crossed as a female haploid inducer. However, the inconsistency of our actual haploid induction rate (HIR) with the predicted HIR ranking through the severity of *cenh3* alleles, as well as seed death scoring when the haploid inducers were crossed as a male warrants the phenomenon to be further investigated. In doing so, not only we could improve the requirements in selecting efficient alleles to induce haploids, but we could also improve the efficiency of male haploids inducers as per our objective in employing reciprocal crosses.

CHAPTER 3

ISOLATION OF EARLY *ARABIDOPSIS* EMBRYO UNDERGOING CENTROMERE-MEDIATED GENOME ELIMINATION

3.1 Abstract

The advancement in haploid induction technology is desirable in plant breeding because homozygous lines created by doubled haploids effectively bypass several generations of inbreeding. Centromere-mediated genome elimination has the ability to induce haploids up to 45%, far surpassing the efficacy of inducing haploids compared to other methods such as the phospholipase-mediated haploid inducer. However, the cellular response of embryos undergoing centromere-mediated genome elimination still has not been characterized. Therefore, we seek to profile the gene expression of embryos undergoing genome elimination by establishing a protocol to extract early embryos from genome elimination and non-genome elimination crosses for gene expression profiling experiments. Comparing the expression profiles of embryos undergoing genome elimination and non-elimination would enable us to describe pathways that lead to centromere-mediated genome elimination. Although we laid the groundwork for extracting embryos, the current method described here did not yield sufficient material in both sets of crosses to enable the extraction of low-input RNA for expression profiling. Improvements can be made to the current embryo isolation methods, which can also be adapted for other haploid induction systems from other plant species.

3.2 Introduction

Haploid plants are desirable as doubled haploids create instantaneous homozygous individuals (Chase, 1969; Chase, 1949). Haploids can be induced through cultured gametophyte cells although most plants are resistant to this method (Forster et al., 2007). Another method to

induce haploid is through postzygotic genome elimination. The route of haploid induction is often referred to as uniparental genome elimination and occurs after the fusion of egg cell and sperm, leaving only one parental genome. In monocot crop plants such as *Zea mays* (maize), a distinct pathway mediated through the *MATRILINEAL* (*MTL)* gene has been found to be the cause of haploid induction through the traditional Stock 6 haploid inducer line (Gilles et al., 2017; Kelliher et al., 2017; Liu et al., 2017; S. Wang et al., 2019; Yao et al., 2018). This mechanism appears to be distinct from centromere-mediated genome elimination, and is more closely allied with the parthenogenic process.

Haploids induced via crossing were thought to occur in rare interspecific crosses (*Nicotiana* spp., *Hordeum* spp., wheat x maize and maize x oat) in the wild (Clausen & Mann, 1924; Laurie & Bennett, 1988; Rines & Dahleen, 1990; Subrahmanyam & Kasha, 1973). Studies show that centromeric dimorphism between the two interspecific parents almost always results in genome elimination of the parent with the "smaller" or "weaker" centromeres (Ishii et al., 2016; Wang & Dawe, 2018). It was not until 2010 when the study published by Ravi & Chan (2010) demonstrated that intraspecific crosses of *Arabidopsis* can induce haploids as well. As with interspecific crosses, the centromere plays a prominent role in genome elimination specifically through the manipulation of CENH3 in haploid inducers.

Since the discovery of the role of altered CENH3 and its contribution to mediate intraspecific haploid induction through genome elimination (Britt & Kuppu, 2016; Karimi-Ashtiyani et al., 2015; Kuppu et al., 2020; Maheshwari et al., 2015), much efforts have been made to improve this method for haploid induction. CENH3 is a histone H3 variant and consists of a conserved Histone Fold Domain (HFD) and a highly variable N-terminal tail, and when loaded on nucleosomes plays an integral role in epigenetically determine the centromeric region. This is

critical for chromosomes segregation as CENH3-containing nucleosomes bind to kinetochore complexes which attach to the microtubules (Darlington, 1939; Gairdner & Darlington, 1931; Malik & Henikoff, 2009; Xiaoyan et al., 2019).

In *Arabidopsis,* when native CENH3 were completely replaced by GFP-tailswap and crossed to wild-type, up to 32% haploids were induced. The embryo lethal null mutation *cenh3-1* allows complete replacement of native CENH3 with transgenic variant is the result of single nucleotide change of G-to-A at 161 relatives to ATG = +1 of *AT1G01370,* causing the first exon to be incorrectly spliced to the middle of the first intron. The mutation resulted in frameshift that encoded stop codon prematurely after just 46 amino acids instead of 178 amino acids. GFP-tailswap is a transgenic construct containing green fluorescent protein-tagged (GFP) with its hypervariable amino-terminal domain of CENH replaced by the H3.3 variant while GFP-CENH3 has GFP tagged at the tail-end of the wild-type CENH3. As altered CENH3 leads to the defective centromere and fully-functional centromere is central to successful chromosomes segregation, the varying degree of haploids induced with different CENH3 alterations are due to the defective centromeres which are considered unequal on the mitotic plate when crossed to wild-type CENH3. When *GFP-CENH3* plants were crossed to another *GFP-CENH3,* no haploids nor aneuploids were produced despite containing "defective" centromeres (Ravi & Chan, 2010). Zero production of haploids were also observed in *GFP-tailswap* when crossed to another *GFP-tailswap* plants, although aneuploids accounted up to 8% in these crosses. These suggest two parental genomes with "defective" centromeres compete equally on mitotic plate and further support that the manipulation of CENH3 caused the genome to be "weak" when crossed to parent with "normal" genome. However, the cellular and molecular mechanism that led to centromere-mediated genome elimination is still under investigation.

A study by Marimuthu et al. (2021) showed that both wild-type CENH3s as well as the GFP-tailswap variant were preferably loaded to the parental genome that carried the wild-type allele just before the first mitosis after fertilization. Interestingly, the parental genome of the haploid inducer that carries the mutant allele can persist and the loading of the wild-type CENH3s and the GFP-tailswap only recovered between 2 – 6 days after pollination (DAP). Therefore, genome elimination in *Arabidopsis* is thought to occur during the pre-globular stage of embryogenesis. Gene profiling on early pre-globular embryos undergoing genome elimination will give insight into the pathway that leads to genome elimination and therefore for the "weak" parental genome to be eliminated. In the animal system, cells that are treated with drugs that prevent the proper formation of mitotic microtubules lead to genome instability giving rise to chromothripsis, which is often associated with cancer and other genetic diseases (Forment et al., 2012; Kloosterman et al., 2014; Korbel & Campbell, 2013). As genome instability often shares common pathways, this work may also provide insight on chromothripsis and the associated genome instability observed during genome elimination, which has been shown to lead to karyotypic change (Comai & Tan, 2019). Therefore, this study seeks to profile the gene expression of *Arabidopsis* embryos during centromere-mediated genome elimination and will bring us closer to dissecting the genetic pathway that governs centromere-mediated uniparental genome elimination in *Arabidopsis*.

3.3 Materials and methods

3.3.1 Plants materials and growth conditions

Arabidopsis ecotypes Landsberg *erecta gl1* (L*er gl1*) and Columbia (Col-0) were selected as plant materials to profile the expression of pooled embryos undergoing genome elimination and compare it to those in a non-genome elimination cross. L*er gl1* has wild-type

CENH3/CENH3 while the haploid inducer was derived from Col-0 with embryo-lethal null mutant *cenh3-1/ cenh3-1* was rescued by transgenic *GFP-tailswap/GFP-tailswap*.

Seeds were sown in PRO-MIX BX, an all-purpose soil mix containing peat moss, peat hummus, perlite and limestone. The seeds were stratified for three days in a dark 4°C fridge to induce germination before being grown under controlled environment in a growth chamber. The growth chamber environment was set to 16 hours of light/8 hours of dark at 21°C with relative humidity of 65%. Within the first week of inflorescence emergence, control and genome elimination crosses were carried out.

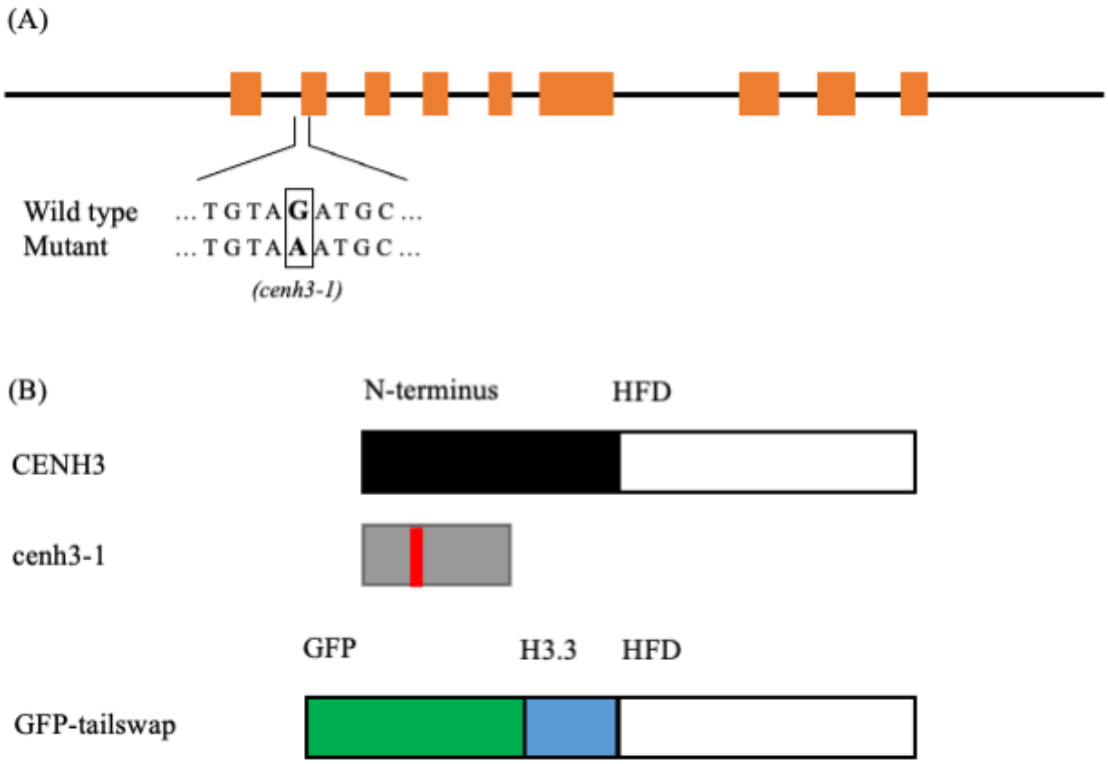


Figure 3.4 Cartoon of *CENH3* and mutant allele *cenh3-1* sequences and resulting CENH3 protein used in this study. All cartoons are not to scale. (A) The sequences in *Arabidopsis* wild-type *CENH3* and its embryo-lethal null mutant *cenh3-1* alleles. Orange boxes denoted exons. (B) The protein structure of wild-type CENH3 and the GFP-tailswap variant. The product from the *cenh3-1* allele makes a non-functional protein.

3.3.2 Genomic DNA extraction and genotyping for *cenh3-1* mutant allele

One small *Arabidopsis* leaf was submerged in 125µl quick DNA extraction buffer (200 mM Tris HCl pH 8.0, 250 mM NaCl, 25 mM EDTA pH 8.0, 0.5% SDS) to extract the genomic DNA and incubated at 99°C for 10 minutes. 100 µl of isopropanol was added to the 100 µl of extracted supernatant and incubated for 15 minutes under room temperature to precipitate the DNA. To pellet the DNA, the mixture was centrifuged for 15 minutes upon which the supernatant was aspirated. 200 µl 70% ethanol was used to wash the pellet and the supernatant was removed again before the pellet was left to dry at room temperature for 15 minutes. 50 µl Tris-EDTA (TE) buffer (10 mM Tris HCl pH 8.0, 0.5 mM EDTA pH 8.0) was added to the pellet and incubated at 55°C to elute the DNA.

Polymerase chain reaction (PCR) was employed to amplify extracted genomic DNA. Each reaction contained 15 µl volume of 7.5 µl Promega's 2X GoTaq® G2 Green Master Mix, 1 µl of 10% forward and reverse primers, 5.5 µl nuclease-free H_20 and 1 µl of DNA per PCR reaction were used. The forward and reverse primers were specially designed derived Cleaved Amplification Polymorphic Sequences (dCAPS) primer sets (Table 3.1). *cenh3-1* dCAPS XbaI forward and reverse primer set targeted 60 bp upstream from the *cenh3-1* SNP and 148 bp downstream. The PCR reactions were carried out in the BIO-RAD T100^{TM} thermal cycler programmed to 5 minutes denaturation at 95°C; 40 cycles of 15 seconds denaturation at 95°C, 15 seconds of primer annealing at 65°C and 30 seconds of primer extension at 55°C. Next, the amplified DNA was incubated overnight at 37°C for the XbaI digestion. This enzyme cleaved the designated sequence in the wildtype allele at 5'…T^CTAGA…3' for XbaI while leaving the mutant allele intact. The final PCR product was then resolved in a 2% agarose gel, stained in ethidium bromide and visualized under UV illumination.

Table 3.3 Forward and reverse dCAPS primers design and Xbal digestion site for *cenh3-1* allele. The carrot sign shows the digestion site of the enzyme.

Allele	Primer	Forward Sequence	Annealing Temperature	Enzyme Digestion Site
cenh3-1	*cenh3-1* XbaI dCAPS forward	AGAATTTTAGGTTTTTTATTTC GATTTTGTTAACCCTAGATTT CGAATCTGAAATTTcTA	66.0°C	5'…T^CTAGA …3' 3'…AGATC^T …5'
	cenh3-1 XbaI dCAPS reverse	GCCTCTCCTTGTCGGGGTCTT CA		

3.3.3 Genome elimination cross

Genome elimination cross was conducted with *cenh3-1/cenh3-1 GFP-tailswap/ GFP-tailswap* plant as the female haploid inducer crossed to the male wild-type L*er-gl1*. No emasculation was needed for genome elimination cross as the *cenh3-1/cenh3-1 GFP-tailswap/ GFP-tailswap* plant is male sterile (Ravi & Chan, 2011). For the control cross (non-genome elimination cross), female *CENH3/CENH3 GFP-tailswap/GFP-tailswap* was crossed to the wild-type male L*er-gl1*. Altered *CENH3* such as *GFP-tailswap* is recessive and plant expressing both the variant and wild-type *CENH3* have wild-type phenotype. For the non-genome elimination crosses, unopen *Arabidopsis* buds were emasculated under a magnifier and pollinated immediately to prevent contamination from self-sterilization. At least 40 pollinations were conducted for each session of embryo extraction.

Table 3.4 Control and genome elimination crosses. *Arabidopsis* plants and their corresponding alleles used in the control and genome elimination crosses. Col-0 with *CENH3/CENH3 GFP/tailswap/GFP-tailswap* exhibited wild-type phenotype while *cenh3-1/cenh3-1 GFP-tailswap/ GFP-tailswap* is male sterile.

Type of crosses	Crosses
Non-genome elimination cross	Col-0 (*CENH3/CENH3 GFP/tailswap/GFP-tailswap*) ♀ x L*er-gl1* (*CENH3/CENH3*) ♂
Genome elimination cross	Col-0 (*cenh3-1/cenh3-1 GFP/tailswap/GFP-tailswap*) ♀ x L*er-gl1* (*CENH3/CENH3*) ♂

3.3.4 Embryo extraction

Embryos that were undergoing genome elimination as well as from non-genome elimination crosses were isolated for RNA extraction. Embryos were extracted at 48 hours after pollination (HAP) for both genome elimination and control crosses as the first mitotic division is at 26 – 36 HAP (Lermontova et al., 2006). A protocol developed by Raissig et al. (2013) was adapted to rapidly isolate, screen, wash and elute *Arabidopsis* embryos. For each session of embryos isolation, seeds were removed from 20 – 30 siliques using an insulin needle and were immersed in 20 μl Tris-EDTA (TE) buffer (10 mM Tris HCl pH 8.0, 0.5 mM EDTA pH 8.0). Seeds were then gently crushed with a plastic pestle to release the embryos. Additional 300 μl TE buffer was used to rinse and dilute the sample which then was filtered using a 30 μm nylon mesh. Embryos were screened with an inverted microscope, isolated using a micromanipulator and washed in Invitrogen's RNAlater Solution.

3.4 Results

We performed genome elimination and non-elimination crosses (Table 3.2) for the experiment. Plants with *CENH3/CENH3 GFP-tailswap/GFP-tailswap* were genotyped and crossed to the wild-type L*er gl1* in the non-elimination cross while plants with *cenh3-1/cenh3-1 GFP-tailswap/GFP-tailswap* were genotyped and crossed to the wild-type L*er gl1* in the genome elimination cross. Figure 3.2 shows the genotyping results conducted on Col-o, *Ler* gl1, *Arabidopsis* with *CENH3/CENH3 GFP/tailswap/GFP-tailswap* and *cenh3-1/cenh3-1 GFP/tailswap/GFP-tailswap.* XbaI restriction enzyme is designed to cleave the wild-type sequence and mutant *cenh3-1* allele is resistant to cleavage. Col-0, L*er gl1* and *Arabidopsis* with *CENH3/CENH3 GFP/tailswap/GFP-tailswap* were identified by the cleaved product of 148 bp. *Arabidopsis* with homozygous mutant *cenh3-1/cenh3-1 GFP/tailswap/GFP-tailswap* was identified by the undigested product of 208 bp in the presence XbaI enzyme.

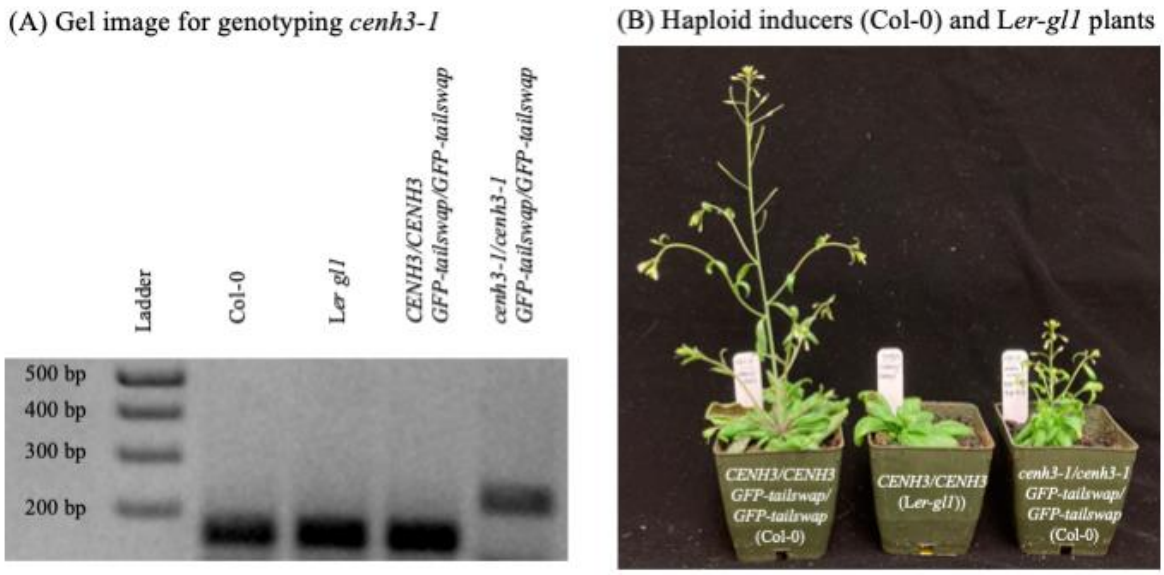


Figure 3.5 Electrophoretic gel image of haploid inducer genotype. (A) Electrophoretic gel image to genotype the *Arabidopsis* used in control and genome elimination crosses. From left to right: control for XbaI digestion, L*er gl1*, *Arabidopsis* with *CENH3/CENH3 GFP-tailswap/GFP-tailswap* and *cenh3-1/cenh3-1 GFP-tailswap/GFP-tailswap*. (B) Picture showing *CENH3/CENH3 GFP-tailswap/GFP-tailswap*, wild-type L*er gl1* and *cenh3-1/cenh3-1 GFP-tailswap/GFP-tailswap Arabidopsis* used in control and genome elimination crosses.

3.4.1 Isolated embryos

Embryos from both the control and genome elimination crosses were extracted for RNA extraction. Figure 3.3 is showing the pre-processed embryos at the scale of 200 μm and 400 μm as pointed out by the red arrows, extracted 48 hours after pollination (HAP). However, we were unable to extricate sufficient pre-globular *Arabidopsis* embryos in both the control and genome elimination crosses through the adapted embryo extraction method. The amount of materials we collected was insufficient to enable us to extract sufficient low-input RNA for the expression profiling of early *Arabidopsis* embryos undergoing genome elimination.

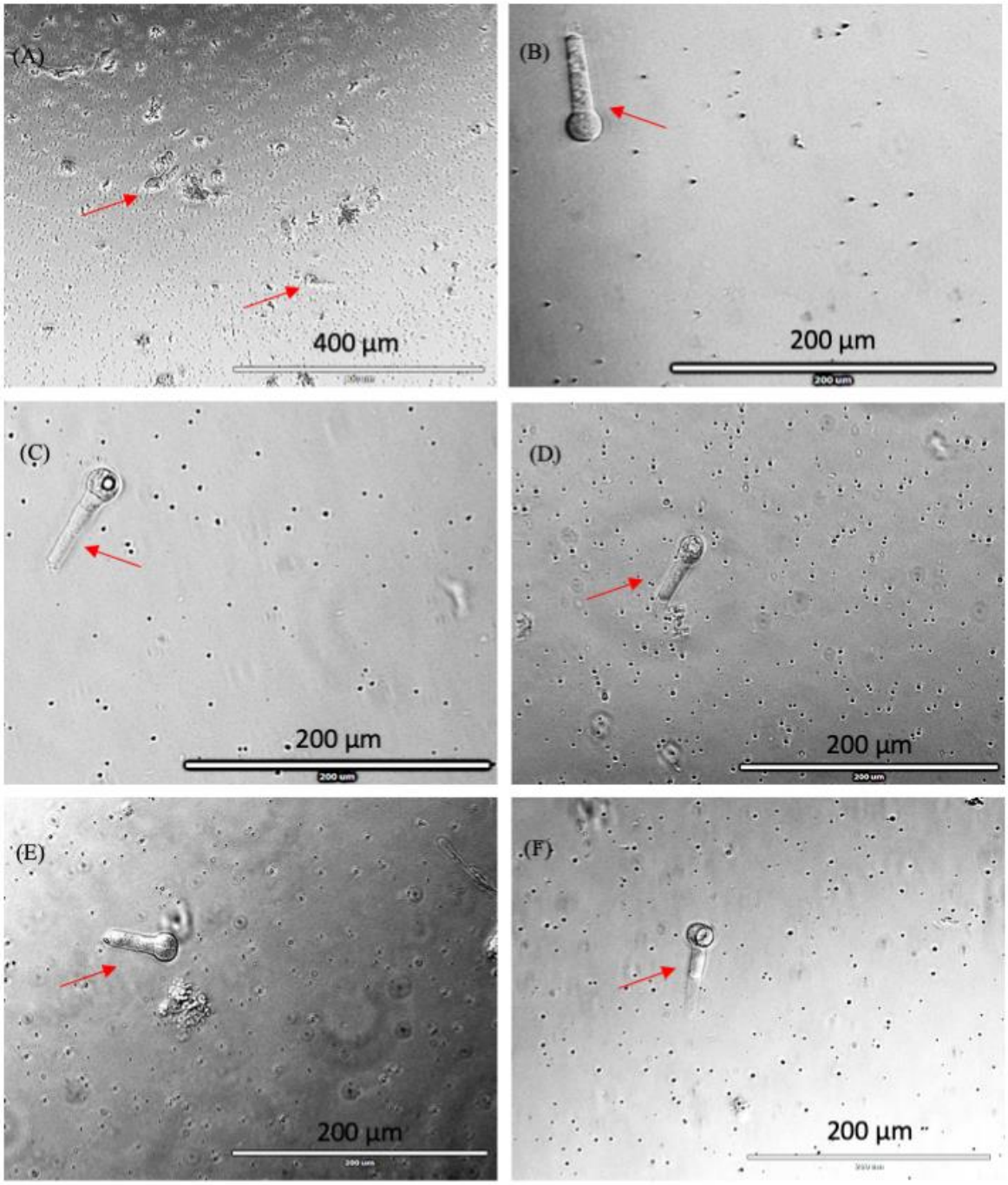


Figure 3.6 Pre-processed early *Arabidopsis* embryos at 48 HAP. The embryos pointed by red arrows, are to scale (A) at 400 µm (B – F) and at 200 µm.

3.5 Discussion and conclusion

Although we were not able to profile the gene expression of early *Arabidopsis* embryos undergoing elimination, we remain optimistic to pursue this study through improving the current technique as well as exploring new techniques to efficiently extract early *Arabidopsis*

embryos. As the embryos extraction work was mostly carried out during the COVID-19 lockdown, collaborative efforts to extract the embryos could also greatly improved the quantity of extracted embryos. Illuminating the spatio-temporal gene expression changes of an embryo undergoing centromere-mediated genome elimination is the next step to understanding this system more and can help improve this haploid induction technology. As centromere-mediated genome elimination is associated with genome instability, elucidating this pathway may also help us further understand the molecular response in genome fate maintenance.

REFERENCES

Blakeslee, A. F., Beling, J., Farnham, M. E., & Bergnier, A. D. (1922). A haploid mutant in the jimson weed, datura stramonium, Blakeslee et al., 1922. *Science*, *LV*(1433), 646–647.

Bohanec, B., Jakše, M., Ihan, A., & Javornik, B. (1995). Studies of gynogenesis in onion (Allium cepa L.): induction procedures and genetic analysis of regenerants. *Plant Science*, *104*(2), 215–224. https://doi.org/10.1016/0168-9452(94)04030-K

Britt, A. B., & Kuppu, S. (2016). Cenh3: An emerging player in haploid induction technology. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2016.00357

Chase, Sherret S. (1969). Monoploids and monoploid-derivatives of maize (Zea mays L.). *The Botanical Review*, *35*(2), 117–168. https://doi.org/10.1007/BF02858912

Chase, SHERRET S. (1949). Spontaneous doubling of the chromosome complement in monoploid sporophytes of maize. *Proc Iowa Acad Sci*, *56*(1), 113–115.

Chaudhary, H. K., Dhaliwal, I., Singh, S., & Sethi, G. S. (2003). Genetics of androgenesis in winter and spring wheat genotypes. *Euphytica*, *132*(3), 311–319. https://doi.org/10.1023/A:1025094606482

Clausen, R. E., & Mann, M. C. (1924). Inheritance in Nicotiana Tabacum: V. The Occurrence of Haploid Plants in Interspecific Progenies. *Proceedings of the National Academy of Sciences*, *10*(4), 121–124. https://doi.org/10.1073/pnas.10.4.121

Coe, E. H. (1959). A LINE OF MAIZE WITH HIGH HAPLOID FREQUENCY. *The American Naturalist*, *XCIII*(873), 9–25.

Comai, L., & Tan, E. H. (2019). Haploid Induction and Genome Instability. *Trends in Genetics*, *35*(11), 1–14. https://doi.org/10.1016/j.tig.2019.07.005

Darlington, C. D. (1939). Misdivision and the genetics of the centromere. *Journal of Genetics*, *37*(2), 341–364. https://doi.org/10.1007/BF02982733

Dwivedi, S. L., Britt, A. B., Tripathi, L., Sharma, S., Upadhyaya, H. D., & Ortiz, R. (2015). Haploids: Constraints and opportunities in plant breeding. *Biotechnology Advances*, *33*(6), 812–829. https://doi.org/10.1016/j.biotechadv.2015.07.001

Forment, J. V., Kaidi, A., & Jackson, S. P. (2012). Chromothripsis and cancer: Causes and consequences of chromosome shattering. *Nature Reviews Cancer*, *12*(10), 663–670. https://doi.org/10.1038/nrc3352

Forster, B. P., Heberle-Bors, E., Kasha, K. J., & Touraev, A. (2007). The resurgence of haploids in higher plants. *Trends in Plant Science*, *12*(8), 368–375. https://doi.org/10.1016/j.tplants.2007.06.007

Gairdner, A. E., & Darlington, C. D. (1931). Ring-formation in diploid and polyploid campanula persicifolia. *Genetica*, *13*(1), 113–150. https://doi.org/10.1007/BF01725041

Geiger, H. H., & Gordillo, G. A. (2009). Doubled haploids in hybrid maize breeding. *Maydica*, *54*(4), 485–499.

Gilles, L. M., Khaled, A., Laffaire, J., Chaignon, S., Gendrot, G., Laplaige, J., … Widiez, T. (2017). Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. *The EMBO Journal*, *36*(6), 707–717. https://doi.org/10.15252/embj.201796603

Henry, I. M., Dilkes, B. P., Young, K., Watson, B., Wu, H., & Comai, L. (2005). Aneuploidy and genetic variation in the Arabidopsis thaliana triploid response. *Genetics*, *170*(4), 1979–1988. https://doi.org/10.1534/genetics.104.037788

Ishii, T., Karimi-Ashtiyani, R., & Houben, A. (2016). Haploidization via Chromosome Elimination: Means and Mechanisms. *Annual Review of Plant Biology*, *67*(1), 421–438. https://doi.org/10.1146/annurev-arplant-043014-114714

Javornik, B., Bohanec, B., & Campion, B. (1998). Second cycle gynogenesis in onion, Allium cepa L., and genetic analysis of the plants. *Plant Breeding*, *117*(3), 275–278. https://doi.org/10.1111/j.1439-0523.1998.tb01939.x

Johnston, S. A., den Nijs, T. P. M., Peloquin, S. J., & Hanneman, R. E. (1980). The significance of genic balance to endosperm development in interspecific crosses. *Theoretical and Applied Genetics*, *57*(1), 5–9. https://doi.org/10.1007/BF00276002

Karimi-Ashtiyani, R., Ishii, T., Niessen, M., Stein, N., Heckmann, S., Gurushidze, M., … Houben, A. (2015). Point mutation impairs centromeric CENH3 loading and induces haploid plants. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(36), 11211–11216. https://doi.org/10.1073/pnas.1504333112

Kelliher, T., Starr, D., Richbourg, L., Chintamanani, S., Delzer, B., Nuccio, M. L., … Martin, B. (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature Publishing Group*. https://doi.org/10.1038/nature20827

Kelliher, T., Starr, D., Wang, W., McCuiston, J., Zhong, H., Nuccio, M. L., & Martinf, B. (2016). Maternal haploids are preferentially induced by cenh3-tailswap transgenic complementation in maize. *Frontiers in Plant Science*, *7*(MAR2016), 1–11. https://doi.org/10.3389/fpls.2016.00414

Kloosterman, W. P., Koster, J., & Molenaar, J. J. (2014). Prevalence and clinical implications of chromothripsis in cancer genomes. *Current Opinion in Oncology*, *26*(1), 64–72. https://doi.org/10.1097/CCO.0000000000000038

Korbel, J. O., & Campbell, P. J. (2013). Criteria for inference of chromothripsis in cancer genomes. *Cell*, *152*(6), 1226–1236. https://doi.org/10.1016/j.cell.2013.02.023

Kuppu, S., Ron, M., Marimuthu, M. P. A., Li, G., Huddleson, A., Siddeek, M. H., … Britt, A. B. (2020). A variety of changes, including CRISPR/Cas9-mediated deletions, in CENH3 lead to haploid induction on outcrossing. *Plant Biotechnology Journal*, *18*(10), 1–13. https://doi.org/10.1111/pbi.13365

Kuppu, S., Tan, E. H., Nguyen, H., Rodgers, A., & Comai, L. (2015). *Point Mutations in Centromeric Histone Induce Post-zygotic Incompatibility and Uniparental Inheritance*. 1–18. https://doi.org/10.1371/journal.pgen.1005494

Laurie, D. A., & Bennett, M. D. (1988). The production of haploid wheat plants from wheat x maize crosses. *Theoretical and Applied Genetics*, *76*(3), 393–397. https://doi.org/10.1007/BF00265339

Lermontova, I., Koroleva, O., Rutten, T., Fuchs, J., Schubert, V., Moraes, I., … Schubert, I. (2011). Knockdown of CENH3 in Arabidopsis reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. *Plant Journal*, *68*(1), 40–50. https://doi.org/10.1111/j.1365-313X.2011.04664.x

Lermontova, I., Schubert, V., Fuchs, J., Klatte, S., Macas, J., & Schubert, I. (2006). Loading of Arabidopsis centromeric histone CENH3 occurs mainly during G2 and requires the presence of the histone fold domain. *Plant Cell*, *18*(10), 2443–2451. https://doi.org/10.1105/tpc.106.043174

Li, Yifan, Li, D., Xiao, Q., Wang, H., Wen, J., Tu, J., … Yi, B. (2022). An in planta haploid induction system in Brassica napus. *Journal of Integrative Plant Biology*, *64*(6), 1140–1144. https://doi.org/10.1111/jipb.13270

Li, Yuan, Lin, Z., Yue, Y., Zhao, H., Fei, X., Lizhu, E., … Song, W. (2021). *Loss-of-function alleles of ZmPLD3 cause haploid induction in maize*. https://doi.org/10.1038/s41477-021-01037-2

Lindsey, B. E., Rivero, L., Calhoun, C. S., Grotewold, E., & Brkljacic, J. (2017). Standardized method for high-throughput sterilization of Arabidopsis seeds. *Journal of Visualized Experiments*, *2017*(128), 2–5. https://doi.org/10.3791/56587

Liu, C., Li, X., Meng, D., Zhong, Y., Chen, C., Dong, X., … Chen, S. (2017). A 4-bp Insertion at ZmPLA1 Encoding a Putative Phospholipase A Generates Haploid Induction in Maize. *Molecular Plant*, *10*(3), 520–522. https://doi.org/10.1016/j.molp.2017.01.011

Liu, H., Wang, K., Jia, Z., Gong, Q., Lin, Z., Du, L., … Ye, X. (2020). Efficient induction of haploid plants in wheat by editing of TaMTL using an optimized Agrobacterium-mediated CRISPR system. *Journal of Experimental Botany*, *71*(4), 1337–1349. https://doi.org/10.1093/jxb/erz529

Maheshwari, S., Tan, E. H., West, A., Franklin, F. C. H., Comai, L., & Chan, S. W. L. (2015). Naturally Occurring Differences in CENH3 Affect Chromosome Segregation in Zygotic Mitosis of Hybrids. *PLoS Genetics*, *11*(1), 1–20. https://doi.org/10.1371/journal.pgen.1004970

Malik, H. S., & Henikoff, S. (2003). Phylogenomics of the nucleosome. *Nature Structural Biology*, *10*(11), 882–891. https://doi.org/10.1038/nsb996

Malik, H. S., & Henikoff, S. (2009). Major Evolutionary Transitions in Centromere Complexity. *Cell*, *138*(6), 1067–1082. https://doi.org/10.1016/j.cell.2009.08.036

Marimuthu, M. P. A., Maruthachalam, R., Bondada, R., Kuppu, S., Tan, E. H., Britt, A., … Comai, L. (2021). Epigenetically mismatched parental centromeres trigger genome elimination in hybrids. *Science Advances*, *7*(47), 1–19. https://doi.org/10.1126/sciadv.abk1151

Mishra, R., & Rao, G. J. N. (2016). In-vitro Androgenesis in Rice: Advantages, Constraints and Future Prospects. *Rice Science*, *23*(2), 57–68. https://doi.org/10.1016/j.rsci.2016.02.001

Niazian, M., & Shariatpanahi, M. E. (2020). In vitro-based doubled haploid production: recent improvements. *Euphytica*, *216*(5), 1–21. https://doi.org/10.1007/s10681-020-02609-7

Pal, A., & Negi, V. S. (2019). Plant Cenh3 evolution is congruent with the phylogeny of plant species. *International Journal of Scientific and Technology Research*, *8*(11), 1473–1476.

Raissig, M. T., Gagliardini, V., Jaenisch, J., Grossniklaus, U., & Baroux, C. (2013). Efficient and rapid isolation of early-stage embryos from Arabidopsis thaliana seeds. *Journal of Visualized Experiments*, (76). https://doi.org/10.3791/50371

Ravi, M., & Chan, S. W. (2011). *Centromere-Mediated Generation of Haploid Plants*. 169–181.

Ravi, M., & Chan, S. W. L. (2010). Haploid plants produced by centromere-mediated genome elimination. *Nature*, *464*(7288), 615–618. https://doi.org/10.1038/nature08842

Ravi, M., Marimuthu, M. P. A., Tan, E. H., Maheshwari, S., Henry, I. M., Marin-Rodriguez, B., … Chan, S. W. L. (2014). A haploid genetics toolbox for Arabidopsis thaliana. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms6334

Riera-Lizarazu, O., Rines, H. W., & Phillips, R. L. (1996). Cytological and molecular characterization of oat x maize partial hybrids. *Theoretical and Applied Genetics*, *93*(1–2), 123–135. https://doi.org/10.1007/BF00225737

Rines, H. W., & Dahleen, L. S. (1990). *CELL BIOLOGY & MOLECULAR GENETICS Haploid Oat Plants Produced by Application*. 1073–1078.

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. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33). https://doi.org/10.1073/pnas.1103190108

Subrahmanyam, N. C., & Kasha, K. J. (1973). Selective chromosomal elimination during haploid formation in barley following interspecific hybridization. *Chromosoma*, *42*(2), 111–125. https://doi.org/10.1007/BF00320934

Tan, E. H., Henry, I. M., Ravi, M., Bradnam, K. R., Mandakova, T., Marimuthu, M. P. A., … Chan, S. W. L. (2015). Catastrophic chromosomal restructuring during genome elimination in plants. *ELife*, *4*(MAY), 1–16. https://doi.org/10.7554/eLife.06516

Tulecke, W. (1964). A haploid tissue culture from the female gametophyte of Ginkgo biloba L. [31]. *Nature*, *203*(4940), 94–95. https://doi.org/10.1038/203094a0

Valero, M., Richerd, S., Perrot, V., & Destombe, C. (1992). Evolution of alternation of haploid and diploid phases in life cycles. *Trends in Ecology and Evolution*, *7*(1), 25–29. https://doi.org/10.1016/0169-5347(92)90195-H

Wang, N., & Dawe, R. K. (2018). Centromere Size and Its Relationship to Haploid Formation in Plants. *Molecular Plant*, *11*(3), 398–406. https://doi.org/10.1016/j.molp.2017.12.009

Wang, N., Gent, J. I., & Dawe, R. K. (2021). Haploid induction by a maize cenh3 null mutant. *Science Advances*, *7*(4). https://doi.org/10.1126/sciadv.abe2299

Wang, N., Xia, X., Jiang, T., Li, L., Zhang, P., Niu, L., … Lin, H. (2022). In planta haploid induction by genome editing of DMP in the model legume Medicago truncatula. *Plant Biotechnology Journal*, *20*(1), 22–24. https://doi.org/10.1111/pbi.13740

Wang, S., Jin, W., & Wang, K. (2019). Centromere histone H3- and phospholipase-mediated haploid induction in plants. *Plant Methods*, *15*(1), 1–10. https://doi.org/10.1186/s13007-019-0429-5

Wang, W. (2022). *acquisition of sperm fusion competence*. 1–11. https://doi.org/10.1073/pnas.2207608119/-/DCSupplemental.Published

Xiaoyan, C., Marshall, J. B., Topp, C., Mroczek, R., Zhong, C. X., Marshall, J. B., … Dawead-, R. K. (2019). *Centromeric Retroelements and Satellites Interact with Maize Kinetochore Protein CENH3 Akio Kato , Kiyotaka Nagaki , James A . Birchler , Jiming Jiang and R . Kelly Dawe Published by : American Society of Plant Biologists (ASPB) Stable URL : https://www*. *14*(11), 2825–2836. https://doi.org/10.1105/tpc.006106.fraction

Yao, L., Zhang, Y., Liu, C., Liu, Y., Wang, Y., Liang, D., … Kelliher, T. (2018). OsMATL mutation induces haploid seed formation in indica rice. *Nature Plants*, *4*(8), 530–533. https://doi.org/10.1038/s41477-018-0193-y

Zhang, J., Yin, J., Luo, J., Tang, D., Zhu, X., Wang, J., … Huang, S. (2022). Construction of homozygous diploid potato through maternal haploid induction. *ABIOTECH*, *3*(3), 163–168. https://doi.org/10.1007/s42994-022-00080-7

Zhong, Y., Chen, B., Li, M., Wang, D., Jiao, Y., Qi, X., … Chen, S. (2020). A DMP-triggered in vivo maternal haploid induction system in the dicotyledonous Arabidopsis. *Nature Plants*, *6*(5), 466–472. https://doi.org/10.1038/s41477-020-0658-7

Zhong, Y., Chen, B., Wang, D., Zhu, X., Li, M., Zhang, J., … Chen, S. (2022). In vivo maternal haploid induction in tomato. *Plant Biotechnology Journal*, *20*(2), 250–252. https://doi.org/10.1111/pbi.13755

Zhong, Y., Liu, C., Qi, X., Jiao, Y., Wang, D., Wang, Y., … Chen, S. (2019). Mutation of ZmDMP enhances haploid induction in maize. *Nature Plants*, *5*(6), 575–580. https://doi.org/10.1038/s41477-019-0443-7

Zhong, Y., Wang, Y., Chen, B., Liu, J., Wang, D., Li, M., … Chen, S. (2022). Establishment of a dmp based maternal haploid induction system for polyploid Brassica napus and Nicotiana tabacum . *Journal of Integrative Plant Biology*, *64*(6), 1281–1294. https://doi.org/10.1111/jipb.13244

BIOGRAPHY OF THE AUTHOR

Diana Spencer was born in Sarawak, Malaysia on December 31st, 1981. She was raised in Kuching, Sarawak and graduated high school in 1998 from SMK Sungai Tapang. Thereafter, she pursued her Bachelor of Engineering with Honours in Electronics and Telecommunications at Universiti Malaysia Sarawak. Upon graduation, she worked with Malaysia's leading conglomerate of vocational and technical institutes and colleges. Her first posting as a technical instructor in 2005 was at the island-state of Penang, where she was part of the pioneering technical instructors of a newly created Diploma of Engineering Technology in Mechatronics program at Kolej Kemahiran Tinggi MARA Balik Pulau. Then, she was posted to her hometown Kuching in 2007 to serve as the head of department for Electronics Technology Department at Institut Kemahiran MARA Kuching before she was internally transfer to lead Modular Program/Civil Engineering Technology department in 2012. In the period of 2012 – 2013, through serendipity, Diana met different groups of people that introduced her to the wonder of life sciences and community-based conservation and her interest and focus started to shift. As her academic and professional background were in the opposite extremes of the spectrum from her new-found interest, Diana enrolled in Master of Environmental Science in Land Use and Water Management at University Malaysia Sarawak in 2016 to help her to contemplate career change. Upon completion of the master program, she was convinced that she was in the right direction. She enrolled in the Master of Science in Botany and Plant Pathology at the University of Maine in the fall of 2019. She is a candidate for the Master of Science degree in Botany and Plant Pathology from the University of Maine in December 2022.