Haploid-inducer development by CRISPR/Cas9 mediated CenH3 gene modification in *Helianthus annuus* (Sunflower)

by

Londiwe M Mabuza

18258566

Supervisor: Dr Bridget Genevieve Crampton

Co-supervisor: Dr Zacharias Hendrik Swanevelder

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor

In the faculty of Natural and Agricultural Sciences
Department of Plant and Soil Sciences
University of Pretoria
Pretoria
April 2023

The financial assistance of the National Research Foundation (NRF) toward this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

TABLE OF CONTENTS

DECLARATION ... v
ACKNOWLEDGEMENTS ... vi
PREFACE ... vii
SUMMARY ... ix
LIST OF ABBREVIATIONS ... x
LIST OF FIGURES ... xii
LIST OF TABLES ... xiv

CHAPTER 1 ... 1

Opportunities and bottlenecks in sunflower breeding through doubled haploid and haploid inducer line development in the era of targeted genome editing ... 1

1.1 INTRODUCTION: SUNFLOWER ... 2
1.2 DOUBLED HAPLOIDS: INDUCTION METHODS AND THEIR ROLE IN PLANT BREEDING ... 3
1.2.1 Haploid induction by Centromeric Histone Protein 3 (CENH3) modification ... 7
1.3 METHODS OF GENOME MODIFICATION ... 11
1.3.1 Epigenetic inheritance and new trait development ... 12
1.3.2 Randomly induced mutagenesis as a tool towards trait diversification: Physical and chemical mutagenesis ... 13
1.3.3 New plant breeding technologies: Targeted gene modification ... 14
1.4 GENOME EDITING OUTCOMES - DOUBLE STRAND-DNA BREAK REPAIR MECHANISMS (DSBS) ... 20
1.5 DELIVERY METHODS FOR GENOME EDITING REAGENTS 24
1.5.1 *Agrobacterium*-mediated transformation ... 24
1.5.2 Particle bombardment ... 25
1.5.3 Viral vectors for plant genome editing ... 26
1.6 CONCLUSION ... 28
1.7 REFERENCES ... 30

CHAPTER 2 ... 42

CRISPR/Cas9 genome editing of the *CenH3* gene in sunflower (*Helianthus annuus*) using *Agrobacterium*-mediated transformation ... 42

2.1 ABSTRACT ... 43
2.2 INTRODUCTION ... 44
2.3 MATERIALS AND METHODS ... 46
2.3.1 *CenH3* target sequence, histone fold domain identification and gRNA design . 46
2.3.2 Donor repair template (DRT) design ... 50
2.3.3 Vector cloning ... 51
2.3.4 *Agrobacterium tumefaciens*-mediated plant transformation ... 54
2.3.5 Transformation of CRISPR/Cas9 system vectors into *Agrobacterium tumefaciens* ... 54
2.3.6 Seed disinfection ... 55
2.3.7 Plant transformation ... 55
2.3.8 Cas9 and sgRNA functional analysis ... 57
2.3.9 Mutation analysis ... 58
2.4 RESULTS ... 59
2.4.1 *CenH3* target sequence identification and design ... 59
2.4.2 Vector cloning ... 65
2.4.3 *Agrobacterium tumefaciens* -mediated plant transformation ... 67
2.4.4 Cas9 and sgRNA functional analysis ... 70

2.4.5 Mutation analysis using next generation sequencing 71
2.5 DISCUSSION..76
2.6 REFERENCES..81
2.7 SUPPLEMENTARY TABLES..88
2.8 SUPPLEMENTARY FIGURES..89

CHAPTER 3 98

Targeted modification of the *CenH3* gene in sunflower using a Tomato yellow leaf curl (IL-60-BS) viral vector system for the delivery of CRISPR/Cas9 components............. 98

3.1 ABSTRACT..99
3.2 INTRODUCTION.. 100
3.3 MATERIALS AND METHODS...102
3.3.1 Viral delivery vector design and construction 102
3.3.2 Evaluation of reporter gene expression using the IL-60-BS-2A delivery system.... 104
3.3.3 Addition of CRISPR/Cas9 gene cassette to the viral delivery vector............. 105
3.3.4 Plasmid extractions and storage 106
3.3.5 Sunflower lines and growth conditions 106
3.3.6 Plant inoculation with IL-60-2A and IL-60-2A-DsRed constructs 106
3.3.7 Leaf inoculation of sunflower with IL-60-BS-2A-Cas9-1185 vector................. 107
3.3.8 Confirmation of construct presence in inoculated plants (T0) and progeny (T1 and T2) 107
3.3.9 Transcription analysis of Cas9 and associated sgRNA 108
3.3.10 CRISPR/Cas9 mutation analyses in sunflower.. 109
3.4 RESULTS 112
3.4.1 IL-60-BS-2A and IL-60-BS-2A-DsRed replication, systemic movement and reporter gene expression in sunflower... 112
3.4.2 Transformation of CRISPR/Cas9 system and testing of systemic movement of the sgRNA and Cas9 transcripts in sunflower leaves 116
3.4.3 cDNA analyses to confirm Cas9 and sgRNA gene transcription 119
3.4.4 CRISPR/Cas9 mutation analyses in sunflower plants 120
3.5 DISCUSSION..124
3.6 REFERENCES.. 129
3.7 SUPPLEMENTARY FIGURES.. 133

CHAPTER 4 142

Haploid production potential of transient *CenH3* recombinant sunflower plants...... 142

4.1 ABSTRACT.. 143
4.2 INTRODUCTION..144
4.3 MATERIALS AND METHODS...146
4.3.1 SIFT predictions 146
4.3.2 Plant material and morphological analysis 146
4.3.3 Ploidy Evaluation using Flow cytometry 146
4.3.4 Chromosome analysis (microscopy) 147
4.4 RESULTS... 148
4.4.1 SIFT predictions 148
4.4.2 Morphological characterization 149
4.4.3 Flow cytometric and microscopic ploidy analysis 149
4.5 DISCUSSION..154
4.6 REFERENCES.. 157
4.7 SUPPLEMENTARY FIGURES.. 161

CHAPTER 5 166

General discussion and future prospects 166

5.1 GENERAL DISCUSSION 167
5.2 REFERENCES 174

DECLARATION

I, Londiwe M Mabuza declare that the dissertation which I hereby submit for the degree *Philosophiae Doctor* at the University of Pretoria, contains my own independent work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Londiwe M Mabuza

27 July 2023

Date

ACKNOWLEDGEMENTS

The completion of the work presented in this dissertation would not have been possible without the following people and institutions:

The financial contribution of the National Research Council (NRF) of South Africa through their professional development programme. The Agricultural Research Council (ARC) of South Africa, particularly the Biotechnology Platform.

FOAR project (with the ARC, University of Pretoria, and INTA) which allowed me to study sunflower tissue culture protocols at the Instituto Nacional de Tecnología Agropecuaria (INTA) (Buenos Aires, Argentina) tissue culture laboratory under their experts on sunflower *in vitro* culturing and transformation, Dr Laura Radonic and Dr Marisa Lopez Bilbao.

My supervisors, Dr Dirk Swanevelder and Dr Bridget Crampton, for their kindness, endless dedication, guidance, support and motivation throughout my study.

Mr. Andrew Mokhele for providing us with sunflower seeds.

My mentor Dr Nokuthula ‘Noks’ Mchunu for her endless support, guidance and gift of friendship.

My friends and sisters, Dr Madeleine Ramantswana, Dr Nokuthula Mchunu, Dr Siphiwokuhle Shandu, Vinolia Danki, and Ms Maria Ramapholo for their gift of friendship, support and just being a soft-landing place.

My dad Mr B.M Mabuza, brothers Zithulele and Uvile Mabuza and my sister Londeka Lungiswa Thwala for their love and support.

My late mom Ms Lindiwe Mary Ntombizodwa Ngwenya (1967 – 2011) and grandmother Ms Selina Ngwenya (1930 – 2020), to whom I dedicate this research - I am because you were.

And lastly, God, my creator.

PREFACE

The research and findings of this dissertation are based on experimental work conducted at the Biotechnology Platform (Ondersterpoort Veterinary Campus, Onderstepoort) of the Agricultural Research Council, South Africa. The thesis is presented in five individual chapters, each in publication format, and therefore some degree of repetition between chapters was unavoidable. We acknowledge the recent species name change of "*Agrobacterium tumefaciens*" to "*Rhizobium radiobacter*". The previous name "*Agrobacterium tumefaciens*" will however be maintained through the remainder of the thesis for consistency and cross referencing with cited literature which cannot be changed. This follows examples of numerous recent publications still using the previous species name.

Sunflower is the third most important oilseed crop in the world and has no available doubled haploid induction procedure/technique that can be efficiently used in breeding programs. A reproducible and efficient doubled haploid induction method would be a valuable tool in accelerating the breeding of new elite sunflower varieties. Although several attempts have been made, the establishment of a sunflower doubled haploid induction protocol has remained a challenge owing to recalcitrance to *in vitro* culture regeneration of the species. Approaches for haploid development in other crops are often cultivar specific, difficult to reproduce, and rely on available tissue culture protocols, which on their own are also cultivar and/or species specific. The absence of a doubled haploid system impacts sunflower breeding and associated improvement processes. Significant molecular advances targeting genes, such as the centromeric histone 3 (*CenH3*) and *Matrilineal* (*MTL*) genes with CRISPR/Cas9, and the successful use of viral vectors for the delivery of CRISPR/Cas9 components into plant cells thus eliminating the *in vitro* culture bottleneck, have the potential to improve doubled haploid technology in sunflower.

The aims of this PhD thesis: were to develop a haploid induction system in sunflower by editing the *CenH3* gene using the CRISPR/Cas9 technology and to develop a viral-based CRISPR/Cas9 delivery system to mitigate the tediousness and low efficiencies accompanied by *Agrobacterium* transformation.

Chapter 1 reviews haploid induction methods in respect to sunflower and highlights hindrances and opportunities for attaining an efficient sunflower haploid induction method in this new era of targeted genome editing. A special focus on the modification of the *CenH3* gene for the production of haploid inducer lines, means of genome modification and constraints and opportunities associated with these, are reviewed. This chapter was published

as: Mabuza, L.M.; Mchunu, N.P.; Crampton, B.G.; Swanevelder, D.Z.H. Accelerated Breeding for *Helianthus annuus* (Sunflower) through Doubled Haploidy: An Insight on Past and Future Prospects in the Era of Genome Editing. Plants 2023, 12, 485.
https://doi.org/10.3390/plants12030485

Chapter 2 focuses on the confirmation of the sunflower *CenH3* gene, the development and design of CRISPR/Cas9 constructs for homology directed repair targeting of the sunflower *CenH3* gene in sunflower for the purpose of haploid induction. This chapter also reports on the *Agrobacterium tumefaciens* transformation of sunflower mature cotyledons with the designed CRISPR/Cas9 construction for modification of the *CenH3* gene.

Chapter 3 focused on the suitability of a geminiviral replicon (IL-60-BS) for sunflower *CenH3* gene targeting. It includes results of successful homology directed repair based CRISPR/Cas9 editing of sunflower T0 seedlings through injection with the replicon plasmid DNA. This chapter further reports on whether geminiviral delivered CRISPR/Cas9 mutations are heritable in the T1 and T2 generations.

Chapter 4 is the final experimental chapter of the thesis and, it involves the evaluation of the haploid inducer potential of *CenH3* mutant lines obtained in Chapter 3 through crosses with wildtype sunflower plants. Flow cytometric analysis of ploidy status of T1 and T2 lines obtained from crossing with T0 and wildtype are evaluated.

Chapter 5 concludes the work and provides insights on the major findings and setbacks reported in this study. This chapter also presents future research prospects which will help the improvement of genome modification in sunflower and the development of a haploid inducer technique.

Research outputs from this PhD thesis to date include:

Publication: Mabuza, L.M.; Mchunu, N.P.; Crampton, B.G.; Swanevelder, D.Z.H. Accelerated Breeding for *Helianthus annuus* (Sunflower) through Doubled Haploidy: An Insight on Past and Future Prospects in the Era of Genome Editing. Plants 2023, 12, 485.
https://doi.org/10.3390/plants12030485

Gene constructs developed. The *CenH3* constructs developed during this work have been requested by the INTA team for testing in their germplasm. A material transfer agreement is currently being negotiated between the parties for refinement and use within the INTA sunflower transformation programme.

SUMMARY

Sunflower is the third most important oil seed crop globally following soybean and rapeseed. Climate change threatens crop production with various limitations, including unpredictable temperature fluctuations, altered rainfall patterns and, novel diseases and pests. Conventional, mutation and accelerated breeding technologies, such as haploidization, have been able to address some of these challenges through the production of resistant cultivars. The production of sunflower doubled haploid lines through *in vitro* approaches such as anther, pollen and unpollinated ovule culturing have been tested with little success, mainly due to sunflower being recalcitrant to tissue culture regeneration. A universal *in vivo* technique based on the modification of the *CenH3* gene leads to uniparental chromosome elimination during hybridisation with a wild type line. This technique has been successfully demonstrated to induce a small number of haploid progenies in *Arabidopsis thaliana*, maize and wheat. Given the universal function and conserved nature of the CENH3 protein, this approach promises to be a breakthrough in haploid induction technology. The aim of the current study was to facilitate the production of a sunflower haploid inducer line through CENH3 protein modification using the CRISPR/Cas9 technology. Two CRISPR/Cas9 constructs were designed to target three different amino acids (P51S, G52E and A55V) in the histone fold domain region of the sunflower CENH3 through homology directed repair (HDR). The individual constructs (pDe-Cas-258-*nptII*, pDe-Cas-1185-*bar*) were introduced to sunflower mature cotyledons by *Agrobacterium*-mediated transformation while the geminiviral construct (IL-60-BS-Cas9-1185) was mechanically introduced into sunflower seedlings. Next generation sequencing revealed that only four of the transgenics obtained via *Agrobacterium* transformation contained two of the targeted amino acid conversions (P51S and A55V) and the other three did not contain any mutations. 20% of the plants inoculated with the geminiviral construct displayed mutations, with one plant displaying complete homology directed repair. Furthermore, a slight increase in HDR was observed with the use of a geminiviral vector when compared to *Agrobacterium* mediated transformation. Sunflower mutant plants were tested for their ability to act as haploid inducers using flow cytometry. Flow cytometric analysis revealed 71% and 12.5% aneuploidy in the T1 and T2 generations, respectively. No haploid progeny was obtained in either generation. The overall results of this study indicate that uniparental chromosome elimination coupled with the use of viral vectors for the delivery of CRISPR/Cas9 components into plant cells, have the potential to improve double haploid technology in sunflower. The research carried out in this study will enhance the improvement of sunflower through genome editing and the production of doubled haploid lines.

LIST OF ABBREVIATIONS

%	percentage
°C	degrees Celsius
bar	*bialaphos* gene
BLAST	basic local alignment search tool
BLASTn	search a nucleotide database using a nucleotide enquiry
bp	basepair (s)
Cas	CRISPR-associated genes
cDNA	complementary DNA
CenH3	centromeric histone 3 gene
cm	centimetre(s)
CRISPR	clustered regularly interspaced short palindromic repeats
ddH_2O	double distilled water
DH	doubled haploid
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
dpi	days post inoculation
DRT	donor repair template
DSB	double-stranded DNA break
DsRed	*Discosoma* sp. red fluorescent protein
g	gram
GFP	green fluorescent protein
GVR	geminiviral vector replicon
HDR	homology-directed repair
HFD	histone fold domain
H_2O	water
IDT	Integrated DNA technologies
kb	kilobase pairs
L	litre
LB	Luria Bertani
mg	milligram (s)
min	minute (s)
mL	millilitre(s)

mM	micromolar
MS	Murashige and Skoog
MUSCLE	multiple sequence comparison by log expectation
NCBI	National Center for Biotechnology Information
ng	nanogram (s)
NGS	next-generation sequencing
NHEJ	non-homologous end joining
nptII	neomycin phosphotransferase
OD	optical density
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
rcf	relative centrifugal force
RNA	ribonucleic acid
RNAi	RNA interference
rpm	revolutions per minute
RT-PCR	reverse transcription PCR
TALEN	transcription activator like nuclease
T-DNA	transfer-DNA
TE	Tris-EDTA buffer solution
s	second(s)
sgRNA	small/single guide RNA
SNP	single nucleotide polymorphism
U	units
USA	United States of America
USD	United States Dollar
UV	ultraviolet
v/v	volume per volume
WT	wildtype
w/v	weight per volume
ZFNs	zinc finger nucleases
µM	micromolar (s)
µL	microlitre (s)

LIST OF FIGURES

Figure 1.1 Comparison of conventional breeding and doubled haploid technology breeding methods. 4
Figure 1.2 Illustration of centromere function during zygote mitosis. 9
Figure 1.3 Typical structure of the plant centromeric histone protein 3 (CENH3). 10
Figure 1.4 Schematic representation of the CRISPR/Cas9 genome editing mechanisms. 17
Figure 2.1 A schematic representation of the sunflower *CenH3* gene. 50
Figure 2.2 Multiple amino acid alignment using MUSCLE of the full CENH3 protein of nineteen (19) plant species. 59-62
Figure 2.3 Multiple sequence alignment of predicted CENH3 histone fold domain (HFD). 64
Figure 2.4 Following colony PCR amplification, the DNA fragments were resolved on 1.5% agarose gels to confirm successful cloning. 66
Figure 2.5 The binary plasmids used for *Agrobacterium* transformation of sunflower mature cotyledons. 67
Figure 2.6 Transformation steps of sunflower (KP328, ARC-GC) cotyledons obtained from mature seeds. 69
Figure 2.7 PCR products for confirmation of presence of T-DNA in putative transgenic plants. 70
Figure 2.8 RT-PCR products confirming the transcription of the sgRNA and Cas9 in sunflower treated plants. 71
Figure 2.9 Targeted mutagenesis in transgenic sunflower of four different T0 shoots analysed by NGS and CRISPresso2. 73
Figure 2.10 Multiple sequence alignment of CENH3 histone fold domain amino acid changes induced by CRISPR/Cas9 HDR in four sunflower plants. 75
Figure 3.1 Schematic map of the IL-60-BS plant expression vector described by Peretz et al. (2007). 103
Figure 3.2 Agarose gel image showing PCR results for the coat protein (CP) and ampicillin (Amp) genes. 113
Figure 3.3 Colony PCR confirming the cloning of the *DsRed* gene into the IL-60-BS-2A vector....... 113
Figure 3.4 Schematic representation of the IL-60-BS-2A-DsRed plasmid. 114
Figure 3.5 DsRed expressing fluorescence in sunflower seedlings inoculated with IL-60-BS-2A-DsRed under blue and red light. 115
Figure 3.6 Colony PCR confirming successful cloning of CRISPR/Cas9 components into the IL-60-BS-2A construct using primers targeting the Parsley ubiquitin promoter and F1 origin of replication. 117
Figure 3.7 Schematic maps for IL-60-BS-2A-Cas9-1185 construct designed for the mutation of the *CenH3* gene in sunflower. 117
Figure 3.8 PCR confirming the presence of the vector by amplification of the U6-26 promoter and sgRNA in T0 plants inoculated with the IL-60-BS-1185 vector. 118
Figure 3.9 cDNA PCR to test sgRNA and Cas9 transcripts in five of the ten T0 sunflower plants inoculated with the IL-60-BS-2A-Cas9-1185. 119

Figure 3.10 Mutations in sunflower seedlings achieved through viral delivery of CRISPR/Cas reagents, sgRNA and donor repair template (DRT). ... 122
Figure 3.11 Multiple sequence alignment of CENH3 histone fold domain amino acid changes induced by CRISPR/Cas9 in sunflower. ... 123
Figure 4.1 Comparison of normal wildtype sunflower line with T2G line obtained from outcrossing T1 generation lines with wildtype sunflower lines. ... 149
Figure 4.2 Histograms obtained from flow cytometric ploidy analysis. ... 151
Figure 4.3 Chromosome spreads of sunflower root tips. ... 153

SUPPLEMENTARY FIGURES

Figure S2.1 Sequence of the sunflower *CenH3* gene obtained from HiSeq next generation sequencing and assembled with SPAdes de novo assembly ... 90
Figure S2.2 Nucleotide sequences of the donor repair templates designed for amino acid changes in the sunflower CENH3. ... 92
Figure S2.3 Partial sequence of the M13 sequence ... 92
Figure S2.4 Partial sequence of *Arabidopsis* U6-26 promoter ... 93
Figure S2.5 Partial sequence of donor sequence ... 93
Figure S2.6 Partial sequence of donor sequence and CaMV promoter ... 94
Figure S2.7 Next generation sequencing and CRISPresso2 analysis results of sunflower plants ... 95
Figure S2.8 Flow diagram of cloning steps for pDe-Cas9-258-*nptII* and pDe-Cas9-1185-*bar* ... 97

Figure S3.1 Full sequence of the IL-60-BS-2A viral vector ... 135
Figure S3.2 DsRed confirmation by Sanger sequencing ... 135
Figure S3.3 CRISPR/Cas9 component cloning confirmation into IL-60-BS-2A-Cas9-1185 by Sanger sequencing ... 135
Figure S3.4 Restriction digestion using the XmaI enzyme for confirmation of successful vector assembly by Infusion® cloning. ... 136
Figure S3.5 PCR confirming the presence of the IL-60-BS-1185 vector by amplification of the viral coat protein from T0 plants inoculated with the vector ... 137
Figure S3.6 PCR confirming Cas9 transcript presence in sunflower T1 and T2 generation plants. .. 138
Figure S3.7 CRISPresso2 analysis of sunflower plants inoculated with IL-60-BS-1185 vector for targeted CRISPR mutations ... 139
Figure S3.8 Next Generation sequencing and CRISPResso2 analysis of T1 plants obtained from KP328-SEVIH ... 140
Figure S3.9 Next Generation sequencing and CRISPResso2 analysis of T2 plants obtained from T1 plants. ... 141
Figure S4.1 SIFT calculated probabilities of amino acid changes in amino acid positions 51 – 75 of the sunflower CENH3 histone fold domain ... 161

Figure S4.2 SIFT predictions of tolerated and non-tolerated amino acid changes in the sunflower CENH3 protein from amino acid position 1 – 100. 163
Figure S4.3 Histograms obtained from flow cytometric analysis showing sunflower lines evaluated for ploidy status 165

LIST OF TABLES

Table 1.1 Doubled haploid induction methods used in oilseed breeding. 6
Table 1.2 Different classes and types of Cas proteins applicable for plant genome editing 19
Table 1.3 List of successful CRISPR/Cas edited homology directed repair systems in crop and model plants 23
Table 1.4 Viruses used for complete CRISPR/Cas reagent delivery 28

Table 2.1 List of primer sequences used in this chapter and study. 48
Table 2.2 Nucleotide and amino acid substitutions made to homology directed repair donor sequence for use in CRISPR/Cas9 mutation of the sunflower CenH3 gene. 51
Table 2.3 Transformation and mutation efficiency of sunflower T0 shoots. 70
Table 2.4 Mutation frequency in the four transgenic lines analysed with CRISPResso2 (Clement et al., 2019) 72
Table 3.1 Primer sequences used in this study (Chapter 3). 111
Table 3.2 Table showing transformation, expression of transgenes and CRISPR/Cas9 mutation efficiency in sunflower seedlings inoculated with two TYLCV-based delivery systems. 121
Table 4.1 SIFT scores for amino acid substitutions introduced by CRISPR/Cas9 genome editing and predicted haploid inducer frequency in sunflower. 148

SUPPLEMENTARY TABLE

Table S2.1 Composition of different medium used in the transformation and tissue culture of sunflower. 88

CHAPTER 1

OPPORTUNITIES AND BOTTLENECKS IN SUNFLOWER BREEDING THROUGH DOUBLED HAPLOID AND HAPLOID INDUCER LINE DEVELOPMENT IN THE ERA OF TARGETED GENOME EDITING

This chapter has been published as a review paper (with reviewer corrections): Mabuza, L.M.; Mchunu, N.P.; Crampton, B.G.; Swanevelder, D.Z.H. Accelerated Breeding for *Helianthus annuus* (Sunflower) through Doubled Haploidy: An Insight on Past and Future Prospects in the Era of Genome Editing. Plants 2023, 12, 485. https://doi.org/10.3390/plants12030485.

1.1 INTRODUCTION: SUNFLOWER

The global oilseed market has emerged as one of the most competitive vegetable markets in the world owing to genetic improvement and advancement of cropping systems (Pilorge, 2020). During the period between 2014-2018, *Helianthus annuus* L. (sunflower) accounted for 9% of the global oilseed market (Pilorge, 2020), maintaining its position as the third most important vegetable oilseed in the world following soybean and rapeseed (Feng et al., 2022). Breeding is an essential part of crop improvement, and the majority of sunflower breeding methods focus on increased yield, seed quality, better oil content as well as disease and pest resistance (Dirks et al., 2009; Drumeva et al., 2014; Hübner et al., 2019). Breeding programs for the development of sunflower hybrids have been on-going for over half a century (Vear, 2016). The pioneering sunflower breeding work started in Russia during the 1960s to develop varieties with increased oil content (Vear, 2016). This was followed by the development of cytoplasmic male sterility (CMS) involving a cross between *Helianthus petiolaris* and cultivated sunflower (Leclercq 1969 as referenced by Vear, 2016). This work resulted in the release of the first sunflower hybrid during the 1970s (Meena et al*.,* 2013). Sunflower is believed to have originated from North America approximately 1000 BC (Davey and Jan, 2010). Sunflower has a haploid (2n=2x=34) genome of approximately 3.8 Gb (Badouin et al., 2017; Hübner et al., 2019). The sunflower whole genome and pan genome sequencing projects have recently been completed (Badouin et al., 2017; Hübner et al., 2019), and this may provide new opportunities for sunflower yield and quality improvement.

Climate change threatens to be accompanied by various limitations to crop production including unpredictable temperature fluctuations, novel diseases and pests (Samantara et al., 2022). These pressures will require accelerated breeding pipelines to ensure increased genetic gain in crucial crops. Conventional plant breeding has been able to achieve cultivars with increased yield and disease resistance (Davey and Jan, 2010). Furthermore, technologies, such as genomics which involve the genetic characterization of organisms by identifying, isolating and modifying genes have accelerated crop improvement and the development of novel genotypes (Sarrafi and Gentzbittel, 2004). Unfortunately, commercial sunflower is regarded to have a limited genetic base (Friedt, 1992). This absence of sufficiently diverse genetic resources has been a hindrance in sunflower breeding and improvement (Dagustu, 2018). The primary aim of any breeding program is the homogenous expression of the desired phenotype in the progeny (Britt and Kuppu, 2016). Sunflower is an insect, open-pollinated species, and therefore, undesirable genetic and phenotypic-variation associated heterogeneity due to random crosses may arise (Cvejic et al., 2020). Furthermore, conventional breeding requires considerable space, and resources for plant selection

(Samantara et al., 2022). This can cause breeding programs to take up to a decade to produce a new cultivar of sunflower (Davey and Jan, 2010). Therefore, to eliminate constraints associated with heterogeneity in lines, the establishment of true breeding lines is required. True breeding lines can be achieved by repetitive backcrossing or by developing haploids with subsequent chromosome doubling to form doubled-haploid (DH) lines (Dwivedi et al., 2015). Although sunflower doubled haploids have been previously applied in breeding programs (Blinkov et al., 2022), the absence of sufficiently diverse genetic resources and its recalcitrance to *in vitro* culturing and transformation have been a hindrance in the development of a doubled haploid induction procedure that could be efficiently used in breeding programs (Darqui et al., 2021, Dagustu et al.,2018). As an out-crossing crop, the lack of a doubled haploid system limits sunflower breeding and associated improvement processes, thereby delaying new trait developments. Nevertheless, recent advances in *in vivo* haploid induction technology though mitotic/meiotic process manipulations and genome editing promise to provide solutions for fast tracked sunflower breeding. In this chapter, past and present haploid induction methods in respect to sunflower are investigated, highlighting hindrances and opportunities for attaining an efficient sunflower haploid induction method in this new era of targeted genome editing.

1.2 DOUBLED HAPLOIDS: INDUCTION METHODS AND THEIR ROLE IN PLANT BREEDING

Haploids can be defined as adult plants carrying a genome derived from a single (haploid) gamete (either a sperm or egg cell), i.e. a cell with a single chromosome set (Dwivedi et al., 2015; Ishii et al., 2016). Haploid plants are unable to go through meiosis and are therefore sterile (Murovec and Bohanec, 2012). Fertility of these haploids is achieved through chemically induced or spontaneous chromosome doubling, resulting in 100% homozygosity in a single generation (Britt and Kuppu, 2016; Murovec and Bohanec, 2012). The availability of DH lines saves a lot of time in the production of true-breeding lines since the need for backcrossing over numerous generations, is removed (Figure 1.1) (Ishii et al., 2016; Karimi-Ashtiyani et al., 2015).

Additionally, DHs can be used to accelerate pyramiding multiple mutants, forward mutagenesis screening, downsizing ploidy levels (e.g. tetra- to diploid), the generation of homozygotes for gametophyte-lethal mutations and reducing inbreeding depression associated with self-pollination (Karimi-Ashtiyani et al., 2015, Murovec and Bohanec, 2012).

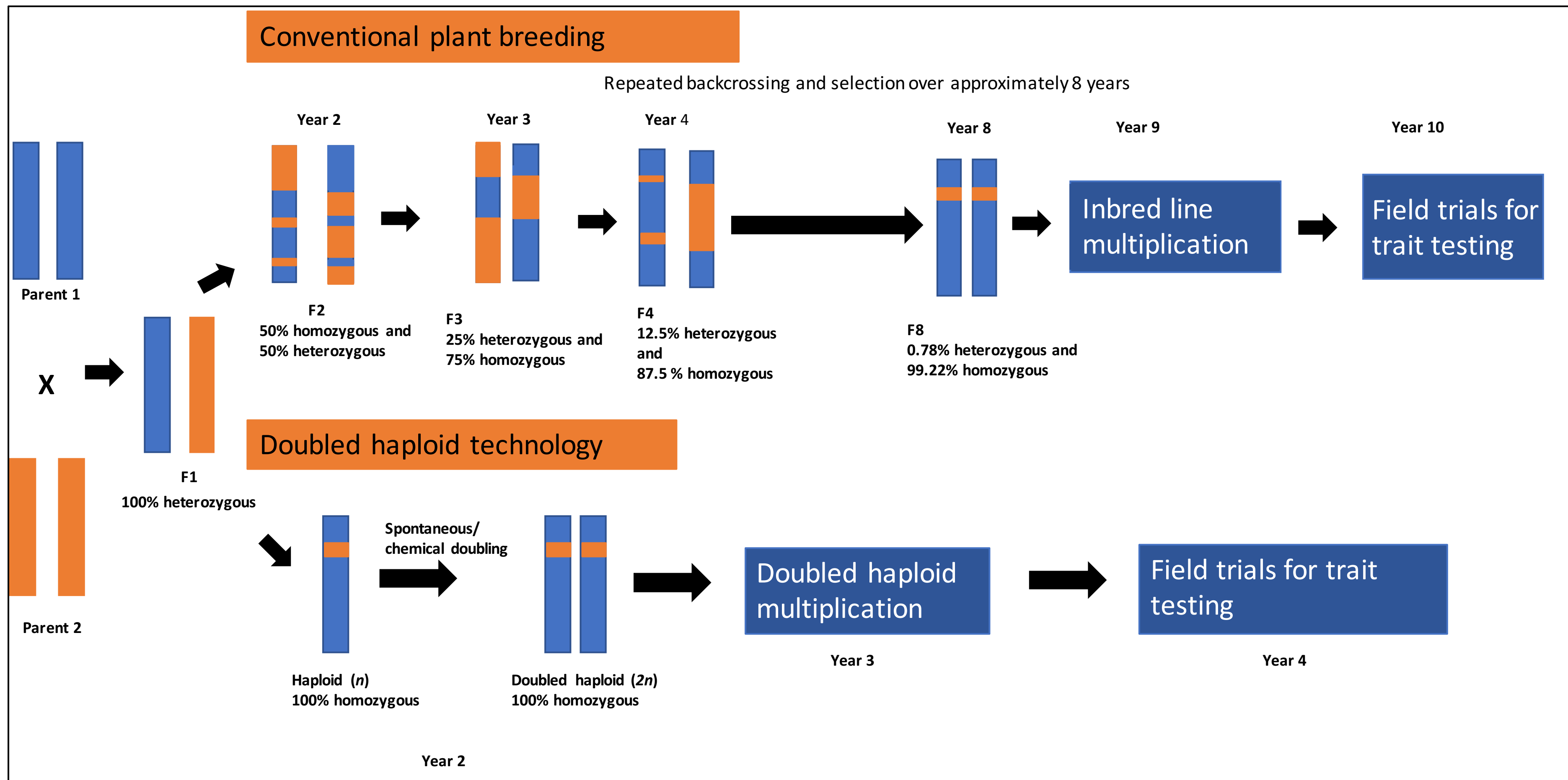


Figure 1.1 Comparison of conventional breeding and doubled haploid technology breeding methods. (Illustration adapted and modified from Eliby et al. (2022).

Doubled haploids can also be employed to rapidly generate mapping populations e.g. chromosome substitution lines (Ishii et al., 2016). Naturally occurring haploids have been reported for several cereal crops, including rice, wheat and maize (Dwivedi et al., 2015). The recovery frequency of these is, however, too low for practical applications and therefore difficult to manage in breeding programs. (Britt and Kuppu, 2016). Efforts for the production of doubled haploids and application of DH lines in plant breeding have been made for numerous oilseed crops, including sunflower but with only a few successes (Table 1.1) (Dunwell, 2010; Ren et al., 2017). Technologies for haploid induction are available for fewer than thirty plant species, which include maize, oats, rice, and wheat (Dunwell, 2010; Dwivedi et al., 2015). These technologies include pollen irradiation, twin embryo seed selection, alien cytoplasm, sparse pollination, wide hybridization and microspore culture (Dunwell, 2010; Wang et al., 2019). Both microspore culture and hybridization were demonstrated as the most successful methods and have become the foundation of modern doubled haploid technology in plant breeding (Ishii et al., 2016; Ren et al., 2017).

Table 1.1 Doubled haploid induction methods reported and used in oilseed breeding.

Plant Species	Haploid induction method	Application	References
Sunflower (*H. annuus*)	Parthenogenesis	Resistance to broomrape, phoma, imidazolinone and downy mildew	Drumeva et al. (2014, 2017), Todorova et al. (1997)
	Anther culture	Fertility restoration	Bohorova (1985); Saji and Sujatha (1998); Jonard and Mezzarobba (1990)
Brassica napus L. (Rapeseed)	Spontaneously occurring	Oil yield enhancement	Reviewed by Kucera et al. (2002)
Brassica spp.	Microspore culture/Embryogenesis	Modification of fatty acid profiles, fatty acid levels and drought tolerance	Reviewed by Ferrie et al. (2016) Daurova et al. (2020)

Microspore culture haploid induction technologies include androgenesis which is described as: the generation of haploids from male gametic material (paternal haploid induction) (Murovec and Bohanec, 2012). Androgenesis, involves *in vitro* cultivation of immature anthers under aseptic conditions to allow the conversion of the developmental pathway of immature pollen grains from gameophytic to sporophytic (Murovec and Bohanec, 2012). Androgenesis is established in haploid induction for plant breeding purposes and has been successfully demonstrated in species such as *Brassica napus* (rapeseed). *Solanaceae* species, *Triticum aestivum* (wheat), *Zea mays* (maize) (Murovec and Bohanec, 2012). Gynogenesis (maternal haploid induction) is haploid induction through the production of haploid embryos from unfertilized female gametophytes (Murovec and Bohanec, 2012; Wang et al., 2019). This is achieved through *in vitro* culture of un-pollinated whole flower buds or flower parts such as, ovules, ovaries and the placenta. Although haploid induction through gynogenesis has been achieved in various species including sunflower, its application in plant breeding is limited to *Beta valguris L* (Sugar beet) and *Allium cepa* (Onion) (Murovec and Bohanec, 2012).

Wide hybridization is a widely demonstrated haploid induction method that involves interspecific or intraspecific crosses that result in the selective loss of the parental chromosomes (Dunwell, 2010; Ishii et al., 2016; Wang et al., 2019). The first example of selective hybridization haploid production was first demonstrated in barley (Kasha and Kao, 1970 as referenced by Dunwell 2010) where crosses between *Hordeum vulgare* and *Hordeum bulbosum* resulted in a haploid barley variety after embryo rescue since the hybrid endosperm causes abortion (Wang et al., 2019). Other common examples of hybridization involve crosses between *Nicotiana tabacum* and *Nicotiana sylvestris* which resulted in the loss of *N. sylvestris* chromosomes in the progeny and haploid plants with only *N. tabacum* chromosomes (Ishii et al., 2016; Ren et al., 2017). This approach has also been applied to other cereal crops too, for example a selective cross between wheat and maize (Dunwell, 2010; Kalinowska et al., 2019), where maize chromosomes were eliminated resulting in haploid wheat progeny (Dunwell, 2010). Wheat X pearl millet; pear X apple are also some examples of selective hybridization (Ren et al., 2017); it was revealed that the eliminated chromosome's *centromere histone 3* (*CenH3*) gene was either removed or smaller in size when compared to the retained chromosome, an association that provided insight on haploid induction via hybridization (Dunwell, 2010; Finch, 1983; Wang et al., 2019). The development of haploid embryos through hybridization has been widely achieved in many species, including in lettuce, where embryos were crossed with sunflower and other genetically related species (Piosik et al., 2016). The use of hybridization for haploid induction has not been explored extensively and no data on sunflower haploid induction through hybridization exists in literature (Blinkov et al., 2022).

The majority of these haploid technologies are time consuming, costly, genotype-dependent and have too low efficiencies to be applied in sunflower breeding (Kelliher et al., 2017; Ren et al., 2017). Doubled haploid induction methods, such as microspore culture and irradiation of pollen grains, have been tested on sunflower with very little success (Coumans and Zhong, 1995). Although anther culture (Bohorova et al., 1985) and parthenogenesis (Todorova et al., 1997) have been partially successful in the induction of haploids in sunflower, but they are genotype specific and of limited use (Coumans and Zhong, 1995; Dagustu, 2018). The genotype specificity and limited usage of anther culture and parthenogenesis techniques (Table 1.1) make them inappropriate for large scale haploid induction especially since there are numerous genotypes involved in breeding programmes.

1.2.1 Haploid induction by Centromeric Histone Protein 3 (*CenH3*) modification

The centromere is a crucial locus responsible for efficient and stable transfer of genetic material from gametes to embryos by mediating sister chromatid segregation (Figure 1.2)

(Maheshwari et al., 2017; Zamariola et al., 2014). The centromere serves as a binding site for the kinetochore complex (Britt and Kuppu, 2016; Watts et al., 2017). Spindle fibres bind to the kinetochore complex and pull chromatids towards opposite poles for proper segregation (Watts et al., 2017). This process is facilitated by several centromeric proteins, including the centromeric histone protein (CENH3), which plays a critical role in chromosome segregation (Britt and Kuppu, 2016). The CENH3 protein acts as an assembly point for other kinetochore proteins (Nagaki et al., 2015). The CENH3 protein is exclusively localized to functional centromeres, which is a key feature in distinguishing the centromere from surrounding pericentromere (Ishii et al., 2016; Maheshwari et al., 2017). It has been suggested that the CENH3 protein is the only protein required for fully functional centromeres (Watts et al., 2017).

Ravi and Chan, (2010) demonstrated a centromere modification method in *Arabidopsis thaliana* through which haploid inducer plants were produced. In this method, *a centromere histone 3* null mutant (*Cenh3*), with a mutated version of the CENH3 protein was crossed with a wild type plant that resulted in only the wild type plant's chromosomes, with functional CENH3 proteins, being transferred to the progeny and none of the mutant's (Figure 1.2) (Plohl et al., 2014). Incomplete uniparental chromosome elimination can occur however, leading to a percentage of aneuploid progeny (Figure 1.2)

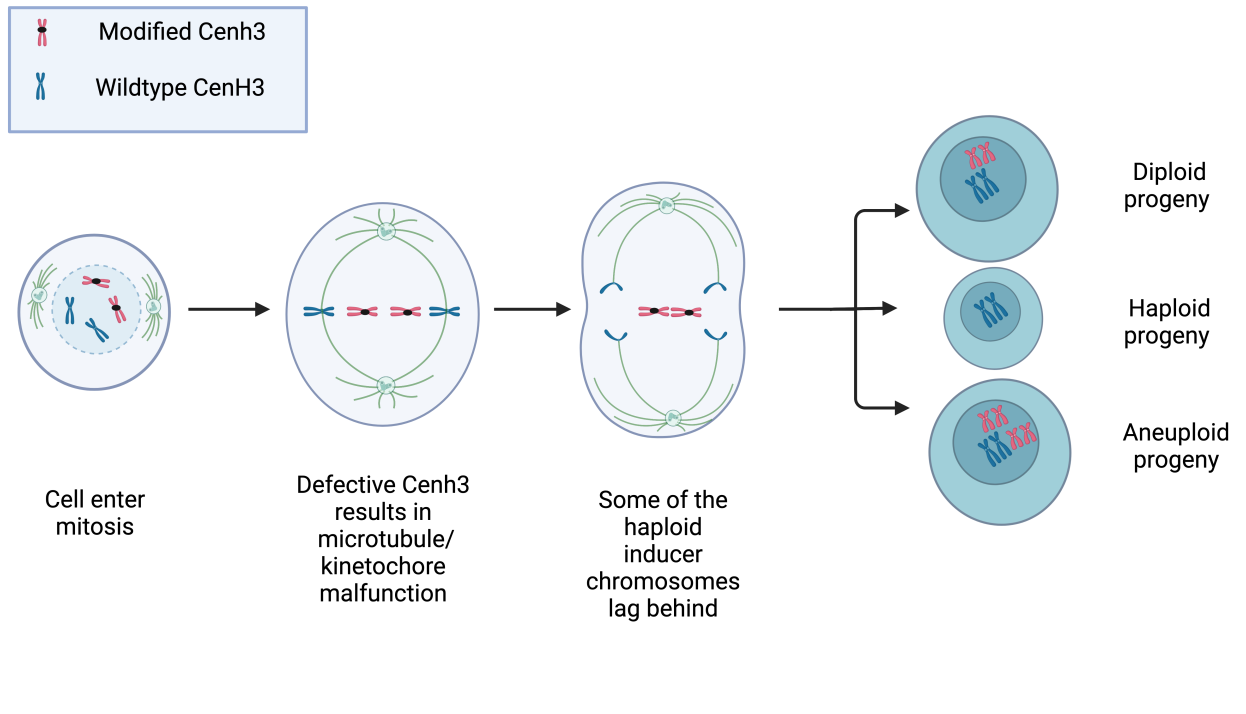


Figure 1.2 Illustration of centromere function during zygote mitosis. A plant with modified *CenH3* gene is crossed with a wild type plant resulting in uniparental chromosome elimination leading to haploid progeny and/or incomplete elimination of haploid inducer genetic material leading to aneuploid progeny (illustration adapted from Chan, 2010). Created with BioRender.com. (Accessed 11 January 2023).

The CENH3 protein consists of two domains: these are a highly variable N-terminal and a highly conserved C-terminal Histone Fold Domain (HFD). The N-terminal is variable in both sequence and length and may even vary within species (Britt and Kuppu, 2016; Kuppu et al., 2015), while the HFD is conserved across species (Ravi et al., 2014). The N-terminal contains a single alpha helix (Figure 1.3), while the HFD region contains three alpha helices separated by two loops (Figure 1.3). Loop 1 and the alpha helix play a role in centromere targeting and this region is known as the centromere targeting domain (CATD). The HFD part of the CENH3 has been identified as the essential part of centromere localization during mitosis, even in the absence of the N-terminal. However, the variable N-terminal tail was observed to play a critical role in meiosis (Lermontova et al., 2011; Ravi and Chan, 2010; Zamariola et al., 2014). The N-terminal tail is also responsible for recruitment and stabilization of centromere complex proteins (Lv et al., 2020).

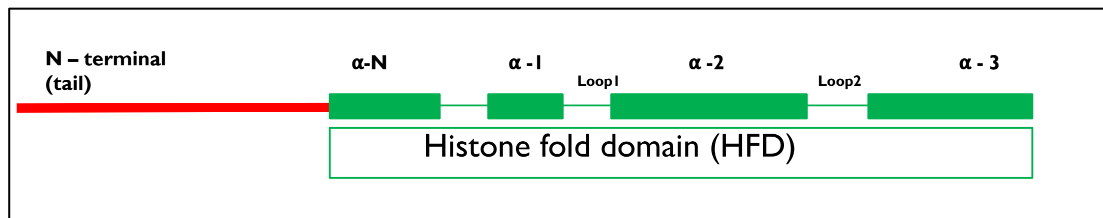


Figure 1.3 Typical structure of the plant centromeric histone protein 3 (CENH3). The CENH3 protein is comprised of a highly divergent tail (N-terminal in red) and highly conserved histone fold domain (green) containing three alpha helices and loops.

There are currently two approaches for applying the haploid inducer method. The first method is based on a two-step approach where a 'mutant tailswap' is used to rescue modified CENH3 transgenes described by Ravi and Chan (2010). In this approach, genetically modified CENH3 variants must be cloned and partially expressed to complement the knockout mutation (Britt and Kuppu, 2016; Kalinowska et al., 2019). The complementation of the hypervariable N-terminal with histone H3.3 and fusion with the green fluorescent protein (GFP) in *Arabidopsis thaliana* resulted in sterile plants with meiotic defects and crossing with wildtype plants resulted in chromosome missegretation and elimination of the mutant's genetic material (Ravi and Chan, 2010). The "tailswap" method has resulted in haploid progeny between 25 - 45% when the inducer line was maternal in *Arabidopsis*. This has been tested in some crop species but had relatively low efficiencies, i.e. 0.065 - 0.86% haploid progeny in maize, 0.2 - 2.3% in tomato and 0.3 - 1% in rice (Wang et al., 2019). In hexaploid wheat containing three copies of the gene namely *CenH3*α-A, B and D, mutations with amino acid changes in the N-terminal tail called the restored frameshift (RFS) in *CenH3*α-A coupled with a knockout of *CenH3*α-B and *CenH3*α-D were induced. This triggered paternal haploid induction with frequencies of up to 8% (Lv et al., 2020). This approach has been used in cereal crops such as maize, wheat and *Brassica* (Lv et al., 2020; Wang et al., 2019).

The second approach is a one-step method, based on the targeting of the endogenous *CenH3* gene by *CenH3* gene silencing using RNAi and gene knockouts via CRISPR/Cas9 (Britt and Kuppu, 2016). Kuppu et al. (2015) and Karimi-Ashtiyani et al. (2015) presented a haploid induction method that studied a variety of ethyl methane sulfonate (EMS) induced mutations in the conserved Histone fold domain (HFD) regions of the *CenH3* gene in *Arabidopsis,* barley and sugar-beet which resulted in haploid inducers (when out-crossed with wild type *CenH3* plants). Point mutations targeting five amino acids; P82S; G83E; A132T; A136T; and A86V produced paternal haploids at a rate of 0.61% to 12.2% that were normal in appearance and fully fertile when self-fertilized (Karimi-Ashtiyani et al., 2015; Kuppu et al., 2015; Wang et al.,

2019). This indicates that the CENH3 from the mutant plants remained functional but incapable to compete with the normal 'wildtype' CENH3 proteins (Kuppu et al., 2015; Ren et al., 2017). Further approaches have demonstrated complete deletions of the α-N helix of the HFD region resulting in haploids upon crossing with wildtype in *A. thaliana* (Kuppu et al., 2020). This single step method provides ease to the previously described 'tailswap' method of combining a chimaeric transgene with a gene knockout which may be difficult to engineer and the applicability may be reduced (Kuppu et al., 2015; Ren et al., 2017). The idea behind the CENH3 modification method is that interference with the centromere structure may result in a less competitive or weak centromere when confronted with a non-modified or wildtype centromere (Britt and Kuppu, 2016). The CENH3 haploid inducer methods result in non-transgenic haploid plants as the chromosomes of the transgenic haploid inducer are eliminated (Kalinowska et al., 2019). Due to the universal structure and function of the CENH3 protein in plants, the use of gene editing haploid induction method promises wide applicability and success when compared to previously described haploid production methods (Britt and Kuppu, 2016). Various gene altering tools have been utilised to achieve modification of the *CenH3* gene for haploid inducer line production, including RNAi, chemical mutagenesis and more recently, CRISPR/Cas9 (Britt and Kuppu, 2016; Karimi-Ashtiyani et al., 2015; Ravi and Chan, 2010; Wang et al., 2021).

1.3 METHODS OF GENOME MODIFICATION

As the demand for agricultural output increases with an increase in the human population and environmental extremities, the need to accelerate the development of crops with invaluable traits will also increase. Among the available technologies used in crop improvement, mutations have been the driving force behind plant improvement, often manipulated by breeders to increase yields and enhance crop quality (Songstad et al., 2017). Mutations are heritable changes in the DNA sequence of a living organism that do not result from genetic recombination or genetic segregation (Pathirana, 2011). Mutation breeding generates crops with desired traits by natural, chemical, physical or biological induction (Pathirana, 2011).

Spontaneously occurring mutations resulting in within-species phenotypic variation are quite a common and necessary phenomenon for the maintenance of the evolutionary process in living organisms (Alonso-blanco et al., 2009). Plants are often exposed to environmental stressors, such as UV radiation, dehydration, pollution from phytochemicals, amongst others, that may generate DNA damage and chromosomal alterations (Zhao and Wolt, 2017). Scientists have also been able to mimic mutations that generate both random and targeted genetic variation for the development of traits in crop breeding. There are several genome

modification methods utilized in crop species including epigenetic modification, randomly induced mutagenesis and targeted new breeding technologies.

1.3.1 EPIGENETIC INHERITANCE AND NEW TRAIT DEVELOPMENT

A large portion of naturally occurring variation in a population can be attributed to epigenetic mechanisms of gene regulation (Alonso-blanco et al., 2009). Epigenetic modifications can be referred to as heritable changes in the chromatin that do not affect the DNA sequence order in a cell (Ikeuchi et al., 2015). Epigenetic modifications, such as DNA methylation, histone variants and histone modifications, have been suggested to play a pivotal role in cellular differentiation necessary for plant development and reproduction regulation (Feng and Jacobsen, 2011; Ikeuchi et al., 2015). The single genome in a plant cell has the ability to produce multiple epigenomes as a developmental and environmental stress response mechanism (Chinnusamy and Zhu, 2009). Epigenetic marks enhance mechanisms to allow for rapid adaptation to changing environmental conditions and are therefore critical and necessary for gene regulation (He et al., 2011). Some epigenetic marks have the ability to be inherited across generations, e.g. epigenetic transgenerational inheritance of environmentally induced phenotypic changes (Skinner, 2015). It is therefore evident that epigenetic inheritance plays a major role in microevolution (Skinner, 2015).

DNA methylation is quite a common phenomenon in plants and can be described as the random addition of a methyl group to the fifth position of the cytosine ring of the DNA molecule resulting in the formation of a 5-methylcytosine (Diez et al., 2014; Feng and Jacobsen, 2011; He et al., 2011). Methylation is conserved in most plant species and is generally restricted in the regions of Cytosine (C) Guanine (G), CHG and CHH (H = Adenine (A), Cytosine(C) or Thymine (T) (He et al., 2011). Repetitive genomic regions, such as centromeric and pericentromeric regions, are highly methylated in plants while promoter regions are hardly methylated (He et al., 2011). In plants *de novo* methylation is governed by RNA dependent methylation (RdDM) (Diez et al., 2014). Methylation within repeat regions is a mechanism to limit the transcription and proliferation of transposable elements (TEs) thereby providing defense against potentially deleterious mutations (Diez et al., 2014). Changes in DNA structure resulting from methylation can potentially affect gene expression during plant development and stress (Diez et al., 2014). Active DNA methylation in plants is also responsible for development, DNA repairand gene silencing (Zhao and Wolt, 2017). Although DNA methylation is a natural occurrence in plants, there are technologies that allow scientists to induce artificial DNA methylation with methods such as viroids and laser irradiation amongst others (Pacher and Putcha., 2017).

1.3.2 RANDOMLY INDUCED MUTAGENESIS AS A TOOL TOWARDS TRAIT DIVERSIFICATION: PHYSICAL AND CHEMICAL MUTAGENESIS

Plant breeders are dependent on randomly occurring mutations for the generation of new traits (Songstad et al., 2017). Over 3000 novel cultivars have been developed using induced mutagenesis, with a range of traits including increased yield, early maturity being enhanced through physical and chemical mutation breeding (Petolino, 2015). Mutations have been successfully induced in plants by exposure of certain plant parts or whole plants to physical and chemical mutagenic agents; these mutations introduce changes in DNA sequences of the target organisms resulting in changes in the characteristics of the said organisms (Zhao and Wolt., 2017).

Physical mutagenesis is usually through electromagnetic radiation, such as gamma rays, X-rays and UV light. Induced mutagenesis gained traction in plant breeding following pioneering studies on the physical mutagenesis of *Hordeum vulgar* L. (barley) by exposure to X-rays and radium (Songstad et al., 2017; Mba et al., 2010). Mutants obtained using X-rays, gamma-rays, beta and ultraviolet radiation then slowly gained popularity and became part of the breeder's toolbox (Songstad et al., 2017). Radiation mostly results in structural mutations such as chromosome breaks and rearrangements (Pathirana, 2011). This method has been successfully used in several plant species including crops such as wheat and maize (Mba et al., 2010).

Chemical mutagens, making use of alkylating agents and intercalating agents, were also identified for use in genetic and breeding research (Zhao and Wolt., 2017). Alkylating agents make up 80% of the total number of mutagens used in plants (Pacher and Puchta, 2017). Alkylation involves the substitution of an alkyl group (C_2H_5) with hydrogen in the nitrogen bases. Alkylating agents normally give rise to 'base-pair' substitutions where alternate bases are incorporated during replication (Pathirana, 2011). These agents' bond with guanine (G) in the DNA sequence to form an abnormal alkyl-guanine base that is recognized as an adenine (A) instead of a G which results in a transition mutation (Pacher and Puchta, 2017). Ethyl methane sulfonate (EMS) is the most widespread alkylating agent since it is easy to use, handle and is effective (Pathirana, 2011). Nitroso compounds are also popular, but require more precautionary measures (Pathirana, 2011). Chemical mutagens are generally considered to be more efficient and specific than physical mutagens (Mba et al., 2010), but both could be used simultaneously to enhance mutation frequency (Songstad et al., 2017). Induced and natural mutagenesis methods are both non-specific in nature and can be disadvantageous because screening for usable mutations can be time consuming and costly. The main advantage of using induced mutagenesis technologies however, is that they are not

considered genetically modified (GM) and as such, they do not exist under the GM regulatory and risk assessment frameworks (Zhao and Wolt, 2017).

1.3.3 New plant breeding technologies: Targeted gene modification

Target specific modifications driven by customizable direct nucleases provide an effective tool to overcome the limitations that arise from random genome modification (Yidong et al*.,* 2017). Zinc Finger Nucleases (ZFNs), Transcription Activator Like Effector Nucleases (TALENs) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) editing systems are the three main approaches used for target specific mutagenesis by inducing double stranded DNA breaks in various organisms (Khatodia et al., 2016). These genome editing approaches provide simplicity and precision, and this offers a range of opportunities for the development of novel phenotypes (Petolino, 2015). There are currently two kinds of sequence-specific genome editing tools: Protein-binding complexes which comprise of ZFNs, TALENs and meganucleases and RNA and endonuclease guided systems including CRISPR and targetrons (Guha et al., 2017). ZFNs, TALENs and CRISPR are however the most commonly applied in research and will be discussed further in this review.

Ribonucleic acid interference (RNAi)

Ribonucleic acid interference (RNAi) is a naturally occurring mechanism used by eukaryotic organisms for regulation of transposable elements, destruction of invading foreign viruses and prevention of homologous chromosomes (Muiruri, 2015). RNAi for artificial gene silencing is based on the production of homologous double stranded RNA (dsRNA) homologous to the target gene with hairpin formation (Senthil-Kumar and Mysore., 2010). This system is processed into small interfering RNAs (SiRNAs, 21 – 24 nucleotides) which provide target sequence specificity of an endonuclease that facilitates the degradation of homologous RNA sequences (Senthil-Kumar and Mysore., 2010). RNAi was one of the first methods utilized for CENH3 suppression for the purpose of haploid induction in plants and has been used for CENH3 silencing in *A. thaliana* (Ravi and Chan., 2010; Lermontova et al., 2011) and banana (Muiruri., 2015). RNAi has also been previously used for gene silencing in sunflower to induce resistance against the Tobacco streak virus (Pradeep et al., 2012). RNAi is however inefficient and often results in incomplete gene silencing (Muiruri 2015), furthermore, RNAi does not result in permanent mutations (Rajput et al., 2021).

Zinc finger nucleases (ZFNs)

Zinc fingers are artificial, sequence-specific DNA-binding proteins that are fused with a non-specific nuclease domain derived from the type II FokI enzyme (Lee et al., 2016; Petolino, 2015). The Cys2His2 type zinc fingers are the widest spread DNA binding protein domains (Razin et al., 2011). These proteins are then customized to cut at specific DNA sequences for the purpose of introducing double strand breaks (DSBs) in DNA of living cells (Marton et al., 2010). A functional FokI nuclease domain is a dimer, therefore two zinc finger nuclease (ZFNs) monomers, one on the forward and one on the reverse strand, are required for successful DNA cleavage to occur (Lee et al., 2016). These monomers need to be properly aligned for efficient dimerization and nuclease activity (Marton et al., 2010). This dimerization of ZFNs extends the length of the recognition site, which further increases ZFN specificity (Lee et al., 2016). A tethered array of between 4-6 zinc finger protein domains are responsible for recognizing about 3 bp of DNA (Petolino, 2015). The specificity of zinc finger proteins can be manipulated by mutagenesis thereby creating a designed nuclease to target a sequence of your choosing (Lee et al., 2016). Zinc finger proteins with desired specificities can be constructed by modularly assembling pre-characterized zinc fingers (Lee et al., 2016). Zinc fingers have been used for target specific gene modification in various crop and model species for both transgenic and transient expression (Lee et al., 2016). Zinc finger nucleases have been delivered through *Agrobacterium*- transformation in numerous of crop species (Petolino, 2015).

Several plant-virus vectors have also been utilised for transient expression, including a Tobacco rattle virus-based expression vector to achieve expression of ZFNs in model *Nicotiana benthamiana* plants and commercially significant crop *Petunia hybrida* (Marton et al., 2010). Zinc-finger nucleases can however be attributed to some disadvantages compared with some of the newly developed direct nucleases (Lee et al., 2016). Zinc-fingers unlike TALENs have a narrow target availability and the construction process is quite tedious and expensive (Abdallah et al., 2015; Lee et al., 2016), some ZFNs are often unable to cause DSBs as efficiently in regions with high guanine repeat sequences (Kim et al., 2009). Zinc finger nucleases have also shown low sequence specificity and affinity resulting in a great deal of off-target mutations (Zhao and Wolt, 2017).

Transcription activator like effector nucleases (TALENs)

Transcription activator like effector nucleases (TALENs) are similar to ZFNs. A FokI derived domain is paired with a sequence specific, DNA binding domain – in this case transcription activator-like effectors (TALEs) (Joung and Sander, 2013). These TALEs are derived from

Xanthomonas spp. plant pathogenic bacteria (Lee et al., 2016). In nature, TALEs are introduced into infected plant cells by a type III secretion system and bind to genomic DNA to affect transcription and facilitate bacterial colonization (Joung and Sander, 2013; Wood et al., 2011). Unlike ZFNs, TALENs are based on a one-to-one match between protein repeats and nucleotide sequences, thereby significantly decreasing the frequency of off-target mutations (Zhao and Wolt, 2017). TALENs comprise of 33 to 35 amino acid repeats, and the recognition of a specific site is governed by amino acids at positions 12 and 13 – known as repeat variable Di-residues (Lee et al., 2016). During DNA cleaving, two TALEN monomers bind to target DNA sequences, thus allowing FokI domains to dimerize and cleave the target sequence (Zhang et al., 2013). Each repeat in the binding domain recognizes one complementary target nucleotide (Zhang et al., 2013). The required dimerization for cleavage in TALENs also enhances the specificity of the system, making them specific when compared to monomeric enzymes (Guha et al., 2017).

TALENs have been introduced into plant cells for stable expression through *Agrobacterium tumefaciens* mediated transformation and transiently through viral vectors (Guha et al., 2017). TALENs have efficiently induced genome modifications in crop species such as rice and *Brachypodium* (Zhang et al., 2013). The disadvantages are similar to those of ZFNs, with a new engineered TALEN protein, though easier to design than ZFNs, still required for each DNA target site (Zhao and Wolt, 2017). TALENs have however demonstrated more ease in application and functionality when compared to ZFNs (Zhang et al., 2013).

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system is a defense mechanism that protects bacteria and archaea against viruses and other mobile genetic elements of foreign origin (Chen et al., 2019; Khatodia et al., 2016; Rozov et al., 2019). The CRISPR-associated protein (Cas) mediated adaptive defence system in prokaryotes uses stretches of RNA sequences in combination with nucleases that cleave DNA at specific sites to suppress these invasions (Ahmad et al., 2019; Khatodia et al., 2016). This process is divided into three stages: adaptation, expression and interference. During adaptation, an insertion of the invading agent's sequence into spacers known as CRISPR RNA (crRNA) of the bacterial CRISPR locus occurs, thereby providing the guide for targeting the virus with this immunity system (Schiml and Puchta, 2016). A second RNA molecule, trans-activating RNA (tracrRNA) is responsible for recruiting the Cas9 endonuclease and binding onto the target DNA (Ahmad et al., 2019; Feng et al., 2016). In the second stage, these two RNA molecules form a complex with the Cas nuclease and this complex binds to the identified target site

(Schiml and Puchta, 2016). In the third stage, interference, the combined action of crRNA, tracrRNA and Cas protein results in the recognition and degradation of the viral target DNA (Rath et al., 2015).

The bacterial CRISPR/Cas based defense mechanism has been altered and developed to allow targeted double strand DNA breaks in various organisms (Khatodia et al., 2016). In the altered version of the CRISPR/Cas technology applicable for genome editing, the crRNA and tracrRNA have been combined into a single 20-nucleotide molecule known as a small guide RNA (sgRNA or gRNA) (Ahmad et al., 2019). The sgRNA and Cas nuclease form a sgRNA/Cas complex (Figure 1.4), with the sgRNA necessary and designed for target site recognition (Jinek et al., 2012). The sgRNA is associated with a short nucleotide sequence (approximately 3-6 nucleotides) known as the Protospacer Adjacent Motif (PAM). The PAM sequence depends on the bacterial species (Chen et al., 2019), for example NGG in the case of *Streptococcus pyogenes*. The presence of the PAM sequence downstream of the sgRNA activates the two Cas9 domains and leads to cleavage of the target site (Zhang et al., 2019). The Cas enzyme typically induces double stranded breaks approximately 4 nucleotides upstream the PAM sequence (Ahmad et al., 2019).

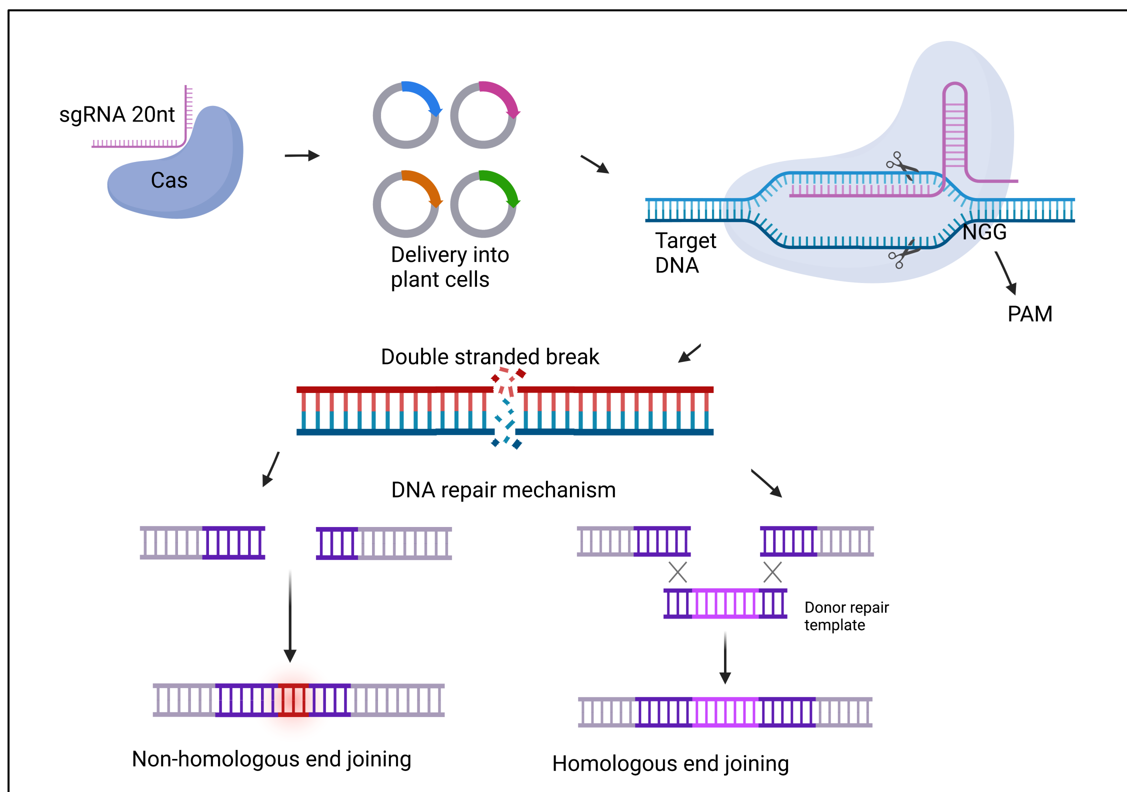


Figure 1.4 Schematic representation of the CRISPR/Cas9 genome editing mechanisms. The Cas9 endonuclease is recruited by a site-specific small guide RNA (sgRNA) to the target site preceded by a

PAM sequence. Once positioned, the Cas9 induces a double stranded break which prompts the plant's natural repair systems, either activating Homology directed repair (HDR) or Non-homologous end joining (NHEJ). Illustration adapted from (Khatodia et al., 2016), Created with BioRender.com. (Accessed 11 January 2022).

The sequence specificity of the Cas9 enzyme permits the targeting of numerous loci and can be easily modified both *in vitro* and *in vivo* by the provision of artificially manufactured guide RNAs (Jinek et al., 2012). CRISPR systems have been categorised into two classes, which have been subdivided into six types based on their Cas genes (Table 1.2) (Chen et al., 2019). Class I systems (type I, III and IV) require multiple-Cas protein complexes for interference while class 2 systems (type II, V and VI) only require single protein complexes with crRNA (Chen et al., 2019). The type II CRISPR/Cas system derived from *Streptococcus pyogenes* was the first system demonstrated for specific gene targeting and cleaving in eukaryotic cells (Chen et al., 2019).

Table 1.2 Different classes and types of Cas proteins applicable for plant genome editing.

Class	Type	Subtypes	Origin species	Signature protein
		I-B	*Clostridium kluyveri*	Cas3, Cas8
		I-C	*Bacillus halodurans*	Cas3, Cas10
		I-D	*Cyanothece* sp.	Cas3, Cas8
		I-E	*Escherichia coli*	Cas3, Cas8
		I-F	*Yersinia pseudotuberculosis*	Cas3, Cas8
		I-U	*Geobacter sulfurreducens*	Cas3, Cas8
	III	III-A	*Staphylococcus epidermidis*	Cas10
		III-B	*Pyrococcus furious*	Cas10
		III-C	*Methanothermobacter thermaautotrophicus*	Csf1
		III-D	*Roseiflexus sp*	Cas9
	IV	IV	*Acidithiobacillus ferrooxidans*	Cas9
Class 2	II	II-A	*Streptococcus thermophilus*	St1Cas9
		II-A	*Streptococcus pyogenes*	SpCas9
		II-B	*Legionella pneumophila*	Cas2
		II-C	*Neisseria lactamica*	Cas9
	V	V	*Francisella cf novicida*	Cas12a/CpF1
	VI	VI	*Leptotrichia shahii*	C2c2

Adapted and modified from: Ahmad *et al*., (2019) and Makarova et al. (2015) to only include Cas proteins previously used for gene editing in plants.

The Cas protein is characterised by two nuclease domains RuvC and HNH which cleave complementary and non-complementary DNA strands respectively (Chen et al., 2019). To expand the scope of gene editing, new variants of the Cas protein have been developed (Khatodia et al., 2016). The Cas9 enzyme has been catalytically deactivated by creating a mutation in the RuvC and HNH nuclease domains (dead Cas9 or dCas9) and is able to act as a gene expression suppressant and activator (Khatodia et al., 2016; Song et al., 2016). Co-expression of dCas9 and a guide RNA targeting the coding region of a gene results in prevention of the transcription elongation process which may completely inhibit gene function (Khatodia et al., 2016; Song et al., 2016). Furthermore, dCas9/sgRNA can be fused with a repressor or activator protein to either reduce or enhance gene expression (Glass et al., 2018; Khatodia et al., 2016). A nickase variant of the Cas9 nuclease was also developed by the substitution of aspartate to alanine in the native Cas enzyme (D10A) (Khatodia et al., 2016), and instead of a double stranded break, this version of the Cas9 enzyme is able to induce a single stranded nick (Songstad et al., 2017). This system has been used for multiplexing

purposes by pairing with two or more sgRNAs (Khatodia et al., 2016). Other alternatives to Cas9 include the type II CpF1(Cas12a); and this is also an endonuclease based on the same principle as Cas9 but uses a different PAM sequence (TTTN) (Glass et al., 2018; Schindele et al., 2018). CpF1 sgRNA sequence is also slightly shorter than that of the Cas9 45 nucleotides for CpF1 compared to 110 nucleotides for Cas9 (Glass et al., 2018; Kelliher et al., 2019; Schindele et al., 2018). This alternative endonuclease has only been tested in mammalian cells and functionality in plants still needs further investigation (Chen et al., 2019). CRISPR has been used to cause a mutation without the breakage of DNA in rice and tomato; this was achieved by 'base editing', which involves the conversion of one base to another (Shimatani et al., 2017; Zong et al., 2017). A high frequency of off-target mutations has been reported for CRISPR applications in mammalian cells but the off-target mutation frequency is almost non-recognizable in plant cells (Yin et al., 2015). CRISPR/Cas9 has been around for almost a decade, with the first successful experiment in plants reported in the generation of mutagenesis events in *Arabidopsis thaliana*, *Nicotiana benthamiana* and *Triticum aestivum* towards the end of 2013 (Feng et al., 2013; Schaeffer and Nakata, 2016).

The CRISPR/Cas technology is more advantageous than both ZFNs and TALENs as it does not require complicated designs and the assembly of DNA binding proteins, it only requires a single Cas9 nuclease that can be programmed by engineering the guide RNA (gRNA) to direct target-specific cleavage (Zaidi and Mansoor, 2017). Additionally, CRISPR has the ability to target methylated DNA, further increasing its applicability (Zhao and Wolt, 2017). The application of CRISPR/Cas9 based genetic editing has increased greatly and rapidly in plants, it has been effectively evaluated in both transgenic and transient plant experiments (Yin et al., 2015).

1.4 GENOME EDITING OUTCOMES - DOUBLE STRAND-DNA BREAK REPAIR MECHANISMS (DSBS)

The targeted genome modification technologies mentioned above achieve precise modifications by inducing DNA double strand breaks (DSBs) linked to a target sequence(s) (Figure 1.4). Lesions caused by DSBs may obstruct replication fork progression resulting in the loss of genome integrity if DSB detection and repair systems are not in place (Ahmad et al., 2019; Ceccaldi et al., 2016). The ability to introduce and repair DSBs in cells is facilitated by recombination between DNA sequences and forms a crucial part of generation of genetic variability and ensuring that the cell's genome remains intact (Petolino, 2015; Puchta, 2005). These DSBs will result in the cell's DNA sequence repair system being manipulated to

generate desired genetic outcomes, i.e. either inducing the non-homologous end joining repair (NHEJ) (Petolino, 2015) or homology directed repair (HDR) (Petolino, 2015).

The non-homologous end joining repair (NHEJ) is based on the rejoining of double stranded DNA breaks either without or with small sequence changes (indels) (Bleuyard et al., 2006; Schmidt et al., 2019). This mechanism occurs throughout the cell cycle and does not require an exogenous repair template (Chen et al., 2019). Two different pathways of NHEJ occur dependant on the presence of the KU heterodimer, namely canonical-NHEJ (cNHEJ), which results from the direct ligation of the broken ends leading to small insertions or deletions. This can result in frameshift mutations and may lead to complete gene function loss (Schmidt et al., 2019). Another NHEJ pathway is the micro-homology-based end joining (MMEJ) repair, which takes over in the absence of the KU heterodimer, this usually requires the presence of 1-20 nucleotide homologies near the break site (Schmidt et al., 2019). The NHEJ system appears to be more effortless and thus prevalent in plants (Bleuyard et al., 2006; Huang and Puchta, 2019), it is however error prone and therefore often causes gene knockouts (Puchta, 2005).

Homology directed recombination (HDR) is a DNA repair mechanism occurring mainly during the S and G_2 phases of the cell cycle (Puchta, 2005; Schmidt et al., 2019). This repair mechanism is based on the use of available sister chromatids, homologous chromosomes or other homologous DNA for synthesis (Petolino, 2015). There are two defined mechanisms for HDR repair in somatic plant cells (Schmidt et al., 2019; Singareddy et al., 2018). The single strand annealing (SSA) which is based on the availability of homologous repetitive sequences that are arranged in close proximity (Steinert et al., 2016) and the synthesis-dependent strand annealing (SDSA) where a homologue's sequence is copied into the break site immediately after a DSB (Schmidt et al., 2019). The SDSA pathway is the predominant HDR repair pathway and is beneficial for genome stability as no genetic information is lost (Steinert et al., 2016). Homologous recombination activity is generally low in eukaryotic cells (Danilo et al., 2019; Sawatsubashi et al., 2018) and is influenced by the availability and position of the homologous matrix; this could either be elsewhere within the genome (ectopic), from the homologue during meiosis (allelic) or from the same chromosome (intrachromosomal) (Puchta, 2005).

HDR allows for more precise insertions, substitutions, or replacement of gene regions, making it more accurate in the preservation of genetic information between the two homologous sequences (Bleuyard et al., 2006). Genome editing in plants via homologous-recombination-mediated DNA repair with CRISPR/Cas9 is inefficient and delivery of DNA repair templates is also challenging (Danilo et al., 2019). Homology directed repair mediated genome editing in

plants is rare, with only a handful of successful reports in crop plants (Table 1.3). Several factors influence the success of HDR experiments. Amongst these is the length of the provided donor sequence, with sequences typically ranging between 100 – 300 bp having reduced efficiencies compared to longer sequences of between 750 – 1000 bp. Sequences larger than this either have no effect or decrease knock in efficiency (Rozov et al., 2019). The structural nature of the donor template is also another factor to take into consideration, e.g. linear, rather than circular sequences, exhibited higher HDR efficiencies (Hahn et al., 2018; Rozov et al., 2019). Single stranded oligonucleotides (ssODNs) are small single stranded donor templates of up to 200 nt and have shown to increase knock-in efficiency – but have only been tested in mammalian cells (Rozov et al., 2019).

Other mechanisms of enhancing HDR in plants using the CRIPSR system involve the manipulation of plant repair enzymes. Endo et al. (2016) demonstrated that ligase 4 knockout resulted in enhanced HDR in rice. The proximity of the donor template to the break site, the presence of the donor template immediately after where the DSB takes place is also a factor to consider for HDR experiments (Gil-Humanes et al., 2017). To address this, Schiml et al. (2014) was able to demonstrate a methodology that allows the delivery of both the CRISPR reagents with the donor template using a single T-DNA vector termed the *in-planta gene targeting* system (IPGT) in *Arabidopsis.* In this method, the Cas9 enzyme cleaves both the target site and donor template, and this results in the stable integration of the donor template into the genome independent of transformation efficiency (Steinert et al., 2016). The use of viral replicons to increase the donor template copy number in CRISPR-based editing in plants cells has also been successfully demonstrated (Baltes et al., 2014; Čermák et al., 2015). However, a recent study comparing the IPGT and viral replicon CRISPR system in tomato, found the IPGT system to be superior with HDR rates of up to 12.7% (Hahn et al., 2018).

Table 1.3 List of successful CRISPR/Cas edited homology directed repair systems in crop and model plants.

Plant species	Target gene (s)	Delivery method	Cas protein	HDR efficiency	Homology arm's length	Reference
Arabidopsis thaliana	*ADH1* gene	*Agrobacterium*	spCas9	-	1347 bp	Schiml et al. (2014)
***Solanum lycopersicum (*Tomato)**	*Antho-cyanin* mutant 1(*ANT1*)	Geminivirus replicons	spCas9	-	1706 bp	Čermák et al. (2015)
***Oryza sativa* (Rice)**	*Acetolactate synthase* (*ALS*) gene	Particle bombardment	spCas9	-	146 bp	Sun et al. (2016)
Arabidopsis thaliana	*TFL1* gene	*Agrobacterium*		0.8%	1558 bp	Zhao et al. (2016)
***Triticum aestivu*m (Wheat)**		Particle bombardment; viral replicons	spCas9	1%	1520 bp/1321 bp	Gil-Humanes et al. (2017)
***Solanum lycopersicum* (Tomato)**	*DFR* gene	*Agrobacterium*	spCas9	1.29%	800 bp	Danilo et al. (2018)
Arabidopsis thaliana	*GL1*	*Agrobacterium*	spCas9	12,7%	1643 bp	Hahn et al. (2018)
***Solanum lycopersicum* (Tomato)**	*Carotenoid isomerase (CRTISO)* and *phytoene synthase 1 (PSY1)* gene	Bean yellow dwarf virus (BeYDV)	spCas9	25%	3796 bp	Dahan-Meir et al. (2018)
***Nicotiana tabacum* L. (Tobacco)**	*ALS* gene	*Agrobacterium*	spCas9	-	-	Hirohata et al. (2019)
***Solanum lycopersicum* (Tomato)**	*ALS* gene	*Agrobacterium*	spCas9	-	494 bp	Danilo et al. (2019)
***Oryza sativa* (Rice)**	*ALS* gene	Particle bombardment	CpF1; crRNA	1,7%	218 bp	Li et al. (2019)
***Nicotiana benthamiana* (Tobacco)**	*Gfp and Yfp* gene	Beet curly top virus replicon (BCTV)	Cas12a	-	238 bp	Eini et al. (2022)

1.5 DELIVERY METHODS FOR GENOME EDITING REAGENTS

Genome editing is steadily becoming the most established technology for crop improvement (Song et al., 2016; Yin et al., 2015). Central to the success of these technologies is the need for genetic material to be stably delivered into target plant cells. Genetic editing of plants can be defined in two stages: the transfer of the genetic material into the cell and the integration of the genetic material into the genome (Altpeter et al., 2005). The genetic material is either transiently expressed, which is only for a short time or stably expressed (Chen et al., 2019). The majority of genome engineering reagents come in the form of DNA plasmids, others however include RNA, proteins (e.g. designed nucleases) and ribonucleoprotein complexes (Yidong et al., 2017). Delivery of genome editing reagents can be challenging in plants (Yin et al., 2015). It is therefore crucial to develop a specific and efficient delivery mechanism for a specific line or species of interest. Given the pace at which routine genome editing is being adopted, more efficient genome editing methods are needed especially in crops that are difficult to transform (Zaidi and Mansoor, 2017).

Most genetic material delivery systems in place focus on the stable transformation of plants resulting in transgenic crops as opposed to transient expression. Delivery of genome editing reagents to intact plant cells is currently limited to particle bombardment, *Agrobacterium*-mediated transformation and to some extent viral vectors (Wolter and Puchta, 2017).

1.5.1 *AGROBACTERIUM*-MEDIATED TRANSFORMATION

The genus *Agrobacterium* consists of plant pathogenic bacteria inhabiting the soil that cause crown gall disease characterized by tumours in infected plants (Michielse et al., 2005). This is achieved through a distinct plasmid that enables them to transfer a particular segment of their DNA (T-DNA) into the plant cells, resulting in the integration of this T-DNA into the plant host genome (Peyret and Lomonossoff, 2015; Ran et al., 2017). Initially, *Agrobacterium* based transformation was achieved by infecting tobacco cells or protoplasts with *Agrobacterium tumefaciens* and regenerating genetically modified shoots from these cells (Peyret and Lomonossoff, 2015). *Agrobacterium* contains a megaplasmid of almost 200kb responsible for tumour induction and opine formation in infected plants (Opabode, 2006, Hellens et al., 2000). This tumour inducing (Ti) plasmid contains, the T-DNA, delimited by repeats known as the left (LB) and right (RB) borders responsible for defining T-DNA boundaries (Hellens et al., 2000). The transfer of this plasmid is mediated by genes in the plasmid virulence regions namely *vir* genes (Michielse et al., 2005; Opabode, 2006). This was further developed by disarming the tumour inducing (Ti) plasmid, through the removal of all the T-DNA genes to allow the delivery

of foreign DNA and direct regeneration of transformed plants (Peyret and Lomonossoff, 2015; Ran et al., 2017). A majority of disarmed *Agrobacterium* T-DNA plasmids typically contain *vir* genes, antibiotic resistance genes, LB and RB borders, and replication origin (*ori*) (Hellens et al., 2000).

Agrobacteriumtumefaciens mediated transient expression assay is referred to as agroinfiltration and has been mainly used for the transformation of dicotyledons (Yidong et al., 2017). Agroinfiltration was used for directly expressing proteins (Peyret and Lomonossoff, 2015). *Agrobacterium*-mediated delivery is one of the most widely adopted plant transformation method. It is relatively affordable, easy to set up in the laboratory due to wide application and demonstration (Yidong et al., 2017).

Agrobacterium mediated transformation still remains the method of choice in plant gene editing experiments, transformation of sunflower however, still remains a challenge (Darqui et al., 2021; Radonic et al., 2015; Sujatha et al., 2012). Sunflower is associated with a host of transformation limitations including low *Agrobacterium* virulence, low transformation efficiencies, lack of stable integration of introduced genes, genotype dependence and the production of chimeric plants amongst others (Darqui et al., 2021; Sujatha et al., 2012). Several studies have focused on the improvement of sunflower transformation including the establishment of genotype wide *Agrobacterium* transformation protocols through testing of various factors affecting plant transformation such as *Agrobacterium* titre, co-culture period and plant vitamin treatments (Sujatha et al., 2012).

In the majority of CRISPR/Cas9 experiments conducted in plants, *Agrobacterium*-mediated transformation, viral delivery or physical methods, such as PEG-mediated transformation of protoplast or biolistic-mediated transformation of callus, were used for the delivery of the Cas9 and sgRNA construct(s) directly into the plant/tissue. With an exception to viral delivery systems, all the mentioned methods require a successful plant regeneration protocol, which can be tedious and/or not established for a specific line or species. The dependence of genome editing on *in vitro* culture therefore explains the delay in adaptation of these technologies in crops considered to be recalcitrant to transformation, such as sunflower.

1.5.2 PARTICLE BOMBARDMENT

Particle bombardment is another widely used technique for the delivery of gene editing reagents in plants (Altpeter et al., 2005; Heim et al., 1995). Particle bombardment, also referred to as “delivery with a gene gun”, is a physical means for the facilitation of nucleic acid

delivery in plants. This method is based on the bombardment of intact plant cells with DNA-coated high velocity microprojectiles “shot” in with enough force to penetrate the target plant cells. The DNA may thereafter integrate into the host genome (Altpeter et al., 2005; Ran et al., 2017). This delivery method targets intact plant material and therefore does not require protoplasts or cloning vectors, although they’re often used (Altpeter et al., 2005; Finer and Mcmullen, 1991). Particle bombardment also allows for the delivery of multiple DNA constructs simultaneously (Ran et al., 2017). A downside of this delivery system is that it often results in multiple copies of the transgene being integrated into the genome (Songstad et al., 2017). This may lead to numerous unwarranted effects, such as gene suppression and changes in gene expression levels in transgenic plants (Ran et al., 2017).

Particle bombardment has been used extensively in the generation of commercial and experimental genetically modified crops (Altpeter et al., 2005). Several research groups have used particle bombardment for sunflower transient transformation (Hunold et al., 1994). Particle bombardment has however, also been used to produce stable sunflower through particle bombardment and co-cultivation with *Agrobacterium* (Knittel et al., 1994; Lucas et al., 2000). These approaches demonstrated low efficiencies in transformation.

Recently, particle bombardment has been used for genome editing using TALENs and CRISPR in rice and wheat. This allowed the editing of the *ALS1* gene in soybean (through CRISPR/Cas mediated HDR recombination) and more recently, transgene-free genome modification in hexaploid wheat (Liang et al., 2017).

1.5.3 VIRAL VECTORS FOR PLANT GENOME EDITING

Viral-based vectors have also been adopted and employed in the delivery of genetic material in plants – although these are more often used in mammalian cells (Chapman et al., 1992; Peretz et al., 2007; Shivprasad et al., 1999). Over the past few decades, modified plant viruses have served as great vehicles for the expression of proteins and RNAs in plants for various functions (Dinesh-Kumar and Voytas, 2020, Kumar and Voytas et al., 2020; Oh et al., 2021). On introduction to a plant, the viral vector system will replicate and express the transgene of interest in the host (Gleba et al., 2007). The efficient machinery and comprehensive genome structure of viruses makes their genomes an excellent choice to be used as vectors (Sanjana et al., 2014; Zaidi and Mansoor, 2017). The tobacco mosaic virus (TMV) was the first viral-based vector to be employed for virus-induced gene silencing (VIGS) in the model plant *Nicotiana benthamiana* (Ran et al., 2017) and numerous viral vectors have

been explored from then on including for use of complete CRISPR/Cas reagent delivery (Table 1.4).

Autonomously replicating virus-based vectors provide alternative means to deliver genome engineering (GE) reagents into plant cells (Zaidi and Mansoor, 2017). Single-stranded (ss) DNA viruses, such as geminiviruses, have also widely been adopted as vectors for diverse crops (Ali et al., 2015). These viruses can be modified to carry heterologous coding proteins, (Zaidi and Mansoor, 2017). Geminiviral vector replicons (GVRs) in particular, have the ability to increase gene targeting frequencies compared to traditional delivery systems, e.g. *Agrobacterium tumefaciens* (Yin et al., 2015). These viruses have a genome size of ±2,8 kb, usually with six overlapping open reading frames (ORFs) (Zaidi and Mansoor, 2017). Begomoviruses (Family: Geminiviridae) are DNA based and have a wide host range, making them the ideal vectors (Ali et al., 2015). They normally consist of two circular single stranded genomes (Hayes et al., 1989), only one of these genomes is however required to initiate replication within a host (Zaidi and Mansoor., 2017). The second genome is required for cell-to-cell movement within the plant host and is dependent on the presence of the former for its replication (Hayes et al., 1989). The use of geminiviruses for gene targeting has long been recognized however, due to size limitations when constructing expression vectors, the viral gene responsible for movement is most often removed to accommodate the insert thereby hampering efficient replication and restricting expression to the affected cells only (Peretz et al., 2007).

Peretz et al. *(*2007) developed a universal gene expression and silencing delivery system with systemic movement abilities for usage in plants. This was developed from deactivating and altering the Tomato yellow leaf curl virus (TYLCV). Tomato yellow leaf curl virus is a whitefly transmitted geminivirus belonging to the begomovirus genus (Cui et al., 2004). Tomato yellow leaf curl virus is made up of six overlapping open reading frames (ORFs) which are bi-directionally transcribed from an intergenic region (IR) (Peretz et al., 2007). The TYLCV genome strand codes for the V1 which is the capsid protein and V2 genes, whereas the virus genome complementary DNA encodes for four genes C1 - C4; with C1 being the replication protein (Czosnek and Laterrot, 1997). The Peretz et al (2007) vector system includes a bacterial plasmid for maintenance in *E. coli*, and a genetically engineered “viral component" that cannot perform rolling circle replication (i.e. single strand DNA replication) or form viral particles. This ensures that no insect or seed transmission could take place. Interestingly, the system does allow double stranded multiplication and still allows systematic movement inside plant host (Peretz et al., 2007). The cell-to-cell movement was attributed to the retention of the N-terminal region of the viral Coat Protein (CP) (Peretz et al., 2007), 20 amino acids were

deleted from the viral CP. In this work, a wide range of monocot, dicot and woody species, including wheat (*Triticum duram*), pepper (*Capsicum annuum*), grapevine (*Vitis vinifera*), citrus, and olive (*Olea europaea*) were injected with the delivery viral system (IL-60-BS) plasmid DNA and tested for systemic protein expression on (Peretz et al., 2007).

Table 1.4 Viruses used for complete CRISPR/Cas reagent delivery.

Plant Species	CRISPR/Cas reagent	Virus and genome size	Systemic movement	Reference
Tobacco	Cas9, sgRNA	Bean yellow dwarf virus	No	Baltes et al. (2014)
Rice	Cas9, sgRNA	Potato virus X	No	Ariga et al. (2020)
Tobacco	Cas9, sgRNA	Sonchus yellow net rhabdovirus	No	Ma et al. (2020)
Tobacco	Cas12a, sgRNA	Potato virus X (PVX), Tobacco etch virus (TEV)	Yes	Uranga et al., (2021)
Tobacco	Cas9, sgRNA	Bean yellow dwarf virus (BeYDV)	Yes	Rezaei et al. (2021)
Tobacco	Cas12a, sgRNA, DRT	Beet curly top virus (BCTV) (2,9-3kb)	Yes	Eini et al. (2022)

1.6 CONCLUSION

The standard aim of any breeding process is to fully express the targeted, superior parent characteristic in the progeny (Ravi et al., 2014). For this dynamic process to occur, homozygous parents – which may require up to eight generations of back crossing (via self-pollination) and selection – are required (Ravi and Chan, 2010). Doubled haploid (DH) technologies can facilitate the production of true breeding lines faster and in a more efficient manner than the traditional back crossing and selection strategies (Kelliher et al., 2017). Haploid induction can significantly decrease the amount of time required to complete such

plant-breeding processes. Furthermore, the induction of doubled haploids holds the potential to increase the genetic variation for selection and accelerate the selection and breeding procedures (Davey and Jan, 2010). At present, sunflower has no available doubled haploid induction procedure that can be efficiently used in breeding programs. Current approaches of haploid development for other crops are often cultivar specific, not reproducible, and rely on available tissue culture protocols – which on their own are also cultivar and/or species specific (Davey and Jan, 2010). As an out-crossing crop, the lack of a doubled haploid system limits sunflower breeding and associated improvement processes, thereby delaying new hybrid and trait developments.

An approach based on the elimination of the *centromere histone 3* (*CenH3*) gene was successfully demonstrated by Ravi and Chan (2010) in *Arabidopsis thaliana*. The approach used a lethal, *Cenh3* mutant and introduced a "rescuing" *CenH3* gene, thus resulting in a haploid inducer line, with only the non-genetically transformed cultivars (non-GM) characteristics being transferred. This approach promised to be a breakthrough haploid induction technology with wide plant applicability due to the universal centromere mechanism based on CENH3 function (Ravi and Chan, 2010). Indeed, the approach type has been used successfully in barley (Karimi-Ashtiyani et al., 2015a), *A. thaliana* (Kuppu et al., 2015), banana (Muiruri, 2015) and maize (Kelliher et al., 2019, 2016).

Typically, plant transformation is mediated by *Agrobacterium* transformation or particle bombardment. Unfortunately, both these methods have reduced efficiencies because they only transform a small portion of the treated tissues, often require a viable plant regeneration tissue culture protocol and are quite tedious to work with (Zaidi and Mansoor, 2017). Reliable and efficient Genetic Engineering (GE) delivery methods are clearly needed if GE is to become routine, especially in economically important crops that are difficult, time consuming and costly to transform with traditional methods (Zaidi and Mansoor, 2017). Researchers have started to explore the use of viral delivery systems, first in animals and recently in plants (Čermák et al., 2015; Yin et al., 2015).

This study aims to develop a haploid induction system in sunflower by editing the *CenH3* gene using the CRISPR/Cas9 technology and to develop a viral-based CRISPR/Cas9 delivery system to mitigate the tediousness and low efficiencies accompanied by *A. tumefaciens* transformation.

1.7 REFERENCES

Abdallah, N.A., Prakash, C.S., Mchughen, A.G (2015) Genome editing for crop improvement: Challenges and opportunities. GM Crops and Food 6, 183-205. https://doi.org/10.1080/21645698.2015.1129937

Ahmad, N., Rahman, M., Mukhtar, Z., Zafar, Y., Zhang, B (2019) A critical look on CRISPR-based genome editing in plants. Journal of Cellular Physiology 235, 666-682. https://doi.org/10.1002/jcp.29052

Ali, Z., Abul-Faraj, A., Li, L., Ghosh, N., Piatek, M., Mahjoub, A., Aouida, M., Piatek, A., Baltes, N.J., Voytas, D.F., Dinesh-Kumar, S., Mahfouz, M.M (2015) Efficient virus-mediated genome editing in plants using the CRISPR/Cas9 system. Molecular Plant 8, 1288-1291. https://doi.org/10.1016/j.molp.2015.02.011

Alonso-blanco, C., Aarts, M.G.M., Bentsink, L., Keurentjes, J.J.B., Reymond, M (2009) What has natural variation taught us about plant development, physiology, and adaptation?. The Plant Cell 21, 1877-1896. https://doi.org/10.1105/tpc.109.068114

Altpeter, F., Baisakh, N., Beachy, R., Bock, R., Capell, T., Christou, P., Daniell, H., Datta, K., Datta, S., Dix, P.J., Fauquet, C., Huang, N., Kohli, A., Mooibroek, H., Nicholson, L., Nguyen, T.T., Nugent, G., Raemakers, K., Romano, A., Somers, D.A., Stoger, E., Taylor, N., Visser, R (2005) Particle bombardment and the genetic enhancement of crops: Myths and realities. Molecular Breeding 15, 305-327. https://doi.org/10.1007/s11032-004-8001-y

Ariga, H., Toki, S., Ishibashi, K (2020) *Potato Virus X* vector- mediated DNA free genome editing in plants. Plant and Cell Physiology 61, 1946-1953

Badouin, H., Gouzy, J., Grassa, C.J., Murat, F., Staton, S.E., Cottret, L., Lelandais-Brière, C., Owens, G.L., Carrère, S., Mayjonade, B., Legrand, L., Gill, N., Kane, N.C., Bowers, J.E., Hubner, S., Bellec, A., Bérard, A., Bergès, H., Blanchet, N., Boniface, M., Brunel, D., Catrice, O., Chaidir, N., Claudel, C., Donnadieu, C., Faraut, T., Fievet, G., Helmstetter, N., King, M., Knapp, S.J., Lai, Z., Le Paslier, M.-C., Lippi, Y., Lorenzon, L., Mandel, J.R., Marage, G., Marchand, G., Marquand, E., Bret-Mestries, E., Morien, E., Nambeesan, S., Nguyen, T., Pegot-Espagnet, P., Pouilly, N., Raftis, F., Sallet, E., Schiex, T., Thomas, J., Vandecasteele, C., Varès, D., Vear, F., Vautrin, S., Crespi, M., Mangin, B., Burke, J.M., Salse, J., Muños, S., Vincourt, P., Rieseberg, L.H., Langlade, N.B (2017) The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature 546, 148-152. https://doi.org/10.1038/nature22380

Baltes, N.J., Gil-Humanes, J., Cermak, T., Atkins, P.A., Voytas, D.F (2014) DNA replicons for plant genome engineering. The Plant Cell 26, 151-163. https://doi.org/10.1105/tpc.113.119792

Bleuyard, J., Gallego, M.E., White, C.I (2006) Recent advances in understanding of the DNA double-strand break repair machinery of plants. DNA Repair 5, 1–12. https://doi.org/10.1016/j.dnarep.2005.08.017

Blinkov, A.O., Varlamova, N. V., Kurenina, L. V., Khaliluev, M.R (2022) The production of *Helianthus* haploids: A review of Its current status and future prospects. Plants 11, 2919.

Bohorova, N (1985) *In Vitro* organogenesis, androgenesis and embryo culture in the genus *Helianthus* L. *Z.* Pflanzenzuchtg 95, 35–44.

Britt, A.B., Kuppu, S (2016) CenH3: An emerging player in haploid induction technology. Frontiers in Plant Science 7, 357. https://doi.org/10.3389/fpls.2016.00357

Ceccaldi, R., Rondinelli, B., Andrea, A.D.D (2016) Repair pathway choices and consequences at the double-Strand break. Trends in Cell Biology 26, 52-64. https://doi.org/10.1016/j.tcb.2015.07.009

Čermák, T., Baltes, N.J., Čegan, R., Zhang, Y., Voytas, D.F (2015) High-frequency, precise modification of the tomato genome. Genome Biology 16. https://doi.org/10.1186/s13059-015-0796-9

Chapman, S., Kavanagh, T., Baulcombe, D (1992) *Potato virus X* as a vector for gene expression in plants. The Plant Journal 2, 549-557. https://doi.org/10.1046/j.1365-313X.1992.t01-24-00999.x

Chen, K., Wang, Y., Zhang, R., Zhang, H., Gao, C (2019) CRISPR/Cas genome editing and precision plant breeding in agriculture. Annual Review of Plant Biology 70, 667-697. https://doi.org/10.1146/annurev-arplant-050718-100049

Chinnusamy, V., Zhu, J.K (2009) Epigenetic regulation of stress responses in plants. Current Opinion in Plant Biology 12, 133-139. https://doi.org/10.1016/j.pbi.2008.12.006

Coumans, M., Zhong, D (1995) Doubled haploid sunflower (*Helianthus annuus*) plant production by androgenesis: fact or artifact? Part 2. *In vitro* isolated microspore culture. Plant Cell, Tissue and Organ Culture 41, 203-309. https://doi.org/10.1007/BF00045083

Cui, X., Tao, X., Xie, Y., Fauquet, C.M., Zhou, X (2004) A DNA β Associated with *Tomato Yellow Leaf Curl China Virus* is required for symptom induction. Journal of Virology 74, 13966-13974. https://doi.org/10.1128/JVI.78.24.13966

Cvejic, S., Radanovic, A., Dedic, B., Jockovic, M., Jocic, S., Miladinovic, D (2020) Genetic and genomic tools in sunflower breeding for broomrape resistance. Genes 11, 152. https:// doi:10.3390/genes11020152

Czosnek, H., Laterrot, H (1997) A worldwide survey of *Tomato yellow leaf curl viruses*. Archives of Virology 142, 1391-1406. https://doi.org/10.1007/s007050050168

Dagustu, N (2018) *In Vitro* tissue culture studies in sunflower (*Helianthus* spp.). Ekin Journal of Crop Breeding and Genetics 4, 13-21.

Danilo, B., Perrot, L., Botton, E., Nogué, F., Mazier, M (2018) The DFR locus: A smart landing pad for targeted transgene insertion in tomato. PLoS ONE 13, e0208395. https://doi.org/10.1371/journal.pone.0208395

Danilo, B., Perrot, L., Mara, K., Botton, E., Nogué, F., Mazier, M (2019) Efficient and transgene-free gene targeting using *Agrobacterium*-mediated delivery of the CRISPR/Cas9 system in tomato. Plant Cell Reports 38, 459-462. https://doi.org/10.1007/s00299-019-02373-6

Darqui, F., Radonic, L., Beracochea, V., Hopp, H., López Bilbao, M (2021) Peculiarities of the transformation of *Asteraceae* family species: The cases of sunflower and lettuce. Frontiers in Plant Science 12, 767459. https://doi.org/10.3389/fpls.2021.767459

Daurova, A.; Daurov, D.; Volkov, D.; Zhapar, K.; Raimbek, D.; Shamekova, M.; Zhambakin, K (2020) Doubled Haploids of Interspecific Hybrids between *Brassica Napus* and *Brassica Rapa* for Canola Production with Valuable Breeding Traits. Oilseeds and Fats, Crops and Lipids 27, 45. https://doi.org/10.1051/OCL/2020041

Davey, M.R., Jan, M (2010) Sunflower (*Helianthus annuus* L.): Genetic improvement using conventional and *in vitro* technologies. Journal of Crop Improvement 24, 349-391. https://doi.org/10.1080/15427528.2010.500874

Diez, C.M., Roessler, K., Gaut, B.S (2014) Epigenetics and plant genome evolution. Current Opinion in Plant Biology 18, 1-8. https://doi.org/10.1016/j.pbi.2013.11.017

Dinesh-Kumar, S.P., Voytas, D.F (2020) Editing through infection. Nature Plants 6, 738-739. https://doi.org/10.1038/s41477-020-0716-1

Dirks, R., Van Dun, K., De Snoo, C.B., Van Den Berg, M., Lelivelt, C.L.C., Voermans, W., Woudenberg, L., De Wit, J.P.C., Reinink, K., Schut, J.W., Van Der Zeeuw, E., Vogelaar, A., Freymark, G., Gutteling, E.W., Keppel, M.N., Van Drongelen, P., Kieny, M., Ellul, P., Touraev, A., Ma, H., De Jong, H., Wijnker, E (2009) Reverse breeding: A novel breeding approach based on engineered meiosis. Plant Biotechnology Journal 7, 837-845. https://doi.org/10.1111/j.1467-7652.2009.00450.x

Drumeva, M., Yonkov, P., Nenova, N., Shindrova, P (2014) Investigation on the resistance of doubled haploid sunflower. Agricultural Science and Technology 6, 11-13.

Drumeva, M.; Yankov, P. Parthenogenetic Responsiveness of Sunflower Hybrid Combinations with Expressed Tolerance to Herbicides. Agricultural Science and Technology. 2017, 9, 190–193. https://doi.org/10.15547/AST.2017.03.034.

Dunwell, J.M (2010) Haploids in flowering plants: Origins and exploitation. Plant Biotechnology Journal 8, 377-424. https://doi.org/10.1111/j.1467-7652.2009.00498.x

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, 812-829. https://doi.org/10.1016/j.biotechadv.2015.07.001

Eini, O., Schumann, N., Niessen, M., Varrelmann, M (2022) Targeted mutagenesis in plants using *Beet curly top virus* for efficient delivery of CRISPR/Cas12a components. New Biotechnology 67, 1-11. https://doi.org/10.1016/j.nbt.2021.12.002

Eliby, S., Bekkuzhina, S., Kishchenko, O., Iskakova, G., Kylyshbayeva, G., Jatayev, S., Soole, K., Langridge, P., Borisjuk, N., Shavrukov, Y (2022) Developments and prospects for doubled haploid wheat. Biotechnology Advances 60, 108007. https://doi.org/10.1016/j.biotechadv.2022.108007

Endo, M., Mikami, M., Toki, S (2016) Biallelic gene targeting in rice. Plant Physiology 170, 667-677. https://doi.org/10.1104/pp.15.01663

Feng, C., Yuan, J., Wang, R., Liu, Y., Birchler, J.A., Han, F (2016) Efficient targeted genome modification in maize using CRISPR/Cas9 system. Journal of Genetics and Genomics 43, 37-43. https://doi.org/10.1016/j.jgg.2015.10.002

Feng, J., Jan, C.-C., Seiler, G (2022) Breeding, production, and supply chain of confection sunflower in China. Oil Seeds and Fats, Crops and Lipids 29, 1-13. https://doi.org/10.1051/ocl/2022004

Feng, J., Liu, Z., Cai, X., Jan, C.-C (2013) Toward a molecular cytogenetic map for cultivated sunflower (*Helianthus annuus* L.) by landed BAC/BIBAC clones. G3 Genes|Genomes|Genetics 3, 31-40. https://doi.org/10.1534/g3.112.004846

Feng, S., Jacobsen, S.E (2011) Epigenetic modifications in plants: An evolutionary perspective. Current Opinion in Plant Biology 14, 179-186. https://doi.org/10.1016/j.pbi.2010.12.002

Ferrie, A.M.R.; Caswell, K.L (2016) Applications of doubled haploidy for Improving industrial oilseeds. In Industrial Oil Crops; AOCS Press: Urbana, IL, USA, 359–378. https://doi.org/10.1016/B978-1-893997-98-1.00013-0.

Finch, R.A (1983) Tissue-specific elimination of alternative whole parental genomes in one barley hybrid. Chromosoma 88, 386-393. https://doi.org/10.1007/BF00285861

Finer, J.J., Mcmullen, M.D (1991) Transformation of soybean via particle bombardment embryogenic suspension culture tissue. *In vitro* Cellular and Developmental Biology – Plant 27, 175-182. https://doi.org/10.1007/BF02632213

Friedt, W (1992) Present state and future prospects of biotechnology in sunflower breeding. Field Crops Research 30, 425-442. https://doi.org/10.1016/0378-4290(92)90009-X

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C. V., Sánchez-León, S., Baltes, N.J., Starker, C., Barro, F., Gao, C., Voytas, D.F (2017) High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. The Plant Journal 89, 1251-1262. https://doi.org/10.1111/tpj.13446

Glass, Z., Lee, M., Li, Y., Xu, Q (2018) Engineering the delivery system for CRISPR-based genome editing. Trends in Biotechnology 36, 173-185. https://doi.org/10.1016/j.tibtech.2017.11.006

Gleba, Y., Klimyuk, V., Marillonnet, S (2007) Viral expression of proteins in plants. Current Opinions in Biotechnology 18, 134-141

Guha, T.K., Wai, A., Hausner, G (2017) Programmable genome editing tools and their regulation for efficient genome engineering. Computational and Structural Biotechnology Journal 15, 146-160. https://doi.org/10.1016/j.csbj.2016.12.006

Hahn, F., Eisenhut, M., Mantegazza, O., Weber, A.P.M (2018) Homology-directed repair of a defective glabrous gene in *Arabidopsis* with Cas9-based gene targeting. Frontiers in Plant Science 9, 424. https://doi.org/10.3389/fpls.2018.00424

He, X.J., Chen, T., Zhu, J.-K (2011) Regulation and function of DNA methylation in plants and animals. Cell Research 21, 442-465. https://doi.org/10.1038/cr.2011.23

Heim, U., Manteuffel, R., Biiumlein, H., Steinbiss, H., Wobus, U (1995) Transient expression of a lysine-rich vicilin gene of *Vicia faba* in barley endosperm detected by immunological tissue printing after particle bombardment. Plant Cell reports 15, 125-128.

Hellens, R., Mullineaux, P., Klee, H (2000) A guide to *Agrobacterium* binary Ti vectors. Trends in Plant Science 5, 446-451.

Hirohata, A., Sato, I., Kaino, K., Iwata, Y., Koizumi, N., Mishiba, K (2019) CRISPR/Cas9-mediated homologous recombination in tobacco. Plant Cell Reports 38, 463-473. https://doi.org/10.1007/s00299-018-2320-7

Huang, T.K., Puchta, H (2019) CRISPR/Cas-mediated gene targeting in plants: finally, a turn for the better for homologous recombination. Plant Cell Reports 38, 443-453. https://doi.org/10.1007/s00299-019-02379-0

Hübner, S., Bercovich, N., Todesco, M., Mandel, J.R., Odenheimer, J., Ziegler, E., Lee, J.S., Baute, G.J., Owens, G.L., Grassa, C.J., Ebert, D.P., Ostevik, K.L., Moyers, B.T., Yakimowski, S., Masalia, R.R., Gao, L., Ćalić, I., Bowers, J.E., Kane, N.C., Swanevelder, D.Z.H., Kubach, T., Muños, S., Langlade, N.B., Burke, J.M., Rieseberg, L.H (2019) Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. Nature Plants 5, 54-62. https://doi.org/10.1038/s41477-018-0329-0

Hunold, R., Burrus, M., Bronner, R., Durret, J., Hahne, G (1994) Transient gene expression in sunflower (*Helianthus annuus* L.) following microprojectile bombardment. Plant Science 105, 95-109. https://doi.org/10.1016/0168-9452(94)04038-I

Ikeuchi, M., Iwase, A., Sugimoto, K (2015) Control of plant cell differentiation by histone modification and DNA methylation. Current Opinion in Plant Biology 28, 60-67. https://doi.org/10.1016/j.pbi.2015.09.004

Ishii, T., Karimi-Ashtiyani, R., Houben, A (2016) Haploidization via chromosome elimination: Means and mechanisms. Annual Review of Plant Biology 67, 421-438. https://doi.org/10.1146/annurev-arplant-043014-114714

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E (2012) A Programmable Dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-822. https://doi.org/10.1126/science.1225829

Jonard, R.; Mezzarobba, A. (1990) Sunflower (*Helianthus* Spp.): Anther Culture and Field Studies on Haploids; Springer: Berlin/Heidelberg, Germany, 485–501. https://doi.org/10.1007/978-3-642-74448-8_23.

Joung, J.K., Sander, J.D (2013) TALENs: A widely applicable technology for targeted genome editing. Nature Reviews Molecular Cell Biology 14, 49-55. https://doi.org/10.1038/nrm3486

Kalinowska, K., Chamas, S., Unkel, K., Demidov, D., Lermontova, I., Dresselhaus, T., Kumlehn, J., Dunemann, F., Houben, A (2019) State of the art and novel developments of *in vivo* haploid technologies. Theoretical and Applied Genetics 132, 593-605. https://doi.org/10.1007/s00122-018-3261-9

Karimi-Ashtiyani, R., Ishii, T., Niessen, M., Stein, N., Heckmann, S., Gurushidze, M., Banaei-Moghaddam, A.M., Fuchs, J., Schubert, V., Koch, K., Weiss, O., Demidov, D., Schmidt, K., Kumlehn, J., Houben, A (2015) Point mutation impairs centromeric CENH3 loading and induces haploid plants. Proceedings of the National Academy of Sciences of the United States of America 112, 11211-6. https://doi.org/10.1073/pnas.1504333112

Kelliher, T., Starr, D., Richbourg, L., Chintamanani, S., Delzer, B., Nuccio, M.L., Green, J., Chen, Z., McCuiston, J., Wang, W., Liebler, T., Bullock, P., Martin, B (2017) *MATRILINEAL*, a sperm-specific phospholipase, triggers maize haploid induction. Nature 542, 105-109. https://doi.org/10.1038/nature20827

Kelliher, T., Starr, D., Su, X., Tang, G., Chen, Z., Carter, J., Wittich, P.E., Dong, S., Green, J., Burch, E., Mccuiston, J., Gu, W., Sun, Y., Strebe, T., Roberts, J., Bate, N.J., Que, Q (2019) One-step genome editing of elite crop germplasm during haploid induction. Nature Biotechnology 37, 287-292. https://doi.org/10.1038/s41587-019-0038-x

Kelliher, T., Starr, D., Wang, W., McCuiston, J., Zhong, H., Nuccio, M.L., Martin, B (2016) Maternal haploids are preferentially induced by CENH3-tailswap transgenic complementation in maize. Frontiers in Plant Science 7, 414. https://doi.org/10.3389/fpls.2016.00414

Khatodia, S., Bhatotia, K., Passricha, N., Khurana, S.M.P., Tuteja, N (2016) The CRISPR/Cas genome-editing tool: Application in improvement of crops. Frontiers in Plant Science 7, 506. https://doi.org/10.3389/fpls.2016.00506

Knittel, N., Gruber, V., Hahne, G., Lenee, P (1994) Transformation of sunflower (*Helianthus annus L*.): a reliable protocol. Plant Cell reports 14, 81 -86. https://doi.org/10.1007/BF00233766

Kucera, V.; Vyvadilov, M.; Klima, M (2002) Utilisation of Doubled Haploids in Winter Oilseed Rape (*Brassica Napus* L.) Breeding. Czech Journal of Genetics and Plant Breeding *38*, 50–54.

Kuppu, S., Ron, M., Marimuthu, M.P.A., Li, G., Huddleson, A., Siddeek, M.H., Terry, J., Buchner, R., Shabek, N., Comai, L., 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, 2068-2080. 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. PLoS genetics 11, e1005494. https://doi.org/10.1371/journal.pgen.1005494

Lee, J., Chung, J.H., Kim, H.M., Kim, D.W., Kim, H (2016) Designed nucleases for targeted genome editing. Plant Biotechnology Journal 14, 448-462. https://doi.org/10.1111/pbi.12465

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

Li, S., Li, J., He, Y., Xu, M., Zhang, J., Du, W., Zhao, Y., Xia, L (2019) Precise gene replacement in rice by RNA transcript-templated homologous recombination. Nature Biotechnology 37, 445-450. https://doi.org/10.1038/s41587-019-0065-7

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., Liu, J., Zhang, H., Liu, C., Ran, Y., Gao, C (2017) Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nature Communications 8, 14261. https://doi.org/10.1038/ncomms14261

Lucas, O., Kallerhoff, J., Alibert, G (2000) Production of stable transgenic sunflowers (*Helianthus annuus* L.) from wounded immature embryos by particle bombardment and co-cultivation with *Agrobacterium tumefaciens*. Molecular Breeding 6, 479-487. https://doi.org/10.1023 /A:1026583931327

Lv, J., Yu, K., Wei, J., Gui, H., Liu, C., Liang, D., Wang, Y., Zhou, H., Carlin, R., Rich, R., Lu, T., Que, Q., Wang, W.C., Zhang, X., Kelliher, T (2020) Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. Nature Biotechnology 38, 1397-1401. https://doi.org/10.1038/s41587-020-0728-4

Maheshwari, S., Ishii, T., Brown, C.T., Houben, A., Comai, L (2017) Centromere location in *Arabidopsis* is unaltered by extreme divergence in CENH3 protein sequence. Genome Research 27, 471-478. https://doi.org/10.1101/gr.214619.116

Makarova, K., Yuri I., Omer, S., Fabrizio, C., Shiraz, A., Shah, S., Saunders, B (2015) An updated evolutionary classification of CRISPR–Cas systems. Nature Reviews Microbiology 13, 722-736.https://doi.org/ i:10.1038/nrmicro3569

Marton, I., Zuker, A., Shklarman, E., Zeevi, V., Tovkach, A., Roffe, S., Ovadis, M., Tzfira, T., Vainstein, A (2010) Nontransgenic genome modification in plant cells. Plant Physiology 154, 1079-1087. https://doi.org/10.1104/pp.110.164806

Mba, C.; Afza, R.; Bado, S.; Jain, S.M (2010) Induced mutagenesis in plants using physical and chemical agents. Plant Cell Culture Essential Methods, 111–130, doi:10.1002/9780470686522.CH7.

Meena, H., Sujatha, M., Varaprasad, S (2013) Achievements and bottlenecks of heterosis breeding of sunflower (*Helianthus annuus* L.) in India. Indian Journal of Genetics and Plant Breeding 73, 123-130. https:// 10.5958/j.0975-6906.73.2.019

Michielse, C.B., Hooykaas, P.J.J., van den Hondel, C.A.M.J.J., Ram, A.F.J (2005) *Agrobacterium*-mediated transformation as a tool for functional genomics in fungi. Current Genetics 48, 1-17. https://doi.org/10.1007/s00294-005-0578-0

Muiruri, S.K (2015) Development of haploid inducers in *Musa* spp. (bananas) by modification of centromere specific histone 3 (CENH3) protein. Ph.D. Thesis, University of Nairobi, Nairobi, Kenya, 2015.

Murovec, J., Bohanec, B (2012) Haploids and doubled haploids in plant breeding. Plant Breeding, IntechOpen, London. https://doi.org/10.5772/29982

Nagaki, K., Tanaka, K., Yamaji, N., Kobayashi, H., Murata, M (2015) Sunflower centromeres consist of a centromere-specific LINE and a chromosome-specific tandem repeat. Frontiers in Plant Science 6, 912. https://doi.org/10.3389/fpls.2015.00912

Opabode, J.T (2006) *Agrobacterium*-mediated transformation of plants: emerging factors that influence efficiency. Biotechnology and Molecular Biology review 1, 12-20.

Pacher, M., Puchta, H (2017) From classical mutagenesis to nuclease-based breeding - directing natural DNA repair for a natural end-product. The Plant Journal 90, 819-833. https://doi.org/10.1111/tpj.13469

Pathirana, R (2011) Plant mutation breeding in agriculture. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources 6. https://doi.org/10.1079/PAVSNNR20116032

Peretz, Y., Mozes-koch, R., Akad, F., Tanne, E., Czosnek, H., Sela, I (2007) A universal expression/silencing vector in plants. Plant Physiology 145, 1251-1263. https://doi.org/10.1104/pp.107.108217

Petolino, J.F (2015) Genome editing in plants via designed Zinc Finger Nucleases. *In Vitro* Cellular and Developmental Biology - Plant 51, 1-8. https://doi.org/10.1007/s11627-015-9663-3

Peyret, H., Lomonossoff, G.P (2015) When plant virology met *Agrobacterium*: the rise of the deconstructed clones. Plant Biotechnology Journal 13, 1121-1135. https://doi.org/10.1111/pbi.12412

Pilorgé, E (2020) Sunflower in the global vegetable oil system: situation, specificities and perspectives. Oilseeds and fats, Crops and Lipids 27, 37. https://doi.org/10.1051/ocl/2020028

Piosik, Zenkteler, E., Zenkteler, M (2016) Development of haploid embryos and plants of *Lactuca sativa* induced by distant pollination with *Helianthus annuus* and *H. tuberosus*. Euphytica 208, 439-451. https://doi.org/10.1007/s10681-015-1578-x

Plohl, M., Meštrović, N., Mravinac, B (2014) Centromere identity from the DNA point of view. Chromosoma 123, 313-325. https://doi.org/10.1007/s00412-014-0462-0

Pradeep, K.; Satya, V.K.; Selvapriya, M.; Vijayasamundeeswari, A.; Ladhalakshmi, D.; Paranidharan, V.; Rabindran, R.; Samiyappan, R.; Balasubramanian, P.; Velazhahan, R (2012) Engineering Resistance against *Tobacco Streak Virus* (TSV) in Sunflower and Tobacco Using RNA Interference. Biologia Plantarum. 56, 735–741, https://doi:10.1007/S10535-012-0111-5.

Puchta, H (2005) The repair of double-strand breaks in plants: Mechanisms and consequences for genome evolution. Journal of Experimental Botany 56, 1-14. https://doi.org/10.1093/jxb/eri025

Radonic, L.M., Lewi, D.M., López, N.E., Hopp, H.E., Escandón, A.S (2015) Sunflower (*Helianthus annuus* L.). Methods in Molecular Biology https://doi.org/10.1007/978-1-4939-1658-0

Rajput, M.; Choudhary, K.; Kumar, M.; Vivekanand, V.; Chawade, A.; Ortiz, R.; Pareek, N (2021) RNA interference and CRISPR/Cas gene editing for crop improvement: Paradigm shift towards sustainable agriculture. Plants 10, 1914. https://doi:10.3390/PLANTS10091914.

Ran, Y., Liang, Z., Gao, C (2017) Current and future editing reagent delivery systems for plant genome editing. Science China Life Sciences 60, 490-505. https://doi.org/10.1007/s11427-017-9022-1

Rath, D., Amlinger, L., Rath, A., Lundgren, M (2015) The CRISPR-Cas immune system: Biology, mechanisms and applications. Biochimie 117, 119-128. https://doi.org/10.1016 /j.biochi.2015.03.025

Ravi, M., Chan, S.W.L (2010) Haploid plants produced by centromere-mediated genome elimination. Nature 464, 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., Urtecho, G., Tan, J., Thornhill, K., Zhu, F., Panoli, A., Sundaresan, V., Britt, A.B., Comai, L., Chan, S.W.L (2014) A haploid genetics toolbox for *Arabidopsis thaliana*. Nature Communications 5, 5334. https://doi.org/10.1038/ncomms6334

Razin, S.V., Borunova, V.V., Maksimenko, O.G. and Kantidze, O.L (2012) Cys2His2 zinc finger protein family: classification, functions, and major members. Biochemistry (Moscow). 77, 217-226.

Ren, J., Wu, P., Trampe, B., Tian, X., Lubberstedt, T., Chen, S (2017) Novel technologies in doubled haploid line development. Plant Biotechnology Journal 15, 1361-1370. https://doi.org/10.1111/pbi.12805

Rozov, S.M., Permyakova, N. V., Deineko, E. V (2019) The problem of the low rates of CRISPR/Cas9-mediated knock-ins in plants: Approaches and solutions. International Journal of Molecular Sciences 20, 3371. https://doi.org/10.3390/ijms20133371

Saji, K.V.; Sujatha, M. (1998) Embryogenesis and Plant Regeneration in Anther Culture of Sunflower (*Helianthus Annuus* L.). Euphytica, 103, 1-7. https://doi.org/10.1023/A:1018318625718.

Samantara, K., Bohra, A., Mohapatra, S.R., Prihatini, R., Asibe, F., Singh, L., Reyes, V.P., Tiwari, A., Maurya, A.K., Croser, J.S., Wani, S.H., Siddique, K.H.M., Varshney, R.K (2022) Breeding more crops in less time: A perspective on speed breeding. Biology (Basel) 11, 275. https://doi.org/10.3390/biology11020275

Sanjana, N.E., Shalem, O., Zhang, F (2014) Improved vectors and genome-wide libraries for CRISPR screening. Nature Methods 11, 783-784. https://doi.org/10.1038/nmeth.3047

Sarrafi, A., Gentzbittel, L (2004) Genomics as Efficient Tools: Example sunflower breeding. Agriculture (Biotechnology in Agriculture and Forestry) 55, 107-119

Sawatsubashi, S., Joko, Y., Fukumoto, S., Matsumoto, T., Sugano, S.S (2018) Development of versatile non-homologous end joining-based knock-in module for genome editing. Scientific Reports 8, 593. https://doi.org/10.1038/s41598-017-18911-9

Schaeffer, S.M., Nakata, P.A (2016) The expanding footprint of CRISPR/Cas9 in the plant sciences. Plant Cell Reports 35, 1451-1468. https://doi.org/10.1007/s00299-016-1987-x

Schiml, S., Fauser, F., Puchta, H (2014) The CRISPR/Cas system can be used as nuclease for *in planta* gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny. The Plant Journal 80, 1139-1150. https://doi.org/10.1111/tpj.12704

Schiml, S., Puchta, H (2016) Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas. Plant Methods 12, 8. https://doi.org/10.1186/s13007-016-0103-0

Schindele, P., Wolter, F., Puchta, H (2018) Transforming plant biology and breeding with CRISPR/Cas9, Cas12 and Cas13. Federation of European Biochemical Societies Letters 592, 1954-1967. https://doi.org/10.1002/1873-3468.13073

Schmidt, C., Pacher, M., Puchta, H (2019) DNA break repair in plants and its application for genome engineering. Methods in Molecular Biology 1864, 237-266. https://doi.org/10.1007/978-1-4939-8778-8_17

Senthil-Kumar, M.; Mysore, K.S. (2010) RNAi in plants: Recent developments and applications in Agriculture. Gene Silencing: Theory, Techniques, and Applications. 183–200.

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., Teramura, H., Yamamoto, T., Komatsu, H., Miura, K., Ezura, H., Nishida, K., Ariizumi, T., Kondo, A (2017) Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nature Biotechnology 35, 441-443. https://doi.org/10.1038/nbt.3833

Shivprasad, S., Pogue, G.P., Lewandowski, D.J., Hidalgo, J., Donson, J., Grill, L.K., Dawson, W.O (1999) Heterologous sequences greatly affect foreign gene expression in *Tobacco Mosaic Virus*-based vectors. Virology 255, 312-323. https://doi.org/10.1006/viro.1998.9579

Skinner, M.K (2015) Environmental epigenetics and a unified theory of the molecular aspects of evolution: A neo-Lamarckian concept that facilitates neo-Darwinian evolution. Genome Biology and Evolution 7, 1296-1302. https://doi.org/10.1093/gbe/evv073

Song, G., Jia, M., Chen, K., Kong, X., Khattak, B., Xie, C., Li, A., Mao, L (2016) CRISPR/Cas9: A powerful tool for crop genome editing. Crop Journal 4, 75-82. https://doi.org/10.1016/j.cj.2015.12.002

Songstad, D.D., Petolino, J.F., Voytas, D.F., Reichert, N.A (2017) Genome editing of plants. Critical Reviews in Plant Sciences 36, 1-23. https://doi.org/10.1080/07352689.2017.1281663

Steinert, J., Schiml, S., Puchta, H (2016) Homology-based double-strand break-induced genome engineering in plants. Plant Cell Reports 35, 1429-1438. https://doi.org/10.1007/s00299-016-1981-3

Sujatha, M., Vijay, S., Vasavi, S., Reddy, P.V., Rao, S.C (2012) *Agrobacterium*-mediated transformation of cotyledons of mature seeds of multiple genotypes of sunflower (*Helianthus annuus* L.). Plant cell, Tissue and Organ Culture 110, 275-287. https://doi.org/10.1007/s11240-012-0149-2

Todorova, M., Ivanov, P., Shindrova, P., Christov, M., Ivanova, I (1997) Doubled haploid production in sunflower (*Helianthus annuus* L.) through irradiated pollen-induced parthenogenesis. Euphytica 97, 249-254. https://doi.org/10.1023/A:1002966824988

Uranga, M., Vazquez-Vilar, M., Orzáez, Di., Daròs, J.A (2021) CRISPR-Cas12a genome editing at the whole-plant level using two compatible RNA virus vectors. The CRISPR Journal 4, 761-769. https://doi.org/10.1089/crispr.2021.0049

Vear, F (2016) Changes in sunflower breeding over the last fifty years. Oilseeds and fats, Crops and Lipids 23. https://doi.org/10.1051/ocl/2016006

Wang, S., Jin, W., Wang, K (2019) Centromere histone H3- and phospholipase-mediated haploid induction in plants. Plant Methods 15, 42. https://doi.org/10.1186/s13007-019-0429-5

Watts, A., Singh, S.K., Bhadouria, J., Naresh, V., Bishoyi, A.K., Geetha, K.A., Chamola, R., Pattanayak, D., Bhat, S.R (2017) *Brassica juncea* lines with substituted chimeric GFP-CENH3 give haploid and aneuploid progenies on crossing with other tines. Frontiers in Plant Science 7. https://doi.org/10.3389/fpls.2016.02019

Wolter, F., Puchta, H (2017) Knocking out consumer concerns and regulator's rules: Efficient use of CRISPR/Cas ribonucleoprotein complexes for genome editing in cereals. Genome Biology 18, 17-19. https://doi.org/10.1186/s13059-017-1179-1

Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X., Zhang, L., Rebar, E.J., Gregory, P.D., Urnov, F.D., Meyer, B.J (2011)

Targeted genome editing across species using ZFNs and TALENs. Science 333, 307. https://doi.org/10.1126/science.1207773

Yin, K., Han, T., Liu, G., Chen, T., Wang, Y., Yu, A.Y.L., Liu, Y (2015) A geminivirus-based guide RNA delivery system for CRISPR/Cas9 mediated plant genome editing. Scientific Reports 5, 14926. https://doi.org/10.1038/srep14926

Zaidi, S.S.-A., Mansoor, S (2017) Viral vectors for plant genome engineering. Frontiers in Plant Science 8, 2012-2017. https://doi.org/10.3389/fpls.2017.00539

Zamariola, L., Tiang, C.L., Storme, N. De, Pawlowski, W., Geelen, D., Maruthachalam, R (2014) Chromosome segregation in plant meiosis. Frontiers in Plant Science 5, 1-19. https://doi.org/10.3389/fpls.2014.00279

Zhang, Yong, Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Yi, Zhang, K., Liu, J., Voytas, D.F., Zheng, X., Gao, C (2013) Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. Molecular Plant 6, 1365-1368. https://doi.org/10.1093/mp/sss162

Zhang, Z., Hua, L., Gupta, A., Tricoli, D., Edwards, K.J., Yang, B., Li, W (2019) Development of an *Agrobacterium*-delivered CRISPR/Cas9 system for wheat genome editing. Plant Biotechnology Journal 17, 1623-1635. https://doi.org/10.1111/pbi.13088

Zhao, H., Wolt, J.D (2017) Risk associated with off-target plant genome editing and methods for its limitation. Emerging Topics in Life Sciences 1, 231-240. https://doi.org/10.1042/ETL S20170037

Zhao, Y., Zhang, C., Liu, W., Gao, W., Liu, C., Song, G., Li, W.X., Mao, L., Chen, B., Xu, Y., Li, X., Xie, C (2016) An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. Scientific Reports 6, 1-11. https://doi.org/10.1038/srep23890

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.L., Wang, D., Gao, C (2017) Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nature Biotechnology 35, 438-440. https://doi.org/10.1038/nbt.3811

CHAPTER 2

CRISPR/Cas9 GENOME EDITING OF THE *CenH3* GENE IN SUNFLOWER (*HELIANTHUS ANNUUS*) USING *AGROBACTERIUM*-MEDIATED TRANSFORMATION

2.1 ABSTRACT

The identification of the *centromere-specific histone* 3 (*CenH3*) gene as a key player in chromosome segregation has been a breakthrough in eukaryotic haploid induction (HI) technology. The ability to target specific conserved regions of the *CenH3* gene using modern gene editing techniques furthers the quest of haploid induction across a variety of species. The CRISPR/Cas9 technique can generate double-stranded breaks at predefined locations in the genome and has become the genome editing platform of choice for most organisms, including mammalian cells and plants. In this study, we identified amino acids in exon 1 of the sunflower *CenH3* gene that were previously associated with haploid inducer lines in *Arabidopsis thaliana.* Two different CRISPR/Cas9 constructs were designed to produce targeted changes in three of the identified amino acids (P51S, G52E and A55V) through homology directed repair (HDR). These constructs were introduced into sunflower mature cotyledons through *Agrobacterium*-mediated transformation. Seven sunflower T0 shoots were identified to be potentially transgenic using PCR targeting the sgRNA and U6-26 promoter. However, next generation sequencing (Illumina) of the exon 1 *CenH3* gene region showed that only four of these transgenics contained two of the targeted amino acid conversions (P51S and A55V) and the other three did not contain any mutations. These results demonstrate the first report of successful CRISPR/Cas9 based genome editing in sunflower and the first crucial step in haploid production.

2.2 INTRODUCTION

Helianthus annuus (sunflower) is one of the most economically important crops in the family *Asteraceae* (Nagaki et al., 2015). Sunflower has a huge impact on a global scale with a variety of uses as an oilseed, animal feed purposes, raw material and biofuel production (Davey and Jan, 2010; Friedt, 1992). It is the main source of vegetable oil for the South African food industry, accounting for 60% of the domestic and 9% of the international oilseed markets (DAFF, 2019; Pilorgé, 2020). Furthermore, the global sunflower seeds market is expected to be valued at 30.11 billion USD by 2025 (www.statista.com; Accessed: 15 April 2022). This crop is characterised by an increased climate change adaptation potential due to its ability to maintain stable yields across multiple environmental conditions (Meena et al., 2013).

The production of pure breeding lines is a crucial part of plant breeding and improvement (Kuppu et al., 2020). Doubled haploid (DH) technologies provide the ability to fast track the production of true breeding lines in a single generation as opposed to traditional back crossing and laborious selection strategies often used in conventional plant breeding (Kelliher et al., 2017; Kuppu et al., 2020). Furthermore, the induction of doubled haploids can expand the genetic variation for selection due to availability of material for selection (Davey and Jan, 2010). Sunflower does not have a doubled haploid induction method that could be used in breeding programs. This absence of an efficient doubled haploid system has resulted in delays in sunflower breeding and improvement processes.

The identification of a more universal approach based on the modification of the *centromere-specific histone 3* (*CenH3*) gene has offered possible ease in haploid induction technology for many crops (Karimi-Ashtiyani et al., 2015; Kelliher et al., 2019; Kuppu et al., 2020; Lv et al., 2020; Ravi and Chan, 2010). The centromere plays a critical role in facilitating the transfer of genetic material from parents to offspring during cell division (Britt and Kuppu, 2016). The CENH3 protein is responsible for the identification of centromere location, and the recruitment of the proteins necessary for chromosome segregation (Dawe, 2020). The CENH3 protein is characterized by a highly divergent N-terminal tail and a conserved C-terminal domain also referred to as the histone fold domain (HFD) (Jiang et al., 2003; Britt and Kuppu, 2016). The HFD comprises of four alpha-helices, i.e. α-N, α-1, α-2, and α-3, that are separated by two loops (Dawe, 2020). The HFD region is conserved between CENH3 proteins of a wide range of crop species whereas the N-terminal tails diverge even between closely related species (Britt and Kuppu, 2016). This makes the HFD region an attractive target for genetic engineering towards producing haploid inducer breeding lines.

The first modification of the CENH3 protein for haploid induction was demonstrated by Ravi and Chan (2010) in *Arabidopsis thaliana*. The approach used a lethal CENH3 mutant and introduced a “rescuing” *CenH3* gene, thus resulting in a haploid inducer line, with only the non-genetically transformed lines characteristics being transferred. This discovery resulted in further studies that have focused on identifying various kinds of CENH3 modifications, including (i) the fusion of parts of the CENH3 proteins with other proteins; (ii) The expression of non-native CENH3 from related species; (iii) Single amino acid substitutions with haploid induction as high as 44%, and (iv) Knock-out of regions of the protein, all which lead to haploid progeny upon crossing with wild type plants (Watts et al., 2017; Ravi and Chan, 2010; Kuppu et al., 2020, Kuppu et al., 2015). Point mutations induced by ethyl methanesulfonate (EMS), in five amino acids located in the α-N helix of the HFD in *Arabidopsis* produced paternal haploids at a rate of 0.61% to 12.2% that were normal in appearance and fully fertile when self-fertilized (Karimi-Ashtiyani et al., 2015; Kuppu et al., 2015; Wang et al., 2019). This indicates that the CENH3 from the mutant plants remained functional but incapable of competing with the normal ‘wild type’ *CenH3* genes (Kuppu et al., 2015; Ren et al., 2017). Sunflower contains a single copy of the *CenH3* gene on chromosome 15 (Nagaki et al., 2015), the *CenH3* based haploid induction approach therefore promises to be a breakthrough in haploid induction technology with wide plant applicability due to the universal centromere mechanism based on CENH3 function (Ravi and Chan, 2010).

The CRISPR/Cas9 technology is currently one of the most efficient targeted genome editing systems with wide application in both plants and animals (Chen et al., 2019). Genome editing with the CRISPR/Cas9 technology involves inducing double-stranded breaks (DSBs) at predefined locations in the genome for recombination (Hirohata et al., 2019). Double stranded breaks are achieved through specific binding of Cas9 on the target DNA, determined by both a single guide RNA-DNA base pairing and a protospacer-adjacent motif (PAM) sequence (NGG for *Streptococcus pyogenes*) downstream of the target (Fauser et al., 2014). The Cas9 enzyme cleaves the DNA at a three-nucleotide distance from the PAM sequence (Xing et al., 2014) and this is either repaired via non-homologous end-joining (NHEJ) or homology directed repair (HDR). Repair of DSBs through homology directed repair (HDR) provides the ability to introduce single to large nucleotide substitutions incorporated in the DNA donor sequence (Fauser et al., 2014; Schiml et al., 2014). However, it has proven very challenging to apply HDR in plants to introduce mutations due to general low frequency of targeted integration (Čermák et al., 2015). A limited number of studies have reported HDR repair systems for CRISPR/Cas9 including tobacco, rice, tomato, and soybean (Hirohata et al., 2019). However, homology directed repair efficiencies differ greatly due to the number of factors influencing success (Zhang et al., 2013; Endo et al., 2016).

Although the CRISPR/Cas technology has been applied in a variety of organisms, it has not been described in sunflower. This could be attributed to the well-documented difficulties in the transformation and delivery of genome editing reagents in sunflower (Darqui et al., 2021; Weber et al., 2003). However, *Agrobacterium tumefaciens*-mediated plant transformation can provide a stable transformation method of delivering transgenes into crops and has been successfully used in the modification of a variety of crop plant species including sunflower, although with little success (Radonic et al., 2015; Sujatha et al., 2012). The ability to genetically modify sunflower in a targeted way would benefit sunflower improvement and breeding. Therefore, the aim of this study was to utilise *A. tumefaciens*-mediated sunflower transformation to deliver CRISPR/Cas9 reagents for targeted editing of the *CenH3* gene to develop haploid inducer lines. It was hypothesized that the CRISPR/Cas9 technique would enable homology directed repair targeting of the sunflower *CenH3* gene.

2.3 MATERIALS AND METHODS

All reagents used for plant transformation were purchased from Merck (Darmstadt, Germany) unless stated otherwise

2.3.1 *CENH3* TARGET SEQUENCE, HISTONE FOLD DOMAIN IDENTIFICATION AND GRNA DESIGN

2.3.1.1 Plant Material

Mature seeds from a sunflower dwarf inbred line, KP328, were obtained from Mr. Andrew Mokhele from the Agricultural Research Council's Grains Crops unit (Potchefstroom, South Africa). This inbred line was selected for its early maturity and was used for all experiments in this study. Early maturing and dwarf lines reduce experiment time and are ideal for space constraints as they don't grow too big, these plants remain similar to normal plants).

2.3.1.2 Sunflower CenH3 DNA sequence confirmation

To confirm the *CenH3* gene sequence in sunflower, sequences obtained from NCBI GenBank (XM_022156616.1) was used to design sequencing primers up and downstream (CenH3Fw and CenH3Rv) of the gene (Table 2.1).. Sunflower plants (KP328) were grown in sterile hydroponic perlite growth mixture (Hygrotech, South Africa) in 20 cm pots placed in the plant GM growth facility (Agricultural Research Council-Biotechnology Platform, Pretoria, South Africa). Sunflower KP328 plants were grown at 32°C:20°C with 16:8 h day/night cycling and

humidity set to 60 - 80%. DNA was extracted from approximately 100 mg leaf material using the Qiagen® Plant mini DNA extraction kit according to the manufacturer's guidelines (Qiagen, Hilden, Germany). A PCR reaction was performed with Q5 High-Fidelity DNA polymerase (New England Biolabs, Massachusetts, USA) in duplicate technical repeats). A 50 µL reaction was set up with the following components, 1X Q5 reaction buffer, 200 µM dNTPs, 0.5 µM of the forward and reverse primers each, 0.02 U/µL High-Fidelity DNA polymerase and 50 ng template DNA. Amplification conditions were as follows: 1 cycle of 3 min at 98°C; 35 cycles of 98°C for 30 s, 65°C for 30 s, 72°C for 3 min; final extension of 72°C for 10 min. nuclease free water was used as a negative control. The resulting PCR product was visualized on 1.5% agarose gel using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA) and the expected band of (2993) bp was obtained (not shown). The PCR product was purified using the AMPure XP beads (Beckman coulter, California, USA). Libraries were prepared using TrueSeq® Nano DNA HT Library prep kit (Illumina, California, USA). The prepared library was sent for Illumina 2500 HiSeq (Illumina, California, USA) sequencing at the ARC-Biotechnology Platform (Onderstepoort, South Africa). The quality of the reads was verified in FastQC. Read trimming was done with Trimmomatic 0.36 (Bolger et al., 2014). FastQC using default parameters, *de novo* assembly of trimmed reads was performed using SPAdes (Bankevich et al., 2012) and the identities of resulting contigs was classified using BLASTn (blast.ncbi.nlm.nih.gov). The *CenH3* contig had an average coverage of 357X. The full-length sequence is in the supplementary data at the end of this chapter (Figure S2.1).

Table 2.1 List of primer sequences used in this chapter and study.

Primer name	Sequence 5'→3'	Function
sgRNA Forward	attgTAGGTTTAAGCCTGGGACAC	Single guide RNA for CRISPR editing
sgRNA reverse	aaacGTGTCCCAGGCTTAAACCTA	
M13 primer	CACAGGAAACAGCTATGAC	For sgRNA confirmation
SS42	TCCCAGGATTAGAATGATTAGG	
(U6-26 Promoter)		
SS43 (*bialaphos* gene)	CGACTAAGGGTTTCTTATATGC	For Cas9 and sgRNA cloning confirmation
SS102 (*nptII* gene)	CACCATGTTATCACATCAATCC	For *nptII* gene confirmation
Donor forward	ATTCCGGCTGCTCCGTTTAT	For donor repair template sequence confirmation
Cas-RT-Forward	TCAACGTACATATCCCTACCG	For Cas9 Transcript confirmation (Gao et al., 2017)
Cas9-RT-Reverse	AGGCTCAAGACTTACGCTCAT	
gRNA-RT-Reverse	CGACTCGGTGCCACTTTTTCAAGTTG	For gRNA transcript confirmation (Ma et al., 2015)
CENH3 Target Forward	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GAGTCCGAGGAAGAATCGAGGTAG	For *CenH3* target for mutation detection. The Illumina 16s barcodes are in red.
CENH3 Target Reverse	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AGTAAAATGGCAGTGTTTTGTTTGACC	
CENH3Rv	GCATGATTAATTAACACACACACACACACAAAC	*CenH3* gene confirmation
CENH3Fw	TGTCTAAAAATACACCGCGCTAA	
CENH3LDFw	AACGCAGTTCAGGCATACCA	*CenH3* gene large donor for mutation detection
CENH3LDRv	AACACGAACACGACCCGTTA	

2.3.1.3 sgRNA target sequence design

The first part of the single guide RNA (sgRNA) design used the conserved region of the *CenH3* gene incorporating the amino acid sequences that were identified by Kuppu et al., (2015) The *CenH3* sequence from this genome (XM_022156616.1) and that of our cultivar showed 100% similarity/alignment (as indicated in Chapter 2, section 2.3.1.2).. To identify these amino acid sequences and positions in sunflower, we used a multiple sequence alignment of the full CENH3 protein of nineteen (19) plant species i.e. *Hordeum bulbosum* (barley; ADB03182.1), *Musa acuminata* (banana; XP_018685258.1), *Zea mays* (maize; NP_001105520.1), *Oryza sativa* (rice; AAR85315.1); *Phaseolus vulgaris* (bean; AGQ21573.1), *Glycine max* (soybean; XP_006598915.1), *Vitis vinifera* (grapevine; RVC10123.1), *Brassica rapa* (rapeseed; ADN92693.1), *Arabidopsis thaliana* (AAL86775.1), *Arabidopsis lyrata* (XP_020870190.1), *Gossypium hirsufum* (cotton; XP_016745574.1), *Helianthus annuus* (sunflower; XP_022005281.1, This study), *Lactuva sativa* (lettuce; XP_023761199.1), *Daucus carota* (carrot; AID21728.1), *Nicotiana tabacum* (tobacco; BAH03515.1), *Solanum lycopersicum* (tomato; XP_010326926.1), *Solanum tuberosum* (potato; XP_006339687.1) and *Brassica juncea* (mustard; BAF49728) followed by a multiple sequence alignment of CENH3 histone fold domains (HFD) protein sequences of five different plant species using MUSCLE (Edgar, 2004): *Zea mays* (NP_001105520.1), *Arabidopsis thaliana* (AAL86775.1), *Arabidopsis lyrata* (XP_020870190.1), *Daucus carota* (AID21728.1), *Lactuca sativa* (XP_023761199.1) against *Helianthus annuus* (XP_022005281.1, This study). A cost matrix BLOSUM with gap open cost 10 and gap extend 0 were used. Once the HFD was identified, the sunflower cultivar KP328 *CenH3* DNA sequence (XP_022005281.1, This study) was analysed and guide RNAs that target these amino acid regions were manually identified by searching for an NGG sequence nearest to the target site (preferably three nucleotides from the target sequence). Based on this criterion, the closest target site was identified on one sgRNA located on exon 1 of the sunflower *CenH3* gene. The selected sgRNA was tested on http://www.rgenome.net/cas-offinder/ (RGENOME FINDER, Accessed 19 July 2018) off target finder (Bae et al, 2014), using default parameters. No other targets were found for this sgRNA in the sunflower genome (HanXRQr1.0).

The sunflower *CenH3* gene is composed of six introns and seven exons of different sizes (Nagaki et al., 2015). A schematic of the *CenH3* gene, the sgRNA DNA target site and target amino acid sequences are shown in Figure 2.1. The first amino acid change required is seven nucleotides upstream of the PAM (P51), while the second one (G52) is 10 nucleotides upstream the PAM, the final change is one nucleotide downstream the PAM sequence (A55).

Figure 2.1 A schematic representation of the sunflower *CenH3* gene (3559 bp). Arrows indicating the exons (7) as boxes and introns (6) as lines. The sgRNA sequence (green) and the protospacer adjacent-motif (PAM) (red) are highlighted below exon 1. The targeted amino acid nucleotide sequences as well as the corresponding DNA sequence are highlighted at the top of the sequence in blue (P51), yellow (G52) and purple (A55).

2.3.2 Donor repair template (DRT) design

The optimal length for double cut donors has been identified to be 600 bp, this is however variable for different target sites (Zong et al., 2017). Studies have shown that for double stranded donors, homology arms as short as 50 bp are sufficient for homology directed repair (Zhang et al., 2017). To test this in sunflower, two donor cassettes of 258 bp and 1185 bp containing three amino acid changes were designed (Table 2.2) using the sunflower *CenH3* sequence (accession: XP_022005281.1, this study).The 258 bp length donor contains a 64 bp left sequence homology flanking the first amino acid change and 134 bp right homology flanking the last amino acid change while the 1185 bp length donor contained a 527 bp left sequence homology and 656 bp right sequence homology (Supplementary Figure S2.2). To prevent the donor sequence from being cleaved by Cas9, three silent mutations were introduced into the donor sequence and PAM site (Table 2.2). The gRNA sequence, together with the PAM sequence were included in the short (258 bp) donor to create a double cut donor, this ensures the release of the donor sequence near the cut site (Schiml et al., 2014). The 258 bp fragment was synthesized by GENEWIZ with XmaI and SpeI cutting sites (Azenta Life Sciences, Massachusetts, USA) while the large one was procured from GeneArt cloned in the pMK-RQ-DRT plasmid with XmaI and SpeI cutting sites (Invitrogen, ThermoFisher Scientific,

Massachusetts, USA). The full donor sequences are listed in the Supplementary material (Figure S2.2).

Table 2.2 Nucleotide and amino acid substitutions made to homology directed repair donor sequence for use in CRISPR/Cas9 mutation of the sunflower *CenH3* gene.

Codon change	**Amino acid position in *Arabidopsis***	**Amino acid position in sunflower**	**Amino acid change**	**Haploid Progeny in *Arabidopsis* (%)**
AAG -> AAA	81	50	Silent mutation to destroy the PAM recognition site	-
CCT -> TCT	82	51	Proline -> Serine	2.4 - 2.7%
GGG-> GAG	83	52	Glycine-> Glutamic acid	9.1 - 12.2%
ACA -> ACC	84	53	Silent mutation to destroy the PAM recognition site	-
CAG -> CAA	85	54	Silent mutation to destroy the PAM recognition site	-
GCG -> GTG	86	55	Alanine -> Valine	2.72%

Target sequences and haploid percentages were obtained from Kuppu et al., (2015).

2.3.3 VECTOR CLONING

Recipes for all bacterial and plant culture are listed in the supplementary data (Table S2.1).

2.3.3.1 Cloning of guide RNA into expression vector pEN-Chimera

Type III RNA polymerases driving sgRNA expression have an affinity to sequences that start with an A, the sgRNA under the U6-26 promoter therefore had ATTG added to the 5' end, the reverse complement of the target sequence had AAAC added to the 5' end. The designed sgRNA oligonucleotides were procured as desalted oligonucleotides by Integrated DNA technologies (IDT, Iowa, USA) through our local supplier (Whitehead Scientific, Johannesburg, South Africa). The sgRNA oligonucleotide was cloned into the pEN-Chimera sgRNA expression vector carrying the *Arabidopsis* U6-26 promoter obtained from Prof. Holger Puchta's lab (Karlsruher Institut fur Technologie, Karlsruhe, Germany) as described by Schiml et al., (2014). The pEN-Chimera vector was digested with BbsI for 2 hours at 22°C and the reaction was purified using the NucleoSpin gel and PCR clean up kit according to the manufacturer's guidelines (Macherey-Nagel, Duren, Germany). The sgRNA oligonucleotides were annealed and ligated into the linearized pEN-Chimera vector using T4 DNA ligase as

described by Schiml et al., (2017). The ligation reaction was transformed in NEB® 5α *E. coli* chemically competent cells according to the manufacturer's guidelines (New England Biolabs, Massachusetts, USA) and plated on ampicillin (100 mg/L) containing Luria-Bertani (LB) Agar plates incubated at 37°C overnight.

A colony PCR was used to identify positive clones by directly picking single a colony from the plate using a sterile pipette and adding it to the PCR mix (Rönspies et al., 2022). The forward sgRNA oligonucleotide was used as a forward primer while the M13 (targeting the M13 sequence of the pUC vector) was used as reverse primer (Table 2.1). A PCR reaction was set up with 12.5 µL Dreamtaq 2X green PCR Master mix (ThermoFisher Scientific, Massachusetts, USA), 0.1 µM of each primer and nuclease free water to a total volume of 25 µL. The cycling conditions were 95°C initial denaturation for 3 mins, 30 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s, extension at 72°C for 1 min and final extension at 72°C for 5 mins. The 375 bp expected PCR product (Figure S2.2A) was visualised on 1.5% TAE agarose gel stained with ethidium bromide using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA). Successful cloning was validated by Sanger sequencing at Inqaba Biotechnical Industries (Pretoria, South Africa) using the SS42 primer (Table 2.1) targeting the U6 promoter and M13 reverse promoter (Sequencing product size 631 bp). This resulting vector was named pEN-Chimera-sgRNA.

2.3.3.2 Cloning of sgRNA expression system (pEN-Chimera-sgRNA) into Cas9 expression system (pDe-Cas9)

The sgRNA, U6-26 promoter and sgRNA scaffold in the pEN-Chimera-sgRNA containing attB1 and attB2 Gateway® sequences were transferred into the binary vector pDe-Cas9 (Fauser et al., 2014; Schiml et al., 2014) codon optimized for *Arabidopsis* expression system by Gateway® cloning (Invitrogen, Massachusetts, USA). The reaction was transformed into in NEB® 5α *E. coli* chemically competent cells as per manufacturer's guidelines (New England Biolabs, Massachusetts, USA) and plated on spectinomycin (100 mg/L) containing LB agar plates and incubated at 37°C overnight. A colony PCR was set up to identify positive clones, a single colony was picked from the overnight culture and suspended in 1 µL of nuclease free water, the SS42 was used as a forward primer and the SS43 (targeting the *bar* gene) as a reverse primer (Table 2.1). A PCR reaction was set up with 12.5 µL Dreamtaq 2X green PCR Master mix (ThermoFisher Scientific, Massachusetts, USA), 0.1 µM of each primer and nuclease free water to a total volume of 25 µL. The cycling conditions were 95°C initial denaturation for 3 mins, 30 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30

s, extension at 72°C for 1 min and final extension at 72°C for 5 min. The 1093 bp expected PCR product (Figure 2.4B) was visualised on 1.5% agarose gel stained with ethidium bromide using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA). Successful Gateway® transfer was validated by sequencing using Sanger sequencing (Inqaba biotechnical industries, Pretoria, South Africa) using the SS42 and SS43 primers (Table 2.1).

2.3.3.3 Synthesized donor cloning into pDe-Cas9 with the sgRNA

For both donor cassettes (258 bp and 1185 bp) (changes to the donor sequences are listed on Table 2.2), the 500 ng donor fragment was resuspended in 50 µL Tris-EDTA (TE) elution buffer resulting in a final concentration of 10 ng/µL. A restriction digestion reaction with 15 µL of nuclease free water, 100 ng of donor fragment DNA for the 258 bp donor and 100 ng of the pMK-RQ plasmid containing the 1185 bp donor, 3 µL CutSmart buffer, 1 µL of XmaI and 1 µL of SpeI was set up resulting in a 30 µL total reaction. This was incubated at 37°C for 1 hour and heat inactivated at 80°C for 20 min. Another digestion for the binary vector pDe-Cas9-gRNA was also set up with 41 µL of nuclease free water, 2 µL (100 ng) of vector DNA, 5 µL CutSmart buffer, 1 µL of XmaI and 1 µL of SpeI was set up resulting in a 50 µL total reaction. This was incubated at 37°C for 1 hour. The 1185 bp donor fragment was extracted from agarose gel and the 258 bp was cleaned up was using the NucleoSpin gel and PCR clean up kit according to the manufacturer's guidelines (Macherey-Nagel, Duren, Germany). A ligation reaction to introduce the individual donors into the pDe-Cas9-sgRNA cassette was set up with 2 µL of the linearized vector (pDe-Cas-gRNA), 3 µL of the donor sequence DNA, 1 µL of T4 ligase, 5 µL of T4 ligase buffer and 3 µL of nuclease free water, performed at 22°C in a thermocycler for 1.5 hours. The ligation reaction was transformed in NEB® 5α *E. coli* chemically competent cells according to the manufacturer's guidelines (New England Biolabs, Massachusetts, USA) plated on LB plates containing spectinomycin (100 mg/L) at 37°C overnight resulting in vectors pDe-Cas9-258-*bar* and pDe-Cas9-1185-*bar*. A colony PCR was set up as described above to identify positive clones. The expected product of 1190 bp was visualised on 1.5% TAE agarose gel stained with ethidium bromide using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA). The cloning was further confirmed by Sanger sequencing with Donor forward primer and primer SS43 (*bialaphos* gene) (Table 2.1).

2.3.3.4 Exchange of **Bialaphos** *gene (***bar***) gene for kanamycin (***nptII***) gene in pDe-Cas9-258*-**bar**

The *nptII* selection marker has been recommended as most suitable selection marker for sunflower transformation (Radonic et al., 2015). Due to the presence of similar restriction sites in both the 1185 donor fragment and *nptII* fragment, this exchange was only possible for the short donor pDe-Cas9-258-*bar* vector. The pDe-Cas9-258-*bar* vector and the pDe-CasD10A (carrying the 1397 bp *nptII* gene and CaMV promoter) vector were digested with HindIII in two separate reactions with 2 µL of Anza 10X buffer, 1 µL of plasmid DNA, 1 µL Anza HindIII and nuclease free water to a final volume of 20 µL. The reaction was incubated at 37°C for 1 hour and heat inactivated at 80°C for 5 min. The target fragments were gel purified using the NucleoSpin gel and PCR clean up kit as previously stated. A ligation reaction was set up with 2 µL pDe-Cas9-258, 3 µL *nptII* gene fragment, 1 µL of T4 ligase, 5 µL of T4 ligase buffer and 3 µL of nuclease free water, performed at 22°C in a thermocycler for 1.5 hours. The ligation reaction was transformed in NEB® 5α *E. coli* chemically competent cells according to the manufacturer's guidelines (New England Biolabs, Massachusetts, USA) containing kanamycin (100 mg/L) and plated on LB plates at 37°C overnight resulting in the vector pDe-Cas9-258-*nptII*. A colony PCR was set up to identify positive clones as described above in section 2.3.3. The resultant PCR product of 501 bp was visualised on ethidium bromide stained agarose gel using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA).Confirmation was done by Sanger sequencing with Donor forward primer and primer SS102 (targeting the *nptII* gene, Table 2.1. This primer spans the ligation site and therefore only binds to correctly orientated plasmids with only one insert). Further confirmation of plasmid molecular weight was done by gel electrophoresis and visualised on ethidium bromide stained agarose gel same as above.

2.3.4 *AGROBACTERIUM TUMEFACIENS*-MEDIATED PLANT TRANSFORMATION

2.3.4.1 TRANSFORMATION OF CRISPR/CAS9 SYSTEM VECTORS INTO *AGROBACTERIUM TUMEFACIENS*

a. Preparation of chemically competent *A. tumefaciens* LBA4404 cells

Agrobacterium tumefaciens LBA4404 strain was used to deliver the pDe-Cas-258-*nptII* and pDe-Cas9-1185-*bar* into the sunflower genome. Chemically competent *A. tumefaciens* cells were prepared by streaking bead cultures of *A. tumefaciens* into LB agar plates containing 50 mg/mL rifampicin (Rif50) and 50 mg/mL spectinomycin (Spec50) or kanamycin 50 mg/mL (Kan50) incubated at 27°C inverted for 48 hours. A single colony was selected from the plate

and was inoculated into 10 mL yeast-mannitol (YM) medium (composition of all media used for transformation are listed in Table S2.1) supplemented with Rif50, and grown overnight at 27°C under constant aeration to allow the cells to reach an OD_{600} of 0.5 to 1.0. A flask containing 100 mL YM medium supplemented with Rif50 and Spec50 was inoculated with the 10 mL of the overnight culture. Subsequently the cells were grown overnight at 27°C under constant aeration. The cultures were chilled on ice for 15 min followed by centrifugation at 4000 rcf for 10 min at 4°C in a bench top centrifuge. The supernatant was decanted and the pellet resuspended in 2.5 mL of ice cold 20 mM calcium chloride ($CaCl_2$). Aliquots of 300 µL were made from the cell suspension and flash frozen in liquid nitrogen and stored at -80°C.

b. Transformation of plasmids into chemically competent *Agrobacterium tumefaciens* LBA4404

The direct *A. tumefaciens* (freeze-thaw) (Xu and Li, 2008) method was used for transformation of pDe-Cas-258-*nptII* and pDe-Cas9-1185-*bar* DNA into *Agrobacterium tumefaciens* strain LBA4404. The competency of the *Agrobacterium tumefaciens* was tested by transforming the pDe-Cas9-gRNA vector. The growth of colonies on selection media indicated *Agrobacterium* competency (Results not shown). Transformation into *Agrobacterium* was confirmed by a colony PCR by directly picking a single colony from the plate and directly adding it to the PCR mix (Rönspies et al., 2022) using the sgRNA reverse primer and SS42 (U6-26 promoter) primer as described previously. On the day of transformation, fresh *Agrobacterium tumefaciens* culture was inoculated into 100 mL AIM medium (Table S2.1) (*Agrobacterium tumefaciens* infection medium) supplemented with 300 µM acetosyringone and incubated at 28°C in low speed shaker until it reached an OD_{600} of between 0.2 – 0.5.

2.3.5 Seed disinfection

For all transformation experiments, mature seeds (ARC-GC inbred line KP328) were surface sterilized with 70% ethanol for 1 min, followed by surface sterilization with mercuric chloride ($HgCl_2$) (0.1% w/v) for 10 min (Sujatha et al., 2012). Seeds were then rinsed with nuclease free water five times for 5 min to remove excess $HgCl_2$. Seeds were left soaking in 1% bleach with a drop of Triton X for overnight shaking at 22°C in low speed. The bleach and Triton X were removed by rinsing seeds three times with sterile water for 5 min each time.

2.3.6 Plant transformation

Disinfected seeds were dehusked using a sterilized scalpel and tweezers. Dehusked seeds were placed in a petri dish containing 20 mL half MS medium (Musharige and Skoog, 1962)

and cultured at 22°C in the dark overnight (Radonic et al., 2015). Seeds that developed a growing radicle were used for transformation. The seeds were heat shocked for 15 min at 42°C and the radicle was cut and discarded immediately. This was followed by a longitudinal cut along the embryo axis with a scalpel and cotyledons containing the embryo axis were placed in sterile petri dishes (approximately 10 - 15 seeds per plate) containing 15 mL *Agrobacterium tumefaciens* culture containing pDe-Cas-258-*nptII* and pDe-Cas9-1185-*bar* construct diluted to 0.6 OD_{600} and 100 µM acetosyringone and subjected to vacuum infiltration for 30 min. The cotyledons in *Agrobacterium tumefaciens* culture were later incubated in the laminar flow in the dark for 2 hours with occasional manual stirring to increase transformation efficiency. The explants were blotted dry with a sterile filter paper and transferred to co-culture medium (Table S2.1) supplemented with acetosyringone (100 µM), 1 mg/mL 6-Benzylaminopurine, 0.2 mg/L naphthalene acetic acid and co-cultivated in the dark for 2 days at 22°C in the dark.

After co-cultivation explants were transferred to regeneration/selection medium number 1 (RE1) (Table S2.1) supplemented with 150 mg/L cefotaxime, 0.2 mg/mL 6-Benzylaminopurine, 0.01 mg/L naphthalene acetic acid, 0.82 µg/mL $AgNO_3$ and 1 mg/L kanamycin/1 mg/L glufosinate ammonium. These were incubated under 16 h photoperiod, at 70 µmol/m^2/s provided by fluorescent bulbs at a constant temperature of 20ºC. After 10 - 14 days of incubation, shoots were transferred to regeneration/selection medium number 2 (RE2) supplemented with 150 mg/L cefotaxime, 0.2 mg/mL 6-Benzylaminopurine, 0.82 µg/mL $AgNO_3$ and 10 mg/L kanamycin/glufosinate ammonium. After 7 - 10 days of incubation, shoots were detached from the originating cotyledon transferred to regeneration/selection medium number 3 (RE3) supplemented with 150 mg/L cefotaxime, 0.2 mg/mL 6-Benzylaminopurine, 0.82 µg/mL silver nitrate ($AgNO_3$) and 50 mg/L kanamycin/glufosinate ammonium. After 7 - 10 days, shoots were transferred to selection medium with the same composition. After 10 - 14 days, shoots were maintained in rooting medium supplemented with 150 mg/mL cefotaxime, 0.1 mg/L naphthalene acetic acid and 50 mg/L kanamycin/glufosinate ammonium or half strength MS medium supplemented with 1 mg/L naphthalene acetic acid and 250 mg/L cefotaxime until they developed roots. Rooted shoots were transferred to the GM growth facility (ARC Biotechnology Platform, Onderstepoort, South Africa) and grown under conditions described in section 2.3.1 until maturity.

2.3.6.1 PCR mediated confirmation of transformed sunflower plants

PCR analysis was performed for the confirmation of successful transformation and T-DNA integration in sunflower plants that survived the selection steps. Genomic DNA from transformed and untransformed (WT) plants was extracted using the QIAGEN DNAeasy plant mini kit (Qiagen, Hilden, Germany) as per manufacturer's guidelines. PCR analysis was carried out to amplify the 284 bp fragment of the U6-26 promoter and sgRNA using the SS42 and sgRNA reverse primers (Table 2.1). The PCR mixture contained 1X OneTaq®2X Master mix with standard buffer, 0.2 µM of each primer, 50 ng of template DNA and 9 µL nuclease free water. The PCR thermal cycling conditions were: Initial denaturation at 94°C for 30 s, 30 cycles of denaturation at 94°C for 15 s, annealing temperatures were target specific, with 60°C for 30 s, elongation 68°C for 30 s, and final extension 68°C for 5 min. The PCR products were visualised on agarose gels, with 1.5% using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA).

2.3.7 Cas9 and sgRNA functional analysis

2.3.7.1 RNA extraction and RT-PCR

Young leaf material was collected from putative transgenic sunflower T0 lines. Approximately 100 mg of fresh leaf material was snap frozen in liquid nitrogen and immediately crushed into powder using a plastic micro-pellet pestle (Sigma Aldrich, Massachusetts, United States). Total RNA was extracted from using the Zymo Quick RNA Plant Miniprep kit (Zymo Research, California, United States) according to manufacturer's guidelines. The RNA quality was confirmed using a NanoPhotometer (IMPLEN GmBH, California, USA). Approximately 500 ng of DNase treated RNA was converted into cDNA using the First strand cDNA synthesis kit (ThermoFisher Scientific, Massachusetts, USA) with random hexamer primers as per manufacturer's instructions and subsequently used for RT-PCR reactions. The polymerase reactions were performed using OneTaq® 2X Master mix with standard buffer according to the manufacturer's guidelines (New England Biolabs, Massachusetts, USA). Primers used for this analysis are listed in Table 2.1. Each RT-PCR reaction consisted of 1X OneTaq®2X Master mix, 0.2 µM of each primer, 100 ng of template cDNA and 9 µL nuclease free water. The PCR thermal cycling conditions were: Initial denaturation at 94°C for 30 s, 30 cycles of denaturation at 94°C for 15 s, annealing temperatures were target specific, with 61°C for Cas9 confirmation and 58°C for sgRNA confirmation - both with 30 s duration, elongation 68°C for 30 s, and final extension 68°C for 5 min. The PCR products were visualised on agarose gels,

with 1.5% for Cas9 and 3% for sgRNA products using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA)

2.3.8 MUTATION ANALYSIS

2.3.8.1 Targeted deep sequencing

To detect mutations in transgenic sunflower shoots, genomic DNA was extracted from the leaves of T0 putative transgenic sunflower shoots once they were rooted using the QIAGEN DNAeasy plant mini kit (Qiagen, Hilden, Germany) as per manufacturer's guidelines. For shoots transformed with pDe-Cas9-1185-*bar* an initial PCR reaction with CENH3LDrv and CENH3LDfw primers (Table 2.1) designed to only target the native *CenH3* sequence and avoid targeting the DRT. This was performed using the Q5 high fidelity DNA polymerase. The 1778 bp PCR product was visualised on the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA) with ethidium bromide on 1.5% agarose gel after electrophoresis and cleaned up using the NucleoSpin gel and PCR clean up kit according to the manufacturer's guidelines (Macherey-Nagel, Duren, Germany). The PCR products from plants transformed with pDe-Cas9-1185-*bar* and samples transformed with pDe-Cas-258-*nptII* were subject to a second PCR using *CenH3* target specific primers (Table 2.1). The *CenH3* target specific primers contained Illumina adaptor sequences prepared in accordance with the "16S metagenomic sequencing library preparation protocol" (Illumina, California, USA). The primers were carefully selected to avoid binding to the short donor sequence (258 bp). These primers were used to amplify the *CenH3* target region using Q5 High-Fidelity DNA polymerase (New England Biolabs, Massachusetts, USA). The composition of 25 µL PCR reactions was: 20 ng genomic DNA; 0.02 U/µL of Q5 High-fidelity polymerase; 1X Q5 reaction buffer; 200 µM dNTPs; and 0.5 µM of each primer. The PCR amplification protocol used: initial denaturation for 30 s at 98°C; 35 cycles of denaturation for 5 s at 98°C; annealing for 10 s at 72°C, elongation for 30 s at 72°C and final elongation for 2 min at 72°C. Amplified DNA was visualised on the using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA) with ethidium bromide under UV light after electrophoresis in 1.5% agarose gels. The 370 bp PCR products were purified using the AMPure bead purification, as per manufacturer's guidelines, and subject to an indexing PCR where the amplicons were pooled into a single library using the Nextra XT index kit (Illumina, California, USA). Library preparation and sequencing were performed on the Illumina MiSeq platform using 300 bp PE sequence at the Agricultural Research Council's Biotechnology Platform Core facility (Onderstepoort, South Africa).

To investigate CRISPR/Cas9 induced mutations at the sgRNA target sites, the FASTQ files were loaded on to the CRISPResso2 software (Clement et al., 2019). The CRISPResso2 software analyses the file by quality filtering the reads, trimming of adapters, merging the reads and aligning to reference (wild type) sequence. Default CRISPResso2 settings were used for this analysis.

2.4 RESULTS

2.4.1 *CENH3* TARGET SEQUENCE IDENTIFICATION AND DESIGN

2.4.1.1 Identification of sunflower histone fold domain

The CENH3 protein is composed of a highly variable N-terminal tail region that can even be variable within species – in both sequence and size (Jiang et al., 2003). Single amino acid changes in the conserved region of the histone fold domain were shown to induce haploid progeny upon crossing of carrier plant with a wild type plant. Multiple sequence alignment of the *CenH3* gene from 19 plants in MUSCLE (Edgar, 2004) indicated that the N-terminal tail had high diversity as expected with considerable differences in both sequence and length even within genetically similar species such as sunflower and lettuce (Figure 2.2). The HFD was identified as the region with the highly conserved sequence across species as highlighted by the red arrow in Figure 2.2. High sequence and size conservation was observed in this region, across all species. Taking this into account, the histone fold domain was identified to begin at position number 51 in sunflower, while it begins at position number 82 in *Arabidopsis,* this is due to *Arabidopsis* having a much longer N-terminal tail in comparison to sunflower. Figure 2.2 sequence accessions are indicated in the legend.

N-terminal tail

```
Label                        . 2 . 4 . 6 . 8 . 10 . 12 . 14 . 16 . 18 . 20 . 22 . 24 . 26 . 28 . 30 . 32 . 34 . 36 . 38 . 40 . 42 . 44 . 46 . 48
Musa_acuminata               MARTKHLSNRSSSR-PRKRFHFGRSP-------GQRTPGPPISPFLLS
betaHordeum_bulbosum]        MARTKKTVAATKRRSPRTRLEPQSQP----------------------
Hordeum_bulbosum             MARTKHPAVRKSKAPPRKKVGSARAPAAAQRRHETDGAGTSETPRRGP
Zea_mays                     MARTKHQAVRKTAEKPKKKLQFERS----------GGASTSATPERAA
Oryza_sativa                 MARTKHPAVRKSKAEPKKKLQFERSPRPSKAQRAGGGTGTSATTRSAA
Phaseolus_vulgaris           MARVKTTPASRKTGKKKAARASTSTP------------RSQQSPGRRR
Glycine_max                  MARVKHTPASRKSGKKKAPRASTSTQ---------PSLSSSQSPATRE
Gossypium_hirsutum           MSRTKHTAAKKPRRKPSA--AAAALP---------ATASPHTRSVTAK
Helianthus_annuus]           MARKKQPAKRSWGTRRDS--RA--SA---------STSTPRKTPRKEP
Lactuca_sativa               MARTKQPAKRSWGKRQSA--GASTST---------STSTPRKSPRKDP
Daucus_carota                MARTKHPAKRTSGHR----------S---------RGPPLSGTPRRRS
Nicotiana_tabacum            MARTKHLALRKQSRPPSRPTATRSAA---------AAASSSAPQSTPT
Solanum_lycopersicum         MARTKHLAKRSRT----------------------TSAAPSATPSTPS
Solanum_tuberosum            MARTKHLAKRSRTKP----------S---------VAAGPSATPSTPT
Vitis_vinifera               MTRTKHLARKSRNRRRQFA-ATPASP---------ASAGPSSAPPRRP
Arabidopsis_thaliana         MARTKHRVTRSQPRNQTDA-AGASSS---------QAAGPTTTPTRRG
Arabidopsis_lyrata           MARTKHFATKSRSENRTD--ANASSS---------QAAGPTTTPTTRG
Brassica_rapa                MARTKHFASRARDRNPTNATASSSAA---------AAAGPSATPTRRG
Brassica_Juncea              MARTKHYASRARDRNRNNATASSTEA---------AAASPSATPTRRG

Label                        . 50 . 52 . 54 . 56 . 58 . 60 . 62 . 64 . 66 . 68 . 70 . 72 . 74 . 76 . 78 . 80 . 82 . 84 . 86 . 88 . 90 . 92 . 94 .
Musa_acuminata               SLVCWWAPSQ--------------------------SKKQPRRRRFR
betaHordeum_bulbosum]        -------------------------------------EKKKRAHRFR
Hordeum_bulbosum             APAADQGAPGEP--------------------------KKRKPHRYR
Zea_mays                     GTGGRAASGGDSVK----------------------KTKPR--HRWR
Oryza_sativa                 GT----SASGTPRQ----------------------QTKQRKPHRFR
Phaseolus_vulgaris           AQQEEPQEAEEEEEEAAAAP----------------ETQGRKKRRNK
Glycine_max                  RSR-----------------------------------RGRKKMRNR
Gossypium_hirsutum           KTGGPGT-----------------------------PRKLKRPHRFR
Helianthus_annuus]           ESSGGG--------------------------------GNKKPHRFK
Lactuca_sativa               GSSGTG------------------------------QRQKQKPHRFK
Daucus_carota                TATPREADA-Q-------------------------GQQQRKPHRFR
Nicotiana_tabacum            RTSQRTAPSTP-------------------------GRTQKKKTRYR
Solanum_lycopersicum         RKSPRSAPATS-------------------------VQKPKQKKRYR
Solanum_tuberosum            RKSPRSAPATS-------------------------VQKPKQKKRYR
Vitis_vinifera               TRTATDASPSTAG-----------------------SQGQRKPFRYR
Arabidopsis_thaliana         GEGGDNTQQTNPTTSPATGTRR---GAKRSRQAMP-RGSQKKSYRYR
Arabidopsis_lyrata           TEGGDNTQQTNPTTSPATGGRR----PRRARQAMP-RVSQNKPYRYK
Brassica_rapa                GSQGGEAQQSAATPPATTTAGRKKGGTKRTKQAMPKSSNKKKTFRYK
Brassica_Juncea              SSSSQGAREAQQSS----------------------NQKKKKSFRYK
```

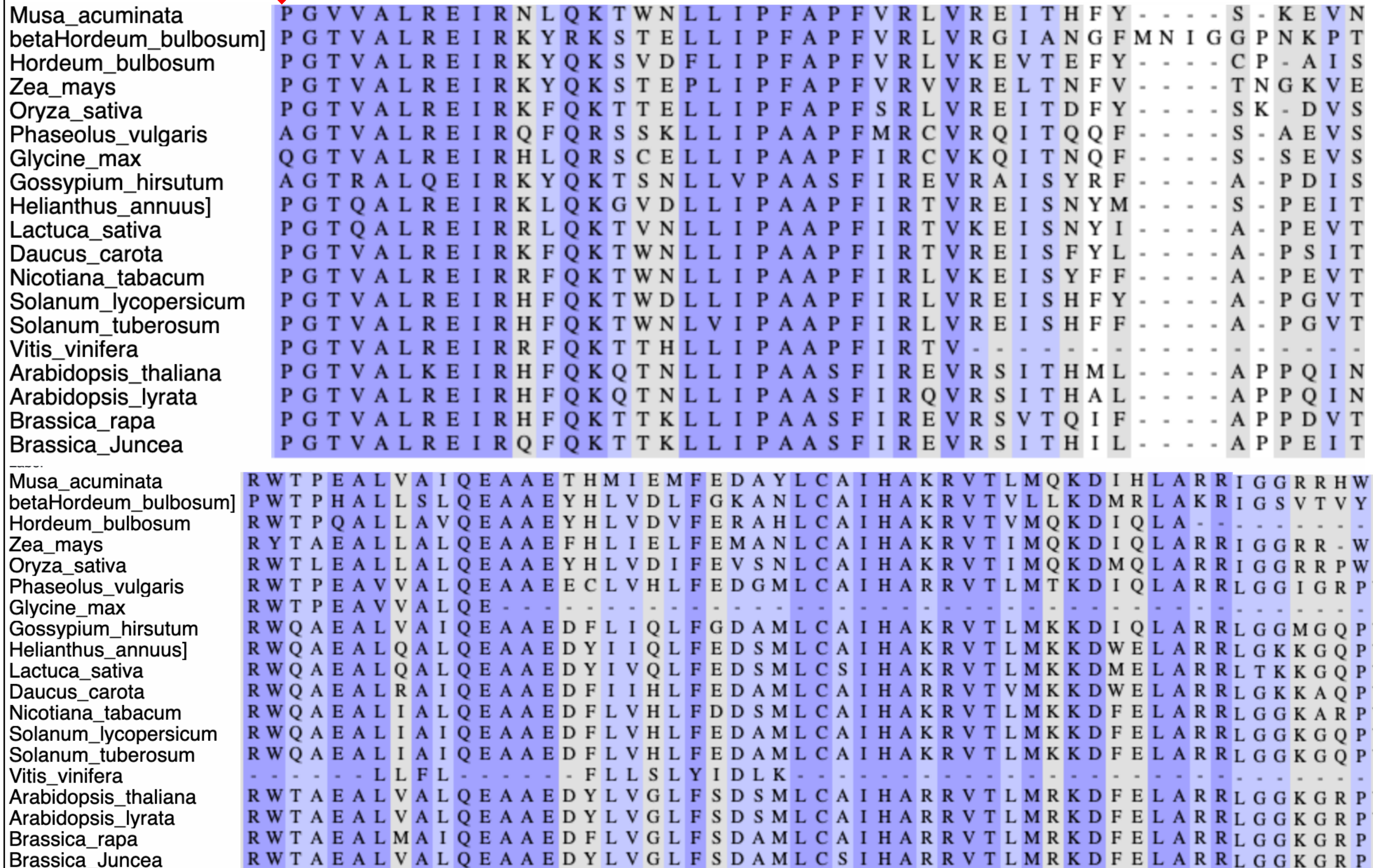

Histone fold domain (HFD)
Musa_acuminata
betaHordeum_bulbosum]
Hordeum_bulbosum
Zea_mays
Oryza_sativa
Phaseolus_vulgaris
Glycine_max
Gossypium_hirsutum
Helianthus_annuus]
Lactuca_sativa
Daucus_carota
Nicotiana_tabacum
Solanum_lycopersicum
Solanum_tuberosum
Vitis_vinifera
Arabidopsis_thaliana
Arabidopsis_lyrata
Brassica_rapa
Brassica_Juncea

Figure 2.2 Multiple amino acid alignment using MUSCLE (Edgar, 2004) of the full CENH3 protein of nineteen (19) plant species i.e. *Hordeum bulbosum* (barley; ADB03182.1), *Musa acuminata* (banana; XP_018685258.1), *Zea mays* (maize; NP_001105520.1), *Oryza sativa* (rice; AAR85315.1); *Phaseolus vulgaris* (bean; AGQ21573.1), *Glycine max* (soybean; XP_006598915.1), *Vitis vinifera* (grapevine; RVC10123.1), *Brassica rapa* (rapeseed; ADN92693.1), *Arabidopsis thaliana* (AAL86775.1), *Arabidopsis lyrata* (XP_020870190.1), *Gossypium hirsufum* (cotton; XP_016745574.1), *Helianthus annuus* (sunflower; XP_022005281.1, This study), *Lactuva sativa* (lettuce; XP_023761199.1), *Daucus carota* (carrot; AID21728.1), *Nicotiana tabacum* (tobacco; BAH03515.1), *Solanum lycopersicum* (tomato; XP_010326926.1), *Solanum tuberosum* (potato; XP_006339687.1) and *Brassica juncea* (mustard;BAF49728). The N-Terminal and Histone Fold Domain are marked at the top of the alignment. Graded blue (high) to grey (low) colour represents the level of conservation among the representative sequences. The red arrow represents the beginning of the histone fold domain (HFD).

A further protein sequence alignment was performed to verify that the amino acids previously identified as targets for haploid induction were conserved in sunflower and other species. The HFD region of four crop species was aligned against *Arabidopsis thaliana* and *Arabidopsis lyrata* (Figure 2.3). The protein sequence had high similarity with all four helix regions present in the histone fold domain aligning to each other. The target amino acids for mutation are highlighted in red (P51), (G52), (A55), (L98) and (A102). The targeted amino acids identified by Kuppu et al., (2015) were found to be conserved in all the plants, excluding carrot in the last amino acid.

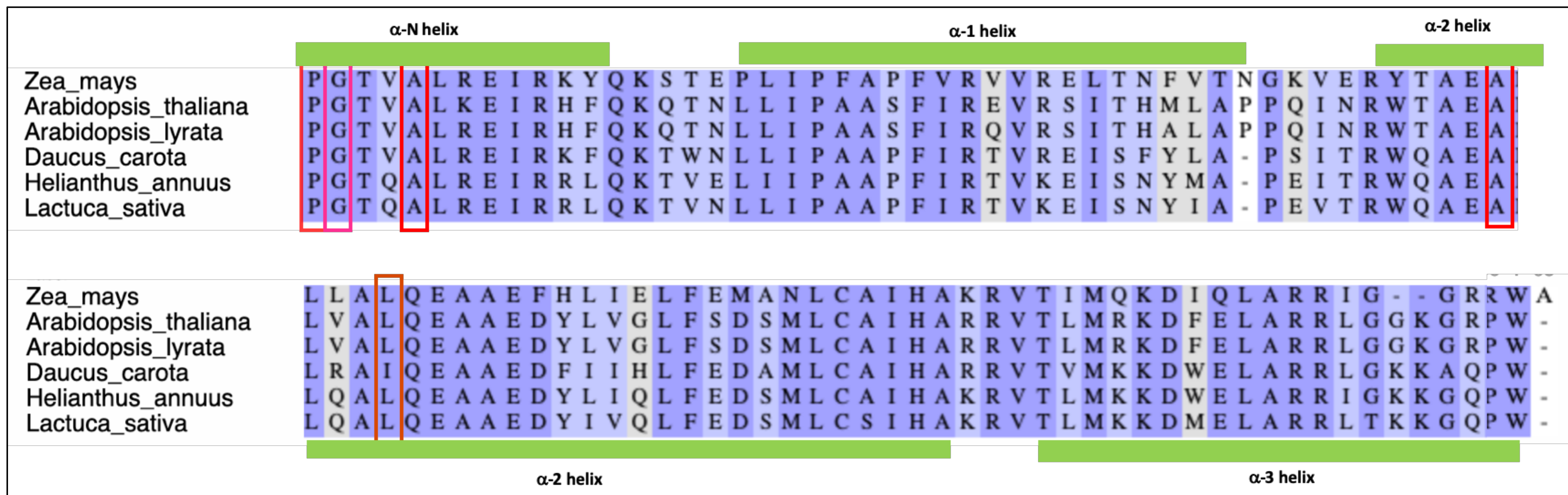


Figure 2.3 Multiple sequence alignment of predicted CENH3 histone fold domain (HFD) of *Zea mays* (NP_001105520.1), *Arabidopsis thaliana* (AAL86775.1), *Arabidopsis lyrata* (XP_020870190.1), *Daucus carota* (AID21728.1), *Helianthus annuu*s (XP_022005281.1, This study) and *Lactuca sativa* (XP_023761199.1). The amino acids previously linked to haploid induction (Kuppu et al., 2015) are within red boxes (P82/P51), (G83/G52), (A86/A55), (A132/L98) and (A136/A102). The alpha (α) helices are indicated as green bars at the top and bottom of the alignment (Karimi-Ashtiyani et al., 2015).

2.4.2 VECTOR CLONING

The aim of this study was to produce haploid inducer lines in sunflower through the modification of the *CenH3* gene. A single gRNA 20 bp target site specific molecule targeting exon 1 of the *CenH3* gene was designed and cloned into the pEN-Chimera plasmid, successful ligation was confirmed by colony PCR (PCR product size 375 bp (Figure 2.4A) and Sanger sequencing (Figure S2.3), this vector was named pEN-Chimera-gRNA. The sgRNA scaffold containing the designed sgRNA and the *Arabidopsis* U6-26 promoter were transferred from the pEN-Chimera-gRNA vector into the pDe-Cas9 plasmid, containing the Cas9 endonuclease, parsley ubiquitin promoter, Gateway® cloning (Invitrogen, Massachusetts, USA). Successful Gateway® transfer was confirmed by colony PCR (Figure 2.4B) and Sanger sequencing (Figure S2.4). Two donor sequences (258 and 1185 bp) with three amino acid changes in the histone fold domain were designed, synthesized by GENEWIZ and GeneArt and ligated into the pDe-Cas9-gRNA plasmid using restriction cloning and named pDe-Cas-258-*bar* and pDe-Cas9-1185-*bar*. For the pDe-Cas-258-*bar*, the *bialaphos* plant selection gene was exchanged with the *nptII* gene obtained from the pDe-CasD10A plasmid. The successful cloning was confirmed by colony PCR (Figure 2.4C and Figure 2.4D) with unsuccessful amplification represented by empty lanes (most likely due to failed cloning) and Sanger sequencing (Figure S2.5 and Figure S2.6). The final T-DNA plasmids (Figure 2.5) were used for *Agrobacterium tumefaciens* transformation in sunflower.

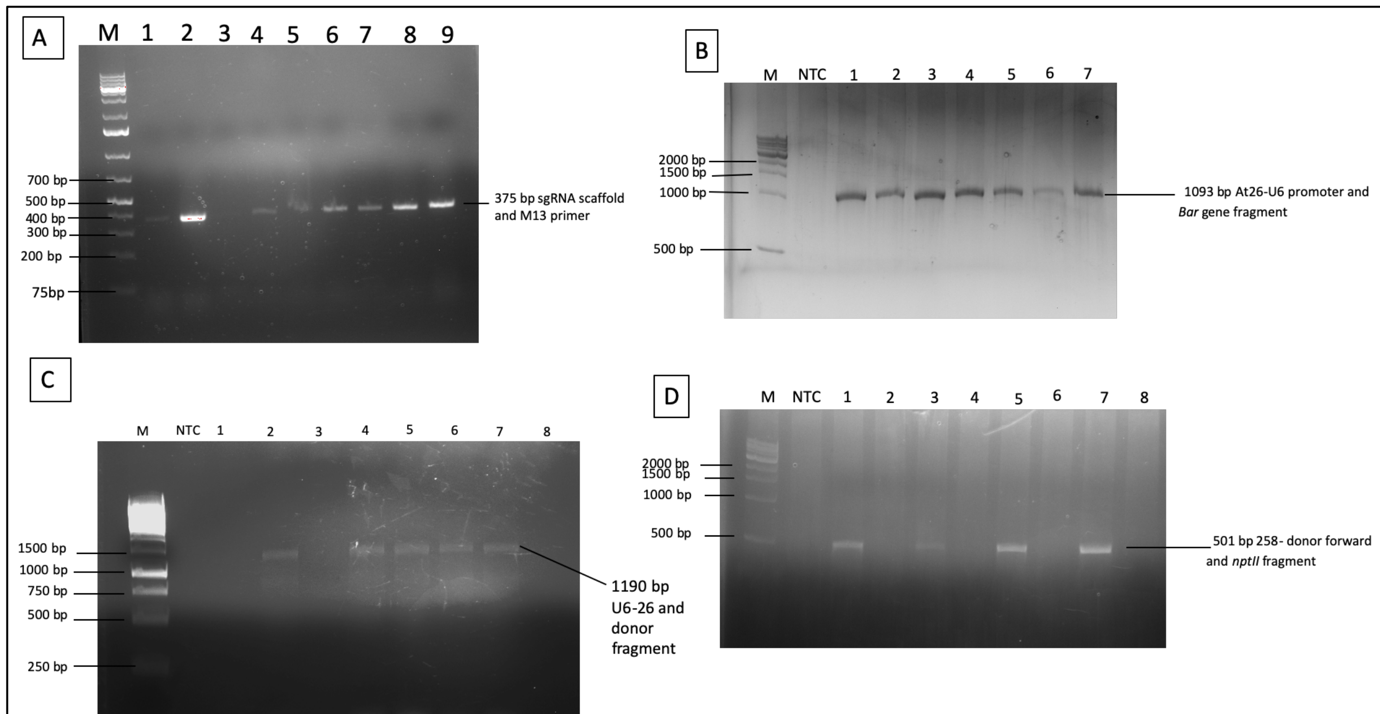


Figure 2.4 Following colony PCR amplification, the DNA fragments were resolved on 1.5% agarose gels to confirm successful cloning of the A). target specific sgRNA: M- Molecular weight DNA marker GeneRuler 1kb plus marker, Lane 2-9 - Randomly selected colonies B). sgRNA, sgRNA scaffold and atU6-26 promoter into pDe-Cas9: M- NEB 1kb molecular, NTC- non-template control, 1-7- Randomly selected colonies C). Large donor (1185 bp) into pDe-Cas9-gRNA: M- GeneRuler 1kb molecular marker, NTC- non-template control, 1-8- Randomly selected colonies and D). Short donor (258 bp) and *nptII* gene confirmation into pDe-Cas9-gRNA: M- NEB 1kb molecular marker, NTC- non-template control, 1-8: Randomly selected colonies. Colony PCR does not provide standardised DNA concentrations which influences the amount of PCR product produced and may thus result in varying gel band intensity as seen in the figure.

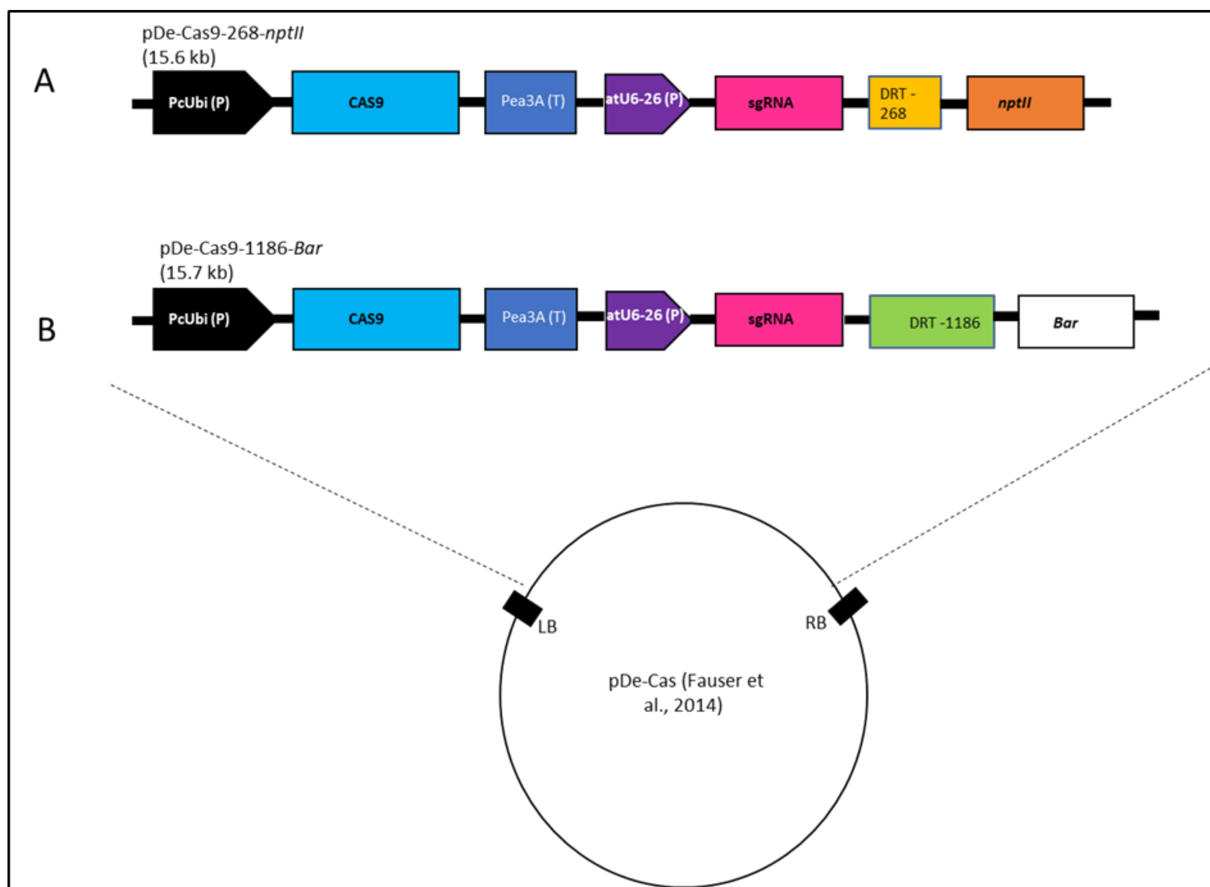


Figure 2.5 The binary plasmids used for *Agrobacterium tumefaciens* transformation of sunflower mature cotyledons with A. kanamycin selection pDe-Cas9-258-*nptII* and B. *bialaphos* resistance gene pDe-Cas9-1185-*bar*. PcUbi (P): Parsley ubiquitin promoter; Cas9: endonuclease Cas9; Pea3A (T): pea RBCS3A terminator; atU6-26 (P): *Arabidopsis* U6-26 promoter; sgRNA: single guide RNA; DRT-258 or DRT-1185: donor sequence DRT; *nptII*: cassette for kanamycin plant selection; *Bar:* cassette for *bialaphos* plant selection.

2.4.3 *AGROBACTERIUM TUMEFACIENS* -MEDIATED PLANT TRANSFORMATION

Sunflower mature cotyledons (Figure 2.6A) were used for transformation with *Agrobacterium tumefaciens* strain LBA4404 containing plasmid constructs pDe-Cas9-258-*nptII* and pDe-Cas9-1185-*bar*. Cotyledons were co-cultured with *Agrobacterium tumefaciens* for 2 days after transformation to enhance infection (Figure 2.6A). Transferral to the regeneration medium (Table S2.1) containing cefotaxime, to suppress *Agrobacterium tumefaciens* growth (Sujatha et al., 2012). The cotyledons produced shoots within 10 days after transformation (Figure 2.6B). Antibiotic sensitivity characterised by bleaching, yellowing or darkening of plants (Figure 2.6C1) was regarded as non-transformed while plants that remained viable were characterised as putative transgenic plants carrying the T-DNA (Figure 2.6C2). Bleaching of

plants was observed from the apex of the shoot and proceeded downwards. Plants that remained viable (green) after the three cycles of selection (approximately 50 - 60 days) were transferred to rooting medium (Table S2.1). Bleached untransformed plants appeared to regenerate quicker and appeared larger, even under antibiotic selection when compared to putative transgenic plants (Figure 2.6 C1 and C2). Root development was poor in several plants (Figure 2.6D). Explants with healthy roots were acclimatized to the growth medium (perlite). Stems that did not develop roots were grafted onto sunflower root stocks (Figure 2.6E) as performed by Radonic et al., (2015). All the T0 plants failed the acclimatization step.

A total of 180 seeds and 360 cotyledons were transformed using the pDe-Cas9-258-*nptII* and pDe-Cas9-1185-*bar* with different length donor sequences for homology directed repair in sunflower. Only 7 plants showed integration of the T-DNA, confirmed by PCR using the U6-26 promoter and gRNA reverse primers (product size 284 bp) (Figure 2.7) and only 4 plants showed mutations (Table 2.3). The observed transformation frequency was 1.5-2.5% for both constructs (Table 2.3).

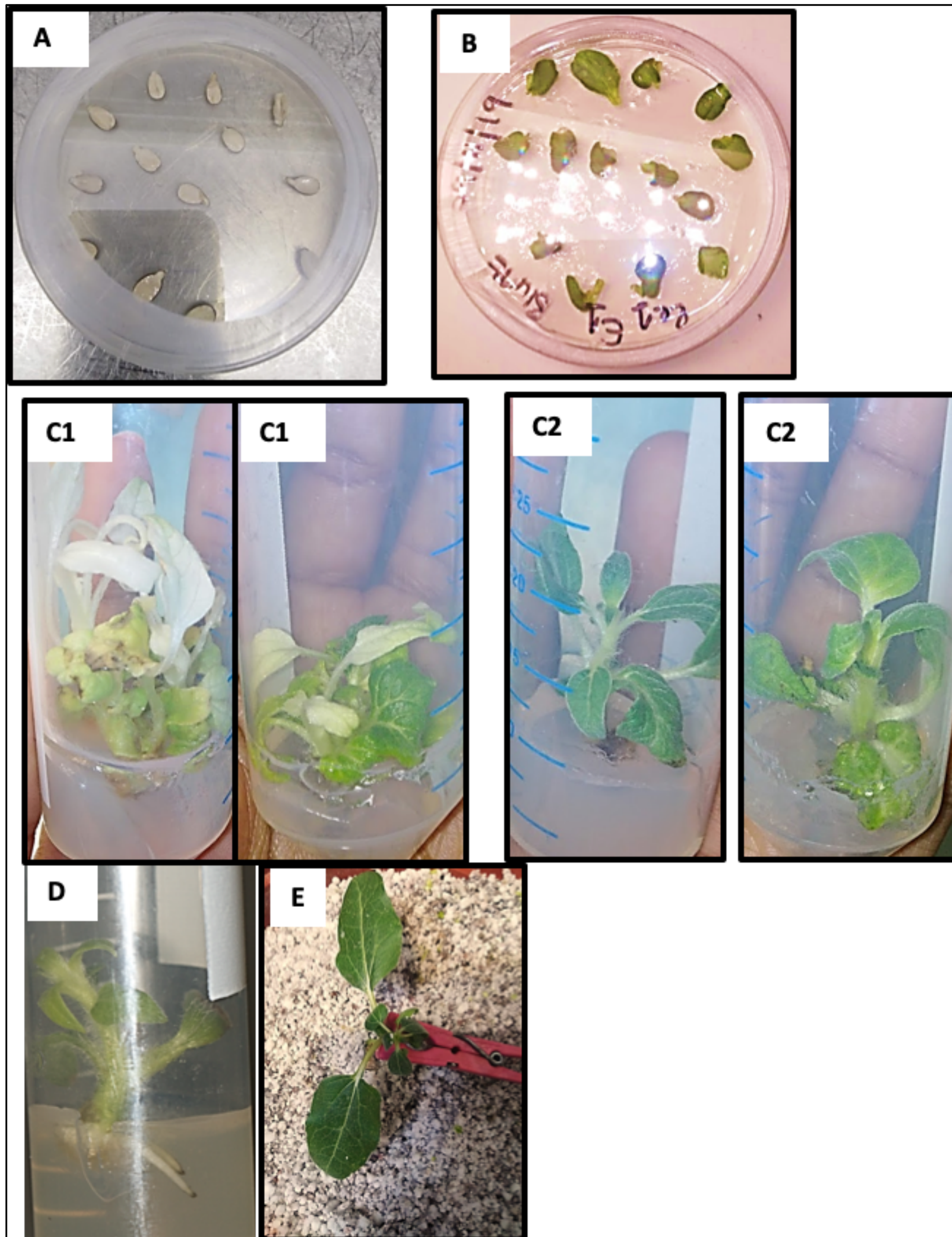


Figure 2.6 Transformation steps of sunflower (KP328, ARC-GC) cotyledons obtained from mature seeds. A. Cotyledons co-culturing with *Agrobacterium tumefaciens* after transformation; B. Shoot regeneration after 2 days of the co-culture period; C. Selection of putative transformed shoots on 50 mg/mL kanamycin or 50 mg/mL glufosinate ammonium and 250 mg/L cefotaxime for suppression of *Agrobacterium tumefaciens* growth; D. Transgenic plantlets on rooting medium and E. Acclimatization of shoots in perlite and C2 shoots that failed to graft onto a sunflower rootstock.

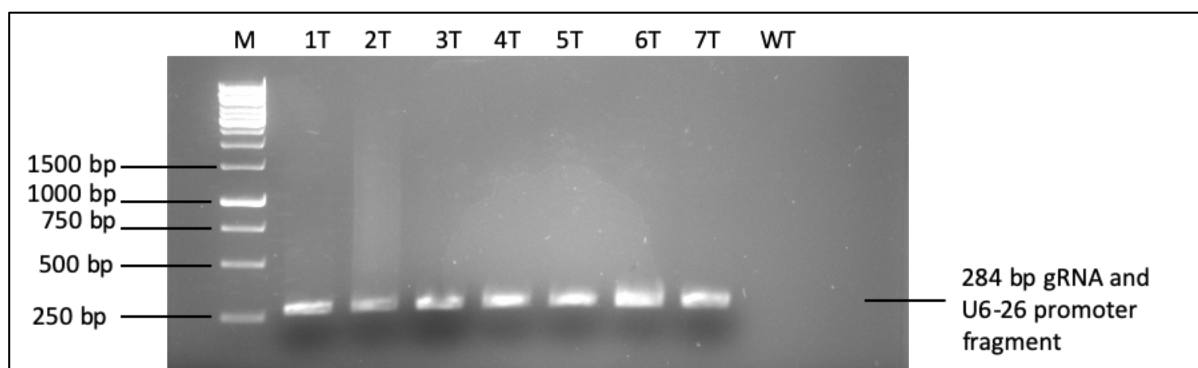


Figure 2.7 PCR products for confirmation of presence of T-DNA in putative transgenic plants using primers amplifying the sgRNA and U6-26 promoter (Table 2.1). M – GeneRuler 1kb molecular marker, 1T- 4T plants transformed with the pDe-Cas9-258-*nptII* vector, 5T-7T plants transformed with the pDe-Cas-1185-*bar* vector, WT- untransformed sunflower plant used as a control.

Table 2.3 Transformation and mutation efficiency of sunflower T0 shoots.

Construct	**Number of transformed cotyledons**	**Number of plants with Cas9/sgRNA transcript**	**Number of plants with mutations**	**Transformation efficiency**	**Mutation frequency**
pDe-Cas-258-*nptII*	160 cotyledons (80 seeds)	4 plants	4 plants	2.5%	100%
pDe-Cas9-1185-*bar*	200 cotyledons (100 seeds)	3 plants	None	1.5%	None

2.4.4 Cas9 and sgRNA functional analysis

Homology directed repair (HDR) requires the Cas9 endonuclease to produce a specific double stranded break in the presence of the donor region. This in turns requires sufficient sgRNA transcription, Cas9 endonuclease expression and copies of the donor region in the host. To test this in putative transgenic sunflower shoots, we performed an RT-PCR with specific primers targeting the sgRNA (Ma et al., 2015) and Cas9 as performed by Gao et al., (2017) (Table 2.1). Plants 1T - 4T were transformed with the pDe-Cas9-258-*nptII* construct while 5T-7T were transformed with the pDe-Cas9-1185-*bar* construct. The expected PCR amplicon of 96 bp (Figure 2.8A) was observed, indicating the presence of the sgRNA molecule in the

complementary DNA and confirming transcription in the putative transgenic lines. All the transgenic plants obtained (Table 2.3) displayed sgRNA expression. Cas9 expression was confirmed by the amplification of the 572 bp region within the Cas9 enzyme (Figure 2.8B). All the plants, except for plant number 4T indicated Cas9 expression. An untransformed control (WT) was used as a negative control in these PCR reactions.

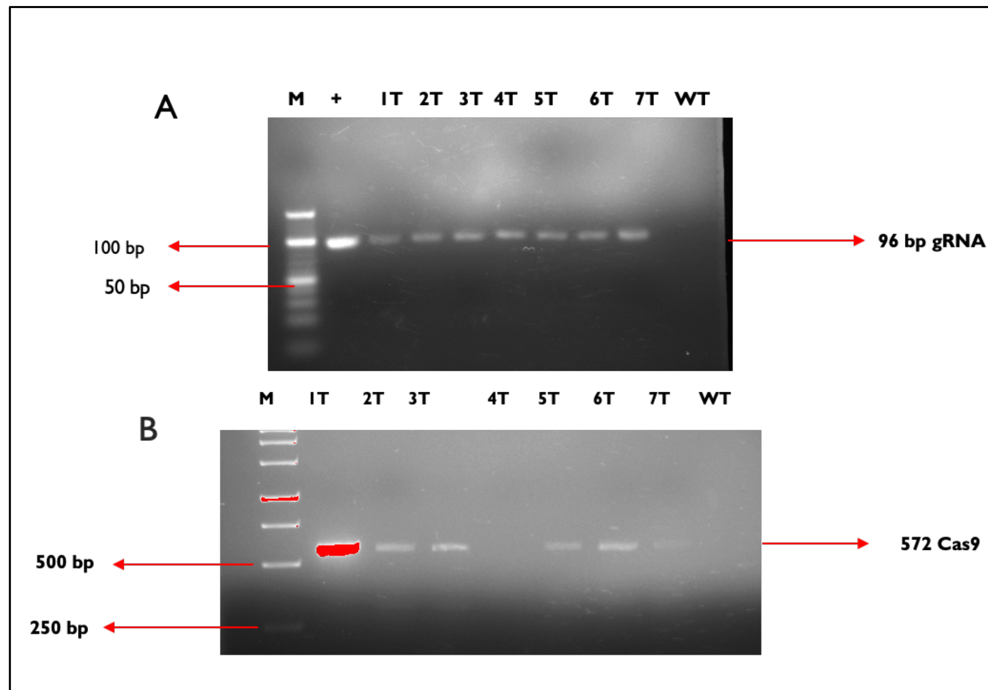


Figure 2.8 RT-PCR products confirming the transcription of the sgRNA and Cas9 in sunflower treated plants with a wild type (WT), untransformed control plant as a negative control. A. Transcription of sgRNA in PCR positive transgenic lines; Lane 1: O'RangeRuler 10 bp Molecular marker (Fermentas, Massachusetts, USA); Lane 2: Plasmid control; Lane 3-9: putative sunflower transgenic lines; Lane 10, Untransformed sunflower line (WT). B. Transcription of Cas9 endonuclease in putative sunflower transgenic lines; Lane 1: GeneRuler 1kb Molecular marker (ThermoFisher Scientific, Massachusetts, USA), Lane 2-8: putative sunflower transgenic lines, Lane 9: untransformed sunflower line (WT).

2.4.5 MUTATION ANALYSIS USING NEXT GENERATION SEQUENCING

The aim of this study was to introduce three amino acid substitutions in sunflower using HDR and the CRISPR/Cas9 system. This was attempted by introducing the Cas9 endonuclease with a single sgRNA target sequence and homology directed repair from an exogenously supplied double stranded donor repair template (DRT) in a single plasmid. Two different donors of different sizes were used, one double cut and one with no cutting sites. To identify mutation induced by CRISPR/Cas9 in sunflower T0 transgenic lines, genomic DNA was isolated from treated plants showing gRNA and Cas9 expression. Target specific sequences

were amplified and used in a next generation sequencing assessment. Mutations around the cleavage site were detected by CRISPResso2 (Clement et al., 2019). The CRISPResso2 analysis displayed the types and numbers of mutations in the different sunflower T0 transgenic lines. Mutations were only observed in the four transgenic plants transformed with the short, double cut donor. 1.77% of the alleles in Line 1T, 1.55% in line 2T, 1.63% in line 3T, and 1.42% of line 4T displayed different substitution mutations (Table 2.4). As shown in Figure 2.9, all the mutations observed were substitutions, with no mutations observed in negative control samples subjected to the transformation process without vectors (Figure S2.7). Nucleotide changes observed were conversions of G to A, C to T, G to T and A to G (Figure 2.9). Two of the targeted nucleotide changes were observed and successfully converted CCT -> TCT, and GCG -> GTG (Figure 2.9). None of the silent mutations to deactivate the recognition of the PAM sequence were observed. Additional changes not included in the donor sequence were however observed in some of the alleles. No allele contained more than a single nucleotide substitution. CRISPresso2 results for the untransformed control and plants 5T, 6T and 7T are displayed in the supplementary material (Figure S2.7), no mutations were observed for any of these.

Table 2.4 Mutation frequency in the four transgenic lines analysed with CRISPResso2 (Clement et al., 2019)

Transgenic plant	Total mutation frequency	Number of different alleles
1T	1.77%	6
2T	1.63%	5
3T	1.42%	4
4T	1.55%	4

Figure 2.9 Targeted mutagenesis in transgenic sunflower of four different T0 shoots (1T - 4T) analysed by NGS and CRISPresso2 (Clement et al., 2019). Distribution of identified alleles around the predicted cleavage site (vertical dash line, 3 bp upstream the PAM sequence) at sunflower *CenH3* target site (nucleotide number 551 – 590, XP_022005281.1). The reference allele is indicated in the first line and the donor repair template is in the second line and alignments are displayed in the low portion of the plot. Substitutions are shown in bold font. Only the target regions of interest are aligned and displayed.

The allele sequences obtained from the CRISPresso2 (Clement et al., 2019) analysis were translated to amino acids using EXPAsy (http://www.expasy.org; Swiss institute of Bioinformatics) and aligned against the wild type sunflower *CenH3* sequence and the donor repair template sequence designed for the gene editing (Figure 2.10). In the donor sequence, three amino acid changes, proline 51 to serine, glycine 52 to glutamic acid and alanine 55 valine were added. The P51S conversion was observed in all four transgenic lines, at varying frequencies of 0.40% - 0.30% (Figure 2.10, green square), while the A55V conversion was observed in two of the four transgenic lines, in frequencies of 0.22% and 0.21% (Figure 2.10). None of the transgenic plants contained the G52E amino acid conversion and none of the amino acid conversions occurred simultaneously in a single allele as designed in the donor. A common nucleotide change of G to A was observed in all the samples in nucleotide number 551. This, however, did not lead to any amino acid changes and was not included in Figure 2.10. Additional amino acid changes, not included in the donor design included Q54R, Q54H and R57H (Figure 2.10, red squares).

AMINO ACID POSITION	51	52	53	54	55	56	57	Occurrence of alleles
WILDTYPE	P	G	T	Q	A	L	R	All four plants
DONOR REPAIR TEMPLATE	S	E	T	Q	V	L	R	None
ALLELE_1	S	G	T	Q	A	L	R	All four plants
ALLELE_2	P	G	T	R	A	L	R	Three plants, 1T;2T and 4T
ALLELE_3	P	G	T	Q	V	L	R	Two plants, 1T and 2T
ALLELE_4	P	G	T	H	A	L	R	Two plants, 3T and 4T
ALLELE_5	P	G	T	Q	A	L	H	Three plants, 1T, 2T and 4T

Figure 2.10 Multiple sequence alignment of CENH3 histone fold domain amino acid changes induced by CRISPR/Cas9 HDR in four sunflower plants. The amino acids in green are the required changes for haploid inducer formation while the red are random changes induced by the CRISPR system through NHEJ. Only the targeted region of interest is aligned and displayed. Amino acid positions of the sunflower CENH3 protein and the occurrence of the alleles in the different plants are listed at the top.

2.5 DISCUSSION

Haploid production is one of the most attractive strategies for breeding programs, thus one of the main objectives of this study was to successfully introduce CRISPR/Cas9 genome editing system into sunflower for haploid inducer line development. In this study, one hundred and sixty sunflower cotyledons were transformed with a CRISPR/Cas9 construct containing a short donor with double Cas9 cutting sites. Two hundred sunflower cotyledons were transformed with a long donor without Cas9 cutting sites. These transformation trials resulted in four and three transgenic explants respectively that could be subjected to genetic testing. Furthermore, two of the target amino acid changes were introduced into sunflower plants transformed with the pDe-Cas9-258-*nptII s*hort donor fragment and no mutations were observed in plants transformed with the pDe-Cas9-1185-*bar* long donor cassette. Transformed plants however failed to acclimatize and reach maturity.

Although there have been improvements in sunflower *in vitro* culture regeneration, it still remains a challenge (Darqui et al., 2021). In this study the transformation frequency was calculated to be 2.5% and 1.5%, respectively for both plasmids, which is of moderate success compared to other reports. The reported transformation efficiencies in other studies range from around 2-7%, with Radonic et al., (2015) reporting a transformation efficiency of between 2.23% to 7.06% while Sujatha et al., (2012) reported an efficiency of 3.0% in sunflower. However, sunflower is generally regarded as recalcitrant to *in vitro* transformation leading to low transformation efficiency compared to other plants. Thus, currently no transgenic sunflower cultivar has been released for commercialization or authorized for cultivation (Darqui et al., 2021). Factors such as the type of promoter used, genotype, and co-culture period have all been attributed to the low transformation efficiency of sunflower (Radonic et al., 2015). The use of the CaMV35S promoter in sunflower with either *bar* or *nptII* selection showed less transformation, rooting issues and instability of transgenes in the T0 and T1 generations when compared to the *chrysanthemum rbcS1* promoter (Radonic et al., 2012, Radonic et al., 2015). The promoter used in this study is a parsley ubiquitin promoter, used for the expression of the Cas9 which has rarely been documented in the transformation of the *Asteraceae* family. The use of the *chrysanthemum rbcS1* promoter could be tested as an alternative in future studies as it has been shown to increase transformation efficiency in sunflower (Radonic et al., 2012, Radonic et al., 2015). After rooting, plants failed the acclimatization step and all the plants failed to graft successfully– a common challenge with sunflower in *in vitro* culturing according to literature (Darqui et al., 2021; Radonic et al., 2015). Grafting often results in a discontinuous vascular system which disrupt water transportation

during graft healing (Zhang and Finer, 2016).It has been widely reported that the success of CRISPR/Cas9 genome editing is largely dependent on target selection. For the sgRNA design, three of the five amino identified by Kuppu et al., (2015) as successful for haploid inducer line development were selected as target sites. The sunflower *CenH3* gene sequence was analysed and guide RNAs that would possibly target three of these amino acids simultaneously were manually identified on the α-N helix of the HFD region by observing for an NGG sequence nearest to the target site (preferably three nucleotides from the target sequence. Sunflower contains a single copy of the *CenH3* gene that is located on chromosome 15 (Nagaki et al., 2015). Based on this criterion, the closest target sites were identified and one guide RNA was located on exon 1. The sgRNA sequence compositions affects efficiency of editing, for example, GC rich sequences (higher than 80%) and GC low sequences (less than 35%) have been shown to have a negative effect on editing efficiency (Ma et al., 2015). The target sgRNA selected for this study had a GC content of 50% which qualifies as an efficient target for Cas9 activity.

The CRISPR/Cas system relies on the expression of the Cas9 endonuclease and sgRNA target specific guide sequence for successful editing (Huang and Puchta, 2019). Although, transcription of the sgRNA and Cas9 molecules to mRNA was confirmed for all transgenic explants in this study, having all the components together with the genomic DNA site is not guaranteed and hence not all plants and/or all cells in the plant contained. In the present study, the *nptII* construct with the short donor sequence (pDe-Cas-258-*nptII*) successfully induced mutations in sunflower while no mutations were observed for plants transformed with the pDe-Cas9-1185-*bar* construct. This could be because expression level of the transgene varies among independent transgenic shoots (Mikami et al., 2015). Alternative methods for screening transgene accumulation and expression such as RT-qPCR and northern blot could be utilized in future studies to accurately quantify and confirm transgene expression levels (Jiang et al., 2021). *Agrobacterium tumefaciens* transformation and expression efficiency has also been attributed to vector sizes. Another possibility could be that the *bialaphos* resistance gene (*bar*) selection marker inhibited efficient transformation and transgene integration as a similar guide RNA sequence and promoters were used for both plasmids. For this study, the 1185 bp donor template could unfortunately not be cloned into a *nptII* gene plasmid as they had similar restriction sites. The *neomycin phosphotransferase* (*nptII*) selection gene still remains the most widely used selection gene in sunflower transformation (Darqui et al., 2021). There are only a few studies on the use of alternative genes such as the *bar* selection gene and it is therefore difficult to draw conclusions on whether this affects transformation efficiency. It would be beneficial for future studies to test the different selection genes to find one that is suitable for sunflower transformation.

The aim of this study was to introduce three different amino acid changes that have been identified to trigger haploid production upon crossing with wild type plants in *Arabidopsis* to the histone fold domain of the sunflower *CenH3* gene (Britt and Kuppu, 2016; Kuppu et al., 2015) through homology directed repair (HDR). The ability to efficiently insert multiple, single nucleotides of your choice in various parts of the genome will significantly improve crop improvement and enhance different breeding programs such as quantitative trait locus mapping techniques by introducing necessary SNPs (Čermák et al., 2015). To implement the homology directed repair mechanism, we used both a short double cut donor repair template (258 bp) and long donor repair template (1185 bp). The idea of a double cut donor is that the Cas9 enzyme will concurrently cut the target site and the plasmid to produce a linear dsDNA fragment available for homologous repair (Schimdt et al., 2019; Putcha et al., 2016).

The outcomes of double stranded breaks are highly variable and unpredictable, from single base changes to chromosome scale gene alterations (van Vu et al., 2021). The most common changes with CRISPR/Cas9 genome editing are small deletions or additions and substitutions (Danilo et al., 2019). In this study, nucleotide substitutions corresponding to the targeted changes were observed together with additional random substitutions (Figure 2.7). Observation of single nucleotide substitutions upstream or downstream of the target region usually indicate activation of the homology directed repair pathway in DSBs (Odipio et al., 2017). The alleles observed in the four edited plants in this study displayed a combination of the targeted amino acid changes and random base substitutions. Most of the observed sequence changes were according to the design made to the donor sequence, an indication that the donor sequence supplied was used as a reference point during the DSB repair. Targeted amino acids conversions Proline 82 (CCT) to Serine (TCT) and Alanine (GCG) 86 Valine (GTG) were observed in different alleles of all four edited sunflower transgenic plants. This could be an indication that HDR occurred in a limited region and in a few cells which has been reported in CRISPR/Cas9 mediated HDR in tobacco (Hirohata et al., 2019) and rice (Endo et al., 2016) where HDR only occurred in one of the two donor homology arms. In this study, the G83E amino acid was not observed in any of the plants or alleles. This could be that any plants with this mutation did not survive, thus an indication that this amino acid is essential for CENH3 function in sunflower. Three silent mutations around the target site were introduced to the donor sequences, to prevent further additional DSBs on the edited and repaired sites. None of these mutations were observed in the T0 plants analysed, instead amino acid changes were observed in those target positions. A similar finding was observed in *Arabidopsis* where only the targeted amino acid changes were observed (Miki et al., 2018). This phenomenon is still however pending an explanation.

Sunflower transformation is known to yield chimaeric primary transformants (Malone Schoneberg et al., 1994). All the transformants that displayed mutations in this study were chimaeric as can be seen in figure 2.9. Numerous reasons have been put forth to explain this phenomenon including that bisection of the embryonic axis negatively affects meristemic cells and the excessive use of hormones in plant regeneration can result in chimeric plants (Malone Schoneberg et al., 1994). Furthermore, all the four plants contained different combinations of alleles, with two different amino acid changes which were not part of the designed replacements. The amino acid changes Q54R, Q54H and R57H were observed in three different alleles, the designed mutation was a silent mutation for the deactivation of PAM recognition. A combination of targeted and illegitimate recombination events leading to unprecedented changes in the host genome have been observed in several of plant species – even when using long homology donors (Puchta, 1998).

Homology directed repair (HDR) remains a challenge, especially in crop plants due to low frequencies in plant cells as well as insufficient donor template availability (Baltes et al., 2014; Li et al., 2021). HDR is usually successful in 0.1 – 1% of recovered plants (Cermak et al., 2021, Schimidt et al., 2019). Biolistic delivery produced 2.5 – 4% HDR with either separate or single donor delivery of donor while *A. tumefaciens* mediated transformation produced none in maize (Svitashev et al., 2015). In general, protoplast transformation seems to yield higher mutation frequencies compared to *Agrobacterium tumefaciens* -mediated transformation (Li et al., 2013; Sauer et al., 2016), since more DNA donor molecules are integrated in the plant cells (Huang and Puchta, 2019). High efficiency homology directed repair was also observed in rice through particle bombardment (Ma et al., 2015). DNA-virus based replicons were also reported to increase HDR in plant species such as soybean, rice, maize, wheat, and cotton (Gil-Humanes et al., 2017). Numerous methods to enhance HDR in plants have been reported, including the use of viral replicons to increase donor copy number in the vicinity of the DSB (Wang et al., 2017; Baltes et al., 2014) blocking the predominant NHEJ pathway through mutations of certain genes (Huang and Puchta, 2019). The use of *Staphylococcus aureus* Cas9 (SaCa9) increased HDR when compared to spCas9 in *Arabidopsis* (Schiml et al., 2014; Wolter et al., 2018). Unfortunately, for this study the PAM sequence required for SaCa9 activity (NNGRRT) was not present in or around the target site. Base editors are also an attractive alternative to homology directed repair (HDR) with higher efficiencies – they are, however, unable to convert all bases and may have target restrictions or produce bystander mutations (Cermak et al., 2021).The system used in this study provides the ability to efficiently insert multiple, single nucleotides of your choice in various parts of the genome and this has the potential to significantly improve crop improvement and enhance different breeding

programs such as quantitative trait locus mapping techniques by introducing necessary SNPs at selected locations.

This experiment was able to successfully induce DSBs with the Cas9 endonuclease and achieve partial HDR repair in sunflower plants resulting in targeted amino acid changes, the plants unfortunately did not regenerate to mature plants beyond the rooting stage. Apart from *Arabidopsis*, maize, carrot, and wheat, reports on the regeneration of whole plants and haploid induction after editing the *CenH3* gene are few (Wang et al., 2022). In carrot, mutations detected in the early stages of plant development in T0 plants, became undetectable during the flowering stage, even with the CRISPR/Cas9 cassette still present in the plant (Dunemann et al., 2019).

A possible explanation for the low editing efficiency in the present study and the inability for plants to acclimatize may be due to them carrying lethal mutations in the *CenH3* gene (Che et al., 2018). The CENH3 is a crucial protein for chromosome segregation and is therefore crucial for plant survival and propagation (Karimi-Ashtiyani et al., 2021). Initially, only a few amino acid changes were observed to produce haploid inducer lines in plants, targeting the more conserved HFD region of the CENH3 protein that could be applied in various plant species. Further studies have however, determined that large nucleotide deletions in the histone fold domain of the CENH3 protein also result in haploid induction (Kuppu et al., 2020). In-frame CRISPR mediated deletion of the HFD region of the CENH3 protein in tomato including deletion of the complete α-N helix resulted in high efficiency haploid induction upon outcrossing with wild type plants in *Arabidopsis* (Kuppu et al., 2020) and deletion of regions of the HFD in wheat (Lv et al., 2020). A wide range of amino acid changes, and double amino acid change combinations have also been explored to enhance haploid induction (Kuppu et al., 2020). These new findings, including results from this study, further enhance the applicability of the CENH3 haploid inducer method, as the HFD is conserved in a majority of plants and a specific knock out mutation can easily be achieved using the CRISPR/Cas9 without the requirement of targeted changes, the difficult design and implementation of low efficiency HDR in crop species (Kuppu et al., 2020; Lv et al., 2020).

The haploid inducer technology resulting from the modification of the *CenH3* has triggered the development of a simplified genome editing platform targeting lines which are recalcitrant to genetic transformation (Kelliher et al., 2019). Haploid inducer mediated genome editing (IMGE) utilizes a haploid inducer line transformed with a CRISPR/Cas9 construct targeting a desired agronomic trait, to pollinate an inbred line and generate haploid lines with the desired modifications (Kelliher et al., 2019). Homozygous pure, transgene free doubled haploid (DH)

lines can therefore be produced within two generations, eliminating the need for the time consuming and strenuous backcrossing (Kelliher et al., 2019). This method has been applied in *Brassica oleracea* (Li et al., 2021) and maize (Kelliher et al., 2019) and removes the limitation of regeneration and tissue culture for editing of recalcitrant lines (Li et al., 2021).

To date, the CRISPR/Cas9 system has been used for a wide range of gene alterations on numerous plant species, to our knowledge, this is the first report of this technology being applied on sunflower genome editing. Although the CRISPR/Cas9 genome editing system has been established and improved for a wide range of crop species, the dependence on tissue culture, i.e. biolistic delivery/*Agrobacterium tumefaciens* transformation, only limits it to crops amenable to such delivery and regeneration systems (Ali et al., 2020). Further technical innovation is required for the transformation and regeneration of crops such as sunflower.

2.6 REFERENCES

Ali, Z., Shami, A., Sedeek, K., Kamel, R., Alhabsi, A., Tehseen, M., Hassan, N., Butt, H., Kababji, A., Hamdan, S, Mahfouz, M (2020) Fusion of the Cas9 endonuclease and the VirD2 relaxase facilitates homology-directed repair for precise genome engineering in rice. Communications Biology 3, 44. https://doi.org/10.1038/s42003-020-0768-9

Bae S., Park, J., Kim, J (2014) Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30, 1473-1475. https://doi: 10.1093/bioinformatics/btu048

Baltes, N.J., Gil-Humanes, J., Cermak, T., Atkins, P, Voytas, D (2014) DNA replicons for plant genome engineering. The Plant Cell 26, 151-163. https://doi.org/10.1105/tpc.113.119792

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V, Nikolenko, S, Pham, S., Prjibelski, A, Pyshkin, A., Sirotkin, A., Vyahhi, N., Tesler, G., Alekseyev, M., Pevzner, P (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19, 455-477. https://doi.org/10.1089/cmb.2012.0021

Bolger, A. M., Lohse, M., Usadel, B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120. https://doi.org/10.1093/Bioinformatics /btu170

Britt, A., Kuppu, S (2016) Cenh3: An emerging player in haploid induction technology. Frontiers in Plant Science 7, 357. https://doi.org/10.3389/fpls.2016.00357

Čermák, T., Baltes, N, Čegan, R., Zhang, Y., Voytas, D.F (2015) High-frequency, precise modification of the tomato genome. Genome Biology 16, 232. https://doi.org/10.1186/s13059-015-0796-9

Čermák, T (2021) Sequence modification on demand: search and replace tools for precise gene editing in plants. Transgenic Research 30, 353-379. https://doi.org/10.1007/s11248-021-00253-y

Che, P., Anand, A., Wu, E., Sander, J., Simon, M., Zhu, W., Sigmund, A., Zastrow-Hayes, G., Miller, M., Liu, D., Lawit, S., Zhao, Z., Albertsen, M., Jones, T (2018) Developing a flexible, high-efficiency *Agrobacterium*-mediated sorghum transformation system with broad application. Plant Biotechnology Journal 16, 1388-1395. https://doi.org/10.1111/pbi.12879

Chen, K., Wang, Y., Zhang, R., Zhang, H., Gao, C (2019) CRISPR/Cas genome editing and precision plant breeding in agriculture. The Annual Review of Plant Biology 70, 667-697. https://doi.org/10.1146/annurev-arplant-050718-100049

Clement, K., Rees, H., Canver, M.C., Gehrke, J., Farouni, R., Hsu, J.Y., Cole, M., Liu, D., Joung, J., Bauer, D., Pinello, L (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nature Biotechnology 37, 224-226. https://doi.org/10.1038/s41587-019-0032-3

DAFF (2019) A profile of the South African sunflower market value chain. Department of Agriculture, Forestry and Fisheries 1, 28.

Danilo, B., Perrot, L., Mara, K., Botton, E., Nogué, F., Mazier, M (2019) Efficient and transgene-free gene targeting using *Agrobacterium*-mediated delivery of the CRISPR/Cas9 system in tomato. Plant Cell Reports 38, 459-462. https://doi.org/10.1007/s00299-019-02373-6

Darqui, F., Radonic, L., Beracochea, V., Hopp, H., López Bilbao, M (2021) Peculiarities of the transformation of *Asteraceae* family species: The cases of sunflower and lettuce. Frontiers in Plant Science 12, 767459. https://doi.org/10.3389/fpls.2021.767459

Davey, M., Jan, M (2010) Sunflower (*Helianthus annuus* L.): Genetic improvement using conventional and *in vitro* technologies. Journal of Crop Improvement 24, 349-391. https://doi.org/10.1080/15427528.2010.500874

Dawe, R. (2020) Charting the path to fully synthetic plant chromosomes. Experimental Cell Research 390, 111951. https://doi.org/10.1016/j.yexcr.2020.111951

Dunemann, F., Unkel, K., Sprink, T (2019) Using CRISPR/Cas9 to produce haploid inducers of carrot through targeted mutations of centromeric histone H3 (CENH3). Acta Horticulturae 1264, 211-219. https://doi.org/10.17660/ActaHortic.2019.1264.26

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32, 1792-1797. https://doi.org/10.1093/nar/gkh340

Endo, M., Mikami, M., Toki, S (2016) Biallelic gene targeting in rice. Plant Physiology 170, 667-677. https://doi.org/10.1104/pp.15.01663

Fauser, F., Schiml, S., Puchta, H (2014) Both CRISPR/Cas-based nucleases and nickases

can be used efficiently for genome engineering in *Arabidopsis thaliana*. The Plant Journal 79, 348-359. https://doi.org/10.1111/tpj.12554

Friedt, W (1992) Present state and future prospects of biotechnology in sunflower breeding. Field Crops Research 30, 425-442. https://doi.org/10.1016/0378-4290(92)90009-X

Gao, W., Long, L., Tian, X., Xu, F., Liu, J., Singh, P., Botella, J., Song, C (2017) Genome editing in cotton with the CRISPR/Cas9 system. Frontiers in Plant Science 8, 1-12. https://doi.org/10.3389/fpls.2017.01364

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C., Sánchez-León, S., Baltes, N.J., Starker, C., Barro, F., Gao, C., Voytas, D. (2017) High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. The Plant Journal 89, 1251-1262. https://doi.org/10.1111/tpj.13446

Hirohata, A., Sato, I., Kaino, K., Iwata, Y., Koizumi, N., Mishiba, K. Ichiro (2019) CRISPR/Cas9-mediated homologous recombination in tobacco. Plant Cell Reports 38, 463-473. https://doi.org/10.1007/s00299-018-2320-7

Huang, T., Puchta, H (2019) CRISPR/Cas-mediated gene targeting in plants: finally, a turn for the better for homologous recombination. Plant Cell Reports 38, 443-453. https://doi.org/10.1007/s00299-019-02379-0

Jiang, J., Birchler, J.A., Parrott, W.A., Dawe, R.K (2003) A molecular view of plant centromeres. Trends in Plant Science 8, 570-575. https://doi.org/10.1016/j.tplants. 2003.10.011

Jiang, Z., Zhao, Q., Bai, R., Yu, R., Diao, P., Yan, T., Duan, H., Ma, X., Zhou, Z., Fan, Y., Wuriyanghan, H (2021) Host sunflower-induced silencing of parasitism-related genes confers resistance to invading Orobanche cumana. Plant Physiology, 185,424-440. https://doi.org/10.1093/plphys/kiaa018

Karimi-Ashtiyani, R., Ishii, T., Niessen, M., Stein, N., Heckmann, S., Gurushidze, M., Banaei-Moghaddam, A.M., Fuchs, J., Schubert, V., Koch, K., Weiss, O., Demidov, D., Schmidt, K., Kumlehn, J., Houben, A (2015) Point mutation impairs centromeric CENH3 loading and induces haploid plants. Proceedings of the National Academy of Sciences of the United States of America 112, 11211-6. https://doi.org/10.1073/pnas.1504333112

Karimi-Ashtiyani, R (2021) Centromere engineering as an emerging tool for haploid plant production: advances and challenges. Doubled Haploid Technology: Volume 3: Emerging Tools, Cucurbits, Trees, Other Species. 2021:3-22. 3, 22. https:// https://dx.doi.org/10.1007/978-1-0716-1331-3_1

Kelliher, T., Starr, D., Richbourg, L., Chintamanani, S., Delzer, B., Nuccio, M.L., Green, J., Chen, Z., McCuiston, J., Wang, W., Liebler, T., Bullock, P., Martin, B (2017)

MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature 542, 105-109. https://doi.org/10.1038/nature20827

Kelliher, T., Starr, D., Su, X., Tang, G., Chen, Z., Carter, J., Wittich, P.E., Dong, S., Green, J., Burch, E., McCuiston, J., Gu, W., Sun, Y., Strebe, T., Roberts, J., Bate, N.J., Que, Q (2019) One-step genome editing of elite crop germplasm during haploid induction. Nature Biotechnology 37, 287-292. https://doi.org/10.1038/s41587-019-0038-x

Kuppu, S., Ron, M., Marimuthu, M.P.A., Li, G., Huddleson, A., Siddeek, M.H., Terry, J., Buchner, R., Shabek, N., Comai, L., 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, 2068-2080. 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. PLoS Genetics 11, e1005494. https://doi.org/10.1371/journal.pgen.1005494

Li, C., Sang, S., Sun, M.D., Yang, J., Shi, Y.Q., Hu, X.Z., Li, Y., Hao, M.Y., Chu, W., Zhang, H.Y., Wang, H., Wang, W.X., Fu, L., Zhang, B.H., Liu, J., Cheng, H.T., Mei, D.S., Fu, S.H., Hu, Q (2021) Direct modification of multiple gene homoeologs in *Brassica oleracea* and *Brassica napus* using doubled haploid inducer-mediated genome-editing system. Plant Biotechnology Journal 19, 1889-1891. https://doi.org/10.1111/pbi.13632

Li, W., Teng, F., Li, T., Zhou, Q (2013) Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. Nature Biotechnology 31, 684-686. https://doi.org/10.1038/nbt.2652

Lv, J., Yu, K., Wei, J., Gui, H., Liu, C., Liang, D., Wang, Y., Zhou, H., Carlin, R., Rich, R., Lu, T., Que, Q., Wang, W.C., Zhang, X., Kelliher, T (2020) Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. Nature Biotechnology 38, 1397-1401. https://doi.org/10.1038/s41587-020-0728-4

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R (2015) A Robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Molecular Plant 8, 1274-1284. https://doi.org/10.1016/j.molp.2015.04.007

Meena, H., Sujatha, M., Varaprasad, S (2013) Achievements and bottlenecks of heterosis breeding of sunflower (*Helianthus annuus* L.) in India. Indian Journal of Genetics and Plant Breeding 73: 123-130. https:// 10.5958/j.0975-6906.73.2.019

Mikami, M., Toki, S., Endo, M (2015). Parameters affecting frequency of CRISPR/Cas9 mediated targeted mutagenesis in rice. Plant Cell Reports 34, 1807-1815. https://doi.org/10.1007/s00299-015-1826-5

Miki, D., Zhang, W., Zeng, W., Feng, Z., Zhu, J.K (2018) CRISPR/Cas9-mediated gene targeting in *Arabidopsis* using sequential transformation. Nature Communications 9, 1967. https://doi.org/10.1038/s41467-018-04416-0

Murashige, T.; Skoog, F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiologia Plantarum 15, 473–497.doi:10.1111/j.1399-3054.1962.tb08052.xS2CID84645704

Nagaki, K., Tanaka, K., Yamaji, N., Kobayashi, H., Murata, M (2015). Sunflower centromeres consist of a centromere-specific LINE and a chromosome-specific tandem repeat. Frontiers in Plant Science 6, 912. https://doi.org/10.3389/fpls.2015.00912

Odipio, J., Alicai, T., Ingelbrecht, I., Nusinow, D.A., Bart, R., Taylor, N.J (2017) Efficient CRISPR/cas9 genome editing of *phytoene desaturase* in cassava. Frontiers in Plant Science 8, 1780. https://doi.org/10.3389/fpls.2017.01780

Pilorgé, E (2020) Sunflower in the global vegetable oil system: situation, specificities and perspectives. Oilseeds and fats, Crops and Lipids 27, 37. https://doi.org/10.1051/ocl/2020028

Puchta, H (1998) Repair of genomic double-strand breaks in somatic plant cells by one-sided invasion of homologous sequences. The Plant Journal 13, 331-339. https://doi.org/10.1046/j.1365-313X.1998.00035.x

Radonic, L.M., López, N.E., Hopp, E., Bilbao, M.L (2012) Analysis of T2 sunflower transgenic plants: high expression level and stability achieved by the rbcS1 promoter regulation. In Proceedings of the 18th international sunflower conference, Mar del Plata.

Radonic, L.M., Lewi, D.M., López, N.E., Hopp, H.E., Escandón, A.S (2015) Sunflower (*Helianthus annuus* L.). Methods in Molecular Biology. https://doi.org/10.1007/978-1-4939-1658-0

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

Ren, J., Wu, P., Trampe, B., Tian, X., Lubberstedt, T., Chen, S (2017) Novel technologies in doubled haploid line development Plant Biotechnology Journal 15, 1361-1370. https://doi.org/10.1111/pbi.12805

Rönspies, M.; Schindele, P.; Wetzel, R.; Puchta, H (2022) CRISPR–Cas9-mediated chromosome engineering in *Arabidopsis thaliana*. In Nature Protocols. https://doi:10.1038/s41596-022-00686-7

Sauer, N. J., Narváez-Vásquez, J., Mozoruk, J., Miller, R. B., Warburg, Z. J., Woodward, M. J (2016) Oligonucleotide-Mediated genome editing provides precision and function to engineered nucleases and antibiotics in plants. Plant Physiology 170, 1917-1928. http://doi: 10.1104/pp.15.01696

Schiml, S., Fauser, F., Puchta, H (2017) CRISPR/Cas-mediated *in planta* gene targeting. Methods in Molecular Biology 1610, 3-11. https://doi.org/10.1007/978-1-4939-7003-2_1

Schiml, S., Fauser, F., Puchta, H (2014) The CRISPR/Cas system can be used as nuclease for *in planta* gene targeting and as paired nickases for directed mutagenesis in

Arabidopsis resulting in heritable progeny. The Plant Journal 80, 1139-1150. https://doi.org/10.1111/tpj.12704

Schmidt, C., Pacher, M., Puchta, H (2019) DNA break repair in plants and its application for genome engineering. Methods in Molecular Biology 1864, 237-266. https://doi.org/10.1007/978-1-4939-8778-8_17

Sujatha, M., Vijay, S., Vasavi, S., Reddy, P.V., Rao, S.C (2012) *Agrobacterium*-mediated transformation of cotyledons of mature seeds of multiple genotypes of sunflower (*Helianthus annuus* L.). Plant Cell, Tissue and Organ Culture 110, 275-287. https://doi.org/10.1007/s11240-012-0149-2

Svitashev, S., Young, J., Schwartz, C., Gao, H., Falco, C., Cigan, M (2015) Targeted mutagenesis, precise gene editing and site-specfic gene insertion in maize using Cas9 and guide RNA. Plant Physiology 169: 931-945. https://doi.org/10.1104/pp.15.00793

Van Vu, T., Thi Hai Doan, D., Kim, J., Sung, Y.W., Thi Tran, M., Song, Y.J., Das, S., Kim, J.Y (2021) CRISPR/Cas-based precision genome editing via microhomology-mediated end joining. Plant Biotechnology Journal 19, 230-239. https://doi.org/10.1111/pbi.13490

Wang, G., Zong, M., Liu, D., Wu, Y., Tian, S., Han, S., Guo, N., Duan, M., Miao, L., Liu, F (2022) Efficient generation of targeted point mutations in the *Brassica oleracea* var. *botrytis* genome via a modified CRISPR/Cas9 system. Horticultural Plant Journal 8, 527-530. https://doi.org/10.1016/j.hpj.2022.01.005

Wang, M., Lu, Y., Botella, J.R., Mao, Y., Hua, K., Zhu, J (2017) Gene targeting by homology-directed repair in rice using a Geminivirus-based CRISPR/Cas9 system. Molecular Plant 10, 1007–1010. https://doi.org/10.1016/j.molp.2017.03.002

Wang, S., Jin, W., Wang, K (2019) Centromere histone H3 and phospholipase-mediated haploid induction in plants. Plant Methods 15, 42. https://doi.org/10.1186/s13007-019-0429-5

Watts, A., Singh, S.K., Bhadouria, J., Naresh, V., Bishoyi, A.K., Geetha, K.A., Chamola, R., Pattanayak, D., Bhat, S.R (2017) *Brassica juncea* lines with substituted chimeric GFP-CENH3 give haploid and aneuploid progenies on crossing with other lines. Frontiers in Plant Science 7. https://doi.org/10.3389/fpls.2016.02019

Weber, S., Friedt, W., Landes, N., Molinier, J., Himber, C., Rousselin, P., Hahne, G., Horn, R (2003) Improved *Agrobacterium*-mediated transformation of sunflower (*Helianthus annuus* L.): Assessment of macerating enzymes and sonication. Plant Cell Reports 21, 475-482. https://doi.org/10.1007/s00299-002-0548-7

Wolter, F., Klemm, J., Puchta, H (2018) Efficient in planta gene targeting in *Arabidopsis* using egg cell-specific expression of the Cas9 nuclease of *Staphylococcus aureus*. The Plant Journal 94, 735-746. https://doi.org/10.1111/tpj.13893

Xing, H.-L., Dong, L., Wang, Z.-P., Zhang, H.-Y., Han, C.-Y., Liu, B., Wang, X.-C., Chen, Q.-

J (2014) A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biology 14, 327. https://doi.org/10.1186/s12870-014-0327-y

Xu, R., Li, Q.Q (2008) Protocol: Streamline cloning of genes into binary vectors in *Agrobacterium* via the Gateway®TOPO vector system. Plant Methods 4. https://doi.org/10.1186/1746-4811-4-4

Zhang, Yong, Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Yi, Zhang, K., Liu, J., Voytas, D.F., Zheng, X., Gao, C (2013) Rapid and efficient gene modification in rice and *brachypodium* using TALENs. Molecular Plant 6, 1365-1368. https://doi.org/10.1093/mp/sss162

Zhang, J.P., Li, X.L., Li, G.H., Chen, W., Arakaki, C., Botimer, G.D., Baylink, D., Zhang, L., Wen, W., Fu, Y.W., Xu, J., Chun, N., Yuan, W., Cheng, T., Zhang, X.B (2017) Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. Genome Biology 18, 35. https://doi.org/10.1186/s13059-017-1164-8

Zhang, Z., Finer J (2016) Use of cytokinin pulse treatments and micrografting to improve sunflower (*Helianthus annuus* L.) plant recovery from cotyledonary tissues of mature seeds. *In vitro* Cellular and Developmental Biology – Plant. 52: 391-399. https:// DOI 10.1007/s11627-016-9770-9

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.L., Wang, D., Gao, C (2017) Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nature Biotechnology 35, 438-440. https://doi.org/10.1038/nbt.3811

2.7 SUPPLEMENTARY TABLES

Table S2.1 Composition of different medium used in the transformation and tissue culture of sunflower.

Name of Medium	Composition
Luria Bertani Broth (LB Broth)	5 g tryptone, 2.5 g yeast extract, 5 g Nacl, 500 mL ddH_2O
Luria Bertani Agar (LB Agar)	5 g tryptone, 2.5 g yeast extract, 5 g Nacl, 7.5 g Agar, 500 mL ddH_2O
Yeast mannitol (YM)	0.4 g yeast extract, 10.0 g mannitol, 0.1 g NaCl, 0.2 g $MgSO_4.7H_2O$ and 0.5 g $K_2HPO_4.3H_2O$ in 1 litre of medium (pH 7.0) (Radonic et al., 2015)
AIM medium (infection medium)	10.25 g/L K_2HPO_4, 2.25 g/L KH_2PO_4, 1 g/L $(NH_4)_2SO_4$, 1 mL/L $MgSO_4.7H_2O$, 2 g/L glucose, 5 ml/L glycerol (PH 6.5 -7)
Half Murashige and Skoog medium	2.15 g/L Murashige and Skoog (MS) salts, 10 g/L sucrose (pH 5.7) (Musharige and Skoog, 1962; Radonic et al., 2015)
Co-Culture (CC) medium	4.3 g/L MS salt medium with vitamins 3.1 g/L KNO_3; 200 mg/L Myoinositol; 0.5 g/L casein; 30 g/L sucrose, 5 g/L glucose; 5 mL/L 2-(n-morpholino) ethanesulfonic acid (MES) 3.2g/400 mL Agar (pH 5.7) (Radonic et al., 2015)
Regeneration/selection (RE) medium	4.3 g/L Murashige and Skoog salt medium with vitamins (2 mg/L glycine; 3.1 g/L KNO_3; 300 mg/L myoinositol; 30 g/L sucrose (pH 5.7) (Radonic et al., 2015)
Rooting medium (RA)	2.15 g/L MS salts with vitamins, 10 g/L sucrose, 4 g/400 mL Agar (5.7) (Sujatha et al., 2012)

2.8 SUPPLEMENTARY FIGURES

ACAAACAAATTCAAAATTATAACACACGGTGCACTTGCAGTCGACCGTTATCACCCCAAA
ACTCCCGGAAGAAGAAGATTCAACAATGGCGAGAACCAAACACCCTGCTAAACGCAGTTC
AGGCATACCAGCAGACGGTGAATATCTTCCTCTCTTGCTCACCCATCCGATTATCATTAT
TATTAGTTTTTTTTTTCTTCAAATTTAGTTCAAAACCCTAATTTTCCCCCTTTTTTTAAA
CTTGCAGGTAGATCATCCACTTCAACAAACACAGTATACTATTTTTATCACTCTGTTGTT
TATTATTAATTTGTTTATATTTTTTTATATAATTCGTTATTTTTTTTTCTGAATTTGGCA
GCCAAGAAAGAGTCCGAGGAAGAATCGAGGTAGTGAGTGTTTAGTGTTAAAATGTGTGAA
TGTGTGTGTTTTTGGTTTTGATGAATTGGAGTGATTTGTGATTCAGTAGGTGGAGAAAAC
AGGAAGCCGCATAGGTTTAAGCCTGGGACACAGGCGCTACGTGAGATTAGGCGTTTGCAG
AAGACGGTTGAACTGATCATTCCGGCTGCTCCGTTTATTCGAACTGTGAGATGCCTTTAT
CTGCCTGGTTTAGTTATGTTGGCTTTTTGTTATTAATGTTTAAGTGGTCAAACAAAACAC
TGCCATTTTACTAATGACACGAGATGAAACAACTAGGTAGATCCTCTAGTCAGATGATCA
TCCTCTTCTAGTGTCGTCTTTCCTTCCTTTGGTCAGATCAGATCATCCTCTAGTGCAGCA
TGATACATCGTCAACTGTTTCTAAAGCGACTCACGCTCAATGTACCATTTTTAGACAAGC
CCTATTAACTGCATTTAATCAAGTCTGAATGACGTGTGAGATCAATGCATTAGTTCTAAG
GCATGATTGTTTTTTCTCGAACAAGCTATACACTTGCTCATAGTTGTTAATAGCGATGAT
AGCGAGCGCTATAGCGAATACCGTAGCGCTGGATCATTGCTATCTGATGTTTAGCGCTCA
ATAGTGTGAAATAGCGACAATTACGGTTTTATTTTTCTTTTTTTTTAAAAAATCTTAAAA
ACCGAATTTTTAGGCATTTTAATCGACCTAACAGCTGTAATCTTTACTAATAGCATCAGA
AAACCTTAAATTTGTGCATAAAACACTAATTTTTCTTTAAAAAAATTCTAATGCATTATC
GCTAAACATTAAATAGCGCTCGCTATCGCAACGAAGCGTATAGGTAGGTTGTCGCTATTT
ACCGCTATTAACAACTATGCACTTGCTAAACTGCAAAGTATATTCCAATCACTAGGAAAT
AACAAACAAATCTCCAACTAATTGTTGTTTCTTTTTTCTACTTCTTTAACTGGGATTGTT
ATATACACTTTTATTGCATATCCCAGATCTTTTGCAAAAAAAAAAAAAAAAAAACATGCG
ATAAATACATCAGTTTTAGGTAAGCTTTCCTCTTTATTTTTGAACTGTCTAATATAGCAA
GTGCATGATTGGTTGAATTCTATGTTACCTGCTTACACATGTTGACTTTCTAGTAAGAAA
TGTTAGTGGACTTATGAGTATCTAGTTTTGGAAATGTTTTTCCGATTTTCAGCTCAAATG
TTAACATCAGAAATATTTAAGAATTGTCGGGTCTCAACTCTTGAAGTACTACTTGTATAT
AGGGGTGGCAATTCTTGACACGACATGAAAACAATCAAGTTATGCTTTGAAATGTCTAAA
CCGTTTAATAAACATGCTGTTTTGGGTTGACCAGTTTAACCCTTTCATTTTATAAGCTCA
ACCCGTCAACCCGTTTATGATTACAACCCACCCCGTCAAGTCATATGACTTGGTTAGAAT
AATGGGTTAACGGGTCGTGTTCGTGTTACATGATGTTAAGTGTGTCCCGGTTAGAGTCAT
GGTTCTGAGATGAACACATATATGGATCGTGTTCCGGTTCAATTGTTCAACCACCAACCC
AAAACATATCAAATAGTCAACCCGATGTGATTGCCACCCCTACTTGTATAATCATTGATC
AATAATAACTTGGTATACTCTTGTGTTACTGTTACATCTCAGATTAAATACTTTTTTTGC
TCTTTTTTTTGTATTTGTTGATGTTAACTAGAGAATAACATCTTGCAGGTAAAGGAGATA
AGCAACTACATGGCCCCTGAAATCACTCGCTGGCAAGCCGAAGCTCTACAAGCCCTTCAA
GAGGTAAATAAATCATAAAAATTTAGCTTCATACTCGTAAAAATCAACTGTTCCACAAAA

TTAAGGTTTTGTAGCATTCTTAATACTGGAAAAGCATATCTTAGACCGATTTCTGGTGAA
CTTTTAACCAAGTAGTCTCATCATTTTTTTATCAGGCAGCAGAAGATTACCTAATTCAGT
TGTTTGAAGACTCAATGCTATGTGCGATTCATGCAAAGCGCGTTACCCTCAGTAAGTTTC
ACAAAACTCTTCTCAACCCAATTACGTATGAACGAACGGGTCGATTTAGGTTGTGTTTTA
TCTTAAAACGGGTCAAAAAAATAGCTAAAATGGAAACATGTTGTAATTTGCCCAATGGGT
CTACTTCAATGCATACGAACTCCTAGATCCTGATACTTTTTTATATAATAGTTTAAAATT
TGGACAAAAAAGTGTTTACGGGTCAACTCAACCTGAACTCACGTTTTTGTTATTTCTGCA
GTGAAAAAGGATTGGGAGTTGGCACGGCGGATTGGTAAGAAAGGGCAGCCATGGTAGTTA
GAAGAAGCAGAGAATTGTTGTACTCTGTAGTGTGTACAAGTATGCATATGTGTGTGATGA
AAATAGGTACATCATTAATTAGCCTATTGAACTTAACTTGGAATGCTAGTTATGTTATTC
TATTAGTGCTTTATAAATGGTCTGCAAAATTTATAATTGGGTTTTGGTGAAAAATATATT
AGCGCGGTGTATTTTTAGACACGGACCGTCATGCATGATTAATTAACACACAC

Figure S2.1 Sequence of the sunflower CenH3 gene (2993 bp) obtained from HiSeq next generation sequencing and assembled with SPAdes *de novo* assembly. The sequence was confirmed by a nucleotide blast (blast.ncbi.nlm.nih.gov). The CRISPR target site is highlighted in green and the protospacer adjacent motif is in purple. sgRNA (491 - 511), PAM (512-514).

a.258 DRT (Synthesized by GENEWIZ)

TAGGTTTAAGCCTGGGACACAGGTTAATTAATTGGAGTGATTTGTGATTCAGTAGGTGG
AGAAAACAGGAAGCCGCATAGGTTTAAATCTGAGACCCAAGTGCTACGTGAGATTAGA
CGTCTGCAGAAGACGGTTGAACTGATCATTCCGGCTGCTCCGTTTATTCGAACTGTGAG
ATGCCTTTATCTGCCTGGTTTAGTTATGTTGGCTTTTTGTTATTAATGTTTACGGACCGTA
GGTTTAAGCCTGGGACACAGG

b.pMK-QR-DRT (1185) (Synthesized by GeneArt)

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCA
TTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAG
ATAGGGTTGAGTGGCCGCTACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAACTGT
TGGGAAGGGCGTTTCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGG
ATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGT
AAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATTGAAGGAA
GGCCGTCAAGGCCGCATCCCGGGGCATGATTAATTAAATGCACACACACACACACAAA
CAAATTCAAAATTATAACACACGGTGCACTTGCAGTCGACCGTTATCACCCCAAAACTC
CCGGAAGAAGAAGATTCAACAATGGCGAGAACCAAACACCCTGCTAAACGCAGTTCAG
GCATACCAGCAGACGGTGAATATCTTCCTCTCTTGCTCACCCATCCGATTATCATTATTA
TTAGTTTTTTTTTTCTTCAAATTTAGTTCAAAACCCTAATTTTCCCCCTTTTTTTAAACTTG
CAGGTAGATCATCCACTTCAACAAACACAGTATACTATTTTTATCACTCTGTTGTTTATTA
TTAATTTGTTTATATTTTTTTATATAATTCGTTATTTTTTTTTCTGAATTTGGCAGCCAAGA
AAGAGTCCGAGGAAGAATCGAGGTAGTGAGTGTTTAGTGTTAAAATGTGTGAATGTGTG

TGTTTTTGGTTTTGATGAATTGGAGTGATTTGTGATTCAGTAGGTGGAGAAAACAGGAA
GCCGCATAGGTTTAAATCTGAGACCCAAGTGCTACGTGAGATTAGACGTCTGCAGAAG
ACGGTTGAACTGATCATTCCGGCTGCTCCGTTTATTCGAACTGTGAGATGCCTTTATCT
GCCTGGTTTAGTTATGTTGGCTTTTTGTTATTAATGTTTAAGTGGTCAAACAAAACACTG
CCATTTTACTAATGACACGAGATGAAACAACTAGGTAGATCCTCTAGTCAGATGATCATC
CTCTTCTAGTGTCGTCTTTCCTTCCTTTGGTCAGATCAGATCATCCTCTAGTGCAGCATG
ATACATCGTCAACTGTTTCTAAAGCGACTCACGCTCAATGTACCATTTTTAGACAAGCCC
TATTAACTGCATTTAATCAAGTCTGAATGACGTGTGAGATCAATGCATTAGTTCTAAGGC
ATGATTGTTTTTTCTCGAACAAGCTATACACTTGCTCATAGTTGTTAATAGCGATGATAG
CGAGCGCTATAGCGAATACCGTAGCGCTGGATCATTGCTATCTGATGTTTAGCGCTCAA
TAGTGTGAAATAGCGACAATTACGGTTTTATTTTTCTTTTTTTTTAAAAAATCTTAAAAACC
GAATTTTTAGGCATTTTAATCGACCTAACAGCTGTAATCTTTACTAATAGCATCAGAAAA
ACTAGTCTGGGCCTCATGGGCCTTCCTTTCACTGCCCGCTTTCCAGTCGGGAAACCTG
TCGTGCCAGCTGCATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTCTCCGC
TTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGGTAAAGCCTGGGGTGCCTAAT
GAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTT
CCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGG
CGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC
GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGG
AAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTC
GCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTAT
CCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGC
AGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
AAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCT
GAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATC
TCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCAC
GTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATT
AAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTATTA
GAAAAATTCATCCAGCAGACGATAAAACGCAATACGCTGGCTATCCGGTGCCGCAATG
CCATACAGCACCAGAAAACGATCCGCCCATTCGCCGCCCAGTTCTTCCGCAATATCAC
GGGTGGCCAGCGCAATATCCTGATAACGATCCGCCACGCCCAGACGGCCGCAATCAA
TAAAGCCGCTAAAACGGCCATTTTCCACCATAATGTTCGGCAGGCACGCATCACCATG
GGTCACCACCAGATCTTCGCCATCCGGCATGCTCGCTTTCAGACGCGCAAACAGCTCT
GCCGGTGCCAGGCCCTGATGTTCTTCATCCAGATCATCCTGATCCACCAGGCCCGCTT
CCATACGGGTACGCGCACGTTCAATACGATGTTTCGCCTGATGATCAAACGGACAGGT
CGCCGGGTCCAGGGTATGCAGACGACGCATGGCATCCGCCATAATGCTCACTTTTTCT

GCCGGCGCCAGATGGCTAGACAGCAGATCCTGACCCGGCACTTCGCCCAGCAGCAGC
CAATCACGGCCCGCTTCGGTCACCACATCCAGCACCGCCGCACACGGAACACCGGTG
GTGGCCAGCCAGCTCAGACGCGCCGCTTCATCCTGCAGCTCGTTCAGCGCACCGCTC
AGATCGGTTTTCACAAACAGCACCGGACGACCCTGCGCGCTCAGACGAAACACCGCC
GCATCAGAGCAGCCAATGGTCTGCTGCGCCCAATCATAGCCAAACAGACGTTCCACCC
ACGCTGCCGGGCTACCCGCATGCAGGCCATCCTGTTCAATCATACTCTTCCTTTTTCAA
TATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

Figure S2.2 Nucleotide sequences of the donor repair templates designed for amino acid changes in the sunflower CENH3. The 258 bp donor (a) was synthesized by GENEWIZ while the (b) 1185 bp (grey) donor with XmaI and SpeI sites was synthesized by GeneArt and cloned into the pMK-QR plasmid. The following colour codes were used for the sequences: sgRNA (1-20), PAM (21-23), Silent mutation(86), Amino acid change (87), Homology arm (24 - 235), XmaI (369 -375), SpeI, pMK-QR vector backbone.

TCCCAGGATTAGAATGATTAGGCATCGAACCTTCAAGAATTTGATTGAATAAAACATCTT
CATTCTTAAGATATGAAGATAATCTTCAAAAGGCCCCTGGGAATCTGAAAGAAGAGAAG
CAGGCCCATTTATATGGGAAAGAACAATAGTATTTCTTATATAGGCCCATTTAAGTTGAA
AACAATCTTCAAAAGTCCCACATCGCTTAGATAAGAAAACGAAGCTGAGTTTATATACAG
CTAGAGTCGAAGTAGTGATTGTAGGTTTAAGCCTGGGACACGTTTTAGAGCTAGAAATA
GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
TTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTAACCCAGCTTTCTTGTACAAA
GTTGGCATTATAAAAAATAATTGCTCATCAATTTGTTGCAACGAACAGGTCACTATCAGT
CAAAATAAAATCATTATTTGATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATA
TGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAAT
AGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTG

Figure S2.3 Partial sequence (631 bp) of the M13 sequence (red, 1 -256) of the pUC vector, sgRNA (green) sgRNA (257 -280) scaffold (280-356, yellow) and *Arabidopsis* U6-26 promoter (blue, 356 -631) by Sanger sequencing to confirm successful cloning of the sgRNA.

TCCCAGGATTAGAATGATTAGGCATCGAACCTTCAAGAATTTGATTGAATAAAACATCTT
CATTCTTAAGATATGAAGATAATCTTCAAAAGGCCCCTGGGAATCTGAAAGAAGAGAAG
CAGGCCCATTTATATGGGAAAGAACAATAGTATTTCTTATATAGGCCCATTTAAGTTGAA
AACAATCTTCAAAAGTCCCACATCGCTTAGATAAGAAAACGAAGCTGAGTTTATATACAG
CTAGAGTCGAAGTAGTGATTGTAGGTTTAAGCCTGGGACACGTTTTAGAGCTAGAAATA
GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT

TTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTAACCCAGCTTTCTTGTACAAA
GTGGTTCGATAATTCCGATCCAGCCTAGGCCCGGGCCTGAGGACGCGTCCATGGTTAA
TTAAGACGTCCGGACCGACTAGTGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCT
TCTTCGTCAACATGGTGGAGCACGACACGCTTGTCTACTCCAAAAATATCAAAGATACA
GTCTCAGAAGACCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCT
CCTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAG
GTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCT
GCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAA
GACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAG
GGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCAT
TTCATTTGGAGAGGACACGCTGAAATCACCAGTCTCTCTCTACAAATCTATCTCTCTCTA
TAATATTGTGTAAGTAGTTCCCAGATAAGGGAATTAGGGTTCTTATAGGGTTTCGCTCAG
CTGTTGAGCATATAAGAAACCCTTAGTCG

Figure S2.4 Partial sequence (1093 bp) of *Arabidopsis* U6-26 promoter (red, 1 - 256), gRNA (green) (257-280) and *bar* gene (356 - 1093) (blue) obtained from Sanger sequencing using primers SS43 and SS42 (Table 2.1).

ATTCCGGCTGCTCCGTTTATTCGAACTGTGAGATGCCTTTATCTGCCTGGTTTAGTTATG
TTGGCTTTTTGTTATTAATGTTTAAGTGGTCAAACAAAACACTGCCATTTTACTAATGACA
CGAGATGAAACAACTAGGTAGATCCTCTAGTCAGATGATCATCCTCTTCTAGTGTCGTC
TTTCCTTCCTTTGGTCAGATCAGATCATCCTCTAGTGCAGCATGATACATCGTCAACTGT
TTCTAAAGCGACTCACGCTCAATGTACCATTTTTAGACAAGCCCTATTAACTGCATTTAA
TCAAGTCTGAATGACGTGTGAGATCAATGCATTAGTTCTAAGGCATGATTGTTTTTTCTC
GAACAAGCTATACACTTGCTCATAGTTGTTAATAGCGATGATAGCGAGCGCTATAGCGA
ATACCGTAGCGCTGGATCATTGCTATCTGATGTTTAGCGCTCAATAGTGTGAAATAGCG
ACAATTACGGTTTTATTTTTCTTTTTTTTTAAAAAATCTTAAAAACCGAATTTTTAGGCATT
TTAATCGACCTAACAGCTGTAATCTTTACTAATAGCATCAGAAAAACTAGTGGATCCTCT
AGAGTCGACCTGCAGGCATGCAAGCTTCTTCGTCAACATGGTGGAGCACGACACGCTT
GTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAATTGAGACTTT
TCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACT
TTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAA
GGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCAC
CCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGA
TTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAG
ACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACACGCTGAAATCACCAG
TCTCTCTCTACAAATCTATCTCTCTCTATAATATTGTGTAAGTAGTTCCCAGATAAGGGA
ATTAGGGTTCTTATAGGGTTTCGCTCAGCTGTTGAGCATATAAGAAACCCTTAGTCG

Figure S2.5 Partial sequence (1190 bp) of donor sequence (yellow, 1- 586) and *bar* gene (green 627 - 1190) obtained from Sanger sequencing using primers Donor forward and SS43 (Table 2.1) to confirm successful donor sequence cloning.

ATTCCGGCTGCTCCGTTTATTCGAACTGTGAGATGCCTTTATCTGCCTGGTTTAGTTATG
TTGGCTTTTTGTTATTAATGTTTACGGACCGTAGGTTTAAGCCTGGGACACAGGACTAG
TGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTCAGCTTGCCAACATGGTGGAG
CACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAAAGGGC
TATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAG
CTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCAT
CATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAAAG
ATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTC
AAAGCAAGTGGATTGATGTGATAACATGGTG

Figure S2.6 Partial sequence (501 bp) of donor sequence (green 1-239) and CaMV promoter (yellow, 240 - 501) obtained from Sanger sequencing using primers Donor forward and SS102 (Table 2.1) to confirm successful *nptII* gene cloning.

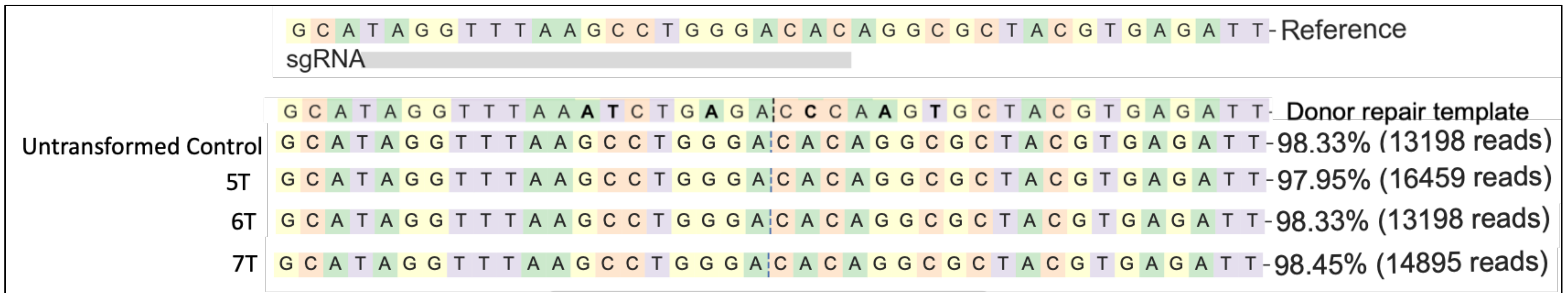
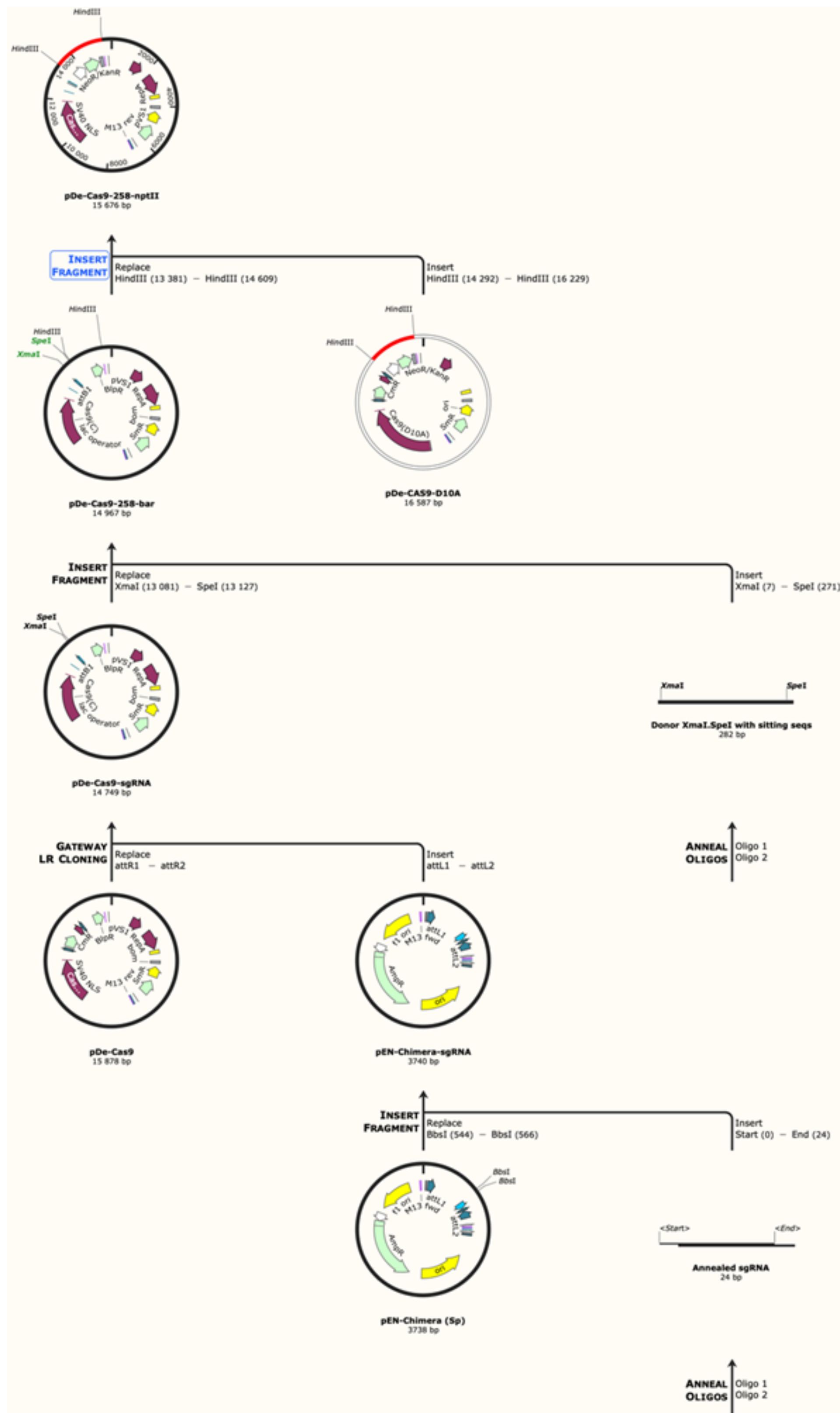


Figure S2.7 Next generation sequencing and CRISPresso2 (Clement et al., 2019) analysis results of sunflower plants: untransformed control, 57-7T plants transformed with the pDe-Cas9-1185-*bar* vector aligned against the sunflower reference sequence and donor repair template. Vertical dotted lines are indicative of Cas9 cutting site (3 bp upstream PAM sequence). None of the targeted mutations or any other mutations were observed.


pDe-Cas9-258-nptII
15 676 bp
Insert Fragment
Replace
HindIII (13 381) – HindIII (14 609)
Insert
HindIII (14 292) – HindIII (16 229)
pDe-Cas9-258-bar
14 967 bp
pDe-CAS9-D10A
16 587 bp
Insert Fragment
Replace
XmaI (13 081) – SpeI (13 127)
Insert
XmaI (7) – SpeI (271)
pDe-Cas9-sgRNA
14 749 bp
Donor XmaI.SpeI with sitting seqs
282 bp
Gateway LR Cloning
Replace
attR1 – attR2
Insert
attL1 – attL2
Anneal Oligos
Oligo 1
Oligo 2
pDe-Cas9
15 878 bp
pEN-Chimera-sgRNA
3740 bp
Insert Fragment
Replace
BbsI (544) – BbsI (566)
Insert
Start (0) – End (24)
pEN-Chimera (Sp)
3738 bp
Annealed sgRNA
24 bp
Anneal Oligos
Oligo 1
Oligo 2

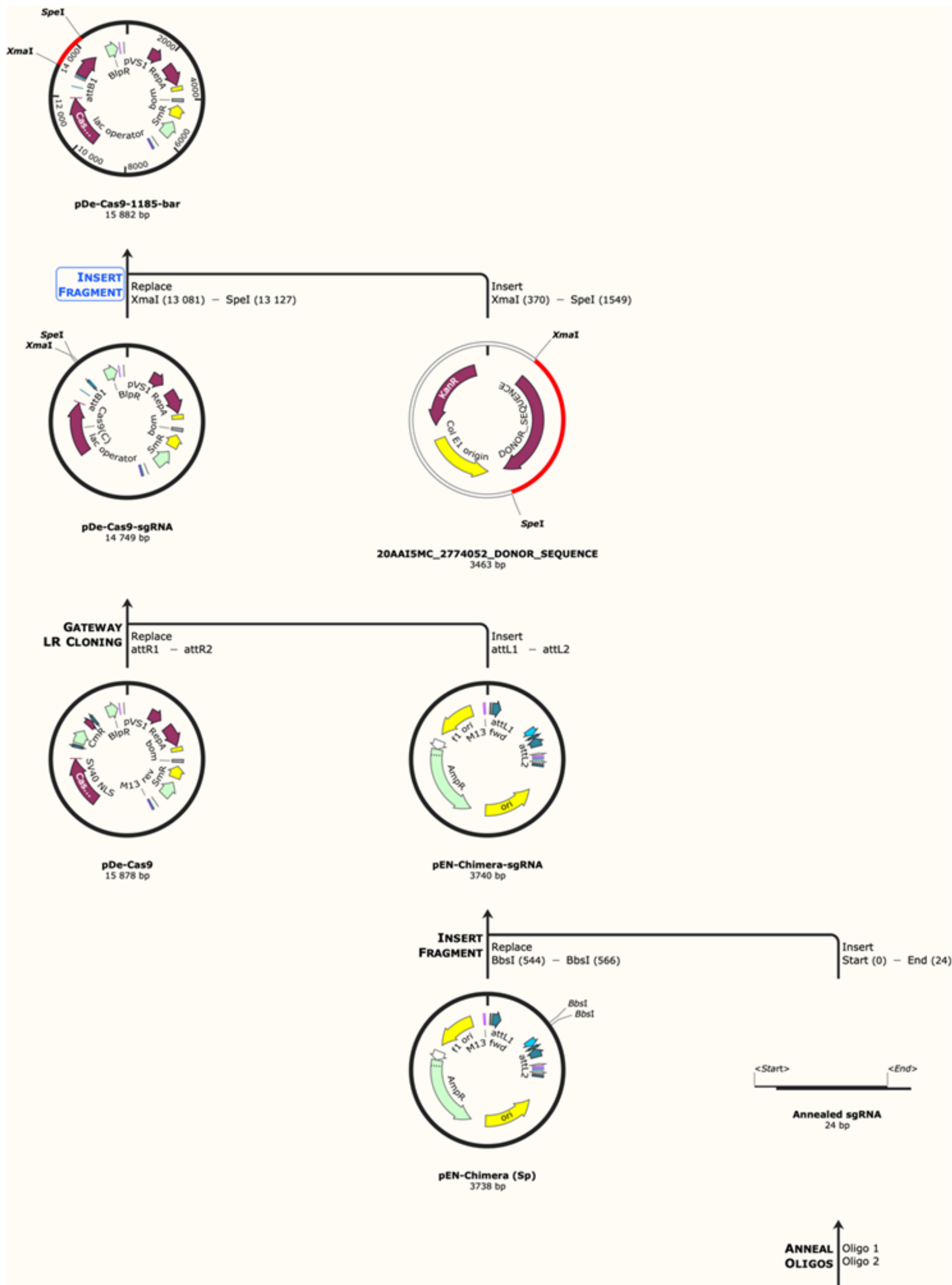


Figure S2.8 Flow diagram of cloning steps for pDe-Cas9-258-*nptII* and pDe-Cas9-1185-*bar* plasmids

CHAPTER 3

TARGETED MODIFICATION OF THE *CENH3* GENE IN SUNFLOWER USING A *TOMATO YELLOW LEAF CURL* (IL-60-BS) VIRAL VECTOR SYSTEM FOR THE DELIVERY OF CRISPR/CAS9 COMPONENTS

3.1 ABSTRACT

Plant viral systems have been utilized for foreign protein expression in a variety of plant species over the years. Recently, several studies have applied the use of viral vectors for delivering gene editing components, specifically those of the CRISPR/Cas system, to edit plant genomes. To test a virus-based delivery system for genome editing in sunflower, plants were mechanically inoculated with a viral vector (IL-60-BS-2A) based on a deactivated tomato yellow leaf curl virus-Israel strain (TYLCV-IL) harbouring the Cas9 endonuclease, guide RNA and a donor repair template. The sunflower *centromere histone 3* (*CenH3*) gene was targeted through homology directed repair (HDR). The CENH3 protein was previously shown to induce haploid progeny in several plant species. PCR amplification of transformed sunflower cDNA confirmed Cas9 transcription in T0, T1 and T2 generations, indicating that the vector is actively expressing components, semi-persistently, and is seed transmissible. Two of the ten inoculated T0 plants contained mutations in the target area: one plant displayed some cells with full homology-directed repair (HDR) converting three amino acids P51S, G52E, and A55V, while the other displayed partial HDR with only two amino acid conversions (P51S and A55V). No mutations were detected in the T1 and T2 generations despite the vector being detectable. This could be due to the transient nature of viral vectors or mutations could possibly have been eliminated by the edited *CenH3* genes during selfing. Previous studies have had viral vector cargo capacity limitations that led to restrictions in systemic movement. This study demonstrates the upper limit of loading capacity of a geminiviral vector to be almost 8kb, while maintaining cell-to-cell movement and gene expression of the CRISPR components leading to successful gene editing in sunflower. The viral vector, harbouring the spCas9, single guide RNA and donor repair template, enhanced HDR when compared to *Agrobacterium tumefaciens* -mediated transformation. This work provides a stepping stone for viral vector employment as a method of choice when delivering genome editing components to plants.

3.2 INTRODUCTION

The CRISPR/Cas genome editing technology has become the method of choice for targeted gene alterations in both plants and animals (Ariga et al., 2020; Chandrasekaran et al., 2021). The CRISPR/Cas technology relies on the expression of the CRISPR-associated (Cas) endonuclease enzyme that is guided to a target site by a recognising RNA, namely the small/single guide RNA (sgRNA) (aka guide RNA, gRNA), prior to double stranded DNA cleavage (Čermák et al., 2015). Cellular DNA repair mechanisms e.g. homology directed repair (HDR) or non-homologous end joining (NHEJ), are activated after the sgRNA-directed Cas9 DNA cleavage, thereby allowing either a template directed (HDR) or random (NHEJ) editing to be incorporated at or near the targeted cut site (Petolino, 2015). Despite the possibilities provided by this technology, it is yet to be routinely used in crop species (Zess and Begemann, 2021). This can be largely attributed to inherent challenges that come with the delivery and expression of recombinant proteins in plants, let alone a large system which includes sgRNA and Cas9 enzyme gene (Ali et al., 2015; Varanda et al., 2021). The dependence on *Agrobacterium tumefaciens* for genetic transformation and delivery of components and tissue culture-based plant workflows and regeneration, including recalcitrance in a number of crops, e.g. sunflower, add a host of limitations (Kumar and Voytas, 2020).

Alternatives to *Agrobacterium tumefaciens* -based transformation include ribonucleoprotein (RNP) technology and particle bombardment (Zess and Begemann, 2021). These methods, however, still have technical barriers, including low efficiencies and the same need for plant *in vitro* culture that limits species or cultivar range (Ariga et al., 2020). Although plant viruses are responsible for significant crop quality and yield losses, given their nature of infection and host interactions, they can be adapted for delivering genetic components for protein production in both the plant-based medicinal and agricultural fields (Liu and Zhang, 2020; Varanda et al., 2021). Furthermore, the replication and genomic structure of viruses make their genomes excellent choices to utilize as expression vectors (Sanjana et al., 2014; Zaidi and Mansoor, 2017). Upon introduction, a plant viral vector will replicate and move systemically within the host which could result in high-level transgene expression (Gleba et al., 2007; Ariga et al., 2021). Viruses have been used before for delivering gene editing systems, including meganucleases, zinc finger nucleases and CRISPR/Cas (Liu and Zhang, 2020).

The initial reports on viral delivery for CRISPR/Cas9 genome editing were on geminivirus based vectors (Ali et al., 2015; Varanda et al., 2021) and since then, numerous studies have

focused on the manipulation of both DNA and RNA viruses for CRISPR/Cas based editing (Varanda et al., 2021). Geminiviral DNA replicons, in particular, have the ability to increase gene targeting frequencies compared to traditional delivery systems such as *Agrobacterium tumefaciens* (Ali et al., 2015; Yin et al., 2015). Furthermore, geminiviruses are DNA based, have a wide host range and some can be transmitted into meristem cells, making them the ideal vectors for genome editing (Ali et al., 2015; Rezaei et al., 2021). Geminiviral replicons (GVRs) have been successfully employed for the delivery of CRISPR/Cas components and donor repair templates in various plant species including tomato (Čermák, et al., 2015; Dahan-Meir et al., 2018), tobacco (Baltes et al., 2014; Yin et al., 2015), wheat (Gil-Humanes et al., 2017) and rice (Wang et al., 2017). However, due to cargo capacity constraints, these geminiviral replicons are usually utilized as “deconstructed” ssDNA viruses and only contain viral elements necessary for replication (Eini et al., 2022). The single stranded genome is converted to a double stranded intermediate by host DNA polymerases (Kujur et al., 2021). As a result, most of the studies on geminiviral replicons could either only deliver the sgRNA molecule and/or required *Agrobacterium tumefaciens* transformation to facilitate systemic infection of the target cells. Recently, two independent studies on two geminiviral replicons, namely the Bean yellow dwarf virus (BeYDV) and Beet curly top virus (BCTV), were employed for the expression of the spCas9, sgRNA, LbCas12a and its corresponding crRNA for successful genome editing and homology directed repair in *Nicotiana benthamiana* while retaining systemic movement after agroinfiltration (Rezaei et al., 2021; Eini et al., 2022).

Peretz et al. (2007) developed a universal gene expression and silencing delivery system with systemic movement abilities for usage in plants. This was developed by deactivating and altering the tomato yellow leaf curl virus, Israeli strain (TYLCV-IL). Tomato yellow leaf curl virus is a whitefly transmitted, DNA monopartite geminivirus, belonging to the begomovirus genus (Cui et al., 2004). It has a single stranded, circular DNA genome of approximately 2.8kb (Yang et al., 2017) that consists of six overlapping open reading frames (ORFs) that are bidirectionally transcribed from an intergenic region (IR) (Peretz et al., 2007). The viral (forward) strand codes for the V1 (capsid) protein and V2 protein, whereas the complementary strand encodes for four genes (C1 to C4), with C1 being the replication protein (Rep) (Czosnek and Laterrot, 1997). The cell-to-cell movement of this virus is governed by the V1 (CP), V2 and C4 proteins (Wang and Zhou, 2016).

The universal vector developed by Peretz et al. (2007) includes a bacterial plasmid for maintaining the construct in *E. coli*, with genetic changes to the C1 (Rep) protein to prevent rolling circle replication (i.e. single strand DNA replication of geminiviruses) and V1 (CP) and V2 genes, preventing viral particles forming to ensure no insect or seed transmission could

take place. Interestingly, the system does allow double stranded multiplication and still allows systematic movement inside the plant (Peretz et al., 2007). Replication was attributed to the viral Intergenic region (IR) and cell-to-cell movement was attributed to the retention of the N-terminal region of the viral coat protein (CP/V1) (Gover et al., 2014, Peretz et al., 2007) after a 20 amino acids deletion in the CP/V1. In their work, the delivery system was tested for systemic protein expression on a wide range of monocot, dicot and woody species, including wheat (*Triticum duram*), pepper (*Capsicum annuum*), grapevine (*Vitis vinifera*), citrus, and olive (*Olea europaea*), but not on sunflower (Peretz et al., 2007). Their vector can express large exogenous sequences, such as entire bacterial operons, (Mozes-Koch et al., 2017) across a variety of plant species (Peretz et al., 2007). This vector presents as an ideal candidate for the delivery and expression of genome editing reagents, such as those for the CRISPR/Cas9 system. Viral vectors have also been observed to increase the occurrence of homology directed repair (HDR) events in plant-based genome editing using the CRISPR/Cas9 system (Gentzel et al., 2022).

Given the recent success using a geminivirus for systemic delivery of Cas9 and sgRNA as a single vector that successfully edited *N. benthamiana*, it is worth investigating other viral vectors for potential CRISPR/Cas delivery for genome editing. It was hypothesized that the TYLCV-based vector developed by Peretz et al. (2007) would successfully deliver CRISPR/Cas9 editing reagents into sunflower cells for *Cenh3* gene modification. DNA-based viral genomes allow for direct infection with the plasmid DNA instead of *in vitro* transformation (Ran et al., 2017). The aim of this study was to demonstrate that a TYLCV-based vector can express the spCas9, sgRNA and donor repair template needed for CRISPR homology directed repair genome editing in sunflower. This chapter further explored the heritability of mutations induced by CRISPR/Cas9 in T1 and T2 sunflower generations.

3.3 MATERIALS AND METHODS

3.3.1 VIRAL DELIVERY VECTOR DESIGN AND CONSTRUCTION

The universal viral delivery system of Peretz et al (2007) (IL-60-BS, Figure 3.1A) was reconstructed (IL-60-BS) and used in this study. Briefly, to obtain the IL-60-BS vector, 60 nucleotides (20 amino acids) were deleted from the viral coat protein (V1/CP) of the TYLCV-IL (GenBank accession no. X15656) and the virus inserted into a pBluescript II KS+ at a SacI site within the replication associated protein (Rep protein), also termed C1 protein, is essential for rolling circle replication of geminviruses, thus causing a frameshift that abolished rolling cycle replication (Peretz et al., 2007). In this study, the IL-60-BS was modified by adding F2A (self-

cleaving peptide derived from the Foot and Mouth Disease Virus (FMDV)) with a 14AA spacer (5'TGGACCAGGGTTAGATTCAACATCTCCAGCAAGCTTCAAAAGATCGAAGTTCAAAAGC TGCTTAACAGGAGCAACAATTTTTTGCTTATGTCTAGCCTC-3') and T2A (self-cleaving peptide derived from the Thosea asigna virus (TaV)) with a 14AA spacer (5'CTGGACCAGGGTTCTCTTCAACATCTCCACATGTCAAAAGAGAACCTCTTCCC TC-3') self-cleaving peptide sequences in frame within the C4 at the SacI site. The self-cleaving peptides were added to reduce the construct size by allowing the in-frame addition of components for expression while using the viral promoter and terminator, i.e. a polycistronic expression cassette similar to Luke et al. (2015). The new vector (IL-60-BS-2A, 5942 bp) (Figure 3.1B) was synthesized and confirmed by sequencing at GenScript Biotech (GenScript, Biotech Corporation, New Jersey, USA). The full nucleotide sequence of the resultant IL-60-BS-2A vector is shown in Figure S3.1.

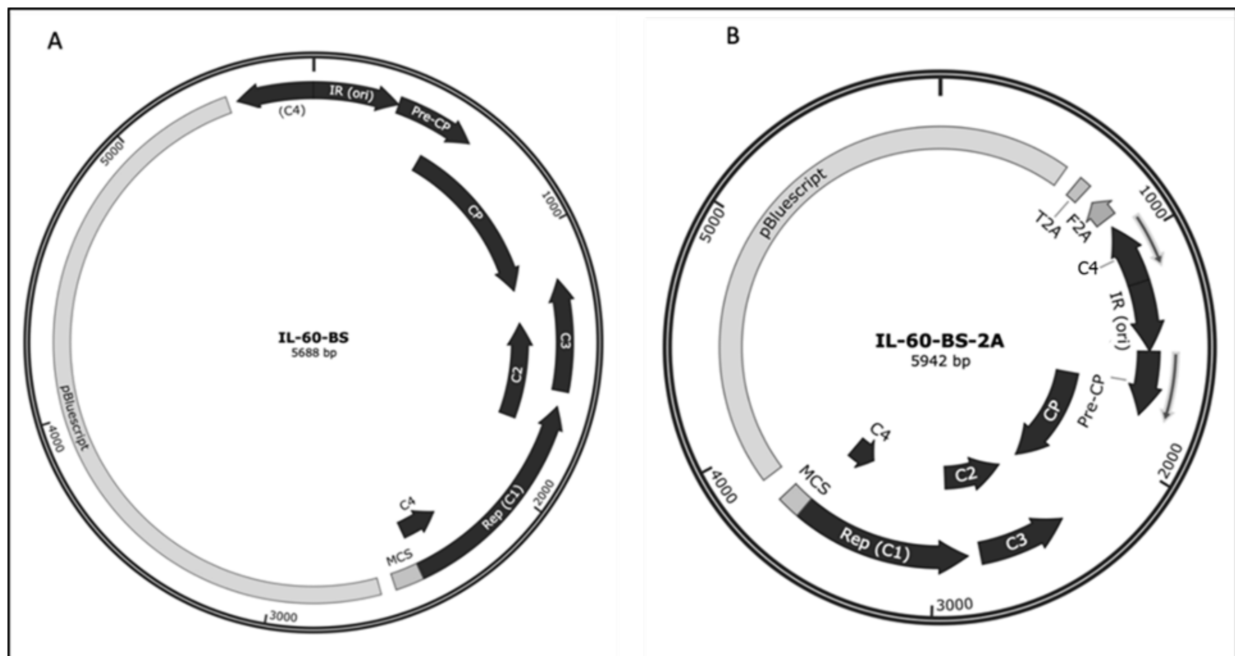


Figure 3.1 Schematic map of the IL-60-BS plant expression vector described by Peretz et al. (2007). (A). The arrows represent viral/plasmid genes and their orientation. The F2A and T2A sites were added to the IL-60-BS in frame with the Rep1 (C1) protein ORF to enable usage of the IR promoter to express the genes of interest (B). The edits in IL-60-BS were also included and the whole construct (IL-60-BS-2A) synthesized, cloned into pBlueScript and sequence verified by Genscript (Genscript Biotech corporation, New Jersey, USA). The plasmid maps were designed on SnapGene software (Insightful Science, www.snapgene.com).

3.3.2 EVALUATION OF REPORTER GENE EXPRESSION USING THE IL-60-BS-2A DELIVERY SYSTEM

The *DsRed* reporter gene was cloned into the IL-60-BS-2A vector to test the vector's ability to deliver and express heterologous genes in sunflower using the In-Fusion®HD cloning kit (Takara Bio, Shiga, Japan). Primers (Table 3.1) were designed using the In-Fusion® primer design platform (www.takarabio.com) to amplify the *DsRed* gene (Primer 1 and 2) from the pKGW-RR-DsRed vector (a gift from Dr Nicky Creux, University of Pretoria, South Africa) and used to linearize (Primer 3 and 4) the IL-60-BS-2A vector. PCR reactions (total volume 25 µL) contained 1X CloneAmp HiFi PCR premix (Takara Bio, Shiga, Japan), 0.2 µM of each primer and 10 ng of plasmid DNA. The PCR amplification protocol included an initial denaturation at 98°C for 10 s; followed by 30 cycles of denaturation at 98°C for 10 s; annealing at 60°C for 5 s; elongation at 72°C for 5 s/kb and a final elongation step at 72°C for 5 min. The amplified PCR products were confirmed on a 1% agarose gel and purified with the NucleoSpin gel and PCR clean up kit (Macherey-Nagel, Germany). The *DsRed* PCR product and the IL-60-BS-2A linearized vector were "fused" between the F2A and T2A sites using the In-Fusion® HD cloning kit according to the manufacturer's guidelines. The reaction was transformed into Stellar™ chemically competent cells (Takara Bio, Shiga, Japan) as per manufacturer's guidelines and plated on ampicillin containing Luria-Bertani (LB) Agar plates incubated at 37°C overnight.

Colony PCR was set up to identify positive clones by amplifying the *DsRed* and F1 origin of replication genes using primers (Primer 2 and 5) listed in Table 3.1. A PCR reaction was set up with 1X Dreamtaq 2X green PCR Master mix (Thermofisher Scientific, Massachusetts, USA), 0.1 µM of each primer and nuclease free water to a total volume of 25 µL. The cycling conditions were 95°C initial denaturation for 3 mins, 30 cycles of denaturation at 95°C for 30s, annealing at 60°C for 30 s, extension at 72°C for 1 min and final extension at 72°C for 5 min. The 1118 bp expected PCR product (Figure 3.2) was visualised on 1.5% agarose gel stained with ethidium bromide using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA).

Plasmids were prepared from positive clones and these were confirmed using Sanger sequencing at Inqaba Biotechnical Industries (Pretoria, South Africa) using the primer DsRed-Fragment Reverse and F1 bacteriophage origin of replication primer (Table 3.1). The final vector was named IL-60-BS-2A-DsRed.

3.3.3 ADDITION OF CRISPR/CAS9 GENE CASSETTE TO THE VIRAL DELIVERY VECTOR

Five amino acids changes in the histone fold domain of the CENH3 protein in *Arabidopsis* were shown to result in haploid progeny upon crossing with wildtype plants (Britt and Kuppu., 2016). The same plasmid pDe-Cas9-1185-*bar* with the donor sequence (1185 bp) designed and synthesised in Chapter 2, section 2.3.2 was used here. Three silent mutations were also added to deactivate the PAM recognition and introduce/eliminate a restriction site needed for screening (Chapter 2, Table 2.2). The designed sgRNA oligonucleotide and DRT sequences were cloned into the pDe-Cas9 vector (a gift from Prof. Holger Puchta's lab, Karlsruher Institut fur technologie, Karlsruhe, Germany) (Fauser et al., 2014; Schiml et al., 2014). The resultant vector pDe-Cas9-1185-*bar*, which was used for *Agrobacterium tumefaciens* mediated transformation in Chapter 2, (section 2.3.2 for fully detailed construction) was used to obtain the whole cassette and which was added to the IL-60-BS-2A delivery vector to deliver CRISPR/Cas9 genome editing components in sunflower. Primers (Table 3.1) were designed to amplify the T-DNA section of the pDe-Cas9-1185-*bar* vector (Primer 6 and 7) excluding the *bialaphos* plant selection gene using the In-Fusion® primer design platform (www.takarabio.com). A PCR reaction to linearize the vector (IL-60-BS-2A) (Primer 3 and 4) and amplify the insert was performed using the CloneAmp HiFi PCR Premix (Clontech, Takara Bio, Shiga, Japan). Each 25 µL reaction contained 1X CloneAmp HiFi PCR premix (Takara Bio, Shiga, Japan), 0.2 µM of each primer and 10 ng of plasmid DNA. The PCR amplification protocol was as follows: initial denaturation at 98°C for 10 s; followed by 30 cycles of denaturation at 98°C for 10s; annealing at 60°C for 5 s; elongation at 72°C for 5 s/kb and a final elongation step at 72°C for 5 min. The PCR product was viewed on 1% agarose gels for size confirmation prior to purification with the NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, Germany). The amplified fragment was transferred into the linearized IL-60-BS-2A vector (as described in 3.3.2.) between the F2A and T2A sites using the In-Fusion®HD cloning kit (Takara Bio, Shiga, Japan) according to the manufacturer's guidelines. The reaction was transformed into Stellar™ *E. coli* chemically competent cells (Takara Bio, Shiga, Japan) as per manufacturer's guidelines and incubated at 37°C overnight.

A colony PCR identified positive colonies using primers targeting the parsley ubiquitin promoter and F1 bacteriophage origin of replication (Primer 5 and 8). PCR reactions containing 1X Dreamtaq 2X green PCR Master mix (Thermofisher Scientific, Massachusetts, USA), 0.1 µM of each primer, with nuclease free water to a total volume of 25 µL, were prepared. The PCR cycling conditions were set up at 95°C initial denaturation for 3 min, 27 cycles of denaturation at 95°C for 30 s, annealing at 68°C for 30 s, extension at 72°C for 1 min and final extension at 72°C for 5 min. PCR products were separated on 1.5% agarose gel

stained with ethidium bromide. The vector was confirmed by Sanger sequencing of the region of the parsley ubiquitin promoter and F1 bacteriophage origin of replication (Table 3.1, Primer 5 and 8) at Inqaba Biotechnical Industries (Pretoria, South Africa). The resulting vector was named the IL-60-BS-2A-Cas9-1185. Successful vector assembly was further confirmed by a restriction digestion with XmaI enzyme (New England Biolabs, Massachusetts, USA) as per manufacturer's guidelines (Figure S3.4).

3.3.4 Plasmid extractions and storage

Plasmid DNA for both the IL-60-BS-2A-DsRed and IL-60-BS-2A-Cas9-1185 constructs were extracted from liquid cultures using the Fermentas GeneJET™ Plasmid mini prep kit (Thermofisher Scientific, Massachusetts, USA). The extracted plasmid DNA was eluted in the supplied Fermentas elution buffer and stored in 50 µL aliquots at -20°C until use.

3.3.5 Sunflower lines and growth conditions

Dwarf lines (KP227, KP328, KP937 and KP1063) were selected from the Agricultural Research Council's Grain Crops' germplasm unit (ARC-GC, Potchefstroom, South Africa) for their fast maturation rates and smaller sizes. Seeds from each line were planted in triplicate in 20cm pots in sterile perlite growth mixture (Chemserve Perlite, South Africa) and grown at 30°C/20°C (±2°C), with 16:8h day/night cycling and humidity set to 60%-80%, in the plant containment and growth facility (BSL-2 Registration no. 39.2/ ARC Biotechnology Platform – 21/0142) of the Agricultural Research Council's Biotechnology Platform (Pretoria, South Africa). The plants were grown until maturity (wilt stage) and the seeds were harvested and stored until use.

3.3.6 Plant inoculation with IL-60-2A and IL-60-2A-DsRed constructs

The four sunflower dwarf lines (section 3.3.5) were used to test for genotype specificity of the IL-60-BS-2A vector. Fourteen (14) day old seedlings were mechanically injected in the leaf veins with approximately 10 µL of the IL-60-BS-2A vector-plasmid DNA (200 ng/100 µL) using standard, sterile 1 mL diabetic syringes (BD Micro-Fine Plus, Becton, Dickinson and Company, NJ, USA) until the viral plasmid solution had fully soaked the inoculated leaf, two leaves were inoculated per plant. The negative control consisted of elution buffer without DNA. Virus movement and persistence were investigated using the Phire Plant Direct PCR kit (Thermofisher Scientific, Massachusetts, USA) to detect the plasmid DNA presence in leaf tissue from 1 to 21 days post inoculation (dpi). Two PCR targets on the vector were selected – one targeting the ampicillin gene (Primer 9 and 10) and the other the coat protein gene

(Primer 11 and 12) – to confirm construct presence in a sample. Approximately 2 mm^2 of leaf material was collected from inoculated and control plants into tubes containing 20 µL of supplied dilution buffer and crushed using a sterile 1 mL pipette tip. The mix was centrifuged for 5 min at 14000 rpm and the supernatant was used as the genomic DNA template. PCR reactions contained 1X Phire Plant PCR buffer, 0.4 µL Phire hotstart II DNA polymerase, 0.5 M of each primer, and 2 µL of the earlier prepared genomic DNA solution, with water up to a final volume of 20 µL. The PCR cycling conditions were: 98°C for 5 min, 30 cycles of denaturation at 98°C for 5 s, annealing at 55°C for 5 s, 72°C for 20 s and final extension at 72°C for 1 min. The amplicons were visualized in 1.5% agarose gel stained with ethidium bromide using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA).

Since sunflower genotype seemed to have no effect on viral movement and persistence (see section 3.4.1), line KP328 was selected for further experiments as this line was used for *Agrobacterium* transformation in Chapter 2.. The reporter gene construct was tested by mechanically inoculating 14-day old sunflower seedlings with approximately 10 µL of the IL-60-BS-2A-DsRed vector (200 ng/100 µL) as described above. Systemic movement of the construct was tested *in planta* by using the youngest leaves that were different from those inoculated, and analysing them for the DsRed signal at 7 dpi. The Zoe Fluorescent imager (BioRad laboratories, California, USA) was used to visualise samples under blue and red light.

3.3.7 Leaf inoculation of sunflower with IL-60-BS-2A-Cas9-1185 vector

For the gene editing experiments, ten test and three control plants of sunflower KP328 line were grown as described in section 3.3.5 Ten 14-day old plants were mechanically inoculated with approximately 10 µL (200 ng/100 µL) of the IL-60-BS-2A-Cas9-1185 vector plasmid DNA as described earlier. Plants were allowed to mature and T0 plants were all self-pollinated to produce T1 plants. T1 plants were crossed with wildtype sunflower (KP 328) lines to produce T2 plants (For haploid induction as described by Britt and Kuppu. (2016). Plants inoculated with plasmid-free TE buffer were used as negative controls.

3.3.8 Confirmation of construct presence in inoculated plants (T0) and progeny (T1 and T2)

Leaf material (approximately 100 mg) was collected from the youngest leaves and snap frozen in liquid nitrogen. DNA was extracted from both inoculated leaves and the youngest leaves of infected T0 plants 7 dpi, and the youngest leaves of the T1 and T2 plants 14 days after germination to detect the presence of the IL-60-BS-2A-Cas9-1185. The Qiagen Plant Mini

DNA Extraction kit (Qiagen, Hilden, Germany) was used for DNA extraction as per manufacturer's guidelines. Successful inoculation was determined by confirming presence of the vector in leaves of the plants using primers flanking the U6-26 promoter and sgRNA reverse (Primer 13 and 14) and the coat protein (Primer 11 and 12) (Table 3.1). PCR reactions containing 1X OneTaq®2X Master mix with standard buffer (New England Biolabs, Ipswich, Massachusetts, USA), 0.2 µM of each primer, 10 ng of template cDNA and nuclease free water to 25 µL. The PCR thermal cycling conditions were: Initial denaturation at 94°C for 30 s, 30 cycles of denaturation at 94°C for 15 s, annealing at 60°C for 30 s (U6-26 promoter/sgRNA reverse) and 55°C for 30 s (coat protein), with the elongation step at 68°C for 30 s, and final extension 68°C or 5 min. PCR products were separated on 1.5% agarose gels stained with ethidium bromide. The agarose gels were visualized using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA).

3.3.9 Transcription analysis of Cas9 and associated sgRNA

To ensure sufficient time for gene editing to take place, approximately 100 mg of fresh leaf material from the youngest leaf (T0, T1 and T2 generations) was collected 21 dpi and snap frozen in liquid nitrogen. Leaf tissue was immediately crushed into powder using a plastic micro-pestle (Sigma Aldrich, Missouri, USA). Total RNA was extracted using the Zymo Quick RNA Plant Miniprep kit (Zymo Research, California, USA) according to manufacturer's guidelines. The RNA quality was confirmed by nanophotometer (IMPLEN GmBH, California, USA) and treated with DNAse (Zymo Research, California, USA).

DNase treated RNA (±500 ng) was converted into cDNA using a first strand cDNA synthesis kit and the random hexamer approach (Thermofisher, Waltham, Massachusetts, USA). The cDNA samples generated were used in PCR reactions to confirm Cas9 and sgRNA transcription by targeting regions of the Cas9 endonuclease (Primer 15 and 16), sgRNA and sgRNA scaffold (Primer 17 and 18, Table 3.1). Each PCR reaction consisted of 12.5 µL OneTaq®2X Master mix with standard buffer, according to the manufacturer's guidelines (New England Biolabs, Ipswich, Massachusetts, USA), 0.2 µM of each primer, 2 µL of template cDNA and 9 µL nuclease free water. The PCR thermal cycling conditions were: Initial denaturation at 94°C for 30 s, 30 cycles of denaturation at 94°C for 15 s, annealing at 61°C for 30 s (Cas9) and 58°C for 30 s (sgRNA), with the elongation step at 68°C for 30 s, and final extension 68°C for 5 min. PCR products were separated on 1.5% (Cas9) and 3% (sgRNA) agarose gels stained with ethidium bromide. The agarose gels were visualized using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA).

3.3.10 CRISPR/CAS9 MUTATION ANALYSES IN SUNFLOWER

Mutations in sunflower seedlings were analysed by extracting genomic DNA with QIAGEN DNAeasy plant mini kit (Qiagen, Hilden, Germany) from putative mutants T0, T1 and T2 leaves 21 dpi. An initial PCR reaction with primers 19 and 20 (Table 3.1) designed to only target the native *CenH3* sequence and avoid targeting the donor repair template (DRT). The PCR reactions were performed with the following composition: 0.02U/µL of Q5 High-fidelity polymerase and 1X Q5 reaction buffer (New England Biolabs, Massachusetts, USA), 200µM dNTPs and 0.5 µM of each primer in a total volume of 25 µL. The PCR amplification protocol consisted of an initial denaturation step at 98°C for 30 s, followed by 35 cycles of denaturation at 98°C for 5 s, annealing at 68 °C for 30 s and elongation at 72°C for 30 s. The final elongation step was 2 min at 72°C. The 1778 bp PCR product was visualised on the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA) after electrophoresis on 1.5% agarose gel with ethidium bromide and cleaned up using the NucleoSpin gel and PCR clean up kit according to the manufacturer's guidelines (Macherey-Nagel, Duren, Germany). This was followed by targeted sequencing of the *CenH3* region using next generation sequencing (NGS).

Primers flanking the CRISPR/Cas *CenH3* target region were designed with adaptor sequences as presented in the 16S metagenomic sequencing library preparation kit protocol of Illumina (Illumina, California, USA) (primer 21 and 22) (Table 3.1). These primers allowed direct indexing and NGS sample preparation using the "standard 16S rRNA Illumina workflow" but for the *CenH3* PCR products (primers listed in Table 3.1), thus reducing NGS preparation costs. Targeting PCR reactions were used to amplify the *CenH3* target region in each T0, T1 and T2 plants using 20 ng extracted genomic DNA, 0.02 U/µL of Q5 High-fidelity polymerase and 1X Q5 reaction buffer (New England Biolabs, Ipswich, MA), 200µM dNTPs and 0.5 µM of each primer in a total volume of 25 µL. The PCR amplification protocol consisted of an initial denaturation step at 98°C for 30 s, followed by 35 cycles of denaturation at 98°C for 5 s, annealing at 72°C for 10 s and elongation at 72°C for 30 s. The final elongation step was 2 min at 72°C. Amplicons were visualised under UV light after electrophoresis in 1.5% agarose gel stained with ethidium bromide and using the Biorad Gel Doc™ XR+ documentation system (Biorad laboratories, California, USA). Expected 370 bp PCR products were purified using the AMPure XP beads (Beckman coulter, California, USA), as per Illumina 16S metagenomic sequencing library preparation guidelines, and indexed using the Nextera®XT index kit V2 (Illumina, California, USA). Library preparation was performed as per the 16S metagenomic sequencing library preparation kit protocol of Illumina and sequencing was performed on an

Illumina MiSeq system using paired-end 300 bp V3 sequencing chemistry at the Agricultural Research Council's Biotechnology Platform (Pretoria, South Africa).

FASTQ files were loaded on to the CRISPResso2 software (Clement et al., 2019, https://crispresso.pinellolab.partners.org/submission) to analyse whether the CRISPR/Cas9 system successfully induced mutations at the target site. The CRISPResso2 software analyses the dataset by first performing quality filtering and trimming adapters of the reads prior to merging paired end reads and aligning these to the reference ("wildtype") sequence. Default CRISPResso2 settings were used for this analysis.

Table 3.1 Primer sequences used in this study (Chapter 3).

No	Primer name	Sequence 5'→3'	Purpose
1	DsRed fragment Forward	CTCCAGATCCAGAAGTCCTACAGGAACAGGTGGTGGCGGC	Cloning DsRed into IL-60-BS-2A
2	DsRed Fragment Reverse	TCTAACCCTGGTCCAATGGCGCGCTCCTCC	Cloning DsRed into IL-60-BS-2A
3	IL-60-BS-2A Forward	CTTCTGGATCTGGAGAGGGAAGAG	Linearization of IL-60-BS-2A for cloning of CRISPR construct
4	IL-60-BS-2A Reverse	TGGACCAGGGTTAGATTCAACATCTCCAG	Linearization of IL-60-BS-2A
5	F1 ori Forward	ATTTAGAGCTTGACGGGGAAAGC	Confirmation of cloning into IL-60-BS-2A.
6	In-Fusion Cas Forward	CTCCAGATCCAGAAGGTGGTGATTTTGTGCCGAGCTG	Cloning of Cas9 into IL-60-BS-2A
7	In-Fusion Cas Reverse	TCTAACCCTGGTCCATGGGTGGGGGTCCATCTTTG	Cloning of Cas9 into IL-60-BS-2A
8	PcUBi Reverse	GTGTTTGAGGCGGTGAAGGA	Amplification of parsley promoter for cloning confirmation.
9	Amp Reverse	ACCCAGAAACGCTGGTGAAA	Amplification of the ampicillin gene used in confirmation of viral movement.
10	Amp Forward	TCCGGTTCCCAACGATCAAG	Amplification of the ampicillin gene for confirmation of viral movement.
11	Coat protein Reverse	TTCTTCACGGTTGCGGTACT	Amplification of the coat protein gene used to confirm viral movement.
12	Coat protein Forward	TCCAGTCTTATGAGCAACGGG	Amplification of the coat protein gene .
13	SS42 (U6-26 Promoter)	TCCCAGGATTAGAATGATTAGG	
14	sgRNA Reverse	AAACGTGTCCCAGGCTTAAACCTA	
15	Cas-RT-Forward	TCAACGTACATATCCCTACCG	For Cas9 Transcript confirmation (Gao et al., 2017).
16	Cas9-RT-Reverse	AGGCTCAAGACTTACGCTCAT	
17	sgRNA-RT-Reverse	CGACTCGGTGCCACTTTTTCAAGTTG	For sgRNA transcript confirmation (Ma et al., 2015).
18	sgRNA Forward	ATTGTAGGTTTAAGCCTGGGACAC	Single guide RNA for CRISPR editing.
19	CENH3LDForward	AACGCAGTTCAGGCATACCA	*CenH3* gene large donor for mutation detection.
20	CENH3LDReverse	AACACGAACACGACCCGTTA	
21	CENH3 Target Forward	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GAGTCCGAGGAAGAATCGAGGTAG	For *CenH3* target for mutation detection. The Illumina 16s barcodes are in red.
22	CENH3 Target Reverse	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AGTAAAATGGCAGTGTTTTGTTTGACC	

3.4 RESULTS

3.4.1 IL-60-BS-2A AND IL-60-BS-2A-DSRED REPLICATION, SYSTEMIC MOVEMENT AND REPORTER GENE EXPRESSION IN SUNFLOWER

Replication and systemic movement of the IL-60-BS-2A vector had to be tested in sunflower prior to developing it into a viral based vector for CRISPR/Cas9 mediated genome editing. Sunflower seedlings from four different ARC lines (KP328, KP227, KP1063 and KP937) were mechanically inoculated with the synthesized vectors. Replication and systemic movement in inoculated seedlings were confirmed by PCR testing for the coat protein and ampicillin genes on the vector in the youngest leaves, adjacent to the inoculated leaves. No visual growth defects or abnormal disease symptoms were observed in inoculated plants (not shown). Virus movement and persistence was determined between 1 dpi to 21 dpi using the youngest leaf of the inoculated plants. The viral vector was only detected in one plant at, 1 dpi (lane 10: KP937), but 4 plants, one from each of the tested genotypes, showed the presence of the vector at 21 dpi (Figure 3.4). This confirms that sunflower genotype had no effect on viral replication and movement, but that sufficient dpi should be permitted for movement to occur. It must be however noted that only one out of three plants for each genotype showed vector presence. The absence of a genotype effect allowed the analysis to proceed with only one genotype (KP328).

Reporter genes are a powerful tool for examining protein expression and localization at a subcellular level. *DsRed* is a more suitable reporter gene than *Gfp* since auto-fluorescence, resulting from plant chlorophyll detected at the same wavelength as *Gfp*, is eliminated. The DsRed gene fragment PCR amplified from the pKGW-RR-DsRed vector was successfully cloned into the PCR linearized IL-60-BS-2A vector using the In-Fusion® cloning kit (Takara, Shiga, Japan). A colony PCR with a 1118 bp product obtained from amplifying the F1 bacteriophage origin of replication and *DsRed* gene confirmed successful cloning (Figure 3.3). Successful cloning was further confirmed through Sanger sequencing (Figure S3.2). The final vector map is displayed in Figure 3.4.

The expression of genes in sunflower using the IL-60-BS-2A vector was tested with the DsRed reporter gene. The *DsRed* gene was cloned into the IL-60-BS-2A vector to produce the IL-60-BS-2A-DsRed vector (Figure 3.4). Three (3) plants were inoculated with the IL-60-BS-2A-DsRed vector and seven (7) days after inoculation, the first leaves above the inoculated leaves were analysed for DsRed expression using the Zoe Fluorescent image (Biorad laboratories, CA, USA). The DsRed signal was observed in all tested plants inoculated with the IL-60-BS-

2A-DsRed vector but not in control (WT) plants under both blue and red light (Figure 3.5). The plant virus infected leaves also showed an altered, slightly smaller stomata when compared to wildtype control plants as seen under the blue light (Figure 3.5).

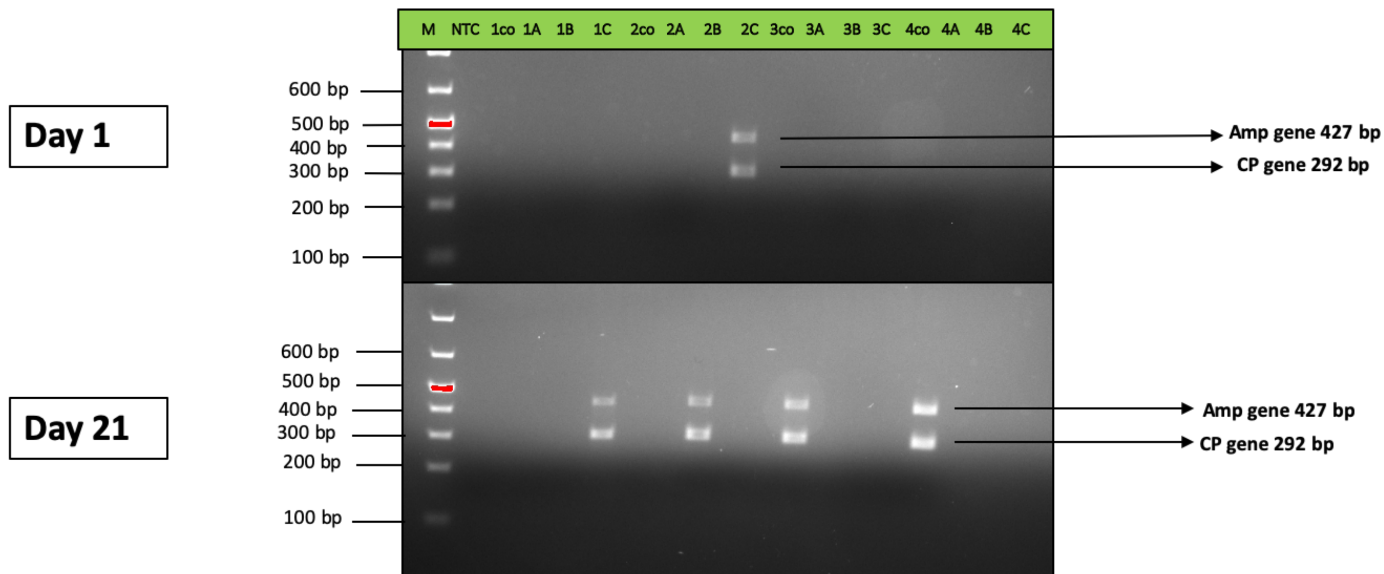


Figure 3.2 Agarose gel image showing PCR results for the coat protein (CP) and ampicillin (*Amp*) genes amplified from young leaves of four different sunflower ARC-inbred lines on 2 different time intervals post inoculation with the IL-60-BS-2A vector. The *Amp* and CP amplicons are 427 bp and 292 bp products respectively. Lane M: 1kb plus molecular marker, Lane NTC: No template control, Lane 1co: Line KP227 negative control, Lanes 1A-1C: Inoculated KP227 lines, Lane 2co: Line KP937 negative control, Lanes 2A-2C: Inoculated KP937 lines, Lane 3co: Line KP328 negative control, Lanes 3A-3C: Inoculated KP328 lines, Lane 4co: Line KP1063 negative control and Lanes 4A-4C: Inoculated KP1063 lines. On day 1, the vector was only detected in lane 2C (KP937) and on day 21, the vector was detected in lane 1C (KP227 inoculated) lane 2B (KP927 inoculated), lane 3A (KP 328 inoculated) and lane 4A (KP1063 inoculated) and not detectable in any of the non- inoculated control in all the tested days.

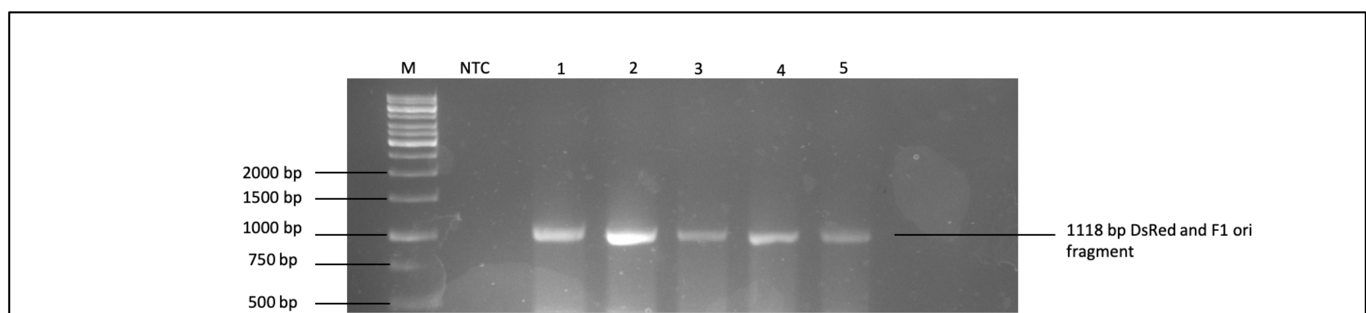


Figure 3.3 Colony PCR confirming the cloning of the *DsRed* gene into the IL-60-BS-2A vector. M - GeneRuler 1kb marker, NTC - none template control, 1-5 randomly selected colonies. All tested 5 colonies were positive for the 1118 amplicon of the F1 bacteriophage origin of replication and *DsRed* gene.

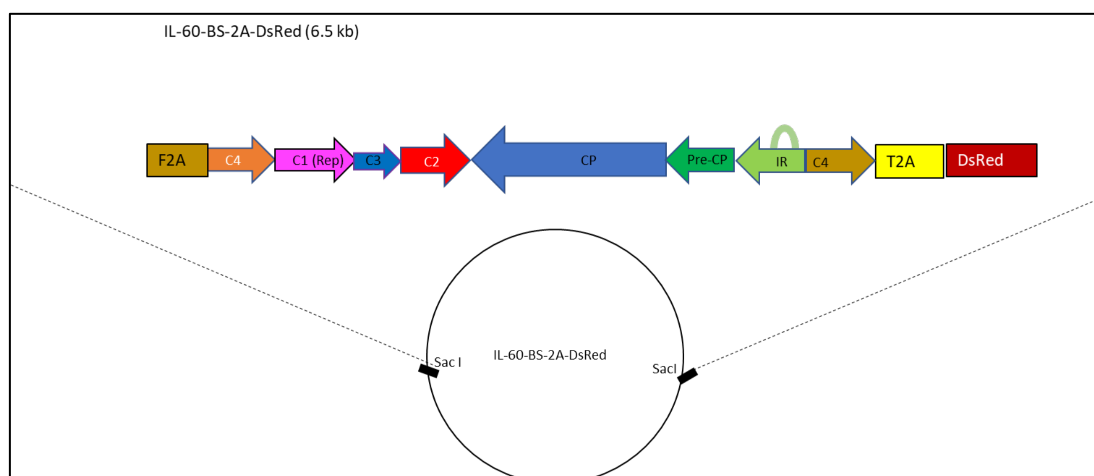


Figure 3.4 Schematic representation of the IL-60-BS-2A-DsRed plasmid. Viral and plasmid genes are shown in the following order: F2A site, C4, C1 *rep* gene C2, Coat protein, Pre-CP, The viral origin of replication/intergenic region (IR), C4, T2A site, and the *DsRed* gene reporter gene. DsRed were inserted in frame within the Rep (C1) reverse strand gene to make use of the Rep1 promoter. By flanking F2A and T2A self-cleaving peptides, a DsRed protein product could be produced from the "Rep1-T2A-DsRed-F2A" RNA during peptide synthesis.

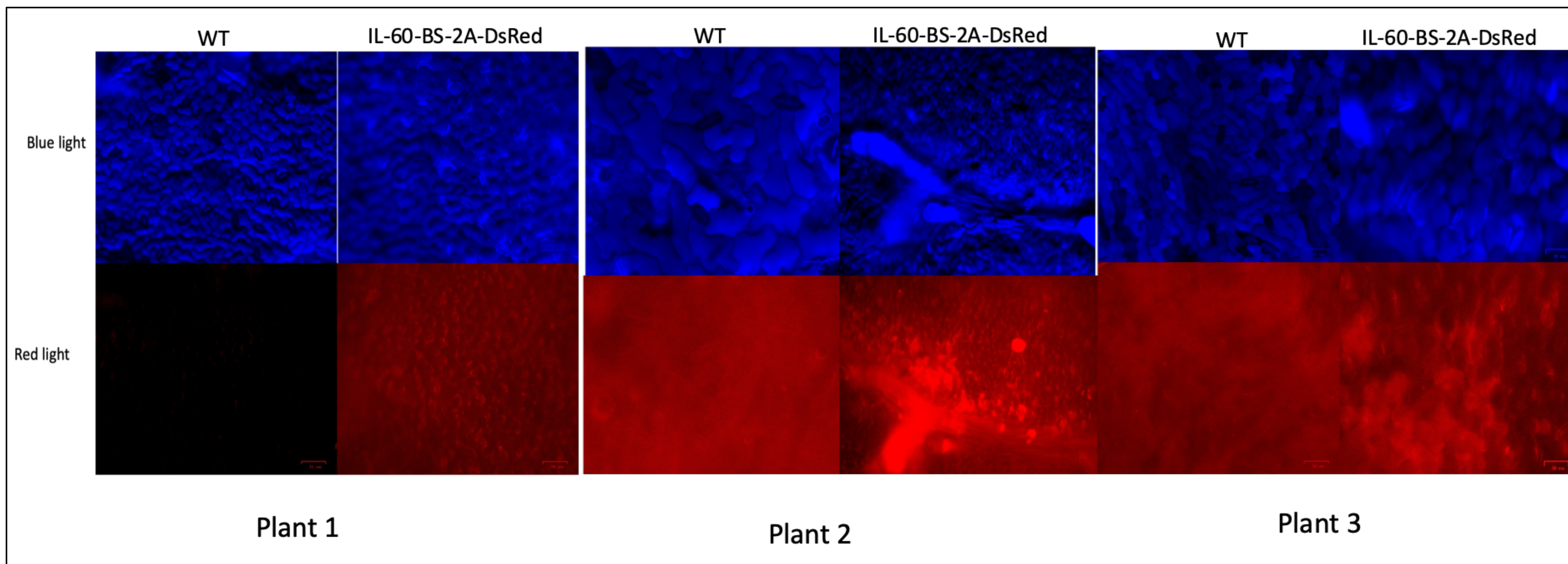


Figure 3.5 DsRed fluorescing in sunflower seedlings inoculated with IL-60-BS-2A-DsRed under blue and red light (50 µm) visualised on the Zoe Fluorescent imager (Bio-Rad laboratories, California, USA). Presence of fluorescence signals was observed in inoculated plants and none was observed in wild type (non-inoculated) plants under blue and red light seven days post inoculation.

3.4.2 Transformation of CRISPR/Cas9 system and testing of systemic movement of the sgRNA and Cas9 transcripts in sunflower leaves

The IL-60-BS-2A-Cas9-1185 vector was successfully assembled through In-Fusion® cloning of the PCR linearized IL-60-BS-2A vector and T-DNA fragment of the pDe-Cas9-1185-*bar* vector. Successful cloning of the vector was initially confirmed by colony PCR using primers targeting the F1 *ori* and parsley ubiquitin promoter and the expected sequence of 1057 bp was observed (Figure 3.6). Successful vector assembly was further confirmed by Sanger sequence (Figure S3.3) and restriction digestion with XmaI (Figure S3.4). The resultant vector was 14 kb (Figure 3.7).

The ability of the IL-60-BS-2A vector to systemically deliver CRISPR/Cas reagents to induce targeted mutagenesis in sunflower was tested by cloning the Cas9 endonuclease, with the PcUBi promoter, Pea (T) terminator, sgRNA with the AtU6-26 promoter, and the donor repair template (DRT) (1185 bp) containing the target amino acid changes. Ten sunflower seedlings were mechanically inoculated with this new vector IL-60-BS-2A-Cas9-1185 for genome editing of the *CenH3* gene in sunflower. DNA from inoculated plants was tested for the presence of the vector using the U6-26 promoter and sgRNA reverse primers as well as the coat protein primers. Of the ten mechanically inoculated plants, the vector was detected in systemically infected leaves of five plants (Figure 3.8). The T0 plants were self-pollinated to obtain T1 plants, and the presence of the vector was detected in all randomly selected eight of the tested T1 plants. The T1 plants which were deemed to be ‘haploid inducers’ were cross-pollinated with wildtype plants as females and the presence of the vector was found in all six (6) randomly selected tested T2 plants using the same primers as above (Figures S3.5).

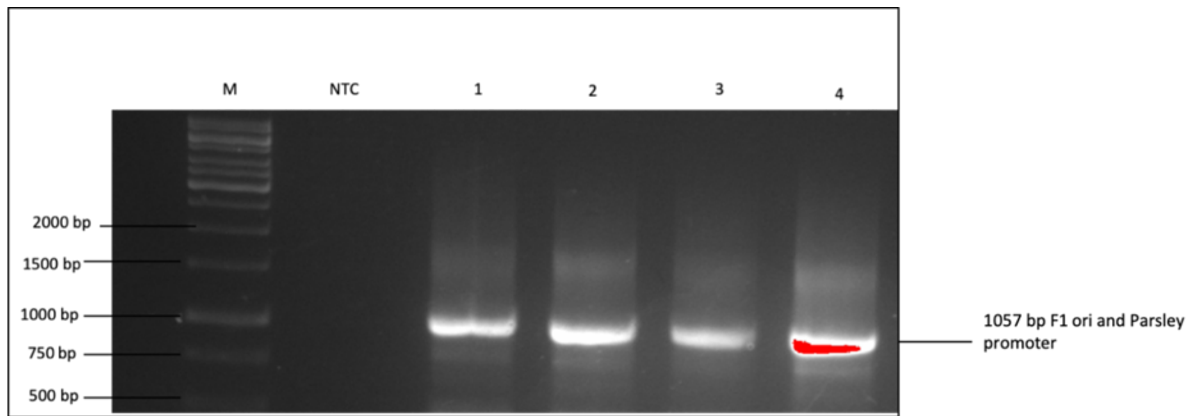


Figure 3.6 Colony PCR confirming successful cloning of CRISPR/Cas9 components into the IL-60-BS-2A construct using primers targeting the Parsley ubiquitin promoter and F1 origin of replication. M – GeneRuler 1 kb molecular marker, NTC- no template control, 1-6 randomly selected colonies. Colonies 1,2,3,4 contained the 1057bp fragment corresponding to the target genes.

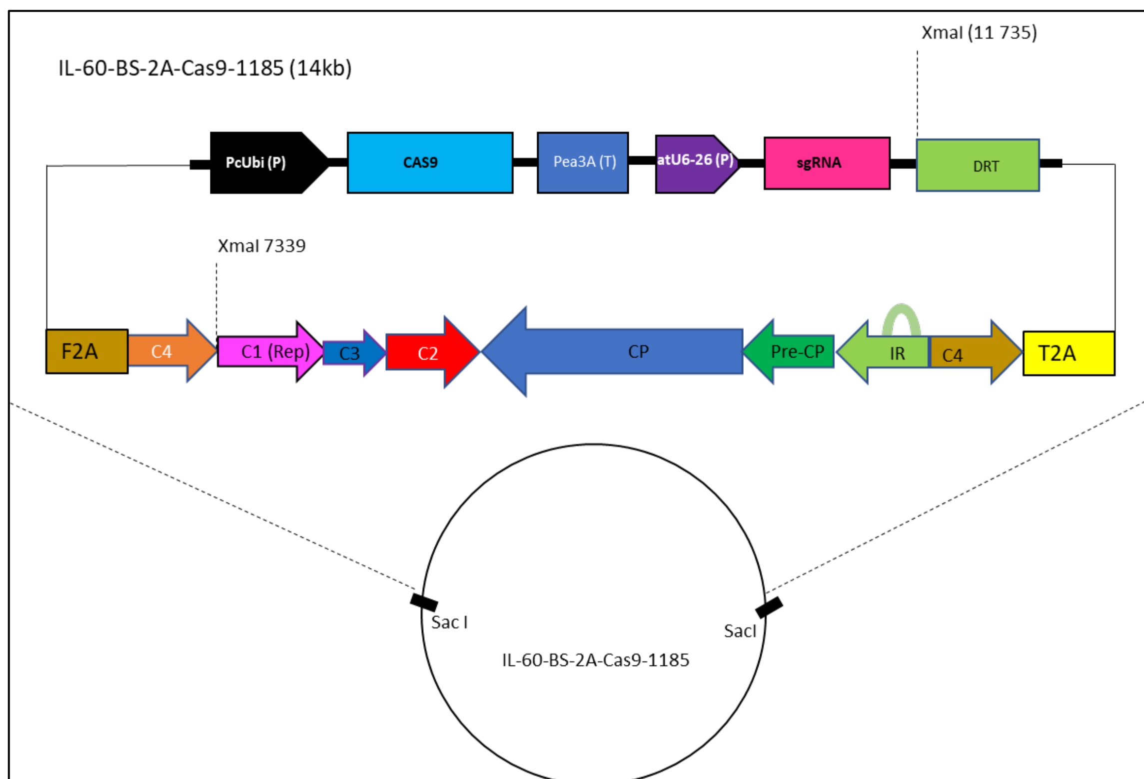


Figure 3.7 Schematic maps for IL-60-BS-2A-Cas9-1185 construct designed for the mutation of the *CenH3* gene in sunflower. The system contains PcUbi (P): The expression of *cas9* was driven by the parsley ubiquitin (PcUBi) promoter; Cas9: *Streptococcus pyogenes* Cas9 coding sequence; Pea3A (T): Pea terminator; sgRNA: small guide RNA was driven by AtU6-26 (P) promoter; DRT: Donor repair template. The coat protein (CP); IR: Intergenic region which is the origin of replication were used for viral replication and systemic movement.

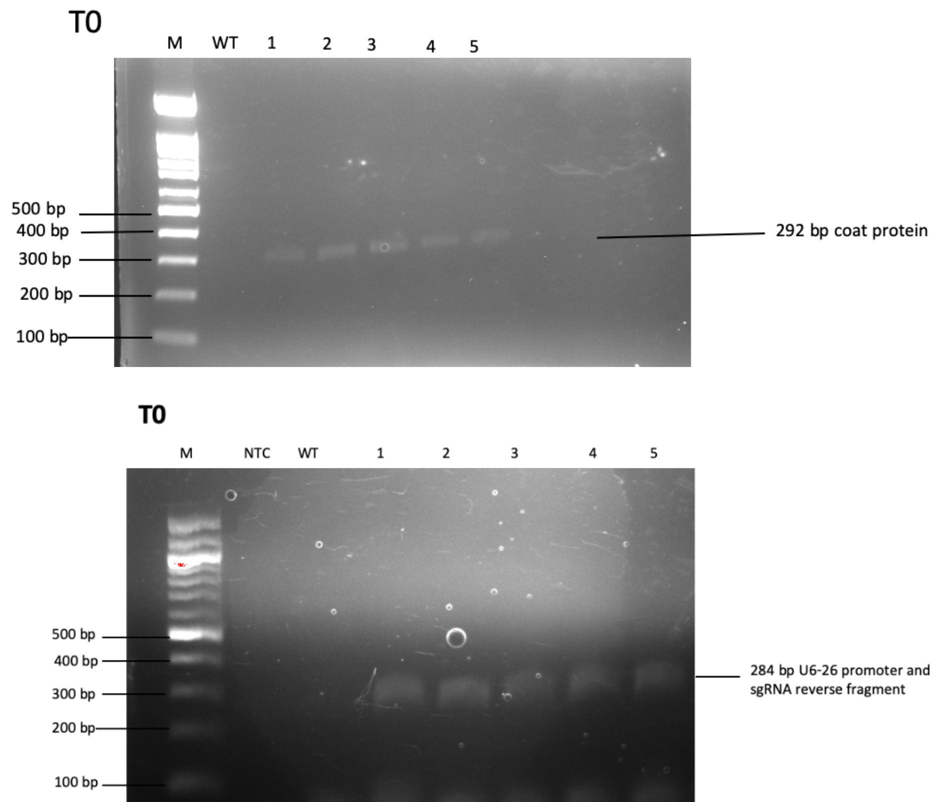


Figure 3.8 PCR confirming the presence of the vector by amplification of the U6-26 promoter and sgRNA in T0 plants inoculated with the IL-60-BS-1185 vector 7 dpi. Five out of the 10 inoculated T0 plants were confirmed to contain the vector through the presence of amplicons targeting the U6-26 promoter (284 bp) and coat protein (294 bp).

3.4.3 cDNA ANALYSES TO CONFIRM CAS9 AND SGRNA GENE TRANSCRIPTION

The CRISPR-Cas9 technology relies on the transcription and expression of the Cas9 endonuclease and the corresponding sgRNA molecule for successful genome editing. A polymerase chain reaction was used to determine if the Cas9 and sgRNA, both under different promoters, were successfully transcribed in sunflower seedlings inoculated with the IL-60-BS-2A-Cas9-1185, 21 days post inoculation. Primers targeting the corresponding fragments of the sgRNA and sgRNA scaffold (96 bp) and the Cas9 endonuclease (572 bp) coding sequences were used to test for transcription of the sgRNA and Cas9 in sunflower cDNA. Five of the ten plants inoculated with the IL-60-BS-2A-Cas9-1185 showed the 96 bp sgRNA fragment (Figure 3.9A) and the 572 bp Cas9 fragment (Figure 3.9B) and no amplification was observed on the wildtype/non-inoculated plants. This confirmed successful sgRNA and Cas9 transcription and systemic movement in leaves 21 days post inoculation. For the T1 generation, eight of the tested plants showed successful transcription of the Cas9 (Figure S3.6A) and all six of the tested T2 generations also demonstrated successful transcription of both Cas9 (Figure S3.6B), sgRNA transcription could, however, not be detected in both T1 and T2 generations.

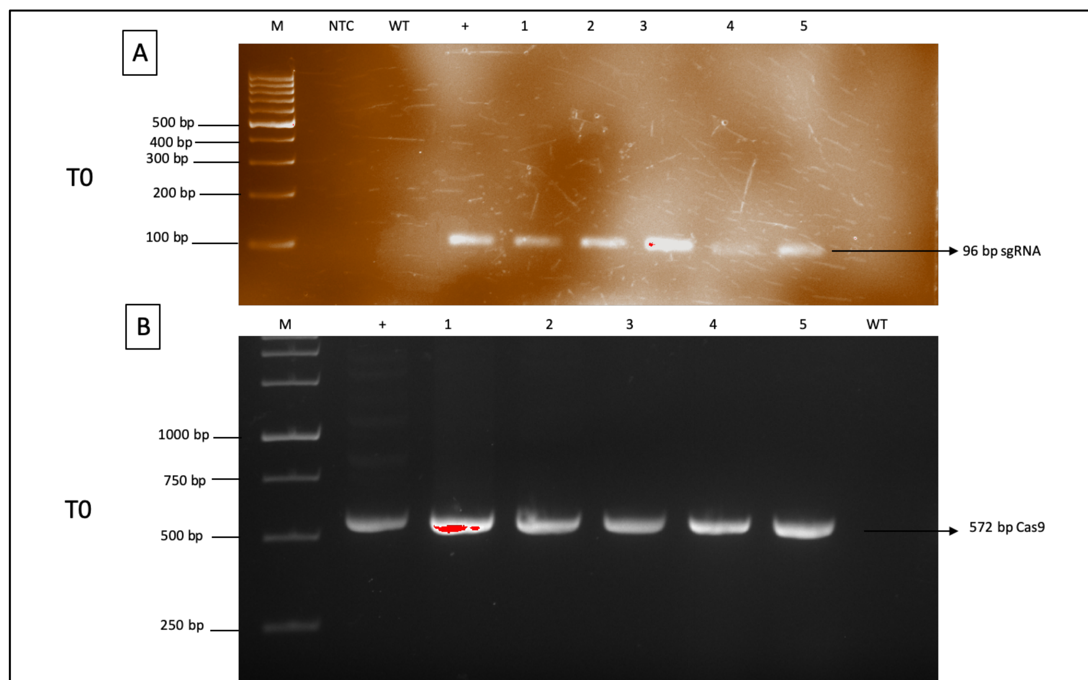


Figure 3.9 cDNA PCR to test sgRNA and Cas9 transcripts in five of the ten T0 sunflower plants inoculated with the IL-60-BS-2A-Cas9-1185 at 21 dpi. A.) M: 100 bp Molecular marker NTC- 2: WT-Wildtype, uninoculated sunflower, +: Positive plasmid control, 1-5- youngest leaves from inoculated plants. B.) M: 1kb Molecular marker, +: Positive plasmid control, 1-5: cDNA from youngest leaves of inoculated plants, WT- Wildtype, uninoculated sunflower control.

3.4.4 CRISPR/CAS9 MUTATION ANALYSES IN SUNFLOWER PLANTS

Successful transcript detection of the Cas9 and sgRNA components in cDNA suggests that genome editing should be possible in sunflower. In combination with the ability of the construct to move systemically through leaves, high genome modification was expected. Therefore, to evaluate if the sunflower genomic *CenH3* target site contained mutations, a targeted PCR was performed on genomic DNA isolated from inoculated plants, as well as resultant T1 and T2 plants. The PCR products were analysed using Illumina MiSeq next generation sequencing. Sequencing results were analysed using Crispresso2 (Clement et al., 2019) (Figure 3.10). Out of the ten plants mechanically inoculated with the IL-60-BS-2A-Cas9-1185, only two (20%) showed mutations around the target site (Table 3.2). For KP328-SEVIH, next generation sequencing and CRISPResso2 analysis showed a total of 7.83% mutations implying heterozygosity (Figure 3.10). For KP328-S-VTH, heterozygous mutations amounting to 3.89% of the alleles were also observed around the target site.

For KP328-SEVIH, results showed only substitutions, 2.84% of the alleles contained a nucleotide substitution of C->A two nucleotides, 2.67% had a nucleotide substitution G->A, six nucleotides downstream the PAM sequence. Of all the alleles, 1.80% contained all the nucleotide changes that corresponded with the target mutations in the DRT including the mutations designed to disable target recognition by deactivating the PAM site. In all inoculated lines, no developmental defects were observed (data not shown). To test the stability/heritability of gene mutation using IL-60-BS-2A-Cas9-1185 for CRISPR/Cas component delivery, T1 plants obtained from self-crossing T0 (KP328-SEVIH and KP328-S-VTH) and T2 plants obtained from crossing T1 and wildtype sunflower plants were screened for mutations using NGS and CRISPResso2 (Clement et al., 2019). None of the plants in both the T1 and T2 generations displayed any mutations. CRISpresso2 (Clement et al., 2019) results of non-modified plants in the T0, T1 and T2 generations are listed in the supplementary data (Figure S3.7, S3.8 and S3.9)

Table 3.2 Table showing transformation, expression of transgenes and CRISPR/Cas9 mutation efficiency in sunflower seedlings inoculated with two TYLCV-based delivery systems.

Construct	Number of transformed/ tested plants	Number of plants expressing Cas9	Number of plants with mutations	Transformation efficiency	Mutation frequency
IL-60-BS-2A-DsRed	3	3/3	-	-	-
IL-60-BS-2A-Cas9-1185 (T0)	10	5/10	2/10	50%	20%
IL-60-BS-2A-Cas9-1185 (T1)	8	8/8	0/8	0%	0%
IL-60-BS-2A-Cas9-1185 (T2)	6	6/6	0/6	0%	0%

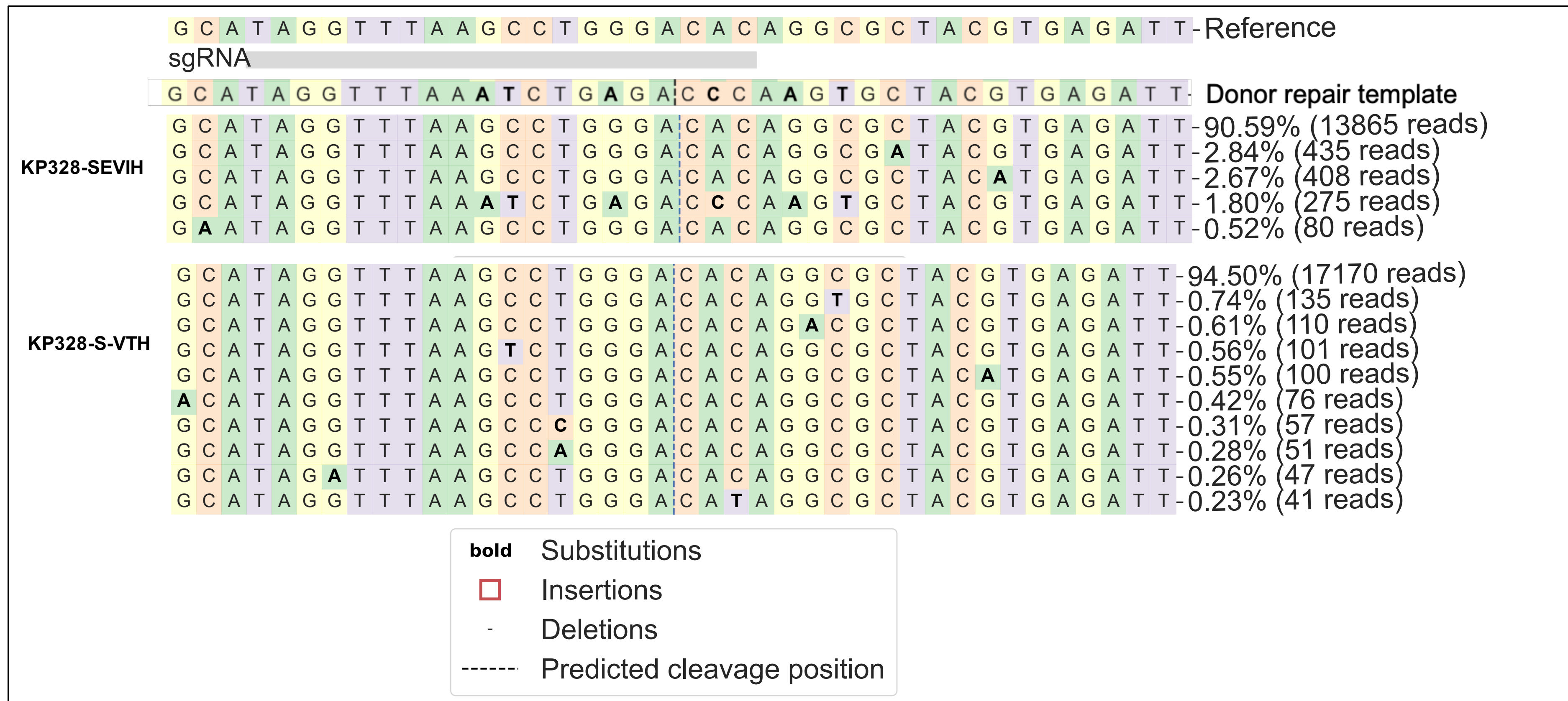


Figure 3.10 Mutations in two sunflower KP328 seedlings achieved through viral delivery of CRISPR/Cas reagents, sgRNA and donor repair template (DRT). The targeted region was analysed with NGS (MiSeq) prior to analysing with CRISPresso2 (Clement et al., 2019). Distribution of identified alleles around the predicted cleavage site (vertical dash line, 3 bp upstream the PAM sequence) on the sunflower *CenH3* targeted gene (nucleotide numbers 551 – 590, XP_022005281.1). The reference (WT) allele is presented first, followed by the DRT template sequence in the second line and an alignment of all the alleles obtained after editing are displayed in the low portion of the plot. Substitutions are shown in bold font. No inserted or deleted nucleotides were observed. The allele frequencies are presented as percentages and read numbers. Only the target region of interest is aligned and displayed.

The alleles obtained through CRISPResso2 (Clement et al., 2019) analysis were translated to amino acids using ExPAsy (http://www.expasy.org; Swiss institute of Bioinformatics) and aligned against the wildtype and DRT (Figure 3.11). In plant one, Allele_3 had a complete conversion of the three targeted amino acids (P51S; G52E; A55V) represented in the DRT sequence, thus confirming complete homology directed repair. Allele 1 and 2 contained two amino acid (L56I; R57H) conversions that were not included in the DRT sequence. For plant two, only two targeted amino acid (P51S, A55V) conversions were obtained, in different alleles. The other alleles contained altered amino acids (A55T; H57H) and a premature stop codon was also introduced in allele_7, KP328-S-VTH.

	AMINO ACID POSITION	51	52	53	54	55	56	57
KP328-SEVIH	WILDTYPE	P	G	T	Q	A	L	R
	DONOR REPAIR TEMPLATE	S	E	T	Q	V	L	R
	ALLELE_1	P	G	T	Q	A	I	R
	ALLELE_2	P	G	T	Q	A	L	H
	ALLELE_3	S	E	T	Q	V	L	R

	AMINO ACID POSITION	51	52	53	54	55	56	57
KP328-S-VTH	WILDTYPE	P	G	T	Q	A	L	R
	DONOR REPAIR TEMPLATE	S	E	T	Q	V	L	R
	ALLELE_1	P	G	T	Q	V	L	R
	ALLELE_2	P	G	T	Q	T	L	R
	ALLELE_3	S	G	T	Q	A	L	R
	ALLELE_4	P	G	T	Q	A	H	R
	ALLELE_7	P	G	T	-	A	L	R

Figure 3.11 Multiple sequence alignment of CENH3 histone fold domain amino acid changes induced by CRISPR/Cas9 in sunflower. The amino acids in green are the required changes for haploid inducer formation while the red are random changes induced by CRISPR. Only regions of interest are aligned and displayed. Amino acid positions in the sunflower CENH3 are listed at the top.

3.5 DISCUSSION

Targeted editing of plant genomes is a promising strategy for enhancing crop production in the face of a growing population and climate change (Uranga et al., 2021). The transformation and delivery of editing components still poses a major hindrance in the adoption and widespread applications of genome editing in crops. Even though some success has been achieved with using *Agrobacterium tumefaciens* for sunflower transformation, it is still considered a highly recalcitrant crop for *in vitro* culturing and modification (Darqui et al., 2021). Recently, two different geminiviral vectors have been successfully utilized as delivery systems for *in planta* delivery of CRISPR- based editing reagents and displayed high cargo capacity and successful genome editing via the homology directed repair mechanism at whole plant level (Rezaei et al., 2021; Eini et al., 2022). This study aimed to circumvent sunflower's *in vitro* restrictions by using a modified previously published viral-based delivery system. Sunflower seedlings were inoculated with a modified IL-60-BS vector containing CRISPR components for targeted genome editing in sunflower. 50% of the inoculated plants showed successful inoculation with the vector, and 20% displayed mutations, with one plant displaying complete homology directed repair. Furthermore, all plants tested in the T1 and T2 generations showed presence of vector and successful transcription of Cas9, no sgRNA transcription was observed. No editing was however observed in both the T1 and T2 generations. The success of the CRISPR/Cas9 gene editing system is highly dependent on the expression of the sgRNA and Cas9, as such the selection of strong promoters is crucial. Pol III-dependent promoters such as the U6 (used in this study) and U3 are regarded as good candidates for the sgRNA expression while Pol II-dependent promoters are recommended for Cas9 expression (Mao et al., 2019). Although the promoters used in this study (U6-26 and PcUBi) are regarded as highly potent and have shown success in a majority of plant species (Mao et al., 2019), there is little information on their use in sunflower. It would be beneficial to explore the different promoters for suitability in sunflower CRISPR/Cas9 gene editing.

The basic process of a productive plant virus infection involves three steps, namely: replication, cell-to-cell movement and long-distance movement through the phloem (Wang and Zhou, 2016). In this study, the IL-60-BS-2A viral system was tested for *in planta* replication and movement, as well as the expression of genes of interest in sunflower seedlings. The detection of the virus in plant tissues was not always consistent using a direct PCR method, most likely due to the small leaf sample size required for the plant direct PCR kit and/or irregular movement of the virus viral vector that results in unequal distribution in the leaf tissue. The use of a reporter gene, *DsRed*, however further confirmed systemic movement and expression of the viral vector in sunflower seedlings' leaves.

Viruses have been widely manipulated to serve as delivery systems for genome editing reagents e.g. CRISPR/Cas9 reagents. However, cargo capacity limitations hamper movement between cells and through the phloem in the majority of these experiments (Kujur et al., 2021). Geminiviruses are the most commonly used virus vectors for CRISPR-based genome editing (Baltes et al., 2014, Čermák, et al., 2015, Eini et al., 2022). In this experiment a deactivated, TYLCV-based delivery vector was used to deliver the CRISPR system, i.e. spCas9 protein, sgRNA and a donor-repair template (DRT) to sunflower seedlings. The results of this study add more evidence for the use of a geminiviral delivery system to deliver all the components, even large constructs containing Cas9, sgRNA and DRT, to recalcitrant plants like sunflower. It also shows that the transient expression of these components can carry out genome editing of the target regions as the construct moves systemically within the sunflower plants, though at low frequencies. This concurred with a study in *N. benthamiana* where it was reported that a beet curly top virus replicon was able to transfer the Cas12a nuclease, DRT DNA, vector backbone, crRNA cassette up to 7kb in size to plant cells and induce successful expression (Eini et al., 2022). However, in contrast to our study, it was shown to be unstable when used for a slightly larger fragment of 7,7 kb (Eini et al., 2022). In other studies, the BeYDV replicon was able to maintain stability and successfully expressed CRISPR-reagents to the excess of 12 kb in tomato (Čermák, et al., 2015) and *Nicotiana tabacum* (Rezaei et al., 2021). As more studies are conducted, better understanding of how to overcome insert size challenges in expression will be demonstrated (Eini et al., 2022). The use of 2A peptides on the vector in this study, could have had a positive effect on vector stability and transgene expression (Luke et al 2015). 2A peptides have been shown to enrich co-expression of transgenes in gene editing experiments thereby increasing gene editing efficiency, including that of the CRISPR/Cas9 system (Mariano et al., 2014, Duda et al., 2014).

To incorporate specific nucleotide changes on a sequence, an exogeneous DNA (donor repair template) is required to enhance HDR activity (Huang and Putcha., 2019). In this study, complete (1.80% alleles, KP328-SEVIH) and partial (0.74% and 0.56% alleles, KP328-S-VTH) homology directed repair was detected, though at a low frequency which was expected. Homology directed repair (HDR) is a rare occurrence in plants as non-homologous end joining (NHEJ) is the main repair mechanism in plants (Eini et al., 2022). The results in this study, support results from Eini et al. (2022) where the use of a geminiviral vector for CRISPR/Cas12a delivery led to both complete and partial conversion of amino acids for the conversion of the green fluorescent protein to the yellow fluorescent protein at frequencies between 2.8% - 23.4% (Eini et al., 2022). Unlike traditional plant transformation methods,

viruses can replicate and reach high titre in plant cells thereby increasing availability of CRISPR/Cas reagents and enhancing editing efficiency (Gentzel et al., 2022).

Compared to the previous observations using *Agrobacterium tumefaciens*-mediated transformation (Chapter 2) where only partial HDR was observed, it can clearly be seen that the type of delivery method used for plant transformation influences both the editing efficiency and HDR efficiency. This is supported by the *N. benthamiana* study by Eini et al (2022), where HDR efficiency was higher where a geminiviral replicon was used for genome editing in comparison to *A. tumefaciens* mediated transformation. DNA viruses, geminiviruses in particular are also said to have the ability to revert cells to the S-phase, where HDR is abundant (Eini et al., 2022). Progression of cells into the S-phase is necessary for geminivirus replication (Baltes et al., 2014). Viruses have been used for the delivery of CRISPR/Cas reagents and donor repair templates for improvement of HDR in plant species including tomato, tobacco, and wheat (Čermák, et al., 2015). Delivery of repair templates using viral replicons improved HDR frequency in wheat and tobacco respectively (Čermák, et al., 2015, Eini et al., 2022). A limited number of studies have reported HDR repair systems for CRISPR/Cas9 including *Nicotiana benthamiana* (Hirohata et al., 2019) and rice (Ma et al., 2015). It was also observed in this study that the lower the allele percentage (editing), the lower the HDR efficiency. This could imply that a high editing efficiency is required for increased successful and complete homology directed repair which was also observed in previous studies (Čermák, et al., 2015, Eini et al., 2022).

Although editing and HDR efficiency (8% and 1.8% respectively) was increased in the present study, compared to the previous chapter (Chapter 2) where *A. tumefaciens* mediated transformation was used (0% complete HDR), it still was much lower when compared to other studies where geminiviral replicons were used for CRISPR/Cas9 reagent delivery (Čermák, et al., 2015; Eini et al., 2022). Previous studies have demonstrated HDR efficiency of up to 23.5% with the use of geminiviral replicons (Eini et al., 2022). In *Agrobacterium tumefaciens*-mediated transformation, transgene copy number has been implicated as the cause of low efficiency in editing (Huang and Puchta., 2019). The retention of rolling circle replication and removal of movement protein genes resulted in high efficiency editing in *Arabidopsis* and tobacco through the delivery of CRISPR components using a geminiviral replicon (BeYDV and BCTV, respectively) (Baltes et al., 2014; Eini et al., 2022). It has been reported that dsDNA is the main form in which the virus moves through the plant (Peretz et al., 2007), however, ssDNA produced through conversion of dsDNA to ssDNA during rolling circle replication may be necessary for high viral DNA accumulation within infected cells. The viral intergenic region (IR) of the TYLCV virus was discovered to be the driver of replication, movement and viral

accumulation even in the absence of the Rep (C1) protein (Gover et al., 2014). Viral replication through dsDNA is also governed by host proteins known as Retino-Blasmota-Related proteins independent of the Rep protein (Zhang et al., 2023). Although the replication of the IL-60-BS-1185 vector used in this study was sufficient to express CRISPR reagents and induce mutations, the possible inhibition of rolling circle replication in the IL-60-BS vector used in study, could have inhibited the accumulation of high copy numbers of virus for efficient genome editing (Baltes et al., 2014; Butler et al., 2016). It may be useful to test a rolling circle replication competent vector in future studies as this may elevate gene editing efficiency-though placing high genetic load on the host plant. The large heterologous sequence (Cas9, sgRNA and DRT) delivered here could have also affected DNA accumulation in plant cells to promote efficient editing. In tobacco, an increase in heterologous DNA size affected DNA accumulation and gene editing efficiency through CRISPR reagent delivery by a geminivirus BCTV replicon (Eini et al., 2022).

The *CenH3* gene, targeted in this work, is crucial for plant development and certain mutations may be too lethal to maintain (Wang et al., 2021). This could also explain the lack of homozygous transformation events observed. In maize, CRISPR/Cas9 editing of the *CenH3* gene also failed to produce homozygous mutations (Wang et al., 2021). Maize, just like sunflower, contains a single copy of the *CenH3* and homozygous mutant alleles may be lethal to plant development. The *CenH3* gene is essential for centromere function and chromosome segregation in plants, modification of this gene has been shown to lead to chromosome segregation errors and lethality (Britt and Kuppu, 2016). The sunflower *CenH3* gene was also found to be highly conserved between different sunflower accessions analysed in this study and this could be a further indication of the level of intolerance of *CenH3* function to sequence variation. The amino acid changes tested in this study could be individually analysed instead of combined as this would possibly reduce the lethality of the mutations. Furthermore, modifications in other regions of the *CenH3* gene such as the N-terminal could be tested as this region is known to be less conserved than the HFD (Ravi and Chan., 2010).

The ability to detect the presence of the vector and expression of Cas9 in the T1 and T2 progeny plants indicates that the TYLCV-IL is seed transmissible to some extent, although in the original study it was indicated not to be (Peretz et al 2007). TYLCV is identified as a phloem-limited virus, and utilises a non-virion form for cell-to-cell movement (Wang and Zhou, 2016). The IL-60-BS vector system described by (Peretz et al., 2007) replaced rolling cycle single-stranded DNA replication with a plasmid-based DNA replicon and this eliminated the phloem restriction of the wildtype TYLCV. Out of 178 progeny plants from parents infected with the IL-60-BS-GUS vector, only one (0.5%) was found to carry the vector and this was

concluded by the researchers as a 'mechanical contamination' due to the low frequency of occurrence (Peretz et al., 2007). The results in this study, however, demonstrate that this transfer might differ between plants species since possible seed transmission of the IL-60-BS-2A-Cas9-1185 vector in sunflower was observed at a higher rate (100%). The vertical transmission of TYLCV-IL from infected plants to offspring via seeds has been described previously (Kil et al., 2016). Seed transmitted TYLCV can infect whole plants, replicate and move systemically (Kil et al., 2016). Different TYLCV stains have exhibited different host interactions for different plant species (Moriones and Castillo et al., 2000). Although the presence of the vector was detectable in T1 and T2 progeny plants in this study, Cas9 activity might be hampered in vertically transferred vector due to low copy number and reduced expression. The lack of sgRNA expression observed in these generations could have also affected gene editing. Transgene instability and silencing in subsequent transgenic line generation is quite a common in sunflower (Darqui et al., 2021). This is evident in that a majority of sunflower transformation studies only show results for T1 and T2 generations. The transgene silencing phenomenon could be attributed to the lack of sgRNA expression in T1 and T2 generations in this study. Furthermore, in tomato, seeds with seedborne TYLCV-IL were subject to qPCR and it was revealed that a high concentration of the TYLCV-IL was located externally, on the seed coat and only a small load was detected internally (Pérez-Padilla et al., 2020). Moreover, plants germinated from infected seed demonstrated no virus symptoms, even with virus detection in 80% of the progeny.

None of the progeny (T1 and T2) plants analysed in the present study had mutations in the *CenH3* gene. This is in correspondence with previous studies, where the use of geminivirus replicon in tomato for CRISPR/Cas9 targeted mutagenesis, resulted in non-heritable mutations in the progeny (Čermák, et al., 2015). The absence of heritable mutations in the T1 and T2 generation could be due to a number of factors including: the fact that the mutations in the T0 cause chromosome segregation errors and the T0 genome is lost and only the wildtype *CenH3* is transferred onto the progeny T0 during selfings, or that, the high percentage of WT allele outperformed the lower copy numbered mutated alleles; or the transient nature and inability to infect germline cells of most viral vector systems (Uranga et al., 2021). The inability to transmit mutations to the next generation is a disadvantage of most viral vector systems (Yin et al., 2015). In most studies, to recover mutations in the next generation, infected plant parts are regenerated via tissue culture (Uranga et al., 2021), this serves as a disadvantage for plant such as sunflower that has a low regeneration efficiency.

From the results in this study and other reports from Eini et al. (2022) and Čermák, et al. (2015) we can conclude that viral vectors increase mutagenesis and HDR in CRISPR-based editing of plant species in comparison to delivery systems based on *Agrobacterium*

tumefaciens. However, the stability of the mutation in progeny is much lower than *Agrobacterium tumefaciens* based system, thus supporting the *A. tumefaciens* as a method of choice for plant transformation. Genome editing experiments in crops struggle to provide consistent results due to limitations observed with the present delivery methods in different species/lines. This study demonstrates the opportunity for *in vitro* recalcitrant crop species, such as sunflower, to be genetically modified and improved in a less complex and laborious manner. Even with the possible persistence of viruses in seeds, efficient seed surface sterilization might eliminate the majority of the virus and viruses are not harmful to humans, genome editing reagents and antibiotic resistance genes carried in these constructs, might be of concern however. This strategy needs to be optimized in plants, to achieve higher rates of homozygous mutated plants.

3.6 REFERENCES

Ali, Z., Abul-Faraj, A., Li, L., Ghosh, N., Piatek, M., Mahjoub, A., Aouida, M., Piatek, A., Baltes, N.J., Voytas, D.F., Dinesh-Kumar, S., Mahfouz, M.M (2015) Efficient virus-mediated genome editing in plants using the CRISPR/Cas9 System. Molecular Plant 8, 1288-1291. https://doi.org/10.1016/j.molp.2015.02.011

Ariga, H., Toki, S., Ishibashi, K (2020) *Potato virus X* vector-mediated DNA-free genome editing in plants. Plant and Cell Physiology 61, 1946-1953. https://doi.org/10.1093/pcp/pcaa123

Baltes, N.J., Gil-Humanes, J., Čermák,, T., Atkins, P.A., Voytas, D.F (2014) DNA replicons for plant genome engineering. The Plant Cell 26, 151-163. https://doi.org/10.1105/tpc.113.119792

Britt, A.B., Kuppu, S (2016) CenH3: An emerging player in haploid induction technology. Frontiers in Plant Science 7, 357. https://doi.org/10.3389/fpls.2016.00357

Čermák, T., Baltes, N.J., Čegan, R., Zhang, Y., Voytas, D.F (2015) High-frequency, precise modification of the tomato genome. Genome Biology 16, 232. https://doi.org/10.1186/s13059-015-0796-9

Chandrasekaran, M., Boopathi, T., Paramasivan, M (2021) A status-quo review on CRISPR-Cas9 gene editing applications in tomato. International Journal of Biological Macromolecules 190, 120-129. https://doi.org/10.1016/j.ijbiomac.2021.08.169

Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., Pinello, L (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nature Biotechnology 37, 224-226. https://doi.org/10.1038/s41587-019-0032-3

Cui, X., Tao, X., Xie, Y., Fauquet, C.M., Zhou, X (2004) A DNA β associated with *Tomato Yellow Leaf Curl China Virus* is required for symptom induction. Journal of Virology 74, 13966-13974. https://doi.org/10.1128/JVI.78.24.13966

Czosnek, H., Laterrot, H (1997) A worldwide survey *of tomato yellow leaf curl* viruses. Archives of Virology 142, 1391-1406. https://doi.org/10.1007/s007050050168

Dahan-Meir T, Filler-Hayut S, Melamed-Bessudo C, Bocobza S, Czosnek H, Aharoni A (2018) Efficient *in planta* gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. Critical Reviews in Plant Science 95, 12-6. https://doi.org/ 10.1111/tpj.13932

Darqui, F., Radonic, L., Beracochea, V., Hopp, H., López Bilbao, M (2021) Peculiarities of the transformation of *Asteraceae* family species: The cases of sunflower and lettuce. Frontiers in Plant Science 12, 767459. https://doi.org/10.3389/fpls.2021.767459

Duda, K., Lonowski, L., Jofoed-Nielsen, Ibarra, A., Delay, C., Kang, Q., Yang, Z., Pruett-Muller, S.,Bennett, E.,Wandall, H., Davis, G.,Hansen, S, Frodin, M (2014) High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and Zinc Finger Nucleases or CRISPR/Cas9 nickase pairs. Nucleic Acids Research 42, 84.

Kujur, S., Senthil-Kumar, M., Kumar, R (2021) Plant viral vectors: Expanding the possibilities of precise gene editing in plant genomes. Plant Cell Reports 40, 931-934. https://doi.org/10.1007/s00299-021-02697-2

Kumar, S.P., Voytas, D.F (2020) Editing through infection. Nature Plants 6, 738-739. https://doi.org/10.1038/s41477-020-0716-1

Eini, O., Schumann, N., Niessen, M., Varrelmann, M (2022) Targeted mutagenesis in plants using *Beet curly top virus* for efficient delivery of CRISPR/Cas12a components. New Biotechnology 67, 1-11. https://doi.org/10.1016/j.nbt.2021.12.002

Fauser, F., Schiml, S., Puchta, H (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*.The Plant Journal 79, 348-359. https://doi.org/10.1111/tpj.12554

Gentzel, I.N., Ohlson, E.W., Redinbaugh, M.G., Wang, G.L (2022) VIGE: virus-induced genome editing for improving abiotic and biotic stress traits in plants. Stress Biology 2, 2. https://doi.org/10.1007/s44154-021-00026-x

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C., Sánchez-León, S., Baltes, N.J., Starker, C., Barro, F., Gao, C., Voytas, D.F (2017) High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. The Plant Journal 89, 1251-1262. https://doi.org/10.1111/tpj.13446

Gleba, Y., Klimyuk, V., Marillonnet, S (2007) Viral expression of proteins in plants. Current opinions in Biotechnology 18, 134-141

Gover, O., Peretz, Y., Mozes-Koch, R., Maori, E., Rabinowitch, H.D., Sela, I (2014) Only minimal regions of tomato yellow leaf curl virus (TYLCV) are required for replication, expression and movement. Archives of Virology. 159, 2263-2274

Hirohata, A., Sato, I., Kaino, K., Iwata, Y., Koizumi, N., Mishiba, K. ichiro (2019) CRISPR/Cas9-mediated homologous recombination in tobacco. Plant Cell Reports 38, 463-473. https://doi.org/10.1007/s00299-018-2320-7

Huang, T.K., Puchta, H (2019) CRISPR/Cas-mediated gene targeting in plants: Finally, a turn for the better for homologous recombination. Plant Cell Reports 38, 443-453. https://doi.org/10.1007/s00299-019-02379-0

Kil, J., Kim, S., Lee, Y. J., Byun, H. S., Park, J., Seo, H., Kim, C. S., Shim, J. K., Lee, J. H., Kim, J. K., Lee, K. Y., Choi, H. S., Lee, S (2016) *Tomato yellow leaf curl virus* (TYLCV-IL): A seed-transmissible geminivirus in tomatoes. Scientific Reports 6, 19013. https://doi.org/10.1038/srep19013

Kujur, S., Senthil-Kumar, M., Kumar, R (2021) Plant viral vectors: expanding the possibilities of precise gene editing in plant genomes. Plant Cell Reports 40, 931-934. https://doi.org/10.1007/s00299-021-02697-2

Liu, H., Zhang, B (2020) Virus-based CRISPR/Cas9 genome editing in plants. Trends in Genetics 36, 810-813. https://doi.org/10.1016/j.tig.2020.08.002

Luke, G.A., Roulston, C., Tilsner, J. and Ryan, M.D (2015). Growing uses of 2A in plant biotechnology. Biotechnology,165-193.

Ma X, Zhang X, Liu H (2020) Highly efficient DNA-free plant genome editing using virally delivered CRISPR-Cas9. Nature Plants 2020, 6:773-779. https//doi: 10.1038/s41477-020-0704-5

Mao, Y., Botella, J.R., Liu, Y., Zhu, J.K (2019) Gene editing in plants: progress and challenges. National Science Review. 6, 421-437. https://doi.org/10.1093/nsr/nwz005

Mariano, A., Xu, L., Han, R (2014) Highly efficient genome editing via 2A coupled co-expression of two TALEN monomers. BMC Research Notes 7, 628. https://doi.org/10.1186/1756-0500-7-628

Moriones, E., Castillo, J (2000) *Tomato yellow leaf curl virus*: An emerging virus complex causing epidemics worldwide. Virus Research 71, 123-134. https://doi.org/10.1016/s0168-1702(00)00193-3

Mozes-Koch, R., Tanne, E., Brodezki, A., Yehuda, R., Gover, O., Rabinowitch, H., Sela, I (2017) Expression of the entire polyhyfroxybutyrate operon of *Ralstonia eutropha* in plants. Journal of Biological engineering 11, 44. https://doi.org/10.1186/s13036-017-0062-7

Peretz, Y., Mozes-koch, R., Akad, F., Tanne, E., Czosnek, H., Sela, I (2007) A universal expression/silencing vector in plants. Plant Physiology 145, 1251-1263. https://doi.org/10.1104/pp.107.108217

Pérez-Padilla, V., Fortes, I.M., Romero-Rodríguez, B., Arroyo-Mateos, M., Castillo, A.G., Moyano, C., de León, L., Moriones, E (2020) Revisiting seed transmission of the type strain of *Tomato yellow leaf curl virus* in tomato plants. Phytopathology 110, 121-129. https://doi.org/10.1094/PHYTO-07-19-0232-FI

Petolino, J.F (2015) Genome editing in plants via designed Zinc Finger Nucleases. *In Vitro* Cellular and Developmental Biology - Plant 51, 1-8. https://doi.org/10.1007/s11627-015-9663-3

Ran, Y., Liang, Z., Gao, C (2017) Current and future editing reagent delivery systems for plant genome editing. Science China Life Sciences 60, 490-505. https://doi.org/10.1007/s11427-017-9022-1

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

Rezaei, A., Farsi, M., Malekzadeh-Shafaroudi, S., Seifi, A (2021) *In planta* removal of *nptII* selectable marker gene from transgenic tobacco plants using CRISPR/Cas9 system. Plant Gene 26, 100288. https://doi.org/10.1016/j.plgene.2021.100288

Sanjana, N.E., Shalem, O., Zhang, F., 2014. Improved vectors and genome-wide libraries for CRISPR screening. Nature Methods 11, 783-784. https://doi.org/10.1038/nmeth.3047

Schiml, S., Fauser, F., Puchta, H (2014) The CRISPR/Cas system can be used as nuclease for *in planta* gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny. The Plant Journal 80, 1139-1150. https://doi.org/10.1111/tpj.12704

Uranga, M., Vazquez-Vilar, M., Orzáez, Di., Daròs, J.A (2021) CRISPR-Cas12a genome editing at the whole-plant level using two compatible RNA virus vectors. The CRISPR Journal 4, 761-769. https://doi.org/10.1089/crispr.2021.0049

Varanda, C.M.R., Félix, M.D.R., Campos, M.D., Patanita, M., Materatski, P (2021) Plant viruses: From targets to tools for CRISPR. Viruses 13, 1-19. https://doi.org/10.3390/v13010141

Wang, A., Zhou, X (2016) Current research topics in plant virology, Current Research Topics in Plant Virology. https://doi.org/10.1007/978-3-319-32919-2

Wang, N., Gent, J., Dawe, K (2021) Haploid induction by a maize cenh3 null mutant. Science Advances 7, eabe2299. https://doi.org/ 10.1126/sciadv.abe2299

Yang, Q., Ding, B., Zhou, X (2017) Geminiviruses and their application in biotechnology. Journal of Integrative Agriculture 16, 2761-2771. https://doi.org/10.1016/S2095-3119(17)61702-7

Yin, K., Han, T., Liu, G., Chen, T., Wang, Y., Yu, A.Y.L., Liu, Y (2015) A geminivirus-based guide RNA delivery system for CRISPR/Cas9 mediated plant genome editing. Scientific Reports 5, 14926. https://doi.org/10.1038/srep14926

Zaidi, S.S.-A., Mansoor, S (2017) Viral vectors for plant genome engineering. Frontiers in Plant Science 8, 2012-2017. https://doi.org/10.3389/fpls.2017.00539

Zess, E., Begemann, M (2021) CRISPR-Cas9 and beyond: What's next in plant genome engineering. *In vitro* Cellular and Developmental Biology – Plant 5, 584-594. https://doi.org/10.1007/s11627-021-10185-1

3.7 SUPPLEMENTARY FIGURES

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCA
TTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAG
ATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTC
CAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCA
CCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGG
GAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGG
GAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGC
GCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGC
CATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACG
CCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTT
TTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTA
TAGGGCGAATTGGAGCTCTGGACCAGGGTTCTCTTCAACATCTCCACATGTCAAAAGA
GAACCTCTTCCCTCTCCAGATCCAGAAGTAGCAACTTCATCAGCAGCATTAGAGGATCC
CCCGGGCTGCAGTGGACCAGGGTTAGATTCAACATCTCCAGCAAGCTTCAAAAGATCG
AAGTTCAAAAGCTGCTTAACAGGAGCAACAATTTTTTGCTTATGTCTAGCCTCTCCAGAT
CCAGAAGTAGCAACTTCATCAGCAGCATTAGAAGCTCTTAGCTGCCTGAATGTTCGGAT
GGAAATGTGCTGACCTGTTTGGGGATACCAAGTCGAAGAACCGTTGGTTCTTACATTGG
TATTTGCCTTCGAATTGGATAAGCACATGGAGATGTGGTTCCCCATTCTCGTGGAGTTC
TTTGCAAACTTTGATGTATTTTTTATTTGTTGGGGTTTCTAGTTTTTTTAATTGGGAAAGT
GCTTCCTCTTTAGAGAGAGAACAATTGGGATATGTTAGGAAATAATTTTTGGCATATATT
TTAAATAAACGAGGCATGTTGAAATGAATCGGTGTCCCTCAAAGCTCTATGGCAATCGG
TGTATCGGTGTCTTACTTATACTTGGACACCTAATGGCTATTTGGTAATTTCATAAATGTT
CATTTCAATTCAAAATTCAAAATTCAAAAATCAAATCATTAAAGCGGCCATCCGTATAATA
TTACCGGATGGCCGCGCCTTTTCCTTTTATGTGGTCCCCACGAGGGTTACACAGATGTT
ATTGTCAACCAATCAAATTGCATTCTCAAACGTTAGATAAGTGTTCATTTGTCTTTATATA
CTTGGTCCCCAAGTTTTTTGTCTTGCAATATGTGGGACCCACTTCTTAATGAATTTCCTG
AATCTGTTCACGGATTTCGTTGTATGTTAGCTATTAAATATTTGCAGTCCGTTGAGGAAA
CTTACGAGCCCAATACATTGGGCCACGATTTAATTAGGGATCTTATATCTGTTGTAAGG
GCCCGTGACTATGTCGAAGCGACCAGGCGATATAATCATTTCCACGCCCGTCTCGAAG
GTTCGCCGAAGGCTGAACTTCGACAGCCCATACAGGCCCATGTACCGAAAGCCCAGAA
TATACAGAATGTATCGAAGCCCTGATGTTCCCCGTGGATGTGAAGGCCCATGTAAAGTC
CAGTCTTATGAGCAACGGGATGATATTAAGCATACTGGTATTGTTCGTTGTGTTAGTGAT
GTTACTCGTGGATCTGGAATTACTCACAGAGTGGGTAAGAGGTTCTGTGTTAAATCGAT
ATATTTTTTAGGTAAAGTCTGGATGGATGAAAATATCAAGAAGCAGAATCACACTAATCA
GGTCATGTTCTTCTTGGTCCGTGATAGAAGGCCCTATGGAAACAGCCCAATGGATTTTG
GACAGGTTTTTAATATGTTCGATAATGAGCCCAGTACCGCAACCGTGAAGAATGATTTG
CGTGATAGGTTTCAAGTGATGAGGAAATTTCATGCTACAGTTATTGGTGGGCCCTCTGG
AATGAAGGAACAGGCATTAGTTAAGAGATTTTTTAAAATTAACAGTCATGTAACTTTATTT
ATATTCATTCAGGAGGCAGCAAAGTACGAGAACCATACTGAAAACGCCTTGTTATTGTA
TATGGCATGTACGCATGCCTCTAATCCAGTGTATGCAACTATGAAAATACGCATCTATTT
CTATGATTCAATATCAAATTAATAAAATTTATATTTTATATCATGAGTTTCTGTTACATTTA
TTGTGTTTTCAAGTACATCATACAATACATGATCAACTGCTCTGATTACATTGTTAATGGA
AATTACACCAAGACTATCTAAATACTTAAGAACTTCATATCTAAATACTCTTAAGAAATGA
CCAGTCTGAGGCTGTAATGTCGTCCAAATTCGGAAGTCGAGAAAACATTTGTGAATCCC
CATTACCTTCCTGATGTTGTGGTTGAATCTTATCTGAATGGAAATGATGTCGTGGTTCAT

TAGAAATGGCCTCTGGCTGTGTTCTGTTATCTTGAAATAGAGGGGATTGTTTATCTCCCA
GATAAAAACGCCATTCTCTGCCTGAGGAGCAGTGATGAGTTCCCCTGTGCGTGAATCC
ATGATTATTGCAGTTGAGGTGGAGGTAGTATGAGCAGCCACAGTCTAGGTCTACACGC
TTACGCCTTATTGGTTTCTTCTTGGCTATCTTGTGTTGGACCTTGATTGATACTTGCGAA
CAGTGGCTCGTAGAGGGTGACGAAGGTTGCATTCTTGAGAGCCCAATTTTTCAAGGAT
ATGTTTTTTTCTTCGTCTAGATATTCCCTATATGAGGAGGTAGGTCCTGGATTGCAGAGG
AAGATAGTGGGAATTCCCCCTTTAATTTGAATGGGCTTCCCGTACTTTGTGTTGCTTTGC
CAGTCCCTCTGGGCCCCCATGAATTCCTTGAAGTGCTTTAAATAATGCGGGTCTACGTC
ATCAATGACGTTGTACCACGCATCATTACTGTACACCTTTGGGCTTAGGTCTAGATGTC
CACATAAATAATTATGTGGGCCTAGAGACCTGGCCCACATTGTTTTGCCTGTTCTGCTAT
CACCCTCAATGACAATACTTATGGGTCTCCATGGCCGCGCAGCGGAATACACGACGTT
CTCGGCGACCCACTCTTCAAGTTCATCTGGAACTTGATTAAAAGAAGAAGAAAGAAATG
GAGAAACATAAACTTCTAAAGGAGGACTAAAAATCCTATCTAAATTTGAACTTAAATTAT
GAAATTGTAAAATATAGTCCTTTGGGGCCTTCTCTTTTAATATATTGAGGGCCTCGGATT
TACTGCCTGAATTGAGTGCTTCGGCATATGCGTCGTTGGCAGATTGCTGACCTCCTCTA
GCTGATCTGCCATCGATTTGGGAAACTCCAAAATCAATGAAGTTTCCGTCTTTCTCCAC
GTAGGTCTTGACATCTGTTGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGA
TCCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGG
GGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATC
ATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACG
AGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA
ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATT
AATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTT
CCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTC
ACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACAT
GTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTT
TTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGG
TGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCG
TGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCG
GGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCG
TTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT
ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCT
TGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCT
GCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACC
ACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAG
GATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAAC
TCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTA
AATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTT
ACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAG
TTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCC
CAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATA
AACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCC
ATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTT
GCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATG
GCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTG
CAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCA
GTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTA
AGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG
GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGA
ACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTT
ACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCAT
CTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA
AAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTAT

TGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAA
AATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

Figure S3.1 Full sequence of the IL-60-BS-2A viral vector synthesised and sequence confirmed by Genescript (GeneScript Biotech Corporation, New Jersey, USA): F1 bacteriophage origin of replication, LacZgene, T2A G*P site, F2A site, C1 rep, Intergenic repeat (IR), ORF V2 ORF V1 coat protein, ORF C3, , Rep C1, RepC4, Multiple cloning site, pUC plasmid origin of replication, Ampicillin, Ampicillin gene promoter.

ATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGC
GAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCA
CCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGC
TGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGG
CGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTC
ACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAA
TTGGAGCTCTGGACCAGGGTTCTCTTCAACATCTCCACATGTCAAAAGAGAACCTCTTC
CCTCTCCAGATCCAGAAGTCCTACAGGAACAGGTGGTGGCGGCCCTCGGTGCGCTCG
TACTGCTCCACGATGGTGTAGTCCTCGTTGTGGGAGGTGATGTCCAGCTTGGAGTCCA
CGTAGTAGTAGCCGGGCAGCTGCACGGGCTTCTTGGCCATGTAGATGGACTTGAACTC
CACCAGGTAGTGGCCGCCGTCCTTCAGCTTCAGGGCCTTGTGGATCTCGCCCTTCAGC
ACGCCGTCGCGGGGGTACAGGCGCTCGGTGGAGGCCTCCCAGCCCATGGTCTTCTTC
TGCATTACGGGGCCGTCGGAGGGGAAGTTCACGCCGATGAACTTCACCTTGTAGATGA
AGCAGCCGTCCTGCAGGGAGGAGTCCTGGGTCACGGTCACCACGCCGCCGTCCTCGA
AGTTCATCACGCGCTCCCACTTGAAGCCCTCGGGGAAGGACAGCTTCTTGTAGTCGGG
GATGTCGGCGGGGTGCTTCACGTACACCTTGGAGCCGTACTGGAACTGGGGGGACAG
GATGTCCCAGGCGAAGGGCAGGGGGCCGCCCTTGGTCACCTTCAGCTTCACGGTGTT
GTGGCCCTCGTAGGGGCGGCCCTCGCCCTCGCCCTCGATCTCGAACTCGTGGCCGTT
CACGGTGCCCTCCATGCGCACCTTGAAGCGCATGAACTCCTTGATGACGTTCTTGGAG
GAGCGCGCCATTGGACCAGGGTTAGA

Figure S3.2 DsRed gene sequence confirmation by Sanger sequencing: partial sequence of the F1 origin of replication (yellow), LacZ gene, T2A G*P site, DsRed1.

ATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGC
GAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCA
CCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGC
TGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGG
CGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTC
ACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAA
TTGGAGCTCTGGACCAGGGTTCTCTTCAACATCTCCACATGTCAAAAGAGAACCTCTTC
CCTCTCCAGATCCAGAAGGTGGTGATTTTGTGCCGAGCTGCCGGTCGGGGAGCTGTTG
GCTGGCTGGTGGCAGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAACTTAATA
ACACATTGCGGACGTTTTTAATGTACTGAATTAACGCCGAATTGCTCTAGCCAATACGC
AAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTC
CCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTA
GGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCG
GATAACAATTTCACACAGGAAACAGCTATGACATGATTACGAATTCAAAAATTACGGATA
TGAATATAGGCATATCCGTATCCGAATTATCCGTTTGACAGCTAGCAACGATTGTACAAT
TGCTTCTTTAAAAAAGGAAGAAAGAAAGAAAGAAAAGAATCAACATCAGCGTTAACAAA
CGGCCCCGTTACGGCCCAAACGGTCATATAGAGTAACGGCGTTAAGCGTTGAAAGACT
CCTATCGAAATACGTAACCGCAAACGTGTCATAGTCAGATCCCCTCTTCCTTCACCGCC
TCAAACAC

Figure S3.3 CRISPR/Cas9 component cloning confirmation into IL-60-BS-2A-Cas9-1185 by Sanger sequencing: partial sequence of the F1 origin of replication (yellow), LacZ gene, T2A G*P site and Parsley Ubiquitin promoter.

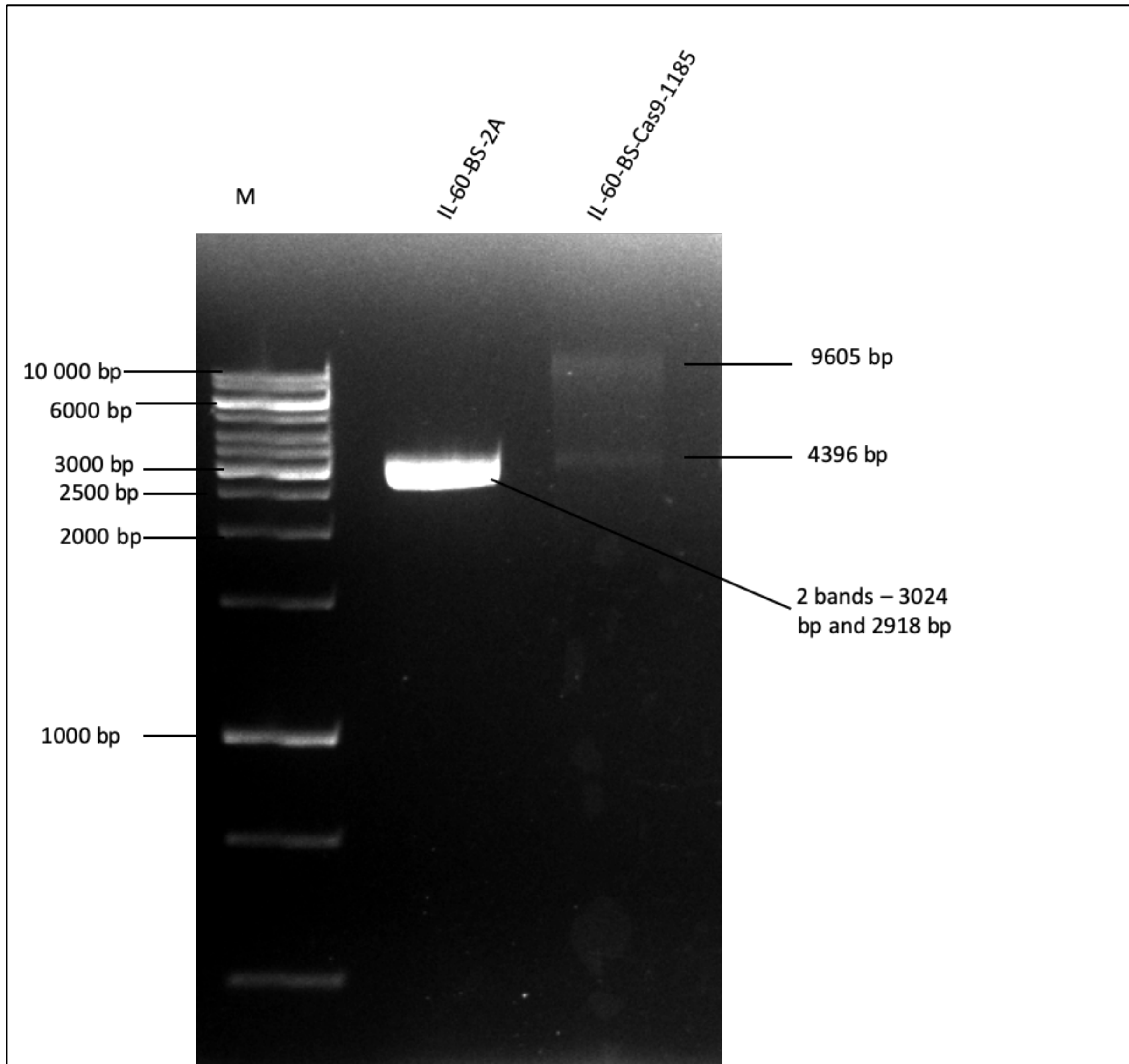


Figure S3.4 Restriction digestion using the XmaI enzyme for confirmation of successful vector assembly by Infusion® cloning ran on 1.5% agarose gel and visualized using the Biorad Gel Doc^{TM} XR+ documentation system (Biorad laboratories, California, USA). M - GeneRuler 1kb molecular marker, The IL-60-2A vector without the insert has an XmaI digestion pattern of two fragments of 3024 bp and 2918, while the IL-60-2A-Cas9-1185 should give two fragments of 9605 bp and 4396 bp. All expected digestion patterns were observed for each plasmid confirming successful cloning/assembly.

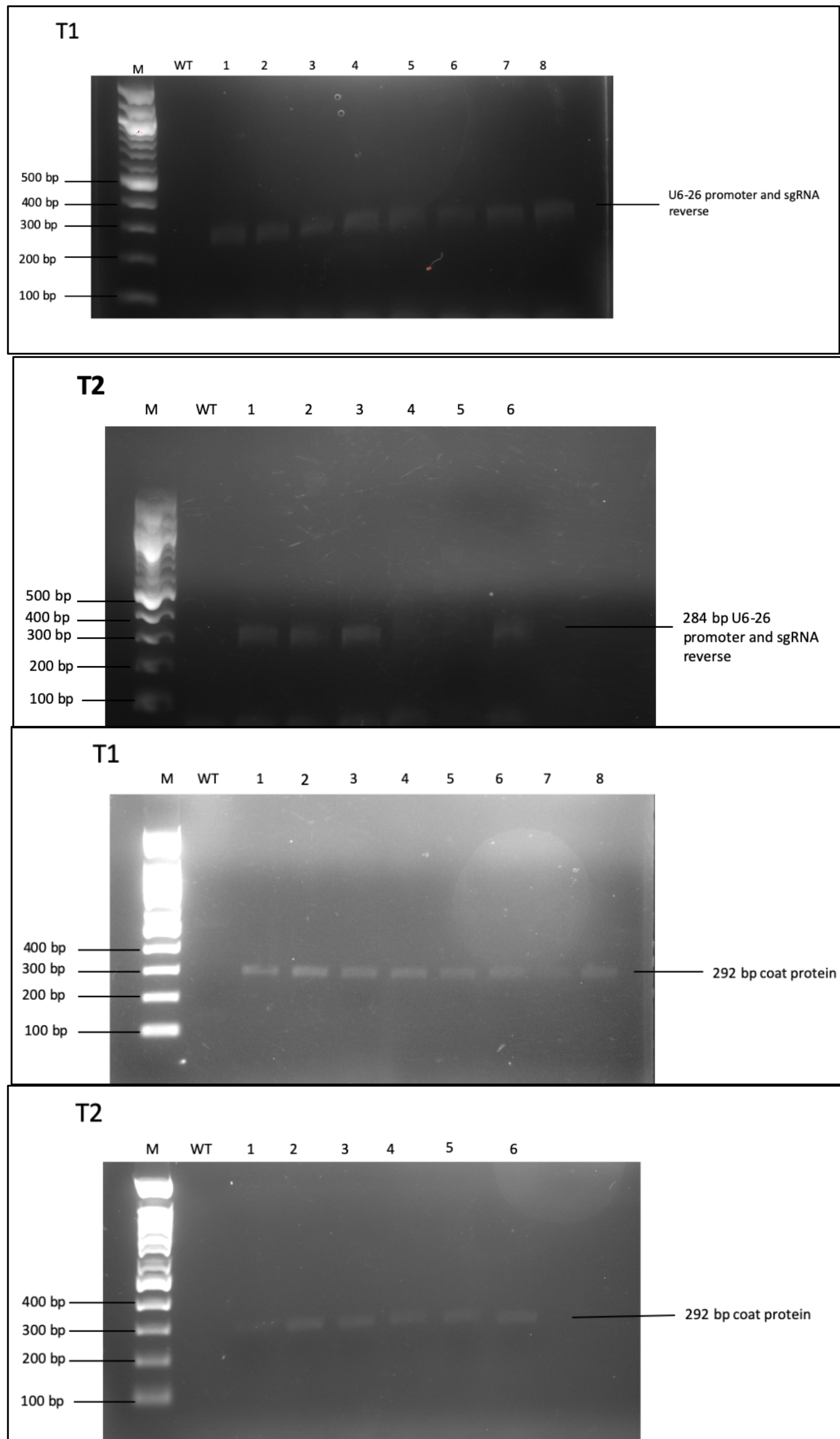


Figure S3.5 PCR confirming the presence of the IL-60-BS-1185 vector by amplification of the viral coat protein from T0 plants inoculated with the vector**.** T1 plants obtained from self-pollinating inoculated T0 plants and T2 plants obtained from crossing T1 plants with wildtype uninoculated plants.

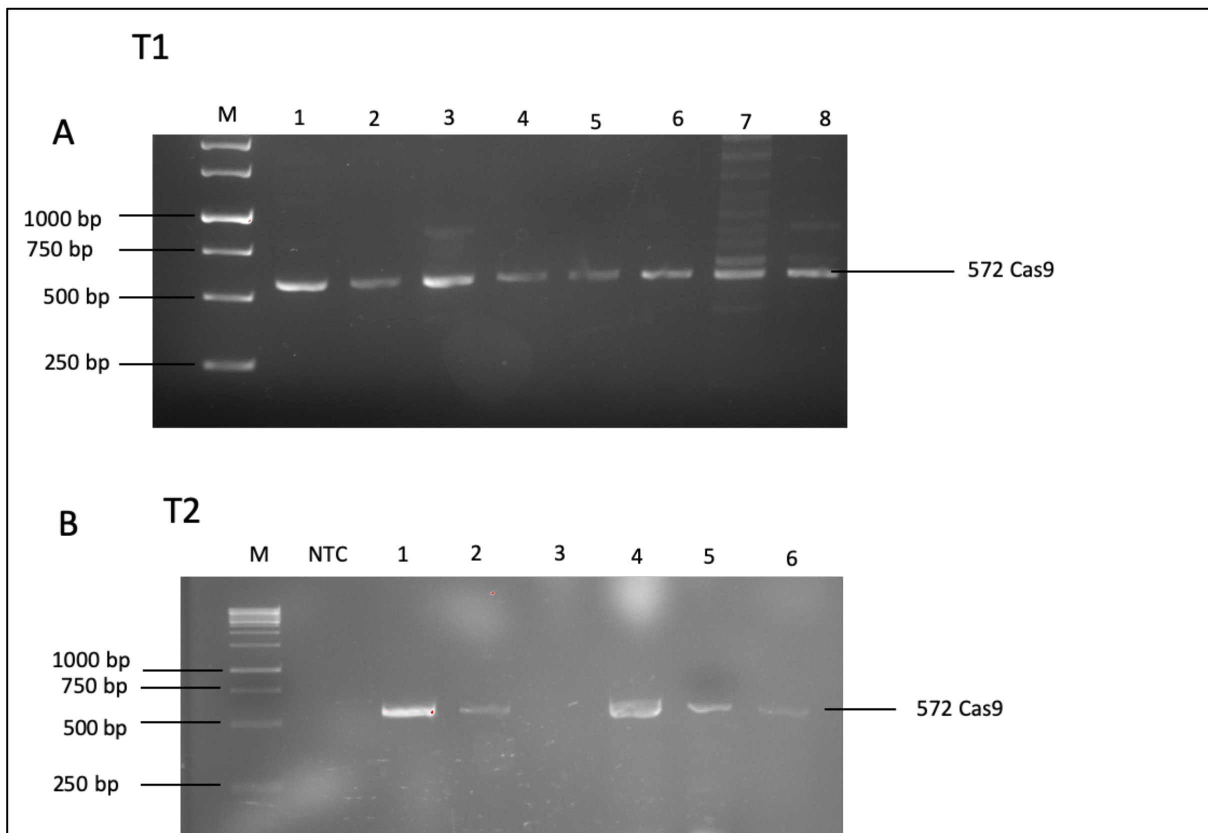


Figure S3.6 RT-PCR confirming Cas9 transcript presence in sunflower T1 and T2 generation plants. M – GeneRuler 1kb DNA marker, WT- Wildtype sunflower, 1-8- cDNA from youngest leaves of T1 plants D. M – GeneRuler 1kb DNA marker, WT- wildtype sunflower, 1-6 – cDNA from youngest leaves of T2 generation plants. All tested plants showed successful Cas9 transcription except for plant number 3 in the T2 generation.

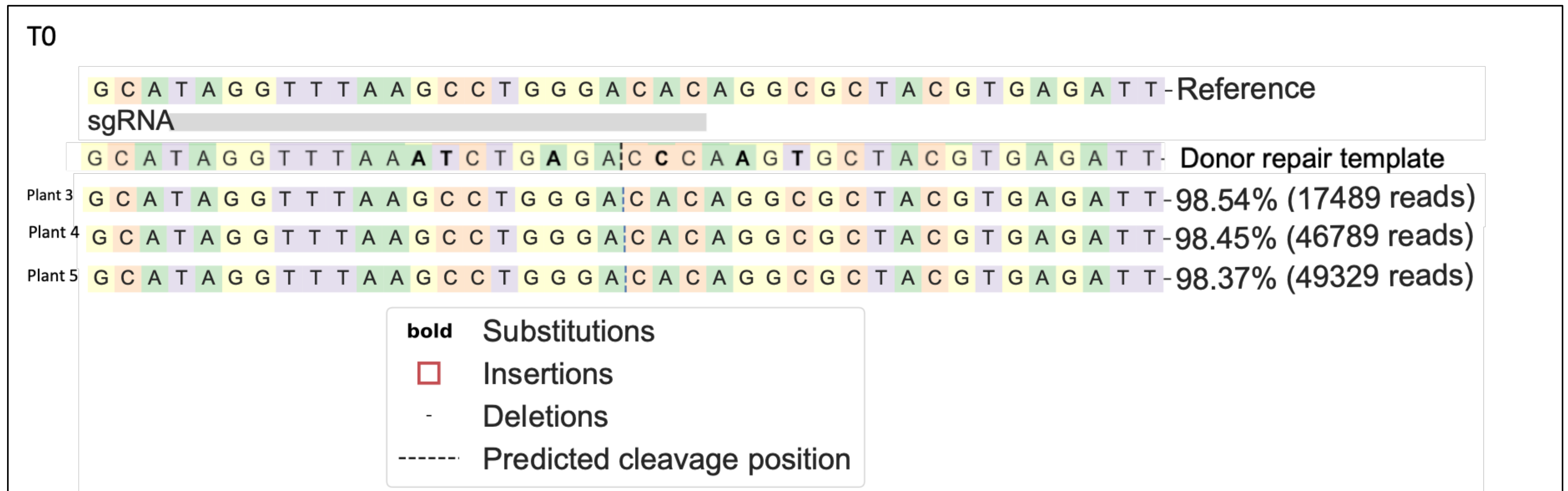


Figure S3.7 CRISPresso2 analysis of sunflower plants inoculated with IL-60-BS-1185 vector for targeted CRISPR mutations. The sunflower *CenH3* reference gene and designed donor repair template were aligned against three of the five T0 plants with successful CRISPR reagent transcription to detect CRISPR induced mutagenesis. No mutation was observed in all analysed plants.

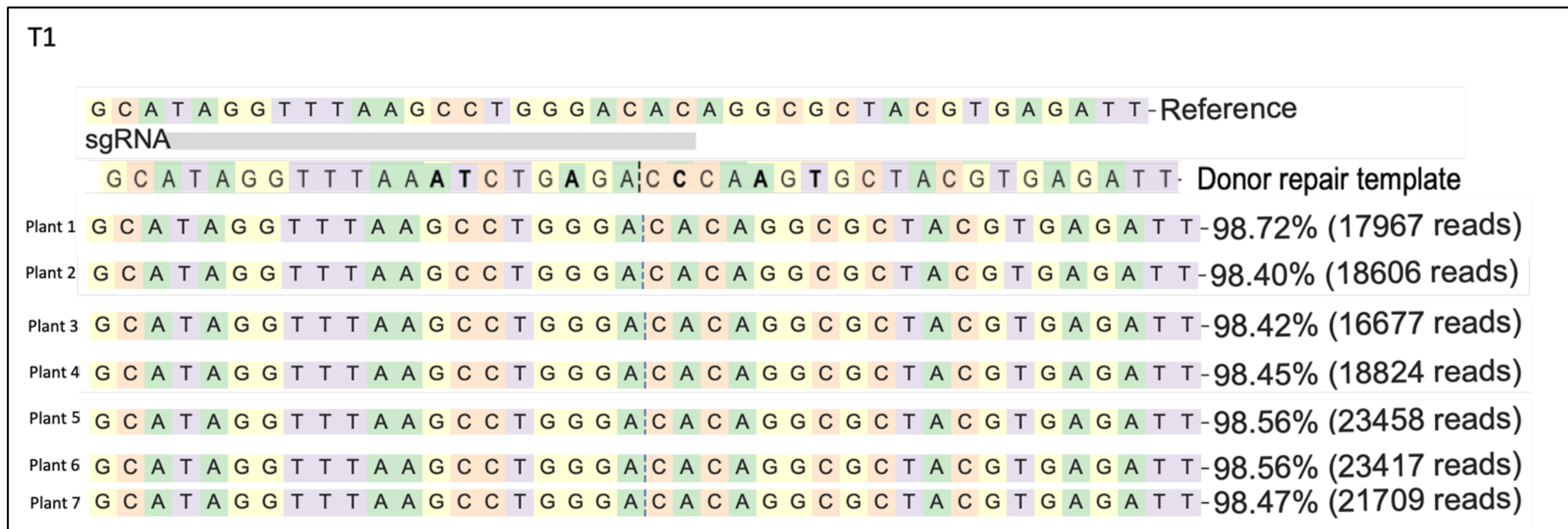


Figure S3.8 Next Generation sequencing and CRISPResso2 analysis of T1 plants obtained from KP328-SEVIH (T0) inoculated with the IL-60-BS-1185 vector. The sunflower *CenH3* reference gene and designed donor repair template were aligned against T1 plants to detect heritable mutagenesis. No mutation was observed in all analysed plants.

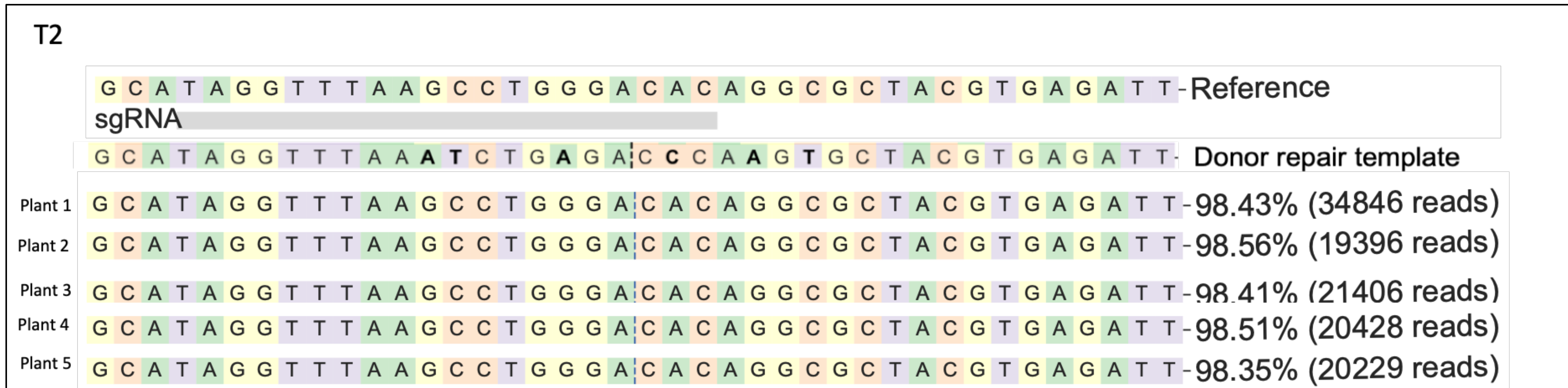


Figure S3.9 Next Generation sequencing and CRISPResso2 analysis of T2 plants obtained from T1 plants. Distribution of alleles around the target site. The sunflower *CenH3* reference gene and designed donor repair template were aligned against T2 plants to detect heritable mutagenesis from the T0 and T1 generations. No mutation was observed in all analysed plants.

CHAPTER 4

HAPLOID PRODUCTION POTENTIAL OF TRANSIENT *CENH3* RECOMBINANT SUNFLOWER PLANTS

4.1 ABSTRACT

The conventional breeding process is extensive and requires multiple generations of backcrossing to achieve an acceptable level of homozygosity (purity). Doubled haploid lines are an asset in the production of pure breeding lines since they are achieved within a single generation, thereby accelerating the breeding process. The centromeric histone 3 (*CenH3*) gene is a key player in facilitating chromosome segregation and centromere localization during cell division. Modification of the CENH3 protein in plant species has been identified to induce uniparental chromosome elimination upon outcrossing with wildtype plants, thereby resulting in a small percentage of haploid progeny. In this study, plants expressing a heterozygous CRISPR/Cas9-modified CENH3 protein were tested for their ability to act as haploid inducer lines in sunflower upon self-fertilization (T0) and subsequent outcrossing of T1 plants with wildtype plants. A high frequency of aborted seeds (seed death) was observed in self-fertilized heteroallelic T1 plants. Flow cytometric analysis revealed (5/7) of the progeny in the T1 generation and (1/8) of the T2 generation to be aneuploids. No haploid progeny were obtained in both generations. Further work is required to elucidate the exact mechanism of this CENH3 based aneuploid production and how to achieve haploid induction in sunflower.

4.2 INTRODUCTION

Homozygosity, typically achieved through back-crossing (inbreeding) and selection, forms a critical part in the development of desired traits in crop breeding (Britt and Kuppu, 2016). The process of achieving and selecting homozygous/pure breeding lines is costly, laborious, lengthy, and tedious (Yoon et al., 2022), thereby severely delaying, usually by years, trait development. Several approaches have been developed to obtain plants carrying a single copy of the parental chromosomes (haploids), which often concludes in spontaneously and/or chemical doubling of chromosomes to form doubled haploids (Kuppu et al., 2020). Doubled haploids are completely homozygous lines containing two identical copies of a single genome and are often produced in a single generation (Kelliher et al., 2019, 2017). Haploids can be induced *in vitro* through culturing plant parts, such as anthers, microspores, and ovaries into embryos or *in vivo* through crosses known as wide hybridization (Dunwell, 2010; Wang et al., 2019). *In vitro* haploid induction methods are expensive, usually cultivar or species specific, and overly labour intensive to use in breeding programs (Kuppu et al., 2020).

Helianthus annuus (sunflower) is one of the main oilseed crops globally (Kaya et al., 2012). Challenges in genetic manipulation and susceptibility to a variety of plant pathogens, have however, compromised its productivity (Vear et al., 2016). Several haploid induction approaches have been evaluated in sunflower, including anther, pollen and the culturing of unpollinated ovules and ovaries, but with little success (Coumans and Zhong 1995). Irradiated pollen-induced parthenogenesis was effectively tested and produced doubled haploids for four different genotypes but with low successful rates of around 10% (Todorova et al., 1997). This laboratory approach could also not be reproduced in industry. Therefore, to date, sunflower does not have a reliable haploid induction method. The lack of a reliable haploid induction method that could efficiently be applied in breeding programs for the enhancement of traits in sunflower has been a major limiting factor in sunflower improvement programs.

The *in vivo* production of haploids through direct manipulation of the centromeric histone 3 (*CenH3*) gene was shown to be a promising alternative in a few plants, including *Arabidopsis thaliana* (Ravi and Chan 2010; Britt and Kuppu 2016, Kuppu, 2020), and more recently wheat (Lv et al., 2020) and maize (Wang et al., 2021). The centromere is a crucial chromosomal locus that ensures the successful transfer of genetic information from parent to daughter cells (Comai et al., 2017). The CENH3 protein is responsible for specific centromere binding and mediates chromosome segregation (Dawe, 2020). It has been shown that certain modifications of the CENH3 can lead to haploid induction, however many have proven to be

detrimental, often leading to chromosome segregation errors and lethality (Britt and Kuppu, 2016).

The CENH3 protein comprises of a highly variable N-terminal tail, responsible for the recruitment and stabilization of centromeric proteins and a highly conserved C-terminal histone fold (Lv et al., 2020). In one of the first CENH3 studies, Ravi and Chan (2010) showed that crossing a wild-type *Arabidopsis thaliana* plant with a genetically modified line where the N-terminal tail domain was swapped with a H3 variant and tagged with the green fluorescent protein (GFP), produced haploid progeny at a rate as high as 25-44%. Other studies have demonstrated the manipulation of the histone fold domain region of the protein through non-targeted single amino acid changes and knockouts could also lead to haploid induction at sufficiently high frequencies (Britt and Kuppu, 2016; Kuppu, 2020). Because the structure, sequence, and function of the CENH3 protein is conserved in most angiosperms (Comai et al., 2017; Kuppu et al., 2020), this technique promised a wide applicability in a range of plant species.

Initially it was proposed that structural differences in the CENH3 protein decreased its activity, resulting in chromosome elimination of the altered centromeres through competition during cell division (Ravi and Chan et al., 2010; Ishii et al., 2016; Karimi-Ashtiyani et al., 2015). Subsequently, centromere size differences between haploid inducer lines and their wildtype counterparts have been highlighted as playing a role in uniparental genome elimination (Wang et al., 2021). The mutant *Cenh3* plants are believed to transfer a smaller, or weaker, CENH3 protein to the progeny and is therefore eliminated through natural clearing mechanisms (Wang et al., 2021). As superior as it appears, this technology had unfortunately not had much success in haploid induction outside of *Arabidopsis thaliana*, until recent positive results by Lv et al. (2020) and Wang et al. (2021) in wheat and maize respectively. It was discovered that crossing lines expressing heterozygous versions of the *CenH3* gene as female parents increased haploid induction to 8% and 5%, respectively (Lv et al., 2021; Wang et al., 2021). Through this approach, it was proposed that *CenH3* dilution to critically low levels during cell division could potentially be another underlying mechanism in haploid induction technology (Wang et al., 2021). Details on the exact mechanism of centromere mediated haploid induction is therefore still unclear and studies to uncover the exact mechanism of CENH3-modification based haploid initiation are still few.

Sunflower has a base chromosome number of 17 and a single gene encoding the CENH3 protein (Nagaki et al., 2015). In preceding chapters of this study, amino acids in the sunflower CENH3 protein identified as possible haploid inducers were partially modified resulting in

chimeric/hetero alleles using the CRISPR/Cas9 technology (Chapter 3). In this chapter, the ability of sunflower plants, carrying a 'heterozygous' CRISPR/Cas9 modified version of the *CenH3* gene (from previous chapters), to act as haploid inducers, is assessed.

4.3 MATERIALS AND METHODS

4.3.1 SIFT PREDICTIONS

Two sunflower plants (KP328) with various mutant alleles in the histone fold domain region of the CENH3 protein were obtained from CRISPR/Cas9 gene editing in sunflower in (Chapter 3). The ability to tolerate the mutant alleles obtained in this study (Chapter 3) were predicted using the SIFT (Sorting intolerant from tolerant, Ng and Henikoff, 2003) (https://sift.bii.a-star.edu.sg/) program. The seven amino acid changes obtained from Chapter 3 in the CENH3 protein (P51S, G52E, A55V, L56I, R57H, A55T, L56H) present in two different lines were tested for lethality using the SIFT database.

4.3.2 PLANT MATERIAL AND MORPHOLOGICAL ANALYSIS

Two independent T0 sunflower plants (KP 328) (KP328-SEVIH – P51S, G52E, A55V – Single allele, L56I, R57H – different alleles) (KP328-S-VTH – L56H, A55T, STOP codon – All in different alleles) (Figure 3.11) expressing heterozygous *CenH3* genes obtained through CRISPR/Cas9 modification, were self-pollinated to produce T1 plants. T1 seeds (50 seeds per plant) were germinated and grown in 20 cm pots in sterile perlite growth mixture (Chemserve Perlite, South Africa). Plant growth conditions were set at 30°C/20°C (±2°C), with 16:8h day/night cycling in the plant containment and growth facility (BSL-2 Registration no. 39.2/ ARC Biotechnology Platform – 21/0142) of the Agricultural Research Council's Biotechnology Platform (Pretoria, South Africa). Maize and wheat heterozygous plants crossed as female resulted in high numbers of haploid progeny (Wang et al., 2021; Lv et al., 2020), therefore selected sunflower T1 plants were allowed to reach maturity, randomly selected, and crossed against wildtype sunflower (KP 328) as female plants (pollen recipients). The pollen was collected and used immediately to pollinate the whole head of recipient plants, plants were covered with brown paper bags to prevent cross contamination. The resulting T2 seeds (20 seeds) were grown under the conditions described above.

4.3.3 PLOIDY EVALUATION USING FLOW CYTOMETRY

The transient nature of viral delivery systems and the nature of haploid induction through CENH3 modification proved to be difficult for haploid testing. The ploidy status for both T1 and T2 generations was therefore tested. In this study, the per cell DNA content was determined

using nuclei extracted by finely chopping fresh young sunflower leaf material (approximately 500 mg) in 1 mL of Otto I nuclei extraction buffer (0.1M citric acid, 0.5% v/v Tween 20), (Otto, 1992).

The extracted nuclei were stained with 0.8 mL of 3:7 Otto I and Otto II (0.4M Na_2PO4) supplemented with 20 μL of 6-diamidino-2-phenylindole (DAPI) (0.1 mg/mL) and 20 μL of ribonuclease A (RNase A) and incubated in the dark for 15 mins. Approximately 0.5 - 1 mL of stained nuclei was loaded onto a 5 mL falcon® polystyrene round bottom tube with a cell strainer cap (Corning Inc, New York, USA) and were analyzed on the BD Accuri C6 flow cytometer (BD Biosciences, New Jersey, USA) at the flow cytometry laboratory at the University of Pretoria, South Africa. Fluorescence was collected through a filter. Evaluation of integral peak, log fluorescence and forward scatter was performed at the FL1 10 000, FL2 1000. At least 10 000 events were collected per sample. The samples were maintained on ice throughout the processing period. Nuclei were gated in a side scatter/forward scatter plot (FSC-A vs SSC-A). Data analysis was conducted on the FlowJo Version v.10 software (BD Biosciences, New Jersey, and USA). Young leaves of wildtype sunflower (KP328) were used as the standard diploid control to determine the G0/G1 peak. DNA content based on 10 000 nuclei per sample and two independent replicates was calculated using the following formula: Sample 1C DNA content = [(Sample G_1 peak mean/ standard G1 peak mean)] x Standard 1C DNA (pg DNA) (Dolezel and Bartos, 2005, Martin et al., 2019).

4.3.4 Chromosome analysis (microscopy)

Roots tips from sunflower plants were harvested, rinsed in tap water and fixed in Carnoy's fluid (3:1 ethanol: acetic acid) solution for 18 hours. Thereafter, the roots were rinsed with 70% ethanol 3-4 times and kept in 70% ethanol at 4°C until analysed. On the day of analysis, the root tips immersed in 1 mol/L HCl at 60°C for 8 – 10 min (Fu et al., 2018). Chromosomes were then counterstained with a 0.1 μg/mL 6-diamidino-2-phenylindole (DAPI) for approximately 10 mins in the dark and washed in phosphate buffered saline (PBS) buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) 3-5 times before visualization and imaging using Zeiss Axio imager.MI florescence microscope (Intelligent Imaging Innovations, Denver, CO, USA) at the University of Pretoria, Microscopy unit. FIJI (ImageJ, National institutes of Health) software was used to analyse the data.

4.4 RESULTS

4.4.1 SIFT PREDICTIONS

The SIFT (Sorting intolerant from tolerant, Ng and Henikoff, 2003) (https://sift.bii.a-star.edu.sg/) program was used to predict the severity of the mutant alleles obtained in this study (Chapter 3). The SIFT program considers the tertiary structure, sequence, and annotation from a variety of species to calculate the level of tolerance. SIFT values range from 1 to zero, where '0' is indicative of not tolerated/deleterious amino acid changes and one is tolerated. Kuppu et al (2020) correlated SIFT scores to determine which amino acid mutations could induce haploids. Haploid inducers with a haploid induction frequency of 5% and greater were found to have a SIFT score of 0.05% or less. The two recombinant plants obtained from Chapter 3 contained a total of seven different amino acid changes in the CENH3 protein. KP328-SEVIH (T0) contained 2.84% alleles with L56I, 2,67% alleles R57H, and 1,80% of alleles with P51S, G52E, A55V, while KP328-S-VTH contained 0,74% alleles with A55V, 0,61% alleles with A55T, 0,56% alleles with P51S, 0,55% alleles L56H and 0,28% alleles with a premature stop codon in amino acid position 54 (Figure 3.12). SIFT scores for the seven different mutant alleles resulting in amino acid changes obtained in Chapter 3 were calculated (Table 4.1). All seven amino acid changes had a SIFT score of less than 0.05% (Table 4.1) and are therefore considered deleterious. SIFT probabilities for CENH3 amino acids 51 – 75 and tolerated and non-tolerated amino acid changes from CENH3 amino acid 1 -100 are listed in Figure S4.1 and Figure S4.2, respectively.

Table 4.1 SIFT scores for amino acid substitutions introduced by CRISPR/Cas9 genome editing and predicted haploid inducer frequency in sunflower.

Amino acid change	SIFT score	Predicted haploid inducer frequency
P51S	0.00	> 5%
G52E	0.00	> 5%
A55V	0.01	> 5%
A55T	0.03	> 5%
L56I	0.04	> 5%
L56H	0.00	> 5%
R57H	0.03	> 5%

4.4.2 MORPHOLOGICAL CHARACTERIZATION

Two individual T0 CENH3 recombinant transient plants obtained in Chapter 3 were tested as possible haploid inducer lines. The plants were self-fertilized and the resulting seeds were grown and plants were outcrossed as females with wildtype sunflower lines. Seeds from KP328-S-VTH (T0) failed to germinate while only 10/50 seeds from KP328-SEVIH (T0) germinated. Plants were visually observed for any morphological characteristics and all plants appeared normal (Figure 4.1) and germinated at the same time. Seed from Plant T2G, however, only germinated approximately 3 weeks after the other seeds and plants appeared to have a smaller stature (Figure 4.1). Plant T2G also exhibited an extended cotyledon during the early developmental stages in comparison to its wildtype counterparts (Figure 4.1), the plant however showed normal growth after 2 weeks of germination and was fully fertile once mature.

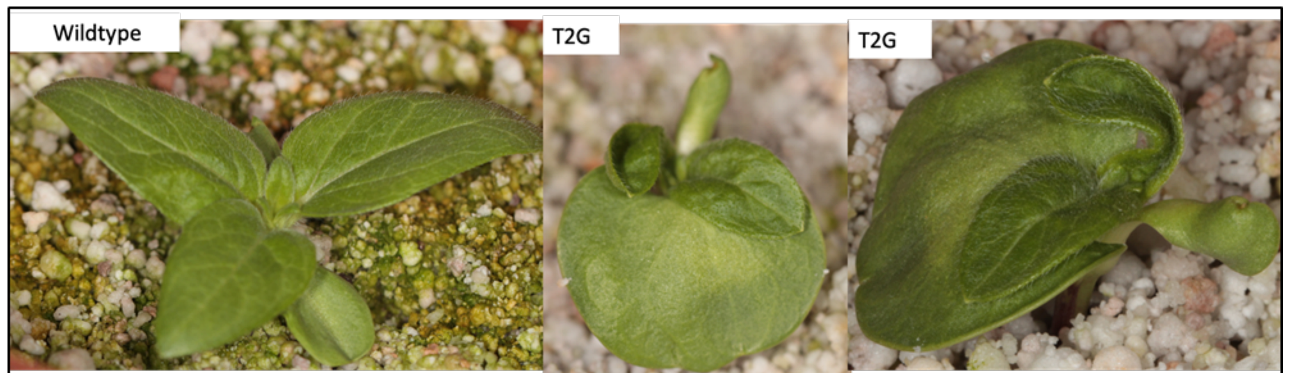


Figure 4.1 Comparison of normal wildtype sunflower line with T2G line obtained from outcrossing T1 generation lines with wildtype sunflower lines. The T2G plant delayed germination and showed an elongated abnormal cotyledon during early developmental stages when compared to wildtype sunflower plants.

4.4.3 FLOW CYTOMETRIC AND MICROSCOPIC PLOIDY ANALYSIS

Ploidy status of sunflower T1 and T2 lines obtained from self-fertilization and outcrossing was determined by using flow cytometry analysis of fresh, young sunflower leaves. Wildtype sunflower (KP 328) was used as a standard for determining peak positioning of the flow cytometric histograms. The first signal peak position at 10^3 -10^4 was considered diploid. Ploidy status was further validated by calculating sample DNA content using the following formula (Dolezel and Bartos, 2005, Martin et al., 2019):

$$\text{Sample 1C DNA content} = \frac{\text{Sample G1 peak mean}}{\text{Standard G1 peak mean}} \times \text{Standard 1C DNA (pg)}$$

Diploid sunflower plants are known to have a 1C value varying between 2.3 and 4.9 with a mean of 3.67 pg. Plants T1A, T1J, T1KA and T1KB had peak positioning in correspondence with the sunflower wildtype control (Figure 4.2) and were considered diploid, calculations however revealed slightly elevated 1C DNA content and suggested that these could be aneuploids (Table 4.2). Plants T1H and T2G contained two different histogram peaks, with one positioned away from control standard indicating the presence of two cell populations with different DNA content in the sample (Figure 4.2). 1C calculations further revealed both elevated and reduced 1C levels for plants T2G and T1H, respectively suggesting aneuploidy in both plants (Table 4.2). All the other samples showed normal histogram positioning comparable to the control (Figure S4.3) and 1C calculations within 3.3 and 5.5 (Table 4.2). Microscopy was performed in an attempt to confirm the ploidy status and all analyzed samples showed the diploid sunflower chromosome number 2n = 34 (Figure 4.3).

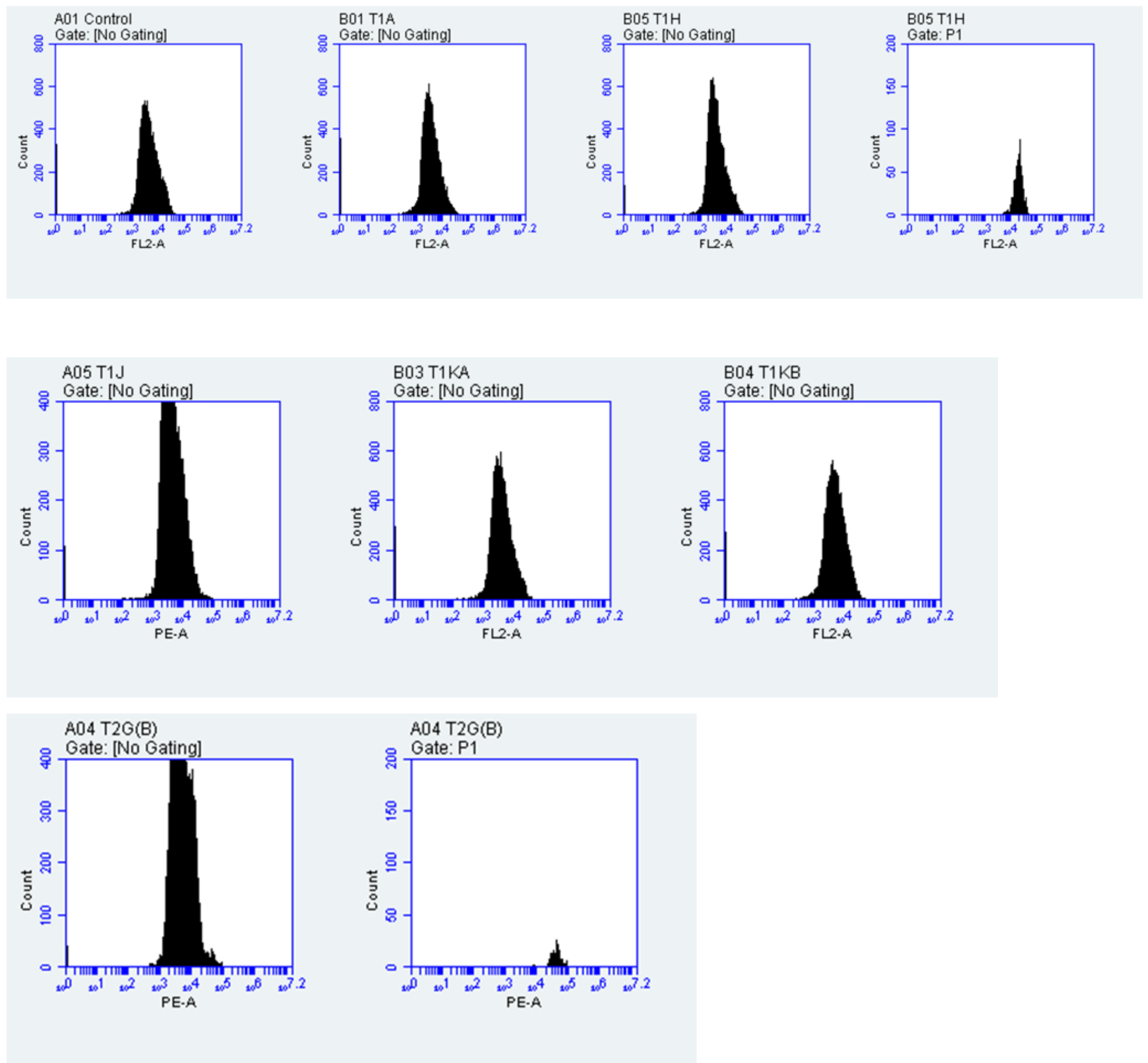


Figure 4.2 Histograms obtained from flow cytometric ploidy analysis showing relative fluorescence intensities of sunflower wildtype (control) and T1 and T2 plants generated from self-fertilization and outcrossing with wildtype plants and stained with 6-diamidino-2-phenylindole (DAPI. The Y-axis represent the number of nuclei while the x-axis represents relative fluorescence intensity.

Table 4.2 Flow cytometric DNA content calculation on sunflower young leaves and predicted ploidy status.

Parent	Sample	Calculated 1C value (pg)	Predicted ploidy status
	Wildtype	3.6	Diploid
T0 (KP328-SEVIH)	T1A	5,9	Aneuploid
T0 (KP328-SEVIH)	T1B	3.6	Diploid
T0 (KP328-SEVIH)	T1G	4,3	Diploid
T0 (KP328-SEVIH)	T1J	7,9	Aneuploid/Euploid
T0 (KP328-SEVIH)	T1H	1,65	Aneuploid
T0 (KP328-SEVIH)	T1KA	9,48	Aneuploid/Euploid
T0 (KP328-SEVIH)	T1KB	7,2	Aneuploid/Euploid
T1	T2A	5,5	Diploid
T1	T2B	3,3	Diploid
T1	T2C	4,7	Diploid
T1	T2D	3,9	Diploid
T1	T2F	4.8	Diploid
T1	T2H	4,0	Diploid
T1	T2G	10,1	Aneuploid/Euploid

Formula used: Sample 1C DNA content = (Sample G_1 peak mean/ standard G1 peak mean) x Standard 1C DNA (pg DNA)

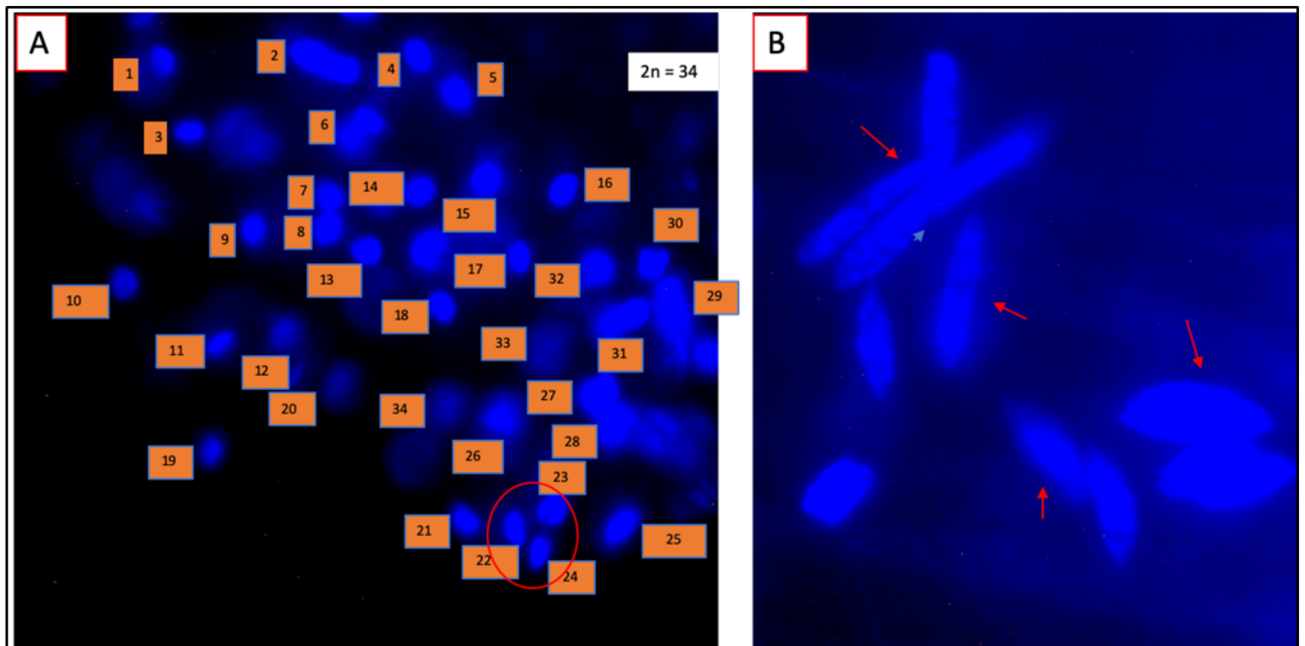


Figure 4.3 Chromosome spreads of sunflower root tips stained with 6-diamidino-2-phenylindole (DAPI, 0.1 μg/mL). and visualized under fluorescence using the Zeiss Axio image MI fluorescence microscope A) Chromosome spreads showing a 2n = 34 (diploid) chromosome number (scale bar 10 μm) and B) Enlarged image (scale bar 2 μm) of enclosed region highlighted in red in image A showing full length chromosomes. The red arrows indicate full length chromosomes.

4.5 DISCUSSION

Doubled haploid technology is essential for accelerating the development of true breeding lines in a more efficient and fast-tracked manner as opposed to conventional breeding techniques (Sharma et al., 2022). Furthermore, the identification of the centromeric histone 3 (*CenH3*) gene as a key player in *in vivo* haploid induction technology promised to minimize the limitations that arise from *in vitro* based methods, such as genotype specificity (Sidhu et al., 2022). Modification of the CENH3 protein through the expression of a non-native protein, amino acid substitutions and knock-out of gene regions, have all been established to induce some haploid, aneuploid and diploid progeny when modified plants are outcrossed with wildtype plants (Kuppu et al., 2020). The CENH3 protein is responsible for kinetochore assembly and chromosome segregation in plants (Lermontova et al., 2011). To determine the potential of sunflower recombinant lines to act as haploid inducers, plants containing a modified heteroallelic *CenH3* gene were self-fertilized to produce T1 plants and the progeny obtained were outcrossed with wildtype plants to produce T2 plants. Plants were phenotypically observed for differences to wildtype plants and flow cytometry was used to analyze ploidy levels through DNA content estimation. Upon evaluation, a high seed mortality rate was observed for the T1 plants and only diploid and putative aneuploids were obtained in both the T1 and T2 generations.

The CENH3 is a crucial developmental and reproductive protein in plants, it is also conserved across a wide range of species and single amino acid changes severely affect function (Kuppu et al., 2020). The success of the CENH3-based haploid induction method is dependent on *CenH3* sequence modifications that cause a significant disruption on CENH3 function (Kuppu et al., 2015). Amino acid changes that result in haploid inducer lines with a frequency of 5% and above were predicted to have SIFT (Sorting intolerant from tolerant) probability of 0.05% or less (Kuppu et al., 2015, Kuppu et al., 2020). The SIFT software was used to determine whether the amino acid changes observed in this study (both targeted and random) could serve as potential haploid inducers. The targeted amino acid changes, proven to be haploid inducers (P51S, G52E and A55V) had the lowest SIFT score while the randomly integrated changes (A55T, L56I, and R57H had a slightly higher SIFT score with the exception of L56H, this could be an indication of the plant repair mechanisms attempting to retain function. All the seven amino acid changes observed in the two T0 plants are, however, predicted to affect CENH3 protein function and therefore could serve as potential haploid inducers.

The CENH3 protein is made up of a highly variable tail domain and an evolutionary conserved histone fold domain (HFD), studies have revealed that small changes in the HFD region affect

function and may potentially result in lethal phenotypes (Kyum et al., 2022). In this study, seeds obtained from self-pollinated T0 KP328-S-VTH failed to germinate, while only 20% of the seeds obtained from KP328-SEVIH were able to germinate successfully. This data correlates with previous studies, where self-pollination of (+/r) *CenH3* heterozygous combination (in wheat) resulted in 46.8% seed set (Lv et al., 2020). Heterozygous *CenH3* has been described to weaken female gamete performance and reduces seed set in self-pollinated plants causing an increased level of seed death (Lv et al., 2020). In *Arabidopsis*, double amino acid substitutions were observed to be better haploid inducers when compared to single amino acid changes, furthermore amino acid changes resulting in high haploid induction frequency were linked to high seed death in *Arabidopsis* (Kuppu et al., 2015, Kuppu et al., 2020). In this study, a single allele in KP328-SEVIH contained triple amino acid changes in the histone fold domain and this could have increased lethality and resulted in higher seed mortality. The plants were, however, infected by powdery mildew fungus during the seed production stages and this could have affected germination of seeds.

Flow cytometry has gained traction as the method of choice for DNA content and ploidy status analysis (Pfosser et al, 1995). Flow cytometry is highly sensitive and can detect minor DNA content deviations (Pfosser et al., 1995). In the present study, flow cytometric analysis was used to measure DNA content of sunflower lines deemed to be potential haploid inducers. Five out of the seven tested T1 plants and 1 out of 8 of the tested T2 plants showed flow cytometric peak positioning and 1C calculations suggesting deviations in the relative fluorescence when compared to wildtype sunflower diploid lines used as controls. Haploid plants generally have lower relative fluorescence when compared to the control (diploid sample) and the histogram peak would be position to the left. Sunflower diploid plants are known to have a 1C DNA content value of between 2.3 – 4.9 pg, and this difference in range is explained as the C-value paradox (Price et al., 1996). In this study, 5/7 T1 plants and 1/8 T2 plants contained elevated DNA content, while 1/7 T1 plants had decreased DNA content. Two of the plants with different DNA content (T1H and T2G) also contained differences in peak positions suggesting a mixture of two populations. The plants with differences in peak positions and DNA content were, therefore, classified as aneuploids and/or euploids. The CENH3-based haploid inducer line technology is often accompanied by a high number of aneuploid progenies triggered by incomplete female genome elimination (Lv et al., 2021, Kuppu et al., 2020, Ravi and Chan, 2010). This is therefore in correspondence with other studies where self-fertilized heteroallelic *CenH3* mutants resulted in aneuploid progeny (Lv et al., 2020, Lermontova et al., 2011, Ravi and Chan, 2010). Conversely, sunflower has been historically reported to contain high variations in DNA content (Price et al., 1996).

Aneusomaty occurring during early mitotic divisions in tissues with a mixture of diploid and aneuploid cells lead to variations in DNA content (Price et al., 1996). Red to far red light, and selfing were found to influence the quantity of nuclear DNA content in sunflower (Price et al., 1996, Cremonini and Cavallini, 1985), the plants used in this study were however grown under similar conditions, however they were inbred lines with a high degree of selfing. The differences in DNA content can therefore be linked to either *CenH3* incomplete uniparental chromosome elimination or inbreeding depression. Although aneuploid characteristics are still being elucidated in plants, aneuploidy was observed to affect plant architecture and phenotype in maize (Makarevitch et al., 2008). Morphological abnormalities were observed in a plant (T2G) in this study which contained the highest (DNA content) indicating aneusomaty. Previous reports indicate that aneusomaty in sunflower is first observed during embryo development and decreases during the course of plant growth disappearing by pre-meiosis (Cremonini and Cavallini, 1986). A similar observation was made in this study where plant T2G showed abnormal morphological characteristics in the early developmental stages, however normal plant growth was observed after early growth stages (+/- 2 weeks after germination). Chromosome counting failed to confirm any differences in chromosome number among the putative aneuploid and diploid plants but requires rigorous testing of a large number of samples to reach a conclusion (Roux et al., 2002). (Roux et al., 2002). Chromosome counting is unable to analyse ploidy in all three histological layers (LI, LII, LIII) as root tip meristems used in chromosome counting only represent LIII (Roux et al., 2002). Root tips are preferable for ploidy evaluation as it is easier to obtain cells undergoing cell division in the root meristem (Kihlman,1971). However, further studies should consider using tissue where the desired phenotype is observed. Furthermore, aneuploidy is only reported for a limited period of time during plant development and this could explain the inability to observe changes in chromosome numbers through chromosome counting in the present study (Cremonini and Cavallini, 1985).

Wheat and maize are currently the only crop plants that have achieved substantial haploid levels (~8% and 5% respectively) through the *CenH3* modification method (Lv et al., 2020; Wang et al., 2021). In both maize and wheat, it was observed that lines heterozygous for the *CenH3* gene resulted in the highest haploid induction rate. This indicates that the presence of the wildtype *CenH3* allele is necessary to achieve high levels of haploid induction (Wang et al., 2021). In this study, sunflower plants expressing a CRISPR/Cas9 modified heterozygous *CenH3* gene (putative haploid inducer lines) were self-fertilized and the seeds from this cross were harvested and grown to produce T1 lines. The T1 lines were outcrossed with wildtype sunflower plants to produce T2 lines. The amino acid changes described by Kuppu et al (2015) and Kuppu et al (2020) in *Arabidopsis*, however, failed to induce haploids in rice (Lv et al.,

2020). In the studies in wheat and maize, an extensive number of plants (702 plants) had to be screened to identify the haploid plants. Unfortunately, due to the high seed mortality rate, such high numbers could not be screened in edited *CenH3* gene sunflower plants.

The CENH3 haploid inducer method has seen the most success in model *A. thaliana* and crop plants such as maize and wheat which are considered model plants. Several crop plants, including tomato, carrot and soybean, have also been evaluated for CENH3-based haploid induction leading to inconclusive results which could be due to the lethal nature of the *CenH3* mutation (Lv et al., 2020). The results obtained in this study are inconclusive as to whether transient recombinant *CenH3* plants can be used as possible haploid inducers in sunflower as variation in chromosome number and DNA content is a common phenomenon in sunflower. Further studies on the exact mechanism of *CenH3* based haploid inducer development are required to better understand and improve *CenH3* based haploid induction. The use of outcrossed starting material for gene editing instead of inbred lines can help eliminate the potential for spontaneous aneuploid occurrence. The SIFT predictions obtained in this study can be used to deduce possible amino acid changes to be used in further studies.

4.6 REFERENCES

Britt, A.B., Kuppu, S (2016) CenH3: An emerging player in haploid induction technology. Frontiers in Plant Science 7, 357. https://doi.org/10.3389/fpls.2016.00357

Comai, L., Maheshwari, S., Marimuthu, M.P. (2017) Plant centromeres. Current Opinions in Plant Biology 36, 158–167.

Coumans, M., Zhong, D (1995) Doubled haploid sunflower (*Helianthus annuus*) plant production by androgenesis: fact or artifact? Part 2. *In vitro* isolated microspore culture. Plant Cell, Tissue and Organ Culture 41, 203-309. https://doi.org/10.1007/BF00045083

Cremonini, R., Cavallini, A (1986) Origin and fate of aneusomaty in sunflower (*Helianthus annuus* L.). Plant Breeding 97, 89-92. https://doi.org/10.1111/j.1439-0523.1986.tb01308.x

Dawe, R.K (2020) Charting the path to fully synthetic plant chromosomes. Experimental Cell Research 111951. https://doi.org/10.1016/j.yexcr.2020.111951

Dolozel, J., Bartos, J (2005) Plant DNA flow cytometry and estimation of nuclear genome size. Annals of Botany 95, 99-110. https://doi.org/doi:10.1093/aob/mci005

Dunwell, J.M., 2010. Haploids in flowering plants: Origins and exploitation. Plant Biotechnology Journal 8, 377-424. https://doi.org/10.1111/j.1467-7652.2009.00498.x

Ishii, T., Karimi-Ashtiyani, R., Houben, A (2016) Haploidization via chromosome elimination: means and mechanisms. Annual Reviews of Plant Biology 67, 421-38.

https://doi.org/10.1146/annurev-arplant-043014-114714

Karimi-Ashtiyani, R., Ishii, T., Niessen, M., Stein, N., Heckmann, S., Gurushidze, M., Banaei-Moghaddam, A.M., Fuchs, J., Schubert, V., Koch, K., Weiss, O., Demidov, D., Schmidt, K., Kumlehn, J., Houben, A (2015) Point mutation impairs centromeric CENH3 loading and induces haploid plants. Proceedings of the National Academy of Sciences of the United States of America 112, 11211-11216. https://doi.org/10.1073/pnas.1504333112

Kaya, Y., Jocic, S., & Miladinovic, D. (2012). Sunflower.Technological Innovations in Major World Oil Crops, 1, 85-129. https://doi.org/10.1007/978-1-4614-0356-2_4

Kelliher, T., Starr, D., Richbourg, L., Chintamanani, S., Delzer, B., Nuccio, M.L., Green, J., Chen, Z., McCuiston, J., Wang, W., Liebler, T., Bullock, P., Martin, B (2017) *MATRILINEAL*, a sperm-specific phospholipase, triggers maize haploid induction. Nature 1-18. https://doi.org/10.1038/nature20827

Kelliher, T., Starr, D., Su, X., Tang, G., Chen, Z., Carter, J., Wittich, P.E., Dong, S., Green, J., Burch, E., McCuiston, J., Gu, W., Sun, Y., Strebe, T., Roberts, J., Bate, N.J., Que, Q (2019) One-step genome editing of elite crop germplasm during haploid induction. Nature Biotechnology 37, 287-292. https://doi.org/10.1038/s41587-019-0038-x

Kihlman, B.A (1971) Root tips for studying the effects of chemicals on chromosomes. In Chemical Mutagens: Principles and Methods for Their Detection. 2, 489-514. Boston, MA: Springer US.

Kuppu, S., Ron, M., Marimuthu, M.P.A., Li, G., Huddleson, A., Siddeek, M.H., Terry, J., Buchner, R., Shabek, N., Comai, L., 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, 2068-2080. https://doi.org/10.1111/pbi.13365

Kyum, M., Kaur, H., Kamboj A., Goyal, L., Bhatia, D (2021). Strategies and prospects of haploid induction in rice (*Oryza sativa*). Plant Breeding 141, 1-11. https://doi.org/10.1111/pbr.12971

Lermontova, I., Koroleva, O., Rutten, T., Fuchs, J., Schubert, V., Moraes, I., Koszegi, D., Schubert, I (2011) Knockdown of *CenH3* in *Arabidopsis* reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. The Plant Journal 68, 40-50. https://doi.org/10.1111/j.1365-313X.2011.04664.x

Lv, J., Yu, K., Wei, J., Gui, H., Liu, C., Liang, D., Wang, Y., Zhou, H., Carlin, R., Rich, R., Lu, T., Que, Q., Wang, W.C., Zhang, X., Kelliher, T (2020) Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. Nature Biotechnology 38, 1397-1401. https://doi.org/10.1038/s41587-020-0728-4

Makarevitch, I., Phillips, R., Springer, N (2008) Profiling expression changes caused by a segmental aneuploid in maize. BMC Genomics 9, 7. https://doi.org/10.1186/1471-2164-9-7

Martin, C., Viruel, M., Lora, J., Hormaza, J (2019) Polyploidy in fruit tree crops of the genus *Annona* (Annonaceae). Frontiers in Plant Science 10, 99. https://doi.org/10.3389/fpls.2019.00099

Nagaki, K., Tanaka, K., Yamaji, N., Kobayashi, H., Murata, M (2015) Sunflower centromeres consist of a centromere-specific LINE and a chromosome-specific tandem repeat. Frontiers in Plant Science 6, 1-12. https://doi.org/10.3389/fpls.2015.00912

Ng, P., & Henikoff, S (2003) SIFT: Predicting amino acid changes that affect protein function. Nucleic acids research, 31, 3812-3814. https://doi.org/10.1146/annurev. genom.7.080505.115630

Otto, F (1992) Preparation and staining of cells for high-resolution DNA analysis. In: Radbruch, A. (ed). Flow cytometry and cell sorting Berlin: Springer-Verlag. pp. 101-104

Price, J., Johnston, S. Influence of light on DNA content of *Helianthus annuus* Linnaeus. Proceedings of the National Academy of Sciences of the United States of America 93, 11264-11267.

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

Roux, N., Toloza, A., Radecki, Z, Zapata-Arias, F., Dolezel, J (2002) Rapid detection of aneuploidy in *Musa* using flow cytometry. Plant Cell Reports 21, 483-490. https://doi.org/10.1007/s00299-002-0512-6

Sharma, D., Gahtyari, N. C., Sharma, P., Khulbe, R. K., Pal, R. S., & Kant, L. (2022). Doubled Haploidy: An accelerated breeding tool for stress resilience breeding in cereals. In Next-Generation Plant Breeding Approaches for Stress Resilience in Cereal Crops 199-240. Singapore: Springer Nature Singapore. https://doi.org/ 10.1007/978-981-19-1445-4_6

Sidhu, G. S., Conner, J. A., & Ozias-Akins, P. (2022). Controlled induction of parthenogenesis in transgenic rice via post-translational activation of PsASGR-BBML.Frontiers in Plant Science,13, 925467.https://doi.org/10.3389%2Ffpls.2022.925467

Todorova, M., Ivanov, P., Shindrova, P., Christov, M., Ivanova, I (1997) Doubled haploid production in sunflower (*Helianthus annuus* L.) through irradiated pollen-induced parthenogenesis. Euphytica 97, 249-254. https://doi.org/10.1023/A:1002966824988

Vear, F (2016) Changes in sunflower breeding over the last fifty years. Oilseeds and Fats, Crops and Lipids 23, D202. https://doi.org/10.1051/ocl/2016006

Wang, S., Jin, W., Wang, K (2019) Centromere histone H3- and phospholipase-mediated haploid induction in plants. Plant Methods 15, 42. https://doi.org/10.1186/s13007-019-0429-5

Wang, G., Zong, M., Liu, D., Wu, Y., Tian, S., Han, S., Guo, N., Duan, M., Miao, L., Liu, F (2022) Efficient generation of targeted point mutations in the *Brassica oleracea* var. *botrytis* genome via a modified CRISPR/Cas9 system. Horticultural Plant Journal.

https://doi.org/10.1016/j.hpj.2022.01.005

Wang, N., Gent, J., Dawe, K (2021) Haploid induction by a maize cenh3 null mutant. Science Advances 7, eabe2299. https://doi.org/ 10.1126/sciadv.abe2299

Yoon, S., Bragg, J., Aucar-Yamato, S., Chanbusarakum, L., Dluge, K., Cheng, P., Blumwald, E., Gu, Y., Tobias, C (2022) Haploidy and aneuploidy in switchgrass mediated by misexpression of CENH3. Plant Genome e20209. https://doi.org/10.1002/tpg2.20209 wileyonlinelibrary.com/journal/tpg.

4.7 SUPPLEMENTARY FIGURES

pos		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
51P	**0.97**	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
52G	**0.98**	0.00	0.00	0.03	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.11	0.01	0.00	0.00	0.00
53T	**0.98**	0.09	0.00	0.09	0.03	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.01	0.00	0.13	0.00	0.12	1.00	0.00	0.00	0.00
54Q	**0.98**	0.05	0.06	0.01	0.04	0.01	0.02	0.02	0.03	0.29	0.26	0.07	0.06	0.02	0.15	0.23	0.03	0.06	1.00	0.01	0.02
55A	**0.98**	1.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.00
56L	**0.99**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	1.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
57R	**0.99**	0.10	0.00	0.00	0.05	0.00	0.00	0.03	0.00	0.29	0.04	0.07	0.01	0.00	0.03	1.00	0.01	0.02	0.00	0.00	0.07
58E	**0.99**	0.00	0.00	0.02	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
59I	**0.99**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
60R	**0.99**	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.05	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
61K	**0.99**	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	1.00	0.00	0.00	0.03	0.03	0.05	0.68	0.00	0.06	0.02	0.00	0.00
62L	**0.99**	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
63Q	**0.99**	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.04	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
64K	**0.99**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.02	0.15	0.03	0.00	0.00	0.00	0.00
65G	**1.00**	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.04	0.00	1.00	0.10	0.00	0.00	0.00
66V	**1.00**	0.04	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	1.00	0.04	0.00	0.00
67D	**1.00**	0.10	0.00	0.82	1.00	0.00	0.05	0.08	0.00	0.01	0.00	0.00	0.15	0.01	0.01	0.07	0.09	0.01	0.00	0.00	0.00
68L	**1.00**	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.02	0.00	1.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
69L	**1.00**	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
70I	**1.00**	0.00	0.00	0.00	0.00	0.03	0.00	0.00	1.00	0.00	0.09	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00
71P	**1.00**	0.04	0.02	0.00	0.01	0.00	0.01	0.04	0.00	0.07	0.01	0.00	0.01	0.11	0.18	1.00	0.14	0.01	0.00	0.00	0.00
72A	**1.00**	0.09	0.04	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.01	0.14	0.00	0.00	0.00	0.00	0.00
73A	**1.00**	0.34	0.00	0.00	0.02	0.01	0.01	0.01	0.08	0.10	1.00	0.08	0.01	0.01	0.06	0.12	0.03	0.01	0.02	0.00	0.01
74P	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
75F	**1.00**	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure S4.1 SIFT calculated probabilities of amino acid changes in amino acid positions 51 – 75 of the sunflower CENH3 histone fold domain. The tolerance probabilities of the amino acid changes and their corresponding amino acid from this study are highlighted in yellow. "SIFT values range from 1 to zero, where '0' is indicative of not tolerated/deleterious amino acid changes and one is tolerated". All the seven amino acid changes observed in sunflower have a score of less than 0.05% and therefore considered deleterious and high haploid inducers.

	Predict Not Tolerated	Position	Seq Rep	Predict Tolerated
y	wtsrqpnlkihgfedca	**1M**	0.79	M
	whyfimnqrdelkcvtpg	**2A**	0.79	SA
y	wtsqpnmlkihgfedca	**3R**	0.79	R
	wfmychilevdqpgnras	**4K**	0.79	KT
y	wtsrqpnmlihgfedca	**5K**	0.79	K
m	wfciyvtlaspnedrkgH	**6Q**	0.79	Q
	wfyhmcreliqgkdvna	**7P**	0.79	PST
y	wtsrqpnmlkihgfedc	**8A**	0.79	A
	cwfdmiyvgphlnateq	**9K**	0.79	SKR
	cwdfmyvgpshnalteqI	**10R**	0.79	RK
	wfyhmcreliqgkdvpN	**11S**	0.79	ATS
	dcepgnkrqs	**12W**	0.79	HaymWiFVLT
w	myfhiqrclekvntdpaS	**13G**	0.79	G
	whyfmrqedclknpIV	**14T**	0.80	SATG
	cwfmiyvphglsatneqD	**15R**	0.79	RK
	wyhfcmgdpnqkl	**16R**	0.71	SIERVTA
	wmifcvlyrhtse	**17D**	0.78	GQNKDAP
	wcfym	**18S**	0.80	hilVGndqaPTESKR
	cwdfmiygpshnlaeq	**19R**	0.81	VTRK
	wyfcmvl	**20A**	0.81	gHnPITRDSEAKQ
		21S	0.81	wcmpIfgVnHdyRQkAETSL
	wcfymhpI	**22A**	0.82	VnrGsTkQDELA
	wfmhivl	**23S**	0.81	CqdYNPRGEKATS
	wyfcmhil	**24T**	0.59	gnVdQPSEARTK
	wfycmhivgd	**25S**	0.59	nrePLQTKAS
	wyfcmhpg	**26T**	0.60	rdVkLINQESTA
	wfm	**27P**	0.60	iyCvdHlnsGeQTAPKR
	wcfmivylhtaeq	**28R**	0.61	nPDSGKR
	wfmyidvphglteC	**29K**	0.86	ANQRSK
	wcmdqnrelyIFH	**30T**	0.86	PKGSVTA
	wcmfyigdqeLH	**31P**	0.84	VNTKRSAP
	cwfdyivgphlne	**32R**	0.35	ASTMQKR
	cwfmyidvgplne	**33K**	0.34	HQASTRK
	wyfc	**34E**	0.34	mHpvlIGndTRQAKES
	wcm	**35P**	0.35	fhydgneQvIkTRSLAP
	wfc	**36E**	0.35	YmhIlgVNPtRQDKASE
	wcmghdqepky	**37S**	0.39	lIRNFVATS
	wyfcmi	**38S**	0.39	vlrHdQNGEAPKTS
	wyfmhlP	**39G**	0.85	InQRCDKVETSGA
	wyf	**40G**	0.86	cMhiVlPNRQDGSATKe
	whyfmnqrdelkct	**41G**	0.68	VIPSAG
	wyfcmhilnkeV	**42G**	0.69	QDTPRSAG
	wcfyh	**43N**	0.91	gIPtLDaRMQSeKNV
	cwdfmyighnsleqAV	**44K**	0.93	TPRK
	cwfdmiygpshnalVTQ	**45K**	0.94	ERK
	wfymhcidleqgn	**46P**	0.93	AVRKTSP
	cwdfmivgpnaleQ	**47H**	0.94	YSTRKH
	wfmiyvdphlgateqCNS	**48R**	0.97	KR
	kqnrhdgepctsmilAV	**49F**	0.97	WFY
	cwdfmiyvgpsnleqTHA	**50K**	0.97	KR
w	lymfinqrcdelkvtsgA	**51P**	0.97	P

	mi f y l c v r h q k e n a P T D	**52G**	0.98	S G
	w f c y m h i l v r g p n E	**53T**	0.98	K A D S Q T
	w d f g y p h s i e a	**54Q**	0.98	N T C M Q R L K V
w	ly f i m n q r d e l k c p s V T G	**55A**	0.98	A
d	h n e c s w y k r p q t a f v M I	**56L**	0.99	L
	c w d f i v g p s n T H Q L E	**57R**	0.99	Y M A K R
c	m f i y v l h r t n g s p a k q D	**58E**	0.99	E
y	w t s r q p n m l k h g f e d c a	**59I**	0.99	I
c	d f m i y v g p s n a l t q H E K	**60R**	0.99	R
	c w d f m i y g s a l e V N P	**61K**	0.99	Q H T R K
	h n k r q d g e p c t s a m v i w	**62L**	0.99	L F Y
c	f m i y v d p g s l a t n e r H K	**63Q**	0.99	Q
	w d f m i y v g p h n a l t e Q S	**64K**	0.99	R K
	w f m y c i l h r v e p k a d Q N	**65G**	1.00	G T S
w	ly f m c r q l d k i g p n E S A V	**66V**	1.00	T
	w m i f c v l y p t q k	**67D**	1.00	G R H S A N D E
	g h n e c s w k r y p q t a v I M	**68L**	1.00	F L
d	lg n e c s r k w p q t y a m i V F	**69L**	1.00	L
	ld w p q e c r s k g y a t m N F V	**70I**	1.00	L I
	w f d m i y v g n l t e C A H	**71P**	1.00	K P S Q R
	w f d m y i v h n p l t g s e q C	**72A**	1.00	A R K
	c w d y g p h f n t e v S	**73A**	1.00	Q M I K R A L
m	y h f i c q n e r l v d k t g a S	**74P**	1.00	P
y	w t s r q p n m l k i h g e d c a	**75F**	1.00	F
	w d y n r f e g k m	**76I**	1.00	p C H l t S I V A Q
c	d f i y v g p h s n l t e q A M K	**77R**	1.00	R
	g w h y d n r q s k c p f I E A	**78T**	1.00	M T V L
w	h d g n r f q k e s p l t i M A C	**79V**	1.00	V
	w d f m i y v g p s n a l t e Q H	**80R**	1.00	K R
	c w m f i y l v r h t p a n q K G	**81E**	1.00	S D E
	h q p w d e n c r k s g y a t m f	**82I**	1.00	L V I
	w h y f r d k g n p l i	**83S**	1.00	M Q E V S C T A
	w y f c i p G a V	**84N**	1.00	S L M H R E N T K D Q
		85Y	0.99	w c p g r h I M V N S a T K L F Q E Y D
	w d p e q n g r t h	**86M**	0.99	C I K V A S M Y L F
	w y f h i p v l M	**87S**	1.00	C N d R Q E G A T S K
	w y f c	**88P**	1.00	m h i l V G N D Q P R A e K S T
	w y c f i p r t V M k L	**89E**	0.99	N H S A Q G E D
		90I	0.99	c w m P k Q E D r N T S a H G f I V Y L
		91T	0.14	w p e C k q g r M D s N a h T I f V Y L
c	d i y v g p h a t e L M S F Q K N	**92R**	1.00	R
	h n k r q d g e p c t s a m v l y	**93W**	1.00	I W F
	w f d m i y v g p s l a n e k H T	**94Q**	1.00	R Q
	w f y m h i c l q e k v d p R T N	**95A**	1.00	G A S
	w c f Y	**96E**	1.00	i v l H P G r N T k D M Q E A S
y	w t s r q p n m l k i h g f e d c	**97A**	1.00	A
	h w q d p e n c r s k g y a t f	**98L**	1.00	M L V I
		99Q	1.00	w c f h y i P M V n R d t G Q L S k E A
w	ly f m i q r n d e l k v p g C T S	**100A**	1.00	A

Figure S4.2 SIFT predictions of tolerated and non-tolerated amino acid changes in the sunflower CENH3 protein from amino acid position 1 – 100. The amino acids changes observed in this study are

highlighted in yellow. This shows tolerated amino acid changes (left) and non-tolerated changes. Non-tolerated changes have been reported to result in haploid progeny.

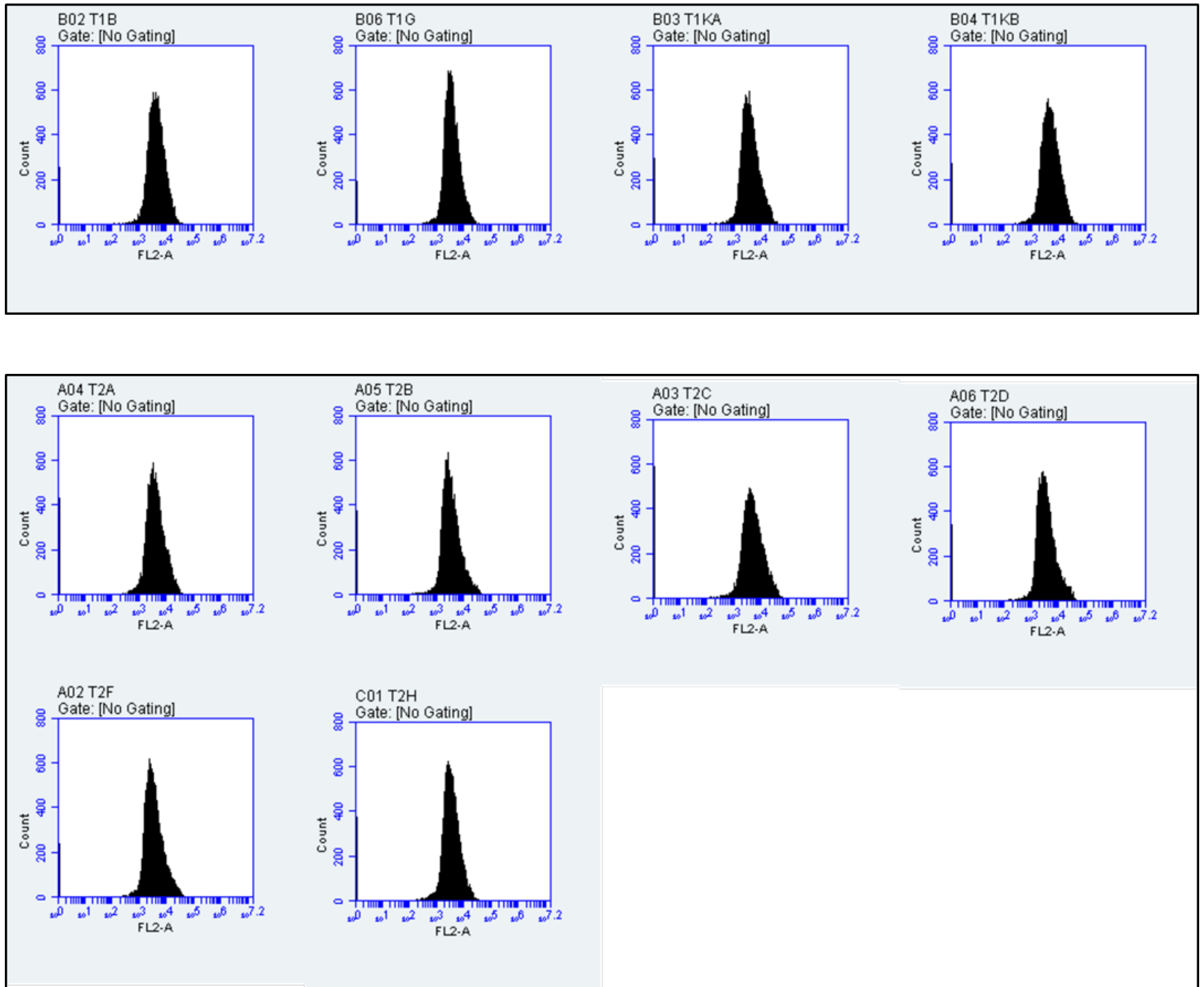


Figure S4.3 Histograms obtained from flow cytometric analysis showing sunflower lines evaluated for ploidy status in the T1 and T2 generations. Plants showed normal diploid histograms of between 10^3 to 10^4.

CHAPTER 5

GENERAL DISCUSSION AND FUTURE PROSPECTS

5.1 GENERAL DISCUSSION

A rapidly increasing global population demands higher yielding and resilient crops – therefore increased efforts and investment in crop improvement are needed. Conventional and molecular plant breeding methods have significantly contributed to the improvement of many crop plants, including oilseed crops such as sunflower (Vear, 2016). Conventional breeding methods are labour intensive, resource and time consuming. Doubled haploid (DH) technology could fast-track the production of elite inbred lines as DH lines contain completely homozygous genomes (Karimi-Ashtiyani et al., 2015). However, the reliance of DH production methods on *in vitro* (tissue) culture has resulted in the lack of a dependable DH production technique for sunflower since the crop is considered recalcitrant to *in vitro* regeneration. Modification of the *CenH3* gene through viral vector delivery systems to produce doubled haploids provides a promising alternative to approaches that require *in vitro* regeneration protocols (Dunemann et al., 2022; Britt and Kuppu, 2016).

Applications in genome editing have grown rapidly in the last few years with the introduction of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas and associated technologies with advancements promising to revolutionize crop breeding and production. The CRISPR/Cas technology has been used in a range of plant improvement applications, including the modification of genomes for the purpose of doubled haploid induction (Kuppu et al., 2020). The research undertaken in this thesis aimed to edit the *CenH3* gene for the purpose of haploid induction in sunflower. Additionally, the development of an alternative genome editing delivery method was investigated to overcome the bottleneck of transformation and *in vitro* regeneration recalcitrance in sunflower.

The introduction of foreign DNA into a genome, frequently achieved through *Agrobacterium tumefaciens* T-DNA transfer, and the regeneration of said transformed cells to obtain intact clones via *in vitro* tissue culture, are the backbone of transgenic plant development and genome editing (Kim, 2020). In a first attempt to explore genome editing in sunflower (Chapter 2), two CRISPR/Cas vectors containing donor repair templates (DRTs) of varying lengths, were developed. Sunflower mature cotyledons were transformed with the above-mentioned constructs using *Agrobacterium tumefaciens*-mediated transformation and regenerated via *in vitro* culture. The genetic transformation efficiency for sunflower via this methodology currently ranges between 1-7%, as demonstrated in this work and previous studies (Radonic et al., 2015; Sujatha et al., 2012). The reported transformed material in these works are PCR positives, but the cefotaxime and/or other antibiotics used to suppress *A. tumefaciens* after plant transformation may not always be effective, thereby resulting in false positives during

PCR amplification. This could explain the presence of 'transformed plants' without the successful genome editing observed in this study. The β-*glucuronidase* (GUS) gene is a common method for successfully detecting transformation in a number of plants including sunflower (Radonic et al., 2015; Darqui et al., 2021). As the main research question for this study was to use CRISPR/Cas for the modification of the *CenH3* gene, it was difficult to include too many variables for testing. It would be ideal for further studies to explore various methods for detecting successful transformation such as the GUS gene, *gfp*. Plants transformed in this study with a vector harboring the *nptII* gene selectable marker resulted in successful genome editing when compared to a vector containing the *bar* gene selectable marker. These two constructs contained two donor repair templates of variable length and cargo capacity which could have affected *A. tumefaciens* stability and therefore transformation efficiency. This, however, seems unlikely as *A. tumefaciens* are known to deliver T-DNA sequences of up to 100 kb (Lee et al., 2019) and the total size of the transformed vectors were approximately 16 kb. The choice of selectable marker could have a potential effect on sunflower transformation but further analysis is required to determine the role it plays in the transformation success of sunflower. The *nptII* gene selectable marker has been observed as the best performing selectable marker in sunflower (pers comm. Dr L. Radonic, INTA, Argentina). Other selection systems such as the *phosphomannose isomerase* (*pmi*) selection system (Reed et al., 2001), may also be a possibility. The *pmi* gene works as a positive selectable marker by enabling plant cells expressing the gene to utilise mannose as a carbon source, thus surviving on culture media amended with it (Reed et al., 2001). This may however require some optimization in sunflower.

Sunflower *in vitro* transformation studies and optimization protocols have been on-going since the early 1980s with minimal improvement (Darqui et al., 2021). The work presented in this thesis concurs with these findings in that sunflower remains a difficult crop to transform and regenerate *in vitro*. Various factors play a critical role in the genetic transformation and *in vitro* generation efficiency in plants, and it may be difficult and tedious to single out individual factors as culprits for decreased transformation efficiency (Darqui et al., 2021). A number of factors have been identified to be responsible for low transformation, transgene instability and rooting issues in sunflower including the type of promoter, co-culture period, unusual sensitivity to antibiotics and genotype (Sujatha et al., 2012; Darqui et al., 2021). For example, the use of the CaMV35 promoter showed decreased transformation and rooting issues when compared to the *rbcS1* promoter (Radonic et al., 2012; Radonic et al., 2015). Other factors such as the choice of selectable marker, size of gene construct may need to be explored in future studies. The low regeneration efficiency of sunflower has also impacted sunflower improvement and is revealed by the lack of a commercially available transgenic sunflower (Darqui et al., 2021)

and the delay in adoption of genome editing techniques such as the CRISPR/Cas9 technology. Given the limitations presented by *A. tumefaciens* transformation in various plant species, analyzing the genomic factors affecting *A. tumefaciens* transformation need to be explored. Recently, a comparative transcriptomic study of *Agrobacterium tumefaciens* infection between tea (*Camellia sinensis* L.) and tobacco revealed the molecular basis for tea's recalcitrance to genetic transformation (Jin et al., 2022). Compounds present in tea known as gamma-aminobutyrate (GABA) and catechins affected plant-pathogen attachment, mineral acquisition and quorum quenching which could potentially impact *Agrobacterium tumefaciens*-mediated transformation of tea (Jin et al., 2022).Rooting is a major problem in sunflower *in vitro* culture. There are however limited studies focusing on sunflower rooting deterrents. It was however shown that improvement in shoot growth improved rooting in sunflower (Sujatha et al., 2012). Studies investigating both shoot regeneration and rooting deterrents in sunflower could offer potential solutions to assist in the optimization of *A. tumefaciens* mediated transformation.

The CRISPR/Cas genome editing tool has revolutionized biology owing to its ease in design and construction, reproducibility, and efficiency (Uranga et al., 2021). Application of CRISPR/Cas9 based genetic editing has increased greatly and rapidly in plants (Wada et al., 2020) and it has been effectively evaluated in both transgenic and transient plant experiments. However, it has not yet been evaluated in sunflower. The ability to target and alter specific genomic sequences has advanced scientific research as it enables specific gene modifications and the introduction of elite alleles from related species directly into commercialised cultivars (Bleuyard et al., 2006; Li et al., 2020). A remarkable challenge observed for the CRISPR/Cas genome editing platform in plants is the successful delivery of the Cas endonuclease, sgRNA and donor repair template (for homology directed repair experiments) into target cells (Uranga et al., 2021). The choice of delivery method has also been observed to influence editing and homology directed repair (HDR) efficiency in plant cells (Eini et al., 2022). Homology directed repair (HDR) mechanisms allow for the precise, error free knock in of desired genetic material into a targeted gene locus (Hahn et al., 2018). Most CRISPR/Cas experiments in plants have demonstrated the use of conventional *A. tumefaciens* T-DNA delivery methods. In this study, partial HDR was obtained at low rates in Agrobacterium tumefaciens transformation (Chapter 2) of sunflower while both partial and full homology repair was obtained at low rates in sunflower transformed with the same constructs delivered by a modified geminiviral based delivery vector (Chapter 3). HDR occurrences are limited in plant species, thus resulting in reduced rates of HDR-based gene editing. Plant viruses have demonstrated the ability to transiently deliver genome editing reagents in various plant species (Baltes et al., 2014; Čermák et al., 2015; Eini et al., 2022). Geminiviruses are

said to have the ability to revert the cell cycle to the S-phase where occurrence of HDR is abundant (Eini et al., 2022). Geminiviral replicons are believed to also increase donor repair template (DRT) copy number in the target cells, thereby increasing chances of HDR mechanism (Čermák et al., 2015). These factors could explain the enhanced editing efficiency observed with the use of a geminivirus based vector when delivering gene editing components to sunflower. However, some studies have also found homology directed repair (HDR) was more enhanced using an *A. tumefaciens* double cut DRT as opposed to a circular DRT delivered via a geminiviral replicon (Hahn et al., 2018; Rozov et al., 2019). These findings could be an indication that the nature/structure of the DRT also plays a role in HDR efficiency. The use of a double cut donor (termed *in planta* gene targeting (IPGT)) was tested in the present study using *Agrobacterium tumefaciens*-mediated transformation and was not tested using the geminiviral vector. Transformation was however at low frequencies and HDR was partial, however the results obtained by Hahn et al. (2018) could imply that the use of the IPGT system together with a geminiviral vector could further increase the occurrence of HDR in plant cells, but further investigation is required in this regard.

The selection of a target site is a crucial part of CRISPR experimental design and is a major determinant of genome editing success (Uranga et al., 2021). Certain target sites have been observed to generally contain low mutation rates when compared to other sites. A single target site was used in this study due to its proximity to the target amino acids. The *CenH3* gene is a functionally essential gene responsible for centromere localization and genetic material transfer during cell division in plant species (Britt and Kuppu., 2015). Mutations in this gene region may therefore result in detrimental effects on plant growth and function. The functional importance of this target gene could have possibly affected mutation frequency and efficiency due to the conserved nature of the *CenH3* gene. It is therefore difficult to conclude whether the low mutation and HDR frequency obtained in this study are representative of the *CenH3* gene mutation or the nature of sunflower genome editing as no reports on sunflower editing exist to our knowledge. In future, further analysis on other target sites, for example non-function essential genes such as the *phytoene desaturase* (*PDS*) gene, can be tested to establish gene editing and HDR efficiency in sunflower. The *PDS* gene mutation gives a visual phenotype i.e. bleached cells and this phenotypic marker makes it easier to confirm successful gene editing. The *PDS* gene has been investigated in various plant studies with high editing and could be a possible optimization test in sunflower.

The size of the genome editing reagents delivered to plant cells have also been linked to the instability of the delivery mechanism utilized (Rezaei et al., 2021). In viruses, particularly DNA viruses, the cargo capacity limitations have contributed to the limited application of whole plant

CRISPR/Cas9 genome editing (Eini et al., 2022; Rezaei et al. 2021). Recent work by Eini et al. (2022), Rezaei et al. (2021) and in this study, have successfully demonstrated the systemic delivery of the CRISPR/Cas construct to induce whole plant genome editing using geminiviral vectors. In this study, self-cleaving (2A) peptides' sequences were used to circumvent cargo constraints. However, other studies have reported that geminiviral vectors can deliver CRISPR-editing reagents comparable to those delivered in this work without the use of linker sequences (Eini et al., 2022, Rezaei et al., 2021). The role of 2A peptides on cargo capacity enhancement, or lack thereof of, in geminiviral vectors could be further investigated.

Various studies, including the present study, have demonstrated the upper limit of geminivirus vector carrying capacity through successful CRISPR reagent delivery and genome editing. However, the lowered gene editing efficiency observed in this study might be linked to vector instability – most likely due to overloading. The SpCas9 endonuclease, encoded by a ±4.2 kb gene, is the main contributor to the large size of the CRISPR/Cas components. Various Cas9 orthologs from other bacterial species have been identified that are smaller than the SpCas9, including *Streptococcus thermophilus* (St1Cas9, 3.2 kb), *Campylobacter jejunii* (CjCas9, 2.95 kb), *Staphylococcus aureus* (SaCas9, 3.2 kb) and have been successfully tested in plant species (Uranga et al., 2022). These smaller variants could provide a solution to limit the Cas9 effect in viral packaging - however the functionality of these Cas9 orthologs are restricted by their complex target specific protospacer adjacent motif (PAM) sequence requirements, e.g. 5' NNGRRT or 5' NNAGAAW for SaCas9 and StCas9 respectively. This keeps SpCas9, with a PAM sequence requirement of NGG, as the preferred nuclease for CRISPR-based genome editing (Uranga et al., 2022).

In this work, slightly increased editing efficiency and successful homology directed repair was observed when a geminiviral delivery system was used as opposed to the conventional *Agrobacterium tumefaciens* T-DNA based delivery. However, none of the mutations observed in parent plants (T0) were present in the progeny. The inability of viruses to infect the apical meristem and reach the germline poses a major setback in genome editing studies. To obtain mutations in the germ-line, modified tissue needs to be regenerated *in vitro* (Uranga et al., 2022). An attempt to culture and regenerate viral delivery system infected sunflower tissue (leaf material) was performed in this study, but the cultured tissues failed to regenerate and produce shoots (data not shown). This further emphasizes the importance of more research efforts on the *in vitro* regeneration of sunflower. Stable transformation by *Agrobacterium tumefaciens* T-DNA transformation is the main strategy used to produce genetically modified plants (Rezaei et al., 2021). Although *Agrobacterium tumefaciens*-mediated transformation provides stable transformation that can be integrated across generations, the introduction of

unwanted plasmid backbone negatively affects transgenic plant quality and brings GMO regulation policies into play (Collier et al., 2018). Several techniques, including traditional plant breeding, zinc finger nucleases (ZFNs) and transcription activation like-nucleases (TALENs), have been tested for the elimination of plasmid backbone sequences, such as selectable marker genes (Rezaei et al., 2021). The removal of the *nptII* selectable marker using the geminivirus delivered CRISPR/Cas9 in transgenic tobacco has been previously demonstrated (Rezaei et al., 2021). Therefore, geminiviruses, including the TYLCV-IL (IL-60-BS-Cas9-1185) vector tested in the present study, could be used for the elimination of unwanted plasmid sequences introduced through the stable transformation of plants, especially crop plants of economic importance. It is important to note that viral vector may be persistent in the progeny of inoculated plants as observed in this study, the IL-60-BS-Cas9-1185 used for plant transformation in Chapter 3 remained detectable in both the T1 and T2 generations tested. The presence of plant virus sequences and CRISPR reagents in crops has however been proven harmless as these exist in nature.

Doubled haploid lines are an invaluable asset in modern plant breeding programs, In the present study, the analysis of T1 and T2 generation plants (Chapter 4) obtained from crossing plants with a modified CenH3 gene generated in this study and a wildtype sunflower line did not result in haploid plants, only putative aneuploids were observed. So far, the *CenH3* method has been successfully described for the model plant *Arabidopsis thaliana* with efficiencies of up to 30%. Crop plants such as wheat and maize have also shown improvement in *CenH3* based haploid induction efficiency to approximately 8.9% using heterozygous alleles (Kelliher et al., 2019, Lv et al., 2020, Wang et al., 2021). At the start of the study, the only CenH3 modification mechanisms identified were the alteration and swapping of the N-terminal domain (Ravi and Chan, 2010), whole gene replacements from other species (Maheshwari et al., 2017) and the use of non-targeted single amino acid changes (Kuppu et al., 2015). These mechanisms have presented challenges including transformation and mutation complementation limitations resulting in delays in the adoption of *CenH3* based haploid induction in crop breeding programs (Lv et al., 2020). A variety of modifications, including CRISPR/Cas9 gene knockouts of *CenH3* regions, have been successfully verified to induce a high rate of haploids in *Arabidopsis* (44%) (Kuppu et al., 2020), maize (5.2%) (Wang et al., 2021) and wheat (8%) (Lv et al., 2020). These newly discovered strategies provide an alternative modification strategy including the use of highly efficient non-homologous end joining and avoiding the complex inefficient homology directed repair (Lv et al., 2020). CenH3 gene knockouts generated through NHEJ can therefore be tested in sunflower in future studies. Heteroallelic *CenH3* mutant plants have been discovered to be better haploid inducers (Wang et al., 2021, Lv et al., 2020), this means that transient expression systems such as the

one used in this study can be used for generating haploid inducers as only the self-fertilization of T0 generation is required for haploid induction as opposed to homologous mutations which can only be obtained in stable transformed T1 and T2 generations. Another challenge observed in this study was the lack of sufficient material to test, as most of the seeds failed to germinate due to the lethal nature of the *CenH3* mutation. In future studies, significant amounts of material should be tested. This can be achieved by inoculating and testing more material in the T0.

The *CenH3* based haploid induction system has seen more failures than successes in crop plants, including tomato, soybean (Lv et al., 2020), carrot (Dunemann et al., 2022) and sunflower (here) failing. Further research on the CENH3 structure and function could assist in the enhancement of haploid induction in non-model crop plants. Modification of phospholipases coding genomic regions; phospholipase-a1 (*PLA1*)/ Matrilineal (*MTL*), Not like dad (*NTL*) and the domain of unknown function 679 membrane protein (*ZmDMP*) in major cereal crops, such as rice and maize, have also been identified as responsible for haploid induction (Kelliher et al., 2019). These genes are responsible for facilitating normal fertilization of egg cells, mutations therefore trigger haploids upon egg fertilization failure resulting in parthenogenesis (Sidhu et al., 2022). These gene homologs have however only been identified in monocotyledonous plants, except for the *ZmDMP* which has also been identified in several dicotyledonous species (Wang et al., 2021). Investigation on the existence of similar gene orthologs in sunflower may assist in accelerating haploid inducer line development. More work needs to be done in crop plants, as opposed to model plants. With the advancement of genomics and genome editing technologies, various gene targets for haploid induction can be explored in various crop species including dicotyledons. Furthermore, the rate of occurrence of haploids using the *CenH3* based approach is low and requires sorting through and screening a large fraction of diploids and aneuploids. A phenotypic marker could assist in screening for haploids.

This study has presented some noteworthy findings on sunflower transformation and genome editing using the CRISPR/Cas9 platform. To our knowledge, this is the first report on CRISPR based genome editing in sunflower. The CRISPR/Cas genome editing platform has made, what was once the dream of many biologists, a reality: pathogen resistant, nutrient rich crops, and doubled haploid crops can now be produced through minute and target specific alterations to DNA. Even with these significant milestones, the CRISPR/Cas technology is not one without bottlenecks viz tissue culture *in vitro* regeneration dependence. Tissue culture remains a labour intensive and time-consuming process even for crops with established regeneration protocols (Kim., 2020). It can take up to a year to obtain a fully developed transformed plant,

the development of a tissue culture free protocol is therefore required for effectively driving plant genome editing. Plant viral vectors hold great potential for accelerated genome editing and crop improvement due to their modest mechanism of delivery. Plants can be directly inoculated with viral vectors harboring genome editing components and it only takes between 3-7 days for successful viral genome editing to take place allowing for rapid screening of a large quantity of plants (Uranga et al., 2022). The work presented in this thesis holds great potential for the development of a *CenH3* based haploid inducer line in sunflower. Further research on viruses that are capable of infecting meristem tissues and advancement of tissue culture techniques is needed, especially for recalcitrant crops. This study provides a steppingstone in sunflower transgenic research, doubled haploid production and potential trait improvement in sunflower.

5.2 REFERENCES

Baltes, N.J., Gil-Humanes, J., Cermak, T., Atkins, P.A., Voytas, D.F (2014) DNA replicons for plant genome engineering. The Plant Cell 26, 151-163. https://doi.org/10.1105/tpc.113.119792

Bleuyard, J., Gallego, M.E., White, C.I (2006) Recent advances in understanding of the DNA double-strand break repair machinery of plants. DNA Repair 5, 1–12. https://doi.org/10.1016/j.dnarep.2005.08.017

Britt, A.B., Kuppu, S (2016) CenH3: An Emerging Player in Haploid Induction Technology. Frontiers in Plant Science 7, 357. https://doi.org/10.3389/fpls.2016.00357

Čermák, T., Baltes, N.J., Čegan, R., Zhang, Y., Voytas, D.F (2015) High-frequency, precise modification of the tomato genome. Genome Biology 16. https://doi.org/10.1186/s13059-015-0796-9

Collier, R., Thomson, JG., Thilmony, R (2018) A versatile and robust *Agrobacterium*-based gene stacking system generates high-quality transgenic *Arabidopsis* plants. The Plant Journal 95, 573-583. https://doi.org/10.1111/tpj.13992

Darqui, F., Radonic, L., Beracochea, V., Hopp, H., López Bilbao, M (2021) Peculiarities of the transformation of *Asteraceae* family species: The cases of sunflower and lettuce. Frontiers in Plant Science 12, 767459. https://doi.org/10.3389/fpls.2021.767459

Dunemann, F., Krüger, A., Maier, K., Struckmeyer, S (2022) Insights from CRISPR/Cas9-mediated gene editing of centromeric histone H3 (CENH3) in carrot (*Daucus carota* subsp. *sativus*).*bioRxiv*, 2022-09.

Eini, O., Schumann, N., Niessen, M., Varrelmann, M (2022) Targeted mutagenesis in plants using *Beet curly top virus* for efficient delivery of CRISPR/Cas12a components. New Biotechnology 67, 1-11. https://doi.org/10.1016/j.nbt.2021.12.002

Hahn, F., Eisenhut, M., Mantegazza, O., Weber, A.P.M (2018) Homology-directed repair of a defective glabrous gene in *Arabidopsis* with Cas9-based gene targeting. Frontiers in Plant Science 9, 424. https://doi.org/10.3389/fpls.2018.00424

Jin, K., Tian N., Ferreira, J., Sandhu, D., Xiao, L., Gu, M., Luo, Y., Zhang, X., Liu, G, Liu, Z., Huang, J., Liu, S (2022). Comparative transcriptome analysis of *Agrobacterium tumefaciens* reveals the molecular basis for the recalcitrant genetic transformation of *Camellia sinensis* L. Biomolecules 12: 68. https://doi.org/10.3390/biom12050688

Karimi-Ashtiyani, R., Ishii, T., Niessen, M., Stein, N., Heckmann, S., Gurushidze, M., Banaei-Moghaddam, A.M., Fuchs, J., Schubert, V., Koch, K., Weiss, O., Demidov, D., Schmidt, K., Kumlehn, J., Houben, A (2015) Point mutation impairs centromeric CENH3 loading and induces haploid plants. Proceedings of the National Academy of Sciences of the United States of America 112, 11211-6. https://doi.org/10.1073/pnas.1504333112

Kim S (2020) The way to true plant genome editing. Nature plants 6, 736-737. http:/ https://doi.org/10.1038/s41477-020-0723-2

Kuppu, S., Tan, E.H., Nguyen, H., Rodgers, A., Comai, L (2015) Point mutations in centromeric histone induce post-zygotic incompatibility and uniparental inheritance. PLoS genetics 11, e1005494. https://doi.org/10.1371/journal.pgen.1005494

Kuppu, S., Ron, M., Marimuthu, M.P.A., Li, G., Huddleson, A., Siddeek, M.H., Terry, J., Buchner, R., Shabek, N., Comai, L., 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, 2068-2080. https://doi.org/10.1111/pbi.13365

Lee, K., Eggenberger, A., Banakar, R., McCaw, M., Zhu, H., Main, M., Kang, M., Gelvin, S., Wang K (2019) CRISPR/Cas9-mediated targeted T-DNA intergration in rice. Plant Molecular Biology 99, 317-328. https:/doi.org/1 0.1007/s11103-018-00819-1

Li, J., Li, H., Chen, J., Yan, L., & Xia, L. (2020). Toward precision genome editing in crop plants. Molecular Plant, 13, 811-813. https://doi.org/10.1016/j.molp.2020.04.008

Lv, J., Yu, K., Wei, J., Gui, H., Liu, C., Liang, D., Wang, Y., Zhou, H., Carlin, R., Rich, R., Lu, T., Que, Q., Wang, W.C., Zhang, X., Kelliher, T (2020) Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. Nature Biotechnology 38, 1397-1401. https://doi.org/10.1038/s41587-020-0728-4

Maheshwari, S., Ishii, T., Brown, C.T., Houben, A., Comai, L (2017) Centromere location in *Arabidopsis* is unaltered by extreme divergence in CENH3 protein sequence. Genome Research 27, 471-478. https://doi.org/10.1101/gr.214619.116

Radonic, L.M., López, N.E., Hopp, E., Bilbao, M.L (2012) Analysis of T2 sunflower transgenic plants: high expression level and stability achieved by the rbcS1 promoter regulation. In Proceedings of the 18th international sunflower conference, Mar del Plata.

Radonic, L.M., Lewi, D.M., López, N.E., Hopp, H.E., Escandón, A.S (2015) Sunflower (*Helianthus annuus* L.). Methods in Molecular Biology https://doi.org/10.1007/978-1-4939-1658-0

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

Reed, J., Privalle, L., Powell, M.L., Meghji, M., Dawson, J., Dunder, E., Sutthe, J., Wenck, A., Launis, K., Kramer, C. and Chang, Y.F. (2001). Phosphomannose isomerase: an efficient selectable marker for plant transformation. In Vitro Cellular & Developmental Biology-Plant, 37,127-132. https://doi.org/:10.1079/IVP2000162

Rezaei, A., Farsi, M., Malekzadeh-Shafaroudi, S., Seifi, A (2021) *In planta* removal of *nptII* selectable marker gene from transgenic tobacco plants using CRISPR/Cas9 system. Plant Gene 26, 100288. https://doi.org/10.1016/j.plgene.2021.100288

Rozov, S.M., Permyakova, N. V., Deineko, E. V (2019) The problem of the low rates of CRISPR/Cas9-mediated knock-ins in plants: approaches and solutions. International Journal of Molecular Sciences 20, 3371. https://doi.org/10.3390/ijms20133371

Sidhu, G. S., Conner, J. A., & Ozias-Akins, P. (2022). Controlled induction of parthenogenesis in transgenic rice via post-translational activation of PsASGR-BBML.Frontiers in Plant Science,13, 925467.https://doi.org/10.3389%2Ffpls.2022.925467

Sujatha, M., Vijay, S., Vasavi, S., Reddy, P.V., Rao, S.C (2012) *Agrobacterium*-mediated transformation of cotyledons of mature seeds of multiple genotypes of sunflower (*Helianthus annuus* L.). Plant cell, tissue and organ culture 110, 275-287. https://doi.org/10.1007/s11240-012-0149-2

Uranga, M., Vazquez-Vilar, M., Orzáez, Di., Daròs, J.A (2021) CRISPR-Cas12a genome editing at the whole-plant level using two compatible RNA virus vectors. The CRISPR Journal 4, 761-769. https://doi.org/10.1089/crispr.2021.0049

Vear, F (2016) Changes in sunflower breeding over the last fifty years. Oilseeds and Fats, Crops and Lipids 23. https://doi.org/10.1051/ocl/2016006

Wada, N., Ueta, R., Osakabe, Y. and Osakabe, K., 2020. Precision genome editing in plants: state-of-the-art in CRISPR/Cas9-based genome engineering. BMC Plant Biology, 20,1-12. https://doi.org/10.1186/s12870-020-02385-5

Wang, S., Jin, W., Wang, K (2019) Centromere histone H3- and phospholipase-mediated haploid induction in plants. Plant Methods 15, 42. https://doi.org/10.1186/s13007-019-0429-5

Wang, N., Gent, J., Dawe, K (2021) Haploid induction by a maize *cenh3* null mutant. Science Advances 7, eabe2299. https://doi.org/ 10.1126/sciadv.abe2299