TARGETED GENE INSERTIONS AND DWARFED STATURE ENGINEERING IN DIPLOID POTATO

By Yordan Nikolaev Dolaptchiev

Student Number: 100312196

A thesis submitted to the University of East Anglia for the Degree of Doctor of Philosophy

The Sainsbury Laboratory

August 2025

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognize that its copyright rests with the author and that use of any information derived there-from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

This work was supported by the UKRI Biotechnology and Biological Sciences Research Council Norwich Research Park Biosciences Doctoral Training Partnership as a CASE Award, [grant number BB/T008717/1], in collaboration with Solynta.

Abstract

Classical crop breeding in tetraploid potatoes is constrained by difficult genetics, resulting in little genetic gain in yield in recent decades and a widespread susceptibility to important pathogens like *Phytophthora infestans*. Diploid potato breeding provides a promising alternative, but adoption requires rapid germplasm improvement. While transgenesis can help, the ability to deliver genes at an operator-defined location, here referred to as a knock-in (KI), could allow for more consistent transgene expression and the repeated stacking of transgenes at the same locus. In this thesis, I worked to establish an efficient method for targeted gene delivery in diploid potato. An anthocyanin KI reporter system was established and used to optimise several aspects of the KI approach. This optimised method was used to deliver the *P. infestans* resistance gene *Rpi-vnt1* at the *Cai1* locus. Based on this progress, I propose a novel two-phase approach for the targeted iterative insertion of genes via recombination (TIIGER). In the first phase a targeted gene replacement is used to exchange the proximal *Cai1* promoter for a cassette with genes of interest, the anthocyanin reporter, and a negative selection marker. In the second phase, the reporter and marker are exchanged for additional genes of interest and the original proximal *Cai1* promoter. Here, I demonstrated the first phase and proposed several approaches to the second phase. Lastly, I used gene editing to target the potato *DELLA* growth regulator gene and generate a dominant mild dwarfism allele which should help improve potatoes' tolerance to high planting density and improve yields per hectare. A phenotypic series of dwarfs was generated and a promising line bearing a Valine deletion was assessed in a field trial, where it compared favourably to controls. The limitations of these approaches, viable solutions, and future work are also discussed.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

List of Contents

Abstract 2
List of Contents 3
List of Figures 9
List of Tables 11
List of Appendices 12
Preface 13
Acknowledgements 13
List of abbreviations 14
CHAPTER 1. GENERAL INTRODUCTION 15
1.1 The state of potato breeding 15
1.1.1 Traditional tetraploid potato breeding 16
1.1.2 Hybrid breeding 18
1.1.3 Diploid hybrid potato breeding 19
1.2 Plant Immunity 20
1.2.2 The rise of Effectoromics 22
1.2.3 Durable field resistance 23
1.3 Horizontal gene transfer 23
1.3.1 Introgression 23
1.3.2 Transgenesis 25
1.3.3 Genome Editing 27
Programmable molecular scissors 27
DNA repair pathways 28
Gene targeting by homologous recombination 29
1.4 Conclusion 32
1.5 Project aims and objectives 32
CHAPTER 2. MATERIALS AND METHODS 33
2.1 Materials 33
Bacterial material 33
Plant material 33

Kits 34
Buffers and media 35
Chemicals and antibiotics 35
2.2 Methods 36
2.2.1 Plant methods 36
Potato seed harvesting 36
Seed germination 36
Diploid potato crossing 36
Microscopy 37
2.2.2 Nucleic acid methods 37
Amplicon sequencing 37
DNA extraction 37
PCR 40
Nucleic acid visualisation 41
2.2.3 Bioinformatics 42
Identifying *Cai1* as *ANT1*'s orthologue 42
crRNA design 43
2.2.4 KI cloning 43
General Golden Gate cloning 43
Cloning of custom L1 modules 46
Level 2 cloning 48
First Knock-In constructs 49
Constructs for exploring alternative KI approaches 51
Cloning alternative nucleases and crRNA expression cassettes 51
Cloning into alternative plasmid backbones 52
Cloning of luciferase constructs 53
2.2.5 TIIGER cloning 53
Cloning a *codA* expression cassette 53
Cloning TIIGER constructs with one or two crRNAs 53
Cloning TIIGER constructs with 0/1/2 R-genes 54

Cloning GTSS constructs 54
Cloning Targeted Replacement + Cre 55
2.2.6 Transformation 55
Transient expression 55
Leaf disk lux assays 56
Stable transformation 56
2.2.7 Breeding lines YD1-3 58
F_1 vigour screen 58
RUBY transformability screen 58
2.2.8 KI tracking 59
KI sector excision and scoring 59
2.2.9 Protein extraction and Western Blots 60
2.2.10 Editing the potato *GAI* gene 61
Cloning Cas9 constructs for *StGAI* editing 61
2.2.11 DELLA dwarf screening 61
GA3 spray screening 61
F_1 genotyping via NGS amplicon sequencing 62
2.2.12 Field trial 62
Field trial design and conditions 62
Field phenotyping via drone 64
2.3 Statistical methods 66
2.3.1 Statistics used in KIs and TIIGER 66
Time-to-Event analysis 67
2.3.2 Field trial statistics 67
Comparing seed tuber weights 67
Field trial manual plant height measurements 68
Comparing field trial plot plant height data 69
Comparing field trial yields 70
Comparing field trial tuber size distributions 72
Comparing field trial tuber shape 73

CHAPTER 3. TARGETED GENE INSERTIONS 74
3.1 Introduction 74
Aims 79
3.2 Results 80
3.2.1 *Cai1*, a functional anthocyanin KI reporter in potato 80
Designating a target locus 80
crRNA and homology arm design 83
Confirming *Cai1* function 84
3.2.2 Gene Knock-Ins in potato are possible 87
KI construct cloning and transformation 87
Characterisation of T_0 KI lines 89
KIs are inherited in a Mendelian fashion 93
T-DNAs can be segregated away 97
3.2.3 Exploring alternative KI approaches 98
3.2.4 Improving KI efficiency 101
Small molecule effectors 101
Breeding a potato for KIs 103
Choosing a better programmable nuclease 107
Testing crRNA expression cassettes 110
Choosing a better plasmid backbone 111
3.2.5 KI of *Rpi-vnt1*, an R-gene against *P. infestans* 113
3.3 Discussion 115
3.3.1 A functional anthocyanin KI reporter in potato 115
3.3.2 Nature of recovered KI events 117
3.3.3 Competition between HDR and NHEJ 121
3.3.4 Frequency of KI events 123
3.3.5 Recovering T-DNA-free lines 126
3.3.6 Segregating away T-DNAs 127
3.3.7 Segregation for transformability 127
3.3.8 Segregation of KI amenability 129

CHAPTER 4. TARGETED ITERATIVE INSERTION OF GENES VIA RECOMBINATION (TIIGER) 132
4.1 Introduction 132
Aims 133
4.2 Results 136
4.2.1 Developing TIIGER 136
Efficient targeted replacements using two crRNAs at the *Cai1* locus 136
Targeted replacement of cassettes up to 12.3 kb 138
Phase 1 T_0 genotyping 140
4.2.2 GT-specific selection (GTSS) 142
How to select for a GT event at any locus but not a T-DNA insertion 142
Using GTSS to recover TIIGER phase 1 events 145
4.2.3 Targeted Replacement + Cre 147
Introducing a LoxP acceptor locus 147
4.3 Discussion 152
4.3.1 TIIGER phase 1 works efficiently 153
4.3.2. GTSS 155
4.3.3. Targeted replacement + Cre 158
CHAPTER 5. ENGINEERING COMPACT GROWTH IN POTATO 159
5.1 Introduction 159
Potato growth and productivity 159
Dwarfism alleles – nature and mechanism 161
Aims 165
5.2 Results 165
5.2.1 Generating DELLA dwarf potatoes 165
Targeting the *StGAI* gene 165
Characterising T_0 lines 167
Line 457 carries a Valine deletion 169
5.2.2 Dwarf performance in the field 172

Selecting F_1 lines for a field trial 172
Field trial design 177
Comparing plant canopies 178
Tuber trait phenotyping 181
5.3 Discussion 187
Generating dwarfism alleles via *StGAI* mutagenesis 187
The ΔVal allele 188
Field performance 189
CHAPTER 6. GENERAL DISCUSSION 192
6.1 Targeted Gene Insertions 192
Future directions 193
6.2 TIIGER 194
Future directions 194
6.3 Engineering Compact Growth into Potatoes 195
Future directions 195
6.4 Concluding remarks 196
BIBLIOGRAPHY 197
APPENDIX 225
Appendix 1. Supplementary Figures 225
Appendix 2. R code analysing KI sector counts 237
Appendix 3. R code output of Non-linear Mixed Effects model 238
Appendix 4. R code for analysis of tuber shape. Box-Cox data transformation and linear modelling. 239
Appendix 5. Primers and oligonucleotides used. 240

List of Figures

Figure 1.1. Lack of genetic gain in potato varieties. 15
Figure 1.2. Late Blight in a potato field 16
Figure 1.3. Maize yields since the introduction of hybrid breeding. 19
Figure 1.4. The Zig-zag-zig model of plant immunity. 21
Figure 2.1. MoClo hierarchical cloning standard. 44
Figure 2.2. Cloning the homology arms. 48
Figure 2.3. Creation of polygenic constructs. 49
Figure 2.4. KI construct structure. 50
Figure 2.5. Level 1 constructs featuring the intron-containing Cas12a cassette. 52
Figure 2.6. KI sectors after excision (A) and one week later (B). 59
Figure 2.7. Field trial climactic conditions. 64
Figure 2.8. Diagnostic plot for weight comparison of tubers used as seed in the field trial. 68
Figure 2.9. Diagnostic plot for manual plant height measurements. 69
Figure 2.10. Diagnostic plots of the Non-linear Mixed Effects Model. 70
Figure 2.11. Descriptive plots of field trial yield data. 72
Figure 2.12. Plots of tuber shape data and linear model residuals. 73
Figure 3.1. Use of Geminivirus replicons for improved KI rates. 75
Figure 3.2. Anthocyanin KI reporter system. 77
Figure 3.3. Cloning *Cai1* alleles. 81
Figure 3.4. The *Cai1* allele sequences. 82
Figure 3.5. B26 and B100 are homo- and heterozygous for *Cai1* ... 83
Figure 3.6. Targeting the *Cai1-long* allele. 84
Figure 3.7. Transient expression in tobacco. 85
Figure 3.8. Stable transformation of *Cai1* and *Ant1*. 86
Figure 3.9. Initial set of KI constructs. 88
Figure 3.10. Purple KI sectors. 89
Figure 3.11. T_0 KI line phenotype and genotype..... 90
Figure 3.12. T_0 T-DNA screen..... 93
Figure 3.13. F_1 KI segregation..... 94

Figure 3.14. KI F_2 phenotypic segregation. 95
Figure 3.15. KI F_2 genotypic segregation. 96
Figure 3.16. T-DNAs and KIs are not genetically linked. 98
Figure 3.17. KI method diagram. 99
Figure 3.18. Alternative approaches to KIs. 100
Figure 3.19. Effects of 3-MB on shoot growth. 102
Figure 3.20. B26 and B100 fertility phenotypes 104
Figure 3.21. Transformability and genotyping of B26 × B100 F_1 hybrids. 105
Figure 3.22. KI rates among potato lines. 107
Figure 3.23. Comparing different nucleases. 108
Figure 3.24. Cas12a protein levels with transient expression. . 109
Figure 3.25. Comparing crRNA expression cassettes. 111
Figure 3.26. Comparing plasmid backbones. 112
Figure 3.27. Transient expression performance of plasmid backbones. 113
Figure 3.28. Phenotype & genotype of T_0 plants carrying *Rpi-vnt1*. 114
Figure 3.29. One-step and Two-step KI process. 119
Figure 3.30. Example of KI constructs with 1 (top) or two (bottom) copies of the crRNA1 target site flanking the KI cassette. 122
Figure 3.31. Model of the KI process. 123
Figure 3.32. Template competition hypothesis. 130
Figure 4.1. Biphasic TIIGER approach. 134
Figure 4.2. Efficient targeted replacements. 137
Figure 4.3. Effect of TIIGER cassettes size on efficiency. 139
Figure 4.4. Genotyping TIIGER T_0 lines with multiple genes delivered. 141
Figure 4.5. GT-specific selection. 143
Figure 4.6. GTSS can be stringent. 144
Figure 4.7. TIIGER events recovered with GTSS. 146
Figure 4.8. Introducing a LoxP acceptor locus. 149
Figure 4.9. LoxP line genotyping. 151
Figure 4.10. Cre-expressing construct for RMCE process. 152
Figure 4.11. TIIGER-GTSS at the *StALS1* locus. 157
Figure 5.1. Relationship between PAR interception & LAI (left) and total dry weight & PAR interception (right) in potato. 160
Figure 5.2. DELLA dwarf tomatoes. 162

Figure 5.3. Knockdown of *StGA20ox1* results in dwarfed potatoes. 164
Figure 5.4. Generating DELLA mutants. 167
Figure 5.5. T_0 dwarf phenotypes..... 169
Figure 5.6. Phenotype and genotype of F_1 dwarfs..... 171
Figure 5.7. F_1 phenotypes in the glasshouse..... 172
Figure 5.8. The non-fluorescent F_1 lines were T-DNA-free..... 174
Figure 5.9. F_1 dwarf and non-dwarf line genotyping..... 176
Figure 5.10. Field trial plan. 177
Figure 5.11. Field trial seed tubers. 178
Figure 5.12. Canopy growth in the field. 180
Figure 5.13. Longest stem length. 181
Figure 5.14. Dwarf and non-dwarf field trial yields. 182
Figure 5.15. Dwarf tuber size distribution. 185
Figure 5.16. Dwarf tuber shape. 186
Supplementary Figure 1. *ANT1* homologue in Solyntus. 225
Supplementary Figure 2. *Cai1* allele in B99, B100, Solyntus. ... 226
Supplementary Figure 3. Sequencing *Cai1* allele amplicons 226
Supplementary Figure 4. KI sectors viewed under a microscope. . 227
Supplementary Figure 5. Mosaicism in some T_0 lines..... 228
Supplementary Figure 6. Sequencing the 4kb bands of pKI1-1, pKI1-3, and pRAD54-3. 229
Supplementary Figure 7. pRAD54-2 KI locus sequencing. 230
Supplementary Figure 8. KI bottom bands sequenced. 230
Supplementary Figure 9. KI T_0 lines screened for the plasmid backbone. 231
Supplementary Figure 10. KI sector emergence over the 7 weeks after co-cultivation. 232
Supplementary Figure 11. Plasmid maps of binary vectors used. . 233
Supplementary Figure 12. The luciferase signal strength in Fig. 3.27 quantified as photon counts. 236
Supplementary Figure 13. Tubers of T_0 line 702..... 237

List of Tables

Table 2.1. Antibiotics used in growing bacteria. 35
Table 2.2. PCR protocols. 40
Table 2.3. PCR product cloning reaction. 40
Table 2.4. Whole insert amplification protocol. 41

Table 2.5. Golden Gate cloning mixtures. 45
Table 2.6. Key plasmids used in cloning. 46
Table 2.7. Selection agents used in Tissue Culture. 57
Table 2.8. Transformed constructs. 58
Table 2.9. DSM correction between dates. 66
Table 3.1. KI rates when treated with 3-MB & SCR7. 102
Table 3.2. Hypothetical inheritance mechanism for transformability. 128

List of Appendices

Appendix 1. Supplementary Figures 225
Appendix 2. R code analysing KI sector counts 237
Appendix 3. R code output of Non-linear Mixed Effects model ... 238
Appendix 4. R code for analysis of tuber shape. Box-Cox data transformation and linear modelling. 239
Appendix 5. Primers and oligonucleotides used. 240

Preface

“The physical forces of inanimate nature are yoked to the service of man; but what of life? How far can we control, how far can we even hope to control the forces which determine the character of living things? […] We cannot, in the brief time at our disposal, do more than indicate some of the lines along which the attack is being led, and some of the successes which have already been achieved.”

Arthur Dendy, 1909

Acknowledgements

I’d like to thank my supervisor Jonathan D. G. Jones for many things: giving me the opportunity to take on this project, affording me the high level of autonomy I desired, his open mindset to my occasionally poorly articulated ideas, and his wise council on all scientific matters. I am grateful to Lila Grandgeorge, who’s work afforded an excellent platform for me to build upon, and to Jasmine Atkins and Ainhoa Valero-Abad for teaching me how to be a mentor and helping me with parts of my project. I’d like to thank the TSL Tissue Culture team whose services I made extensive use of. I am extremely grateful to Mark Youles and Liam Egan from SynBio TSL, to the former for cloning the urgently required module *du jour*, and to the latter for repeatedly providing me with the module I ordered at 16:55 on a Friday. Many thanks to Phil Robinson and Martin Vickers for help with scientific photography and collecting the drone data from my field trial. Thanks to Sara, Justine, and Matt for looking after my oftentimes feeble plants. I am grateful to my contacts at Solynta Ernst-Jan Eggers and Michiel de Vries, who remained thoroughly engaged in my project throughout. Most of all, I am grateful to my partner Lilja – for her unfaltering support and dedication to me these last four hugely stressful months.

List of abbreviations

3’UTR	3’ untranslated region
5’UTR	5’ untranslated region
AR	aspect ratio
BLAST	basic local alignment search tool
Cai1	chlorogenic acid-inducer 1
CTAB	cetyltrimethylammonium bromide
dHA	distal homology arm
dpcc	days post co-cultivation
dpi	days post infiltration
DSM	digital surface model
EBN	endosperm balance number
EDTA	ethylenediaminetetraacetic acid
GCP	ground control point
gDNA	genomic DNA
GTSS	gene targeting-specific selection
HA	homology arm
KI	knock-in
KO	knock-out
Mtf1	MYB transcription factor 1
MYB	*v-Myb* myeloblastosis viral oncogene homolog
ORF	open reading frame
PAR	photosynthetically active radiation
PCR	polymerase chain reaction
PPT	phosphinothricin
PVP	polyvinylpyrrolidone
RGB	red-green-blue
SDS	sodium dodecyl sulphate
SNP	single nucleotide polymorphism
SPXXXXX	seed pack number (e.g. SP10724)
ssDNA	single-stranded DNA
TEs	transposable elements
TF	transcription factor
TIIGER	targeted iterating insertion of genes via recombination

CHAPTER 1. GENERAL INTRODUCTION

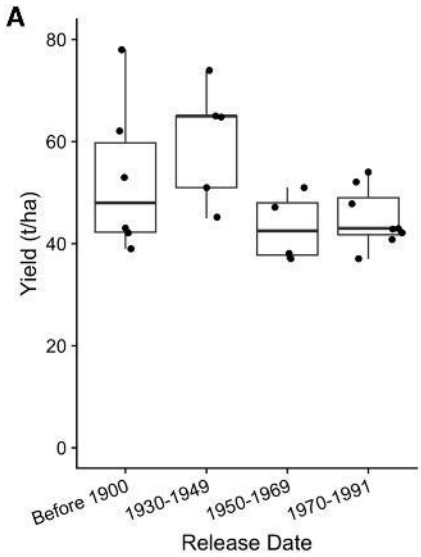


Figure 1.1. Lack of genetic gain in potato varieties.
Source: Agha et al. (2023)

Modern agriculture needs to provide for an increasing world population while confronting the damaging effects of worsening climate change. Potatoes are the third most important food crop after rice and wheat, and are of particular importance to the world's poor which are most exposed to the effects of global warming (FAO, 2019). Worryingly, potato performance has plateaued over the last decades and modern varieties remain vulnerable to widespread and economically important pathogens. Modern biotechnology can help address this and fortify our food systems by leveraging the natural diversity in plant germplasm. In this introductory chapter, I will present the interlocking issues which have plagued potatoes, and the work which has been done to address these. In later chapters, I will describe my efforts to use modern biotechnology to protect the crop from diseases using genetics, rather than agrochemicals, as well as to improve its stagnating productivity.

1.1 The state of potato breeding

Many of the most popular potato varieties are already many decades old, including Maris Piper (1956), Russet Burbank (1873), King Edwards (1902), and Bintje (1904). Over the last century potatoes have seen little genetic gain, at least in terms of yield (Douches et al., 1996) (Fig. 1.1).

This lack of improvement in potato varieties has persisted despite the crop facing considerable challenges. Arguably the main issue for modern potato cultivation is that widely grown varieties are highly vulnerable to a variety of pathogens, the most important of

which is *Phytophthora infestans*. This oomycete is the causative agent behind late blight, a disease which causes more than 1 billion euros in annual losses in the EU alone (Haverkort et al., 2008), and which has been devastating potato crops for more than a century (Fig. 1.2).

The main reason for the crop's lack of progress is that cultivated potatoes are tetraploid and mostly heterozygous on all four haplotypes, which makes the trait selection and recombination fundamental to breeding expensive and laborious, often spanning many years, and requiring the evaluation of thousands of plants resulting from crosses between quadruply heterozygous parents.

Figure 1.2. Late Blight in a potato field
Pictures are of the same field, one week apart.
Source: Fry (2008)

1.1.1 Traditional tetraploid potato breeding

Traditional cultivated potato (*Solanum tuberosum L.*) breeding is made difficult by several compounding issues. Firstly, practically all widely-grown varieties are tetraploid, which makes the mathematics involved in crossing potatoes more complicated (Cai et al., 2012). Secondly, cultivated potatoes are exceptionally heterozygous, which makes the results of a cross difficult to predict. The average number of alleles per locus in tetraploids is 3.2 (Uitdewilligen et al., 2011) and the frequency of single nucleotide polymorphisms is one per 15-30 base pairs (Visser et al., 2014), meaning two chromosomes in a chromosomal pair are four times as genetically distant as the genomes of humans and chimpanzees (Lindhout et al., 2018). Much of this heterozygosity consists of "bad" alleles – versions of genes with deleterious mutations which have been maintained in the tetraploid state due to complementation from the alleles of other haplotypes. Repeated

rounds of self-pollination (i.e. “selfing”) could be used to reduce the degree of heterozygosity but this exposes the “bad” alleles, and inbred lines quickly lose their vigour.

One way to circumvent the issues described above would be to breed potatoes at the diploid level. Diploid landraces do exist, although they aren’t commercially competitive, and it is possible to reduce tetraploids into diploids by pollinating them with specific diploid lines called “haploid inducers”, such as *Solanum phureja* (Peloquin et al., 1996). However, this again results in a loss of heterozygosity and therefore vigour. Furthermore, diploid potatoes are generally “self-incompatible”, meaning they cannot self-fertilize which might allow for the “bad” alleles to be segregated away. This combination of issues has led to potato being described as a “fortified heterozygous out-crosser” (Adams et al., 2022) and it has turned potato breeding into a slow, laborious, and expensive task.

Potatoes’ difficult genetics have indirectly caused further issues, which have impacted the crop’s performance in the field. Since plants cannot be crossed to produce economically viable offspring, potatoes are generally propagated via tubers. These can form uninterrupted lines of somatic tissue with a high genetic load going back decades. In addition, the germplasm available to smallholder farmers is often burdened with a high viral load which lowers yields (Schulte-Geldermann et al., 2012). Using cumbersome and perishable tubers as planting material also imposes significant logistical challenges on the industry, especially in hot climates.

The fact the potato remains an important crop despite the issues described above is a testament to the utility of the crop and potato breeders’ perseverance. More than that, breeders have also been working on adapting potato into a diploid self-compatible crop compatible with a hybrid breeding approach. In the following sections, I will outline the principles of hybrid breeding, and the effort to implement these in potato.

1.1.2 Hybrid breeding

Hybrid breeding is an approach which seeks to benefit from the phenomenon of heterosis, also called hybrid vigour (Shull, 1908). This refers to the fact that the hybrid offspring of two varieties often exhibit superior vigour compared to their parents. In hybrid breeding (recently reviewed by Labroo et al., 2021), highly homozygous parental lines are generated, with desirable traits being introgressed and fixed in at least one of these. Genetically similar parental lines are combined into heterotic groups, which are then intercrossed on a large scale to produce highly heterozygous F_1 seeds, which are grown in the field to capitalize on heterosis. The magnitude of heterosis between two heterotic groups is referred to as their combining ability and this guides the breeder's selection of parents and hybrid combinations (Sprague & Tatum, 1942). Curiously, it is still not entirely clear what the mechanism behind heterosis is (a topic recently reviewed by Hochholdinger & Yu (2025)). The hybrids' superior performance has alternately been ascribed to dominance of one parent's alleles over the deleterious alleles of another (Dominance hypothesis), the superior functionality of heterozygous loci over either homozygous state (Overdominance hypothesis), as well as positive interactions between genes from different parents (Epistasis hypothesis). Notably, these hypotheses are not mutually exclusive and the total impact of heterosis is likely due to some combination of the proposed contributing factors.

The posterchild for hybrid breeding is maize, where the approach was first demonstrated. Maize has proven particularly well-suited for hybrid breeding, as it is wind-pollinated, and the male and female flowers are located on different organs, facilitating control of pollination. As reviewed by Duvick (2001), the practice has transformed maize productivity (Fig. 1.3), with field yield increasing at a rate of ca. 100 kg/ha/year (Cooper et al., 2014; Di Matteo et al., 2016). Notably, Duvick (1999) compared the performance of inbred lines and their hybrids and found the level of heterosis has not increased substantially between 1930 and 1980. This suggests that the gradual improvement seen in maize varieties is the result of purging "bad" alleles from the

germplasm and the introduction of novel “good” ones, rather than the strengthening of the factors underpinning heterosis.

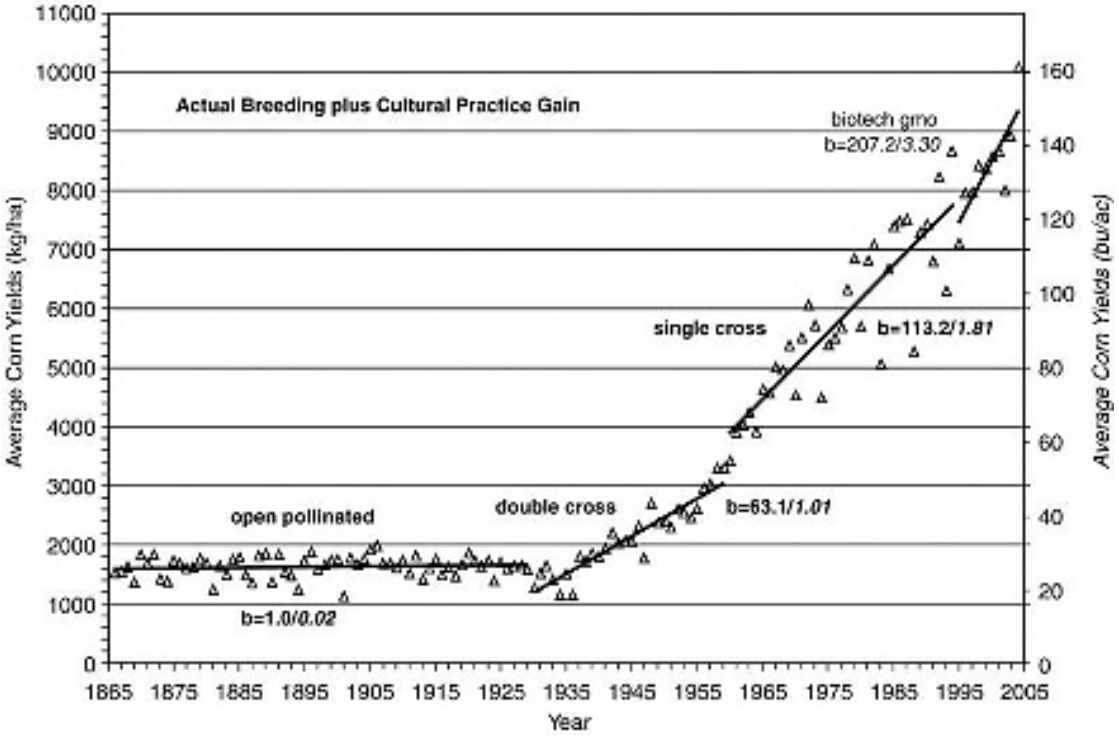


Figure 1.3. Maize yields since the introduction of hybrid breeding.
Double- and single-cross refer to alternative methods of hybrid breeding. b is the rate of improvement. Source: Troyer (2006).

More recently, rice has also benefitted from hybrid breeding. Rice’s propensity to self-fertilise initially precluded the practice, until male-sterile lines were identified (L. Yuan, 1966). Yields have since increased from 3-5 t/ha with traditional rice varieties to 16 t/ha with modern hybrid rice (L. Yuan, 2017). As is discussed in the next chapter, several companies and programs are betting that potato breeding is poised to follow the same trajectory.

1.1.3 Diploid hybrid potato breeding

The idea to transform potato into a diploid self-compatible crop, which might then be used in hybrid breeding, was proposed almost 70 years ago (Hougas & Peloquin, 1958). This was effectively prevented by potatoes’ self-incompatibility until Hosaka & Hanneman (1998) identified a self-compatibility allele (*Sli*). This allowed for the diploid potatoes to be repeatedly selfed,

progressively reducing their level of heterozygosity. Unfortunately, this resulted in major inbreeding depression due to the exposure of deleterious alleles which precluded further breeding (Birhman & Hosaka, 2000; Phumichai et al., 2005). To overcome this, Lindhout et al. (2011) combined the *Sli* allele with elite diploid potato lines which were developed at Wageningen University (Hutten, 1994), resulting in lines which retained reasonable vitality after several rounds of selfing. This work has since been taken forward by a company (Solynta), who have been able to achieve significant improvements in diploid productivity (Adams et al., 2022; Stockem et al., 2020). The new and improved breeding practices have also been used to stack two resistance genes against *P. infestans*, resulting in field resistance (Y. Su et al., 2020). Two established potato breeding companies, HZPC and KWS, are also pursuing diploid hybrid potato breeding programs.

In summary, steady advancements in the last 25 years have successfully addressed the major technical hurdles to implementing diploid hybrid breeding in potatoes. This promises to deliver the gradual yet comprehensive improvements in agricultural performance that have long eluded breeders. However, even in crops with efficient breeding practices, durable disease resistance can be difficult to achieve. In the next chapter, I will outline the central elements of plant immunity and the challenges of creating disease-resistant plants.

1.2 Plant Immunity

The current paradigm of plant disease resistance is based on the Zig-zag-zig model of plant immunity (Fig. 1.4) (Jones & Dangl, 2006). In brief, plant cells detect pathogens via highly conserved molecular hallmarks (called Pathogen-Associated Molecular Patterns, or PAMPs) or tell-tale signs of the damage done during

infection (termed Damage-Associated Molecular Patterns). These are perceived through Pattern Recognition Receptors (PRRs) located on the plant cells' plasma membrane, initiating a broad-spectrum immune response dubbed Pattern-Triggered Immunity (PTI).

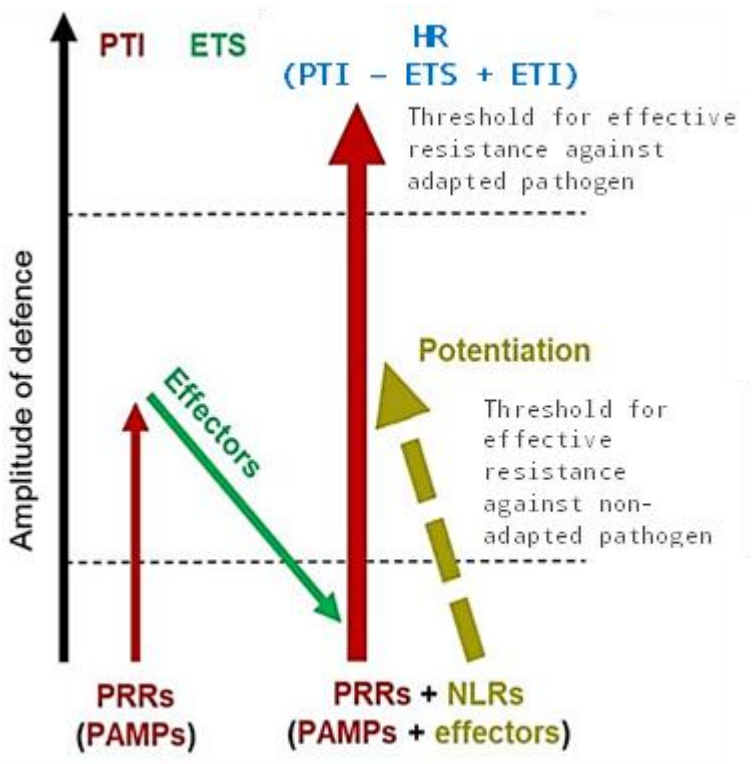


Figure 1.4. The Zig-zag-zig model of plant immunity.
Adapted from: Pok Man Ngou et al. (2021).

PTI is usually enough to deal with non-adapted candidate pathogens, but well-adapted pathogens have coevolved with their host plants and developed means to supress these defences. They do this by secreting Effectors into the host, which disable PTI by targeting key players of the molecular signalling cascades that underpin it (Ceulemans et al., 2021). If successful, these adapted pathogens can induce Effector-Triggered Suppression (ETS) of the host's immune response, and then the infection can develop further, and the plant becomes susceptible.

To prevent this, plants deploy intracellular receptors from the Nucleotide-binding domain and Leucine-rich Repeat (NLR) class, which detect the presence of pathogen effectors using a variety of strategies, including "guarding" key PTI signalling hubs, serving as target decoys, or by directly interacting with the effector molecules (Khan et al., 2016). When triggered, NLRs induce an immune response, termed Effector-Triggered Immunity (ETI). They execute this in one of two main modes: through unleashing an intrinsic NADase activity, which depletes cellular NAD+ and

generates small immunogenic signalling molecules (Ma et al., 2020; Martin et al., 2020), or by forming cation channels in the plasma membrane, enabling Ca^{2+} influx into the cell (Bi et al., 2021; J. Wang et al., 2019). Notably, PTI and ETI do not function independently but engage in mutual potentiation, which results in a reinforced immune response and extensive cell death, a phenomenon called a Hypersensitive Response (HR) (Pok Man Ngou et al., 2021). In this case, the infection is stopped or suppressed, and the plants gain resistance to the pathogen. Genes which confer resistance against a pathogen like this are also called resistance genes (R-genes), and they are valuable in resistance breeding.

1.2.2 The rise of Effectoromics

Since the turn of the millennium, our view of the genes taking part in this molecular “arms race” has rapidly expanded. Next Generation Sequencing allowed whole genomes to be sequenced relatively easily, including those of important *Solanaceae* crops (D. Tang et al., 2022; van Lieshout et al., 2020; X. Xu et al., 2011) and pathogens like *P. infestans* (Haas et al., 2009). Crucially, both immune receptors and Effectors have features which set them apart from other genes. Plant PRRs and NLRs each have typical overall domain structures, while pathogen effectors often share protein motifs, such as the RXLR motif found in the cytoplasmic effectors of *P. infestans* and other oomycetes (Armstrong et al., 2005; Ellis et al., 2009). This enables genome sequences to be scanned for potential R-genes or Effectors, which could then be corroborated with RNA expression profiles and cloned into libraries, where interactions could be established through co-expression (Ellis et al., 2009; Vleeshouwers et al., 2011). This suite of high-throughput approaches was dubbed “Effectoromics”, and it has greatly increased the number of known R-genes and Effectors. A notable advance was the employment of sequence capture techniques to sequence only that part of genomes relevant to immunity (Cronn et al., 2012). This was integrated into three advanced platforms for the high-throughput capture and sequencing of NLRs, PRRs, and Effectors, respectively called RenSeq, RLP/KSeq, and PenSeq (Jupe et al., 2013; Lin, Armstrong, et al., 2020; Lin, Song, et al., 2020). All three of these were

deployed in Solanaceous plants or *P. infestans* and identified genes relevant to potato resistance breeding.

1.2.3 Durable field resistance

Effectoromics has revealed much about the determinants of a crop's resistance to an adapted pathogen. However, it has not provided a comprehensive solution to the issue of crop susceptibility. In some ways, this gain in understanding highlighted a more fundamental unknown: why are some plant species susceptible to a pathogen, while most are not?

Numerous potato relatives are non-hosts for *P. infestans* (Vleeshouwers et al., 2011). The understanding and deployment of such durable "non-host" resistance for crop protection has been a long-standing research goal. Work on the *P. infestans* non-hosts pepper (*Capsicum annum*) and American black nightshade (*Solanum americanum*) has thus far shown that non-hosts' robust resistance is at least partially due a large number of NLRs which recognise *P. infestans* effectors (Oh et al., 2023; Witek et al., 2016, 2020). McLellan et al. (2022) also showed that quantitative resistance can result from substituting a host's endogenous Effector-targeted gene with a homologue from a non-host. Thus, it seems the evolutionary arms race between host and pathogen, which could also be described as an "evolutionary treadmill", could be overcome through sufficiently diverse robust NLR surveillance, Effector-target incompatibility, or most likely a combination of both. Deploying such durable resistance, particularly against more than one pathogen, would require rapid and efficient transfer of numerous immunity-implicated genes between crops and non-hosts species. In the next section, I will discuss the various way this can be achieved and the inherent challenges therein.

1.3 Horizontal gene transfer

1.3.1 Introgression

Introgression is the traditional avenue for horizontal gene transfer between crops and their relatives. Potato relatives from the *Solanum nigrum* complex, including *S. nigrum* itself, *Solanum villosum, Solanum americanum*, and others, are a rich potential

source of R-genes (Horsman et al., 1997; Witek et al., 2016; Witek et al., 2021). However, breeding barriers have made introgressing these resistances challenging (Eijlander & Stiekema, 1994; Horsman et al., 1997), as crosses can fail or produce infertile offspring. Wild relatives can also differ in their ploidy level, necessitating ploidy reduction or increase prior to introgression which can negatively affect fertility (Barandalla et al., 2006; Hougas & Peloquin, 1960). A further complication is that in the *Solanaceae* many of the crosses which should work, as suggested by the ploidy level, do not, and some which should not, do. This phenomenon gave rise the Endosperm Ratio Hypothesis, which posits that each species has a genome-specific effective ploidy, also referred to as an Endosperm Balance Number (EBN), and that normal development of the triploid endosperm requires a 2:1 ratio of maternal to paternal EBNs (Carputo et al., 1999). Crucially, this effective ploidy can differ from the chromosomal ploidy, erecting a breeding barrier between species which might share a ploidy level but differ in EBN, such as the diploid *S. tuberosum* (2× EBN=2) and some wild relatives like *S. cardiophyllum* (2× EBN=1) (Johnston & Hanneman, 1982). To date, the mechanism which underpins the Endosperm Ratio Hypothesis remains unknown but its usefulness in guiding potato breeding efforts has been convincingly demonstrated (Masuelli & Camadro, 1997). The EBNs of many *Solanum* species have been experimentally established and found to differ from those of diploid and tetraploid potatoes (Spooner et al., 2014). This effectively segments the genus' gene pool and necessitates the use of complex bridge crossing schemes to introgress genes from some wild relatives into potato (Jansky & Hamernik, 2009).

Furthermore, even if a hybrid is obtained, recombination rates between the chromosomes can be low or non-existent, requiring the use of complicated breeding schemes and "bridge" parents to recover progeny stably bearing the gene of interest (Hermsen, 1994). Finally, even successfully introgressed R-genes can suffer from "linkage drag" – genetically linked "bad" alleles which can be difficult to segregate away from the desired allele.

Efforts to introgress R-genes against *P. infestans* into tetraploid potatoes have been ongoing for decades. Some relatives, like *S. bulbocastanum* and *S. microdontum* are readily crossable with potato and have been widely employed by the potato breeding industry (Bethke et al., 2017; Rakosy-Tican et al., 2020). However, the R-genes identified in these species have largely been overcome. Other parts of the gene pool where potent R-genes are known to reside have proven less accessible. For example, attempts to cross members of the *S. nigrum* complex, such as *S. americanum*, with potato have failed unless techniques like embryo rescue were employed, but even then the resulting hybrids were sterile and could not be used for further introgression (Eijlander & Stiekema, 1994). Horsman et al. (1997) had more success by generating somatic hybrids between members of the *S. nigrum* complex and potato. These were fertile and backcrosses to both parents were obtained, however the progeny's fertility and vigour was poor and this precluded further introgression (Horsman et al., 1999). Thus far, introgression has failed to deliver durable sources of resistance to *P. infestans* into the potato germplasm (Fry, 2008).

In contrast, introgression into self-compatible diploid potato has seen some success. Su et al. (2020) demonstrated the introgression of four anti-late blight R-genes (*Rpi-avl1, Rpi-tar1, Rpi-chc1*, and *Rpi-vnt1*) and their subsequent combination into double "stacks" via crosses which resulted in resistance in the field. However, the resistant lines produced were experimental lines and delivering these R-genes into elite cultivars would still be challenging. Ultimately, introgression remains a useful but slow and laborious method.

1.3.2 Transgenesis

A straightforward solution to potatoes' complicated genetics is to bypass traditional avenues of gene transfer and use transgenesis to move genes between species. This generally involves delivering a gene of interest alongside a selectable marker, usually an antibiotic-resistance gene. Several delivery methods are routinely used in potatoes, including *Agrobacterium tumefaciens* transformation (Wang et al., 2020), particle bombardment (Romano et al., 2001), and protoplast transformation (Nicolia et al.,

2021). Among these, *A. tumefaciens* transformation is particularly effective. As part of its life cycle, *A. tumefaciens* integrates a DNA segment known as "transfer DNA" (T-DNA) into the plant genome (Schell & Van Montagu, 1977). This natural process has been adapted to efficiently transfer even large sequences into the plant genome, flanked by "Left Border" and "Right Border" sequences 25 bp-long. Cells that receive the specified DNA are selected on antibiotics in tissue culture, and the resulting callus is used to recover Genetically Modified (GM) shoots. In this way, transgenesis allows the simultaneous introduction of multiple genes from different sources with no linkage drag, enabling the stacking of multiple R-genes in elite cultivars (Ghislain et al., 2019; Jacobsen & Schouten, 2007; Jo et al., 2014).

A major limitation of the method is that T-DNA integration occurs at random locations in the genome and this has several consequences. Firstly, this results in a variable genetic context for the inserted genes, which leads to differing levels of expression and necessitates the selection of well-expressing GM lines (Peach & Velten, 1991). Secondly, the insertion can occur within an important gene and disrupt its function, as occurred in one of the lines of Golden Rice created by Bollinedi et al. (2017). Thirdly, multiple T-DNA integrations can occur, resulting in complex events which are prone to transgene silencing (Jones et al., 1987). While all issues can be avoided with careful selection from among sufficient candidate lines, the genes' non-targeted integration restricts what can be done with transgenics, makes the process more laborious and expensive, and is especially limiting when interrogating scientific questions.

Furthermore, GM crops often fall foul of public opinion and face restrictive legislation. Thus, arguably transgenesis' greatest limitation is legal in nature, and limits the deployment of GM plants to field trials in many parts of the world (Dima & Inzé, 2021), although attempts at liberalising the current regulatory regimes are ongoing in the UK (Genetic Technology (Precision Breeding) Act, 2023) and the EU (European Parliament, 2024).

1.3.3 Genome Editing

Genome editing (GE) involves the targeted modification of genomes *in vivo*, including the targeted insertion of genes. However, the efficiency of plant GE varies drastically depending on the desired outcome, and targeted gene insertion has proven challenging.

GE generally involves two sequential processes: (i) the introduction of double-stranded breaks (DSBs) in the cellular DNA and (ii) the leveraging of error-prone endogenous DNA repair mechanisms that create insertions, deletions, substitutions or other changes. Because of this, progress has largely depended on two underlying lines of research: the development of programmable nucleases and detailed understanding of DNA repair mechanisms.

Programmable molecular scissors

The development of nucleases, which can be directed to selectively cut at loci with a particular sequence, has been fundamental to GE. The first two generations of truly programmable endonucleases came in the form of Zinc-Finger Nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). Both of these consist of DNA-Binding Domains (DBD) and a *FokI* nuclease domain; upon binding of two different specified protomers, each to adjacent sequences, the *FokI* monomers on each dimerise, activate, and induce a DSB in the DNA (Christian et al., 2010; Kim et al., 1996). While ZFNs and TALENs enabled scientists to induce specified DSBs, they also have considerable limitations. ZFNs have a complicated multi-domain structure which makes reprogramming difficult, a binding activity which is hard to predict during design (Joung & Sander, 2013), interference between some zinc-finger domains (Isalan et al., 1997), and a degree of cytotoxicity (Cornu et al., 2008). Ultimately, ZFNs have only been used to edit a small selection of crops, and potato is not one of them (Hameed et al., 2018). TALENs are superior in several aspects, being less cytotoxic (Mussolino et al., 2011), and having a simpler structure which makes reprogramming easier, although still not easy (Joung & Sander, 2013). TALENs also have a superior track record, with multiple studies demonstrating efficient gene editing in potato (Forsyth et al., 2016; Nicolia et al., 2015; Yasumoto et al., 2019). Nevertheless, both technologies have now been superseded.

Most GE today is based on Clustered, Regularly Interspaced, Short Palindromic Repeats (CRISPR) and CRISPR-Associated Proteins (Cas). In a seminal paper, Jinek et al. (2012) described a minimal system involving the Cas9 protein from *Streptococcus pyogenes* and a single guide RNA (sgRNA), capable of inducing DSBs at genomic loci complementary to a 20 bp "guide" region of the sgRNA. This elegant system, requiring only the change of the sgRNA sequence to reprogramme the endonuclease, allowed biologists to cut almost any locus, provided that a short 5'-NGG-3' motif is present next to the target site, called a Protospacer-Adjacent Motif (PAM). Notably, if several sgRNAs are provided, the simultaneous targeting and editing of multiple loci is possible (Armario Najera et al., 2019; Nakayasu et al., 2018). Cas9's impressive capabilities also prompted an energetic search for alternative Cas endonucleases which could also be employed in gene editing. One notable example is Cas12a (formerly Cpf1), which Zetsche et al. (2015) found has an analogous structure and function to Cas9 and induces DSBs at loci with a 5'-TTTV-3' PAM if they are complementary to its guide RNA (in Cas12a's case usually referred to as a CRISPR RNA, or crRNA). While both systems were initially demonstrated in mammalian cells, both have also been used in editing plants (Nekrasov et al., 2013; X. Tang et al., 2017).

DNA repair pathways

Regardless of the endonuclease used, all the above technologies ultimately generate a lesion in the genome which needs to be repaired. Below, I will briefly cover the main projects relevant to this project. For a comprehensive overview of DNA repair and its role in GE, see Schmidt et al. (2019).

DSBs in somatic plant cells are predominantly repaired via non-homologous end joining (NHEJ, also called canonical NHEJ), and this is also the case in diploid (Butler et al., 2015) and tetraploid potato (Butler et al., 2015; S. Wang et al., 2015). In NHEJ, the two sides of a DSB are directly ligated, which often results in small-scale deletions or insertions (indels) at the break site. This is a "messy" repair mechanism and when such indels occur within gene coding sequences, they often result in frameshift mutations. This is commonly used to inactivate genes of

interest, also referred to as a gene “knock-out” (KO). To date, this is the only form of GE to have become routine in a broad spectrum of plants (Puchta, 2017).

Another DNA repair pathway is microhomology-mediated end joining (MMEJ, also called alternative NHEJ). In MMEJ, the two strands are resected in a 5’→3’ direction, revealing single-stranded microhomologies which in plants tent to be 6-18 bp in length (Kwon et al., 2010; Maréchal et al., 2009; Parent et al., 2011). Once revealed, these can anneal and reconnect the two strands, after which any gaps are filled in by DNA Polymerase θ. MMEJ is highly error-prone and often results in the deletion of the sequence between the microhomologies. It plays an important role in T-DNA integration (van Kregten et al., 2016), indicating that it is an active pathway in plant cells, but it has been little utilised in GE.

Alternatively, DSBs can be repaired via homology-directed repair (HDR). Crucially, this process differs in that it can use homologous sequences *in trans* to restore the sequence at the point of damage. In HDR, DSB sites are again subject to 5’→3’resection, generating 3’-overhangs. If homologies flank the break site, these can anneal and resolve the DSB in a process called single strand annealing, which is reminiscent of MMEJ. In the absence of such local homologies, the 3’-overhangs can invade distal regions of homology in a process called strand invasion. Here, the invading strand can bind the complementary strand and elongate, thus incorporating the sequence of the homologous locus. After filling in the missing sequence, the invading strand is released and realigns with its original partner strand, thus resolving the DSB in a conservative manner. This process is called synthesis-dependent strand annealing, and due to its utilisation of a repair template, it can be employed to introduce desirable changes at the targeted locus, provided a custom template bearing the modifications is available.

Gene targeting by homologous recombination

Gene targeting (GT) broadly refers to GE which relies on HDR to bring in novel sequence or modify the pre-existing sequence at a specific locus in a controlled and pre-determined manner. This

generally involves introducing a programmable nuclease alongside a repair template bearing the desired end sequence. Depending on the desired outcome, Huang & Puchta (2019) identified several distinct types of GT:

Targeted Insertion: entails introducing genetic elements, such as whole genes, promoters, or tags, at a specific locus. This approach is often referred to as a gene “knock-in” (KI). KIs are particularly relevant to this work as they could allow for the deployment of R-genes across species barriers.

Targeted Substitution: involves changing one or a group of nucleotides at a specific locus to a predetermined set of alternative nucleotides.

Targeted Replacement: refers to replacing a specific DNA sequence of substantial length with an entirely different predefined sequence, effectively swapping one segment for another.

KIs in plants could allow for the versatile deployment of R-genes while avoiding some limitations of GM. If genes can be targeted, they could be stacked at the same locus over multiple transformation events, ensuring concerted inheritance. Also, the genes’ genetic context would remain consistent, resulting in less variable levels of expression. Furthermore, targeted gene delivery could result in a locus without T-DNA borders or other scar sequence, resulting a “cleaner” event.

However, methods relying on HDR face a major challenge, as it rarely occurs in somatic plant cells. The first attempt at GT in plants was reported by Paszkowski et al. (1988), who transformed tobacco protoplasts carrying a disrupted *neomycin phosphotransferase II* (*NptII*) gene with a complementary, partially overlapping DNA fragment. They hypothesised that if homologous recombination occurred, it would restore the functional *NptII* gene, thereby conferring kanamycin resistance to the plant. While successful, the observed frequencies were extremely low ($0.5–4.2 \times 10^{-4}$), far below those reported in yeast, fungi, or certain mammalian cells. A significant improvement came from Puchta et al. (1996), who showed that introducing DSBs at the target locus increased GT frequencies ($2.2–18.8 \times 10^{-3}$). Using a similar

approach to Paszkowski et al. (1988), they cleaved the disrupted *NptII* with the *I-SceI* meganuclease, enhancing recombination efficiency. Despite persistent efforts over the next two decades, further increases in GT rates remained limited.

Due to the lack of progress on boosting HDR rates and GT efficiency, attempts were also made to conduct similar edits via NHEJ. This is especially true in rice, where attempts have been made at targeted insertions (J. Li et al., 2016), targeted substitutions (J. Li et al., 2016; S. Li et al., 2018), and targeted replacements (Weinthal et al., 2013). While these approaches showed some success, efficiencies remained too low for practical use. More recent strategies have attempted to bypass double-strand breaks (DSBs) altogether. Anzalone et al. (2019) introduced Prime Editing, which combines a Cas9 nickase, which cuts only one DNA strand, with a reverse transcriptase which can introduce novel sequence into the cut site. Prime Editing can introduce precise indels and substitutions, but these are short, usually less than 100 bp long, and thus cannot introduce whole new genes. This method has seen multiple refinements and performs well in rice (Choi et al., 2022; Q. Lin et al., 2020, 2021; Nelson et al., 2021), but less effectively in potato (Perroud et al., 2022). Notably, Sun et al. (2023) applied Prime Editing to introduce recombinase recognition sites, which were subsequently used for large gene insertions, though such gene delivery does retain scar sequences. More recently, Liu et al. (2024) accomplished something similar by fusing a catalytically dead Cas9 to a transposase, enabling targeted gene delivery as a ‘domesticated’ transposon, although off-target insertions could also be detected. The pace of innovation in this field illustrates the perceived value and utility of making GT work efficiently in plants. However, each method has limitations, and none have fully realized the potential of plant GT. To date, conducting any form of GT in plants remains a considerable challenge.

1.4 Conclusion

The last 50 years have seen immense progress in the field of biotechnology, with techniques like transgenesis and gene editing promising to push genetics to the forefront of how we protect crops from pathogens. Potatoes have concurrently been developed from a “fortified heterozygous out-crosser” (Adams et al., 2022) with tetraploid genetics, to a diploid self-compatible species well-positioned to benefit from both hybrid breeding schemes and modern bioengineering. Work on this topic has become even more pertinent as of the writing of this thesis, due to the growing acceptance of GM and GE in numerous countries due to growing climactic and populational pressures. In the United Kingdom, the recently passed Precision Breeding Act (Genetic Technology (Precision Breeding) Act, 2023) has made it possible to do field trials with GE crops and ultimately promises to open an avenue for bringing to market plant varieties generated through such methods.

1.5 Project aims and objectives

In this PhD project, I have worked towards several goals, united by the theme of improving diploid potato germplasm. The first was the establishment of an efficient method for the KI of R-genes from wild potatoes and their relatives, across breeding barriers and into elite diploid lines; an approach referred to here as KI breeding. The second goal is to build on the results of the previous line of work and establish an efficient method for the targeted iterative insertion of genes via recombination (TIIGER) at the same locus. Lastly, I used gene editing to generate a potentially valuable compact dwarf allele in potato and to document its performance in the field.

CHAPTER 2. MATERIALS AND METHODS

2.1 Materials

Bacterial material

I used *Escherichia coli* strain DH10B (Grant et al., 1990) for all cloning. This I grew on liquid or solid L media containing appropriate antibiotics at 37°C over 16 hours. I prepared electrocompetent DH10B cells in 50 µL aliquots following a protocol kindly provided by Mark Youles, originally published in Sambrook et al. (1989).

Stable and transient expression of genes in plants was done using *Agrobacterium tumefaciens* strain AGL1 (Lazo et al., 1991). This was grown on liquid or solid L media containing Rifampicin, Carbenicillin, and other appropriate antibiotics at 28°C over 48 hours.

Plant material

Nicotiana benthamiana (LAB isolate, as described by Bally et al. (2015)) and *Nicotiana tabacum* cv. Petite Gerard plants used for transient expression experiments were provided by JIC Horticultural Services. These were grown in *JIC MULTIPURPOSE + GRIT* (90% peat, 10% grit, 4 kg/m^3 dolomitic limestone, 0.75 kg/m^3 PG mix, 1.5 kg/m^3 OSMOCOTE™ Bloom) in controlled environment rooms at 24°C, 16 h photoperiod, 55% relative humidity. Other plants were grown in a glasshouse in pots (3 L) with *JIC MULTIPURPOSE + GRIT* under supplemental lighting.

Seed tubers from potato varieties Swift and Cara were obtained from D.T. Brown Seeds™. Once delivered, these were stored in a refrigeration unit (4°C) until the tubers were planted in the field.

Several packs of diploid potato seed were kindly provided by Solynta, two of them labelled “B26” and “B100”. While this seed’s exact pedigree is unknown, it originated from the early stages of Solynta’s diploid breeding programme and thus segregates for many loci.

Matthew Smoker germinated seed from the B26 seed pack and compared the resulting lines' rate of transformation via *A. tumefaciens* co-cultivation. He identified one line which transformed particularly well and designated this "B26F", referred to henceforth simply as B26. In the glasshouse, B26 grows slowly and not very large. Vegetative growth is light green in colour, lacking any obvious anthocyanin pigmentation. The leaves are small and heart-shaped, with an uneven surface. Strikingly, B26 spontaneously aborts most flower buds, resulting in poor fertility. When buds are retained, they produce white flowers. Berries, which are set very rarely, are small, green, and tend to hold few seeds which germinate poorly. Tubers are set readily, and these have yellow skin and flesh.

Seed from the B100 seed pack was germinated and grown in the glasshouse. One of these plants was designated as line B100, although it is not known exactly what criteria informed this choice. In the glasshouse, B100 grows vigorously. Vegetative growth is again light green in colour, lacking any obvious anthocyanin pigmentation. The leaves are large and elongated, with an even surface – this makes B100 readily distinguishable from B26. B100 flowers abundantly and sets many berries. The berries are large, green, and hold many seeds which germinate well. Tubers are set readily, and these have yellow skin and flesh. B100's rate of transformation via *A. tumefaciens* co-cultivation was also assessed by Matthew Smoker and found to be poor.

Kits

I miniprepped bacterial colonies using the NucleoSpin Plasmid Kit (Macherey-Nagel, REF 740588.50), which is based on the alkali lysis method (Bimboim & Doly, 1979). In brief, bacterial cultures were grown in 10 mL tubes of L broth plus appropriate antibiotics. These were pelleted and resuspended in an isotonic solution with RNAse A. To lyse the bacteria, an alkaline solution containing SDS and sodium hydroxide was added, and the samples were incubated (5 min; RT). The lysed cell mixtures were neutralised with the addition of potassium acetate (pH 5.2), which precipitated out the bacterial proteins and genomic DNA. The precipitates were pelleted by centrifugation and the supernatants were loaded onto silica

membrane columns. The membranes were washed with 70% ethanol, before the plasmids were eluted in the provided AE buffer (5 mM Tris/HCl, pH 8.5).

I cleaned-up PCR reactions and excised agarose gel bands using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, REF 740609.50), which is based on the method for nucleic acid purification described by Boom et al. (1990). In short, a buffer containing the chaotropic salt guanidinium thiocyanate was added to the samples. If the samples were excised gel bands, then they were incubated at 50°C to dissolve the agarose. The samples were then loaded onto silica membrane columns, which bind nucleic acids in the presence of concentrated chaotropic salts. The membranes were washed with 70% ethanol, before the DNA was eluted in the provided NE buffer (5 mM Tris/HCl, pH 8.5).

Buffers and media

L media	LSR Broth	LSR1 Medium	LSR2 Medium	MS Medium	Infiltration buffer
10 g/L Tryptone 5 g/L Yeast extract 5 g/L NaCl 1 g/L D-glucose Solid medium made with 1% agar	1X Murashige and Skoog medium 3% sucrose pH 5.7	1X Murashige and Skoog medium 3% sucrose 2.0mg/L zeatin riboside 0.2mg/l NAA 0.02mg/L GA_3 0.6% agarose pH 5.7	1X Murashige and Skoog medium 3% sucrose 2.0mg/L zeatin riboside 0.02mg/l NAA 0.02mg/L GA_3 0.6% agarose pH 5.7	1X Murashige and Skoog medium 2% sucrose 0.6% agarose 100mg/L casein acid hydrolysate pH 5.7	10 mM $MgCl_2$ 10 mM MES 200 µM pH 5.6 + Sylwet® L-77 (0.1 µL/mL) if vacuum-infiltrating

Chemicals and antibiotics

Antibiotic	Concentration used (mg/L)	
	E. coli	A. tumefaciens
Kanamycin	50	50
Spectinomycin	50	50
Carbenicillin	100	100
Streptomycin	10	-
Rifampicin	-	50

Table 2.1. Antibiotics used in growing bacteria.

Antibiotics were all prepared as aqueous stock solutions at stored at -20°C, except for Rifampicin which was dissolved in methanol and stored at 4°C. Added to media as necessary.

Small molecule effector compounds 3-methoxybenzamide (3-MB; Sigma-Aldrich; M10050) and SCR7 (Sigma-Alrich; SML1546), whose effect on the frequency of gene targeting was assessed in Chapter 3.2.4, were dissolved in DMSO at stock concentrations of 1 M. These were then stored at -20°C and added to media as necessary.

2.2 Methods

2.2.1 Plant methods

Potato seed harvesting

Berries ripened in a paper bag with banana peel until soft. Seeds were then cleaned of fruit flesh, washed with water, soaked in Na_3PO_4 (10% w/v, 10 min), rinsed with water and dried on filter paper. Once dry, the seeds were stored in paper envelopes at room temperature.

Tubers were harvested after desiccating the plant pots for a week. After being dug up, tubers were allowed to air dry for two days before being stored in paper bags at 4°C.

Seed germination

Seeds were treated with GA_3 (2 mg/mL, 48-72 hours). They were then surface sterilized with ethanol (70%, 2 min), then thick bleach (diluted tenfold in water, 20 min, shaking), and finally rinsed with dH_2O in a laminar flow hood and plated on MS media with GA_3 added (0.5 mg/L). Germination was allowed to take place over 2 weeks before moving everything to soil.

Tuber sprouting was encouraged by keeping them on a bright windowsill, at RT, plus a weekly GA_3 spray (100 mg/L) until the tubers had started to sprout.

Diploid potato crossing

I conducted the crosses manually, usually early in the morning. Firstly, the female parent was emasculated by removing the anthers with a pair of forceps. This was done right before the flower opened to avoid self-pollination. Pollen was harvested from the male parent by inserting the still-attached anthers into a 1.5 mL microfuge tube and vibrating the flower with an electric toothbrush. The collected pollen was applied to the stigma of the female parent with a paint brush. Lastly, the pollinated flower

was tagged proximal to the abscission zone at the base of the pedicel, so that should a berry form and fall off it would remain tagged.

I obtained selfed berries from fertile lines like YD1-3 by bagging whole inflorescences and periodically tapping them to release pollen from the anthers. This would usually result in multiple berries forming inside the bag.

Microscopy

I took pollen images with a phone down the eyepiece of a Wilovert stereomicroscope at 100X magnification. Pollen was harvested, germinated *in vitro*, and stained with Lugo's solution as described by Ordoñez (2014).

2.2.2 Nucleic acid methods

Amplicon sequencing

Cai1 amplicons from the Solynta lines were gel-extracted (PCR Clean-up kit, Macherey-Nagel), blunt end-cloned (Zero Blunt® TOPO® PCR Cloning Kit, Invitrogen) and Sanger sequenced (GeneWiz). Later, amplicons were amplified with primers containing Golden Gate-compatible restriction sites and overhangs, thus allowing for their cloning using the protocol described in Section 2.2.4. Finally, amplicons would directly be sent for Nanopore sequencing using Plasmidosaurus' "Premium PCR" service. The raw reads would be obtained as a FASTQ file which was then mapped to the expected locus in Geneious.

DNA extraction

Genomic DNA was extracted using a modified version of the CTAB extraction protocol published by Inglis et al. (2018). That protocol is notable for including a pre-wash step with a sorbitol-containing buffer, which helps remove interfering compounds like polysaccharides and polyphenols. I modified the protocol by using polyvinylpyrrolidone (PVP) with an average molecular weight of 10,000 g/mol, rather than the usual 40,000 g/mol, as some authors have suggested this is less likely to coprecipitate with DNA and thus contaminate it (Schenk et al., 2023). I also used phenol: chloroform: isoamyl alcohol 25:24:1 (Sigma-Aldrich,

P2069), instead of just chloroform: isoamyl alcohol, to better remove protein contaminants.

1. Collect fresh tissue (50 – 100 mg) in 2 mL Eppendorf tubes and add 2 tungsten carbide beads/tube. Flash-freeze and shake in a Genogrinder (1,250 rpm, 1min).

Sorbitol wash buffer	
Reagent	**Final concentration**
Tris-HCl pH 8.0	0.1 M Tris/HCI
EDTA	5 mM EDTA
Sorbitol	0.35 M
PVP (Avg. MW 10,000)	1 % (w/v)
2-mercaptoethanol	1 % (v/v)
Extraction buffer	
Reagent	**Final concentration**
Tris-HCl pH 8.0	0.1 M Tris/HCI
EDTA	20 mM EDTA (pH 8.0)
NaCl	3 M NaCl
CTAB	3% CTAB
PVP (Avg. MW 10,000)	1 % (w/v)
2-mercaptoethanol	1 % (v/v)

2. Put back on ice and add Sorbitol wash buffer (1 mL per tube). Cap tubes and shake/vortex. Inspect to confirm suspension of the powdered material and shake again, if necessary.

3. Centrifuge (2,500 g, 5 min, 4°C).

4. The supernatant is aspirated and discarded.

5. (Optional) Repeat the sorbitol wash if the supernatant from step 3 is highly turbid or viscous. Resuspend and centrifuge as before. Discard supernatant.

6. Add pre-warmed Extraction buffer (800 µL) to the sample tubes and shake/vortex.

Note: The extraction buffer is viscous and can be hard to pipette.

7. The tubes are heated (65 °C, min 1 hour) and mixed occasionally (every 15-20 minutes). Allow the samples to cool down to room temperature for 5 min.

8. Add phenol: chloroform: isoamyl alcohol (800 µL, 25:24:1, pH > 7.5) to the samples and mix vigorously (10s), then centrifuge (2,500 g, 20 minutes).

9. Carefully remove the upper aqueous phase (ca. 750 µL) by pipetting with a cut tip. Transfer to a new tube, carefully avoiding disturbance of the debris between phases.

10. Add RNAse A (1 µL, 20 mg) to each tube, first diluting 15X in dH2O, shake/vortex and incubate (37 °C, min 1 hour).

11. (Optional) Repeat steps 8 & 9 with pure chloroform - this will remove the RNAse A and residual phenol, leading to a better 260/230 nm ratio.

12. Add 0.1 volume (ca. 76 µL) of sodium acetate (3M, pH 5.2) and 0.7 volumes (ca. 530 µL) of cold isopropanol (stored at -20 °C). Mix by inverting ten times and keep at -20 °C for one hour or overnight.

13. Centrifuge (13,000 g, 10 min, RT, soft).

14. Carefully decant off the supernatants and drain tubes by resting inverted on paper towels.

15. Wash pellets by adding 70 % ethanol (1 mL). Centrifuge (13,000 g, 10 min, RT, soft).

16. Carefully remove supernatants by aspiration to avoid loss of the nucleic acid pellet.

17. Dry open tubes at room temperature. DO NOT OVER-DRY!

18. Gently suspend pellets in 25-50 µl of desired buffer, e.g. TE (10 mM Tris-HCl pH 8.0 + 1 mM EDTA).

To measure the concentration of high molecular weight DNA in each prep, 1 µL was visualised on an agarose gel (0.7% TAE) alongside the 1kb+ DNA ladder (NEB), for which the amount of DNA in each band is known. The sample and ladder bands were then quantified using the Band/Peak Quantification plug-in (Ohgane & Yoshioka, 2019) in ImageJ (1.54d) running Java (1.8.0_345).

PCR

I did most PCRs with the Phusion™ High-Fidelity DNA Polymerase (Thermo-Fisher Scientific™, F350S) following the manu-facturer's instructions (Table 2.2), including using the provided online primer annealing temperature calculator. When the use of a high-fidelity polymerase was not warranted, such as when screening colonies on a plate, I used Taq DNA Polymerase (NEB, M0273S) as directed by the manufacturer's instructions (Table 2.2). Primers and other oligonucleotides were ordered from Integrated DNA Technologies®.

Phusion protocol

PCR mix (µL)		Thermal Programme (°C)			
dH_2O	Up to 12.5	Initial Denat.	98	30 s	
5X HiFi buffer	2.5	Melting	98	10 s	X35
dNPTs (10 mM)	0.5	Annealing	Tm	10 s	
Primer F (10 µM)	0.63	Extension	72	30 s/kb	
Primer R (10 µM)	0.63	Final ext.	72	5 min	
Template (~50ng)	Var.	Cool down	16	10 s	
Phu. Polymerase	0.1				

Taq protocol

PCR mix (µL)		Thermal Programme (°C)			
dH_2O	Up to 12.5	Initial Denat.	95	30 s	
10X Taq Buffer	1.25	Melting	95	30 s	X35
dNPTs (10 mM)	0.25	Annealing	Tm	15 s	
Primer F (10 µM)	0.25	Extension	72	60 s/kb	
Primer R (10 µM)	0.25	Final ext.	72	5 min	
Template (~50ng)	Var.	Cool down	16	10 s	
Taq Polymerase	0.7				

Table 2.2. PCR protocols.

When PCR products needed to be cloned, I used the Zero Blunt® TOPO® PCR Cloning kit (Invitrogen, K283020) following the manufacturer's instructions (Table 2.3). After a short incubation (5 min; RT) the reaction was diluted with water 4X and electroporated into lab-grown DH10B cells.

Reagent	Volume (µL)
Purified PCR product	4
Salt Solution	1
pCR™ II-Blunt-TOPO®	1
Final volume	6

Table 2.3. PCR product cloning reaction.

To confirm purple lines carried genuine targeted insertions or replacements, I PCR amplified across the KI locus. Such “whole insert” amplifications required extensive optimisation because any *Cai1* locus with a WT sequence or a small indel produced an amplicon which was significantly smaller than the KI amplicon. PCR reactions have a bias towards amplifying smaller products, as the smaller size allows more of these to be synthesised per reaction cycle. This initially small difference is then magnified by the exponential nature of PCR amplification. To mitigate this, whole insert amplifications were supplied with a large amount of template and polymerase, were run for a reduced number of cycles, and had longer extension times (Table 2.4).

PCR mix (µL)		Therm. Programme (ºC)			
dH_2O	18.1	**Denat.**	98	1 min	
5X HiFi buffer	7.5	**Melting**	98	10 s	X25
dNPTs (10 mM)	1.5	**Annealing**	Tm	20 s	
Primer F (10 µM)	2.25	**Extension**	72	45 s/kb	
Primer R (10 µM)	2.25	**Final ext.**	72	7 min	
Template (~150ng)	5	**Cool down**	16	10 s	
Phu. Polymerase	0.9				
Total	37.5				

Table 2.4. Whole insert amplification protocol.

Whole-insert PCRs often produced faint intermediate bands in addition to the top and bottom bands expected from a plant heterozygous for a KI. These were ascribed to the formation of heteroduplexes between amplicons of the KI and WT products. The formation of such heteroduplexes has previously been reported (Jensen & Straus, 1993) and was addressed by adding ETDA (0.5 M, 6 µL per 37.5 µL reaction) to the PCR product and gradually re-annealing the PCR products (98ºC for 3 min, 22ºC for 10 s with a ramp rate of -0.1ºC/s).

A table of all primers and oligonucleotides used is available in the Appendix (App. 5).

Nucleic acid visualisation

I visualised DNA via agarose gel electrophoresis. To make the gels, I dissolved agarose (0.7-4% w/v) in Tris - Acetic acid – EDTA (TAE) buffer and added ethidium bromide (0.2 µg/mL). I ran gels at 6.7 V/cm for 15-45 minutes, depending on the size of the

products and the percentage of the gel. I imaged gels under UV with a wavelength of 302 nm.

2.2.3 Bioinformatics

Identifying *Cai1* as *ANT1*'s orthologue

I extracted the sequence of the homology arms used by Čermák et al. (2015) to target the *SlANT1* gene and BLAST-searched this against the SL4.0 version (Hosmani et al., 2019) of the published tomato genome (Sato et al., 2012). The top hit (ch10 64287737-64288455) was a perfect match (SFig. 1A) and corresponded to the gene locus *SlANT1* (Gene ID: 543897), which confirmed this was indeed the gene they targeted, and not some highly similar paralogue. I then locally BLAST-searched the associated ORF (NC_015447.3, 65260733-65261744; 1,012 bp in length), plus ca. 2 kb of sequence upstream and ca. 1 kb of sequence downstream, against the Solyntus V1.1 genome (SFig. 1B) (van Lieshout et al., 2020). Of the 100 hits reported, 32 were on chromosome 10 (same chromosome as *SlANT1*). This included the two hits with the highest confidence (E-value = 0 for both) which sat only ca. 20 kb apart on the chromosome. The first of these covered the whole ORF and was superior in length and similarity (82.8% versus 79.4% identity; 2,282 bp versus 1,302 bp hit length) and was thus earmarked as the true potato orthologue of *SlANT1*. This locus contains a MYB gene called *StCai1* (alternatively called *Mtf1*; Gene ID: 102590860; UniProt entry: A9YY81), which has previously been identified as the potato orthologue of *SlANT1* (Powell et al., 2022). I combined this top BLAST hit with two short BLAST hits (259 and 362 bp long) which mapped on either side of it into the Solyntus *Cai1* allele sequence (SFig. 1B highlighted in grey).

I also investigated the other high confidence hit. Syntheny analysis by Powell et al. (2022) of the chromosome 10 region surrounding *StCai1* identified five neighbouring MYB genes: *StANT1*, *StAN2*, *StANT1-like*, *StAN2-like*, and *StANT1-like2*. Submitting the hit to the NCBI BLASTn search service (Boratyn et al., 2013; Z. Zhang et al., 2000) yielded a list of similarities to several genes, including *StCai1*, *StAN1*, *StAN2*, *MYB113*, and others – a fact reflective of the confusing state of potato MYB nomenclature.

Ultimately, I was unable to definitively identify which of these genes gave rise to the other high-confidence BLAST hit.

Alignments, unless otherwise stated, were made with the Geneious Prime software's Geneious Alignment function using the default settings.

crRNA design

I identified LbCas12a cut sites with Geneious Prime's "Find CRISPR Sites" function, using the Solyntus V1.1 genome (van Lieshout et al., 2020) as an off-target database. In *Cai1*-long, this identified two overlapping crRNA target sites with no perfect off-target sites. To help discriminate between the two, I submitted both to DeepCpf1, an AI tool which scores AsCas12a crRNAs based on data from human cells (H. K. Kim et al., 2018). While DeepCpf1 was trained on the performance data of a Cas12a sourced from a different bacterium (*Acidaminococcus sp. BV3L6*) and used in a mammalian context, it was the only tool for the prediction of Cas12a crRNA efficiency available at the time. One target (crRNA1) received a much higher score (83 versus 27, more is better) and taken forward. No further guide optimisation was done.

2.2.4 KI cloning

General Golden Gate cloning

All cloning in this thesis was done using the Golden Gate cloning system (Engler & Marillonnet, 2014) unless stated otherwise. I planned all molecular cloning in the programme Geneious (final version 2025.0.1). Golden Gate cloning makes use of Type IIS restriction enzymes, which cut in a staggered manner outside of their recognition site, leaving behind overhangs several bp-long. These and other necessary reagents were purchased from New England Biolabs (NEB): BsaI-HF®v2 (R3733S), BbsI-HF® (R3539S), Recombinant Bovine Serum Albumin (BSA; B9200S) and T4 DNA Ligase (M0202S). As the T4 ligase buffer contains ATP, which can degrade because of freeze-thaw cycles, it was aliquoted in 10 µL portions on arrival and stored at -20°C.

The Golden Gate cloning was done following the MoClo cloning standard (Weber et al., 2011), which categorises DNA elements into several categories, based on their complexity, and allows for the

combination of simpler modules into complex assemblies. This is possible thanks to the consistent use of particular four bp-long overhangs for each type of genetic element (e.g. promoter, terminator, etc.). Such elements are designated as Level 0 (L0), and the MoClo system allows for their combination into gene expression cassettes (Level 1, L1), and subsequently the combination of these into polygenic constructs (Level 2, L2) (Fig. 2.1). This system was later extended, allowing for the creation of polygenic constructs (Level M, LM) which could then be combined further (Level P, LP), thus enabling the generation of very large constructs with tens of genes (Werner et al., 2012).

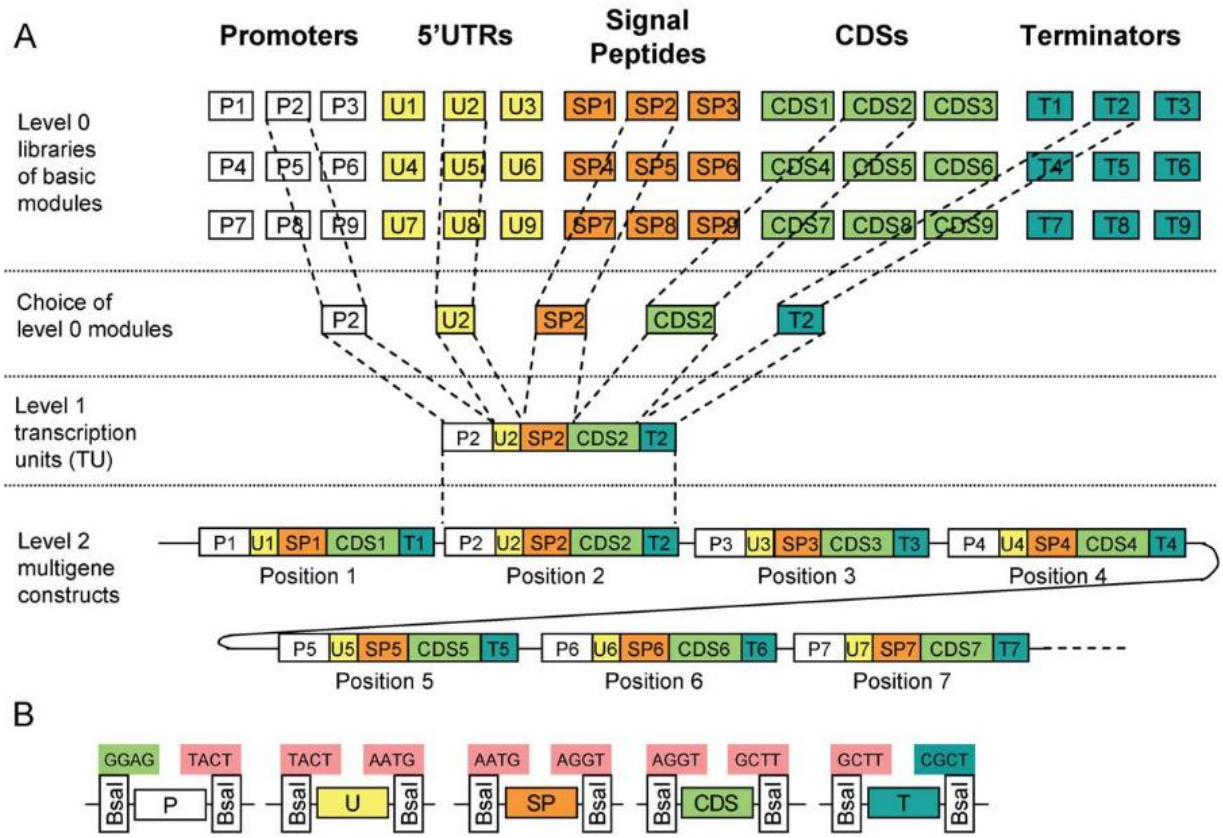


Figure 2.1. MoClo hierarchical cloning standard.
A The assembly genetic elements (L0) into gene expression cassettes (L1) and the assembly of gene expression cassettes (L1) into polygenic constructs (L2). **B** The overhangs of genetic elements defines their function. Source: Weber et al. (2011).

Elements to be cloned together were mixed in a digestion-ligation (diglig) reaction, with an acceptor plasmid, the necessary enzymes and buffers according to the level of the plasmid being assembled (Table 2.3). Each element was added in a 2:1 molar ratio compared

to the acceptor. For more challenging assemblies, particularly ones being assembled into the LP_pRKan290 acceptor (Youles et al., 2024), the cycle number was increased to 50 and the reaction was run overnight.

BpiI (L0, L2 and LM)					
\	Item	Vol (µL)	Mass required (ng)	Conc. (ng/µL)	size (bp)
1	Element 1	0.20	20.0	100	1000
2	Element 2	0.40	40.0	100	2000
3	Element 3	0.60	60.0	100	3000
4	Element 4	0.80	80.0	100	4000
5	Element 5	1.20	120.0	100	6000
Acceptor	**Acceptor 1**	0.50	50.0	100	5000
0.1X BSA (2 mg/mL)		2	Lid = 45°C. The program will take ~ 3hr 40 min to be completed.	Them. prog.	
T4 DNA Ligase Buffer		2		3 min	37°C
BbsI-HF 10U/µL		0.5		4 min	16°C
T4 DNA Ligase		0.5		^26 cycles	
Total before water:		8.70		5 min	50°C
Water up to 20 uL		11.30		5 min	80°C
TOTAL VOLUME		**20**		10s	16°C

BsaI (L1, LP)					
\	Item	Vol (µL)	Mass required (ng)	Conc. (ng/µL)	size (bp)
1	Element 6	1.00	100.0	100	2000
2	Element 7	0.91	100.0	110	2000
3	Element 8	0.83	100.0	120	2000
4	Element 9	0.77	100.0	130	2000
5	Element 10	0.71	100.0	140	2000
Acceptor	**Acceptor 2**	0.50	50.0	100	2000
0.1X BSA (2 mg/mL)		2.00	Lid = 45°C. The program will take ~ 3hr 40 min to be completed.	Them. prog.	
T4 DNA Ligase Buffer		2.00		3 min	37°C
BsaI-HFv2 10U/µL		0.50		4 min	16°C
T4 DNA Ligase		0.50		^26 cycles	
Total before water:		9.73		5 min	50°C
Water up to 20 uL		10.27		5 min	80°C
TOTAL VOLUME		**20**		10s	16°C

Table 2.5. Golden Gate cloning mixtures.

5 µL of each diglig was electroporated into 50 µL electrocompetent DH10B cells using a BioRad Micro Pulser™ electroporator. This was done as described in the protocol by New England Biolabs (2020), except L media was used in place of SOC media.

Cloning efficiency in *E. coli* was generally high enough that no colony PCR was done. Instead, three white colonies were grown up, mini-prepped, and digested with either *EcoRI*, *HaeII*, *HindIII*, or *DraIII* depending on plasmid size and cut site availability. These enzymes were purchased from NEB, opting for "HF" versions when available. Such diagnostic digests of putative cloned plasmids were done according to the manufacturer's instructions to confirm correct overall structure. If a sample's digestion pattern matched expectations, the plasmid would be sent for Sanger sequencing (GeneWiz®) or Whole Plasmid sequencing (Plasmidosaurus™) to confirm its sequence was accurate.

Name	Purpose	Source	Reference
ANT1 ORF+Ter	Tomato ANT1 ORF + terminator	Lila Grandgeorge	(Grandgeorge, 2021)
DWF4 pro L0	*A. thaliana* DWF4 promoter	Hsuan Pai	
EC15191	3x HA+ Gly-Gly linker	SynBio TSL	
LGJJ21	L1P7 F LoxP site	Lila Grandgeorge	(Grandgeorge, 2021)
LGJJ22	L1P2 F LoxP site	Lila Grandgeorge	(Grandgeorge, 2021)
LP_pBIN19	Derivative of pBIN19	SynBio TSL	(Bevan, 1984)

LP_pRKan290	Derivative of pRK290	SynBio TSL	(Ditta et al., 1980)
pAGT6477	UL12-4LF2	Tom Schreiber	(Schreiber et al., 2024)
pAGT8173	NLS-LbCas12ai(D156R)-NLS intronized and codon-optimised	Tom Schreiber	(Schindele et al., 2023)
pICH41414	35S terminator	SynBio TSL	
pICH41432	Ocs terminator	SynBio TSL	
pICH51277	35Spromoter + Ω leader sequence	SynBio TSL	
pICSL00006	P2A ribosomal skipping sequence	SynBio TSL	(J. H. Kim et al., 2011)
pICSL11180	Ubi10:LbCas12a:E9 ter; L1P2 R	SynBio TSL	(Zetsche et al., 2015)
pICSL12015	AtUbi10 pro	SynBio TSL	
pICSL12048	FMV promoter	SynBio TSL	(Mait et al., 1997)
pICSL12063	L0 AtAct2:SSU301 leader	SynBio TSL	
pICSL1P4_crU626	LbCas12a crRNA L1P4 acceptor	SynBio TSL	
pICSL31841	Nos:Npt\ecbe:Nos	Sylvestre Marillonnet	
pICSL32281-LacZ	L2 binary vector acceptor	SynBio TSL	
pICSL4723_P1	LP binary vector acceptor	SynBio TSL	
pICSL50016	turboGFP + St-LS1 intron 2	SynBio TSL	(Evdokimov et al., 2006)
pICSL60004	E9 terminator	SynBio TSL	
pICSL60024	*A. thaliana* U6-26 Terminator	SynBio TSL	
pICSL60026	Extensin-Truncated-EU + NbACT double terminator	SynBio TSL	(Diamos & Mason, 2018)
pICSL60028	AtHSP18.2 terminator	SynBio TSL	
pICSL80001	Firefly luciferase CDS	SynBio TSL	(Ow et al., 1986)
pICSL80021	codA CDS	SynBio TSL	(Stougaard, 1993)
pICSL80094	RUBY L0	SynBio TSL	(He et al., 2020)
pICSL90002	*A. thaliana* U6-26 Promoter	SynBio TSL	
YDJJ061	L1P4 F 35Somega:HA-R (Cai1-long)	This work	
YDJJ101	L1P6 F Cai1-long Distal HA-L blunt-end cloned PCR (4500 bp)	This work	
YDJJ106	L1P6 F Cai1-long Proximal HA-L blunt-end cloned PCR (4500 bp)	This work	

Table 2.6. Key plasmids used in cloning.

Cloning of custom L1 modules

Cloning 35S:Cai1 (YDJJ009) and DWF4:Cai1 (YDJJ007) required the cloning of the Cai1 ORF and terminator. This I amplified from B26 genomic DNA (primers YD3 & B26 Cai1 ter L1 R), with BsaI sites added using primer overhangs. The PCR product was then purified and used in a diglig as usual. The *ANT1* ORF and terminator modules necessary for the cloning of 35S:ANT1 (YDJJ010) and DWF4:ANT1 (YDJJ008) were previously cloned by Lila Grandgeorge.

The two 800 bp-long homology arms were cloned as Level 1s (Fig. 2.2). The Left homology arm was cloned with primers YD1 & YD2. The forward primer (YD1) added a BpiI site and a copy of the crRNA1 protospacer (incl. PAM), with the Cas12a cut site proximal to the homology arm. The reverse primer (YD2) added a BpiI site and mutated the PAM of the natural crRNA1 protospacer, thus ensuring the repair template could not be cut. The finished PCR was purified, blunt end cloned (the Zero Blunt™ TOPO™ kit used does not require post-PCR procedures like amplicon phosphorylation), and used as a L1 module in later digligs.

The Right homology arm was cloned with primers YD3 & YD4. This homology arm included a short 5'UTR and a partial *Cai1* ORF. This would be cloned downstream of promoters like 35S but the MoClo-defined overhang downstream of promoters is "TATG". If this were used, it would introduce an upstream ORF (uORF) in front of *Cai1*. To avoid this, the forward primer (YD3) introduced a BsaI site with a custom "AGTA" overhang. Furthermore, the primer extended into the start of the *Cai1* ORF to introduce a synonymous mutation in a natural BpiI site. The reverse primer (YD4) added a BsaI site and a copy of the crRNA1 protospacer (incl. PAM), with the Cas12a cut site proximal to the homology arm. To accommodate the homology arm's custom "TATG" overhang, I reamplified the 35SΩ promoter from pICH51277 using primers LG14 & LG15, which changed the right overhang to "TATG". Both PCR fragments were purified and used in a L1 diglig.

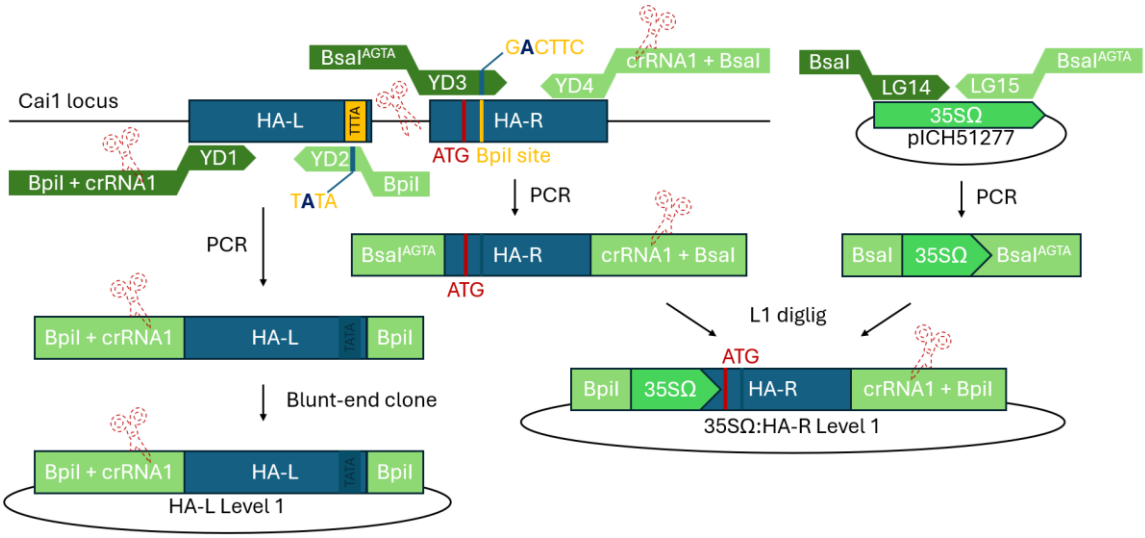


Figure 2.2. Cloning the homology arms.
Primers YD1-4 and LG14-15 were used to PCR amplify and clone homology arm left (HA-L) and right (HA-R), which were necessary to target gene KIs. Red scissors indicate the presence of a protospacer for crRNA1. Sequences in yellow (the "TTTA" PAM sequence and a natural BpiI site in HA-R) needed to be disrupted; this was achieved through single nucleotide changes in YD2 & YD3, as indicated above. A BsaI site with a custom "AGTA" overhang was introduced by primers YD3 & LG15 – this allowed for Golden Gate cloning without the introduction of a novel "ATG" start codon upstream of the native "ATG" (indicated in red). 35SΩ - 35S cauliflower mosaic virus promoter with the Omega leader sequence.

Level 2 cloning

The Level 1 constructs 35S:Cai1 (YDJJ009), DWF4:Cai1 (YDJJ007), 35S:ANT1 (YDJJ010), DWF4:ANT1 (YDJJ008) were paired with the *neomycin phosphotransferase II* (*NptII*) gene into polygenic Level 2 constructs (YDJJ015-018) (Fig. 2.3). As discussed in the section on Golden Gate cloning, every Level 1 element in the MoClo system has a position between one and seven. Selectable markers were always cloned in position one, as T-DNA integration at the left T-DNA border can messy (Kralemann et al., 2022) and having the *NptII* immediately adjacent should help to avoid selecting for partial T-DNA integrations. Each Level 2 also included one of four possible Level 1 elements (YDJJ007-10). As these Level 1

constructs were cloned in position three acceptors, the empty second position was occupied by a “Dummy” – a short placeholder sequence ca. 15 bp in length. Lastly, the final Level 1 element (in this case, position three) was linked to the invariable right-side overhang of the Level 2 acceptor via an End Linker. Subsequent Level 2 constructs, as well as the larger Level M and P constructs discussed earlier (Werner et al., 2012), were cloned in an analogous manner.

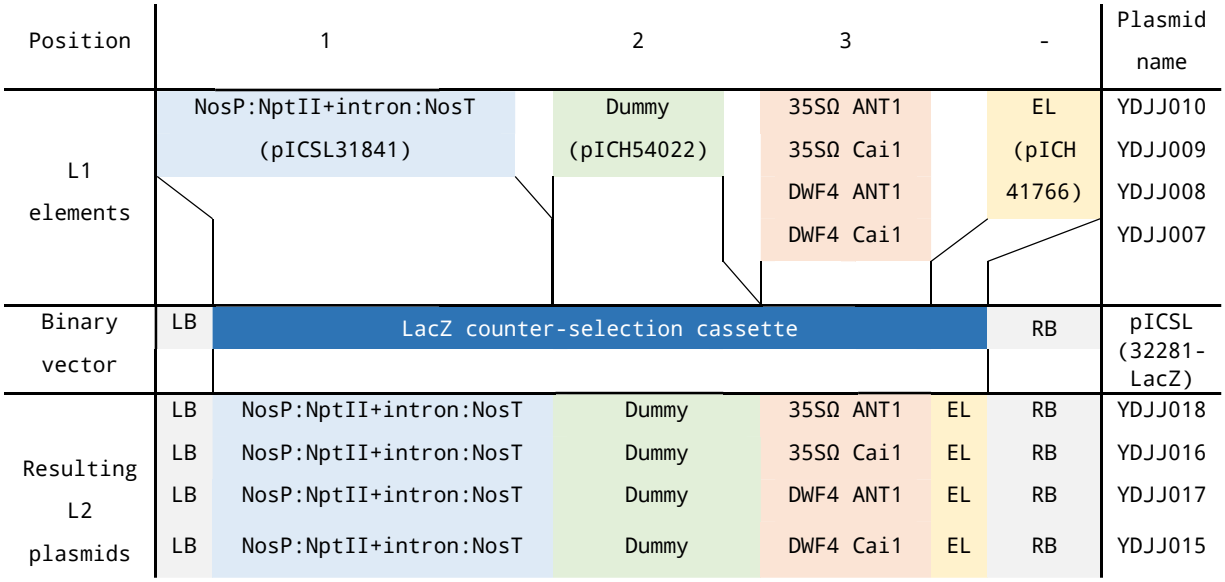


Figure 2.3. Creation of polygenic constructs.
Polygenic Level 2 constructs consisted of the *NptII* selection marker in position 1, a “Dummy” placeholder short sequence in position 2, and one of four possible Level 1 elements in position 3. Lastly, an End Linker bridged the variable final Level 1 (position 3 in this case) with the invariable right-side overhand of the Level 2 acceptor.
L1 – Level 1; L2 – Level 2; EL – End Linker module; LB – T-DNA Left Border; RB – T-DNA Right Border; NosP – Nos promoter; NosT – Nos terminator.

First Knock-In constructs

The cloning of the first KI constructs will be described in detail, while later only substantial differences in the cloning approach will be noted. Because of their large size and complexity, KI construct creation made use of the extended Golden Gate cloning system featuring Levels M and P (Werner et al., 2012). In short, multiple Level 1 constructs can be combined into polygenic Level Ms, which in turn can be combined into even larger

Level Ps. I designed all KI constructs as Level Ps consisting of two Level Ms, as done by Grandgeorge (2021) (Fig. 2.4). This approach allows for a set of KI reagents like a programmable nuclease and crRNAs, which sit on the left Level M (Level M-L), to easily be reused to KI a different gene cassette which sits on the right Level M (Level M-R), and *vice versa*.

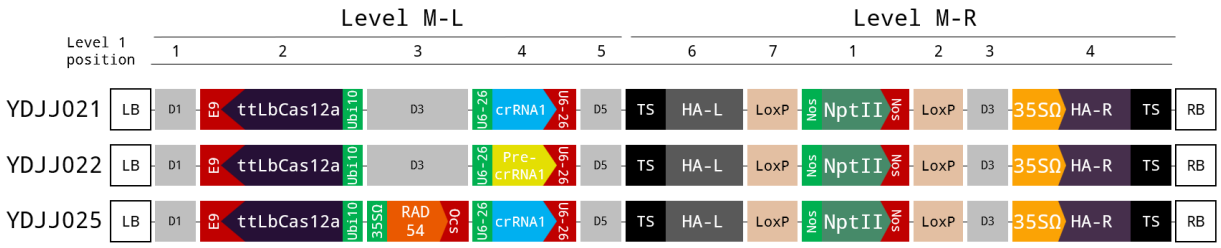


Figure 2.4. KI construct structure.
D1, D3, D5 – dummies; TS – target site, a copy of the crRNA1 protospacer; LB & RB – left and right borders.

A temperature-tolerant LbCas12a (ttLbCas12a) was used (Schindele & Puchta, 2020), expressed in the reverse direction by the *Arabidopsis thaliana* Ubiquitin-10 promoter and the *Pisum sativum rbcS* E9 terminator, as shown to work well by Castel et al. (2019).

The crRNAs were expressed by the *A. thaliana* U6-26 PolIII promoter and terminator. The crRNA was flanked by the ribozymes, as used by Gao & Zhao (2014) to ensure the "clean" transcription of crRNAs without any additional nucleotides on either side. The pre-crRNA was instead flanked by two full-length LbCas12a CRISPR repeat sequences. This alternative approach aimed to replicate the crRNA structure of a natural CRISPR array, which was reported by Xu et al. (2017) to lead to higher editing efficiencies. The protospacer in the crRNAs was 23 bp long, while that of the pre-crRNAs was 28 bp, which is the average length between crRNA knots in natural *Lachnospiraceae bacterium* CRISPR loci (Xu et al. (2017) used a 31 bp protospacer). The protospacer sequences were ordered as a pair of oligonucleotides, diluted to 100 µM, and added directly to digligs (1µL each).

The CDS of *ScRAD54* was previously cloned and had any native BpiI and BsaI cut sites synonymously mutated by Laurance Tomlinson. This I cloned between the 35SΩ promoter and the *octopine synthase* (*Ocs*) terminator for *A. tumefaciens*.

Inside the KI cassettes (Fig. 2.4) sat an *NptII* gene which had an intron inserted and was codon-diverged from the standard *NptII* CDS to avoid plasmid rearrangements with any potential *Npt* genes present on plasmid backbones (pICSL31841); this was kindly provided by Sylvestre Marillonnet. Flanking the *NptII* were two LoxP sites cloned by Lila Grandgeorge as Level 1s in position 7 and 3 (LGJJ21 & LGJJ22). These were part of an idea to use the Cre recombinase to excise the *NtpII* marker after a KI had occurred. This was later abandoned and the LoxP sites were never used; they were not added to later plasmids.

I cloned all of this into the Level P binary vector pICSL4723_P1.

Constructs for exploring alternative KI approaches

The KI of the 35S cauliflower mosaic virus promoter plus the Ω leader resulted in highly pigmented plants but also a developmental phenotype. To test if omitting the Ω leader would result in a milder KI phenotype, I re-amplified 35S promoter from pICH51277 without the Ω leader using primers LG14 & LG266. Similarly, to clone a further-truncated version of the 35S promoter (-106 35S) I re-amplified it using primers LG15 & YD29.

To try and use MMEJ and NHEJ to facilitate KIs, I cloned constructs where the L1 homology arms were substituted with oligonucleotide pairs encoding 25 bp microhomologies (YD40-41 and YD44-45) or a region with homology (essentially just crRNA1 protospacers; YD46-49) (Fig. 3.18).

Cloning alternative nucleases and crRNA expression cassettes

This cloning was completed with the help of TSL MSc student Jasmine Atkins. Tom Schreiber kindly provided several Level 0 modules. One of these contained an intronized temperature-tolerant LbCas12a, as described by (Schindele et al., 2023), with a nuclear localization signal on both ends (pAGT8173). Another contained the UL12 exonuclease plus a 4xLF2 flexible linker, as described by Schreiber et al. (2024) (pAGT6477). Jasmine cloned this new LbCas12a cassette (pAGT8173) into Level 1 modules fused either to the UL12 exonuclease (Fig. 2.5 YDJJ084), or to a “placeholder” 3xHA tag plus Gly-Gly linker (EC15191) (Fig. 2.5 YDJJ083).

Diagrams of the full LP constructs these L1 plasmids were first used in can be found in the main text (Fig. 3.23).

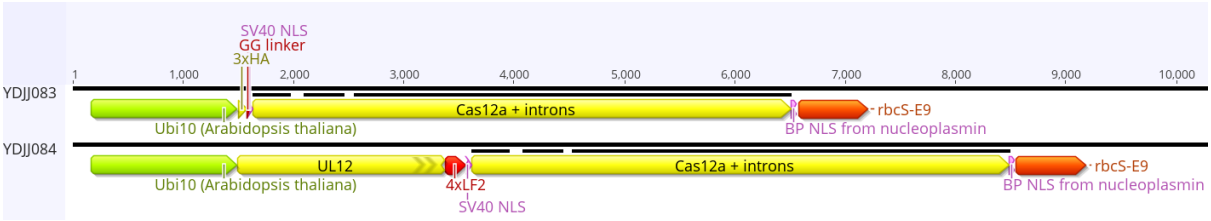


Figure 2.5. Level 1 constructs featuring the intron-containing Cas12a cassette.
Ubi10 – ubiquitin 10 promoter from *A. thaliana*; HA – hemagglutinin tag from human influenza; UL12 – UL12 exonuclease from human simplex virus 1; NLS – nuclear localisation signal; BP NLS - bipartite NLS from human nucleoplasmin; SV40 NLS – NLS from Simian virus 40 large T-antigen; rbcS-E9 - ribulose-1,5-bisphosphate carboxylase small subunit E9 terminator from *Pisum sativum*; Cas12a + introns – intronized version of temperature-tolerant LbCas12a.

Jasmine also cloned the crRNA1 protospacer between tRNA elements PCR-amplified (primers YD61 & YD64-66) from plasmid ASJJ7, which was provided by Aga Alexander. This tRNA-flanked crRNA1 was cloned either between the AtU6-26 promoter and terminator, or between the FMV promoter (pICSL12048) and the Extensin-Truncated-EU + NbACT double terminator (pICSL60026) as described by Diamos & Mason (2018) (for plasmid diagrams see Fig. 3.25).

The KI cassettes of these and later constructs generally contained an R-gene – *Rpi-vnt1* (S. J. Foster et al., 2009; Pel et al., 2009). A L0 plasmid carrying the gene with its native promoter and terminator was provided by Robert Heal. I then cloned this gene into a L1 position 7 acceptor (pICH47791).

All constructs were cloned into a LP acceptor pAGM31195, which had the *isopentenyl transferase (IPT)* gene cloned outside of the Left border by Mark Youles (for plasmid map see SFig. 10).

Cloning into alternative plasmid backbones

This cloning was completed with the help of visiting student Ainhoa Valero-Abad. The T-DNA cassette from pUL12-iCas12a, which

was re-cloned into a LP acceptors LP_pBIN19 and LP_pRKan290 (Youles et al., 2024). Both backbones were modified and adapted to use with Golden Gate cloning by Mark Youles.

Cloning of luciferase constructs

I cloned the firefly luciferase CDS (pICSL80001) under the 35SΩ promoter (pICH51277) and the *Ocs* terminator (pICH41432). The resulting L1 I cloned alongside the *NptII* gene (pICSL31841) into pAGM31195, LP_pBIN19, and LP_pRKan290. The LP diglig into LP_pRKan290 proved less efficient. To mitigate this, the number of cycles in the diglig thermal programme (Table 2.5) was increased to 50.

2.2.5 TIIGER cloning

Cloning a *codA* expression cassette

I cloned the L0 CDS for *cytosine deaminase* (*cod*) from *E. coli* (pICSL80021) under the control of the *A. thaliana actin 2* promoter fused to the 5' UTR of the petunia *SSU301* gene (pICSL12063) (Dean et al., 1988), kindly cloned by Mark Youles, as well as the 35S terminator (pICH41414).

Cloning TIIGER constructs with one or two crRNAs

Suitable protospacers for crRNAs 3 & 5 were identified in the same way as crRNA1 was. crRNA5 was then cloned next to crRNA1, separated and flanked by tRNA elements instead of the usual ribozymes. This was done as tRNAs allow for multiple crRNAs to be efficiently expressed *in cis* (Xie et al., 2015). While crRNA3 was cloned on its own, it was still flanked by tRNAs to facilitate the comparison with the use of two crRNAs. As before, crRNAs were expressed by the *AtU6-26* promoter and terminator.

TIIGER required the cloning of the distal *Cai1* promoter as a homology arm. This 800 bp-long element was cloned in the same way the proximal promoter region was. The region was amplified from genomic DNA from potato line B26 using primers with overhangs. The forward primer (YD27) introduced a BpiI binding site and a copy of the crRNA1 protospacer (incl. PAM), with the Cas12a cut site proximal to the homology arm. The reverse primer (YD28) added a BpiI site. The finished PCR was purified, blunt end cloned, and used as a L1 module in later digligs.

For use with crRNA3, a version of the TIIGER cassette was cloned with the target sites changed to match the new crRNA. This was done by amplifying across the crRNA1-targeted TIIGER cassette with primers which bind the distal ends of the homology arms and introduced crRNA3 target sites (YD168 & YD170).

Cloning TIIGER constructs with 0/1/2 R-genes

A L0 plasmid carrying *Rpi-amr3* with its native promoter and terminator was provided by Rober Heal. I then cloned this gene into a L1 position 1 acceptor (pICH47732). *Rpi-vnt1*, *Rpi-amr3*, both, or neither of the genes were cloned into TIIGER cassettes alongside the *codA* gene. The TIIGER cassette Level M-Rs were then all cloned alongside a Level M-L containing the intronized Cas12a nuclease and crRNAs 1&5 (YDJJ162). These were all cloned into the LP_pRKan290 binary vector acceptor.

Cloning GTSS constructs

To construct the GTSS cassettes, the *Nos* promoter – *bar* CDS region was amplified as a L0 from pICSL11170 using primers which added BsaI recognition sites and removed the CDS' STOP codon (YD158 & YD159). The reverse primer (YD159) was carefully designed to add a "GAGC" overhang in frame so that the "AGC" coded for a Serine. This was then cloned into a complex L1 construct alongside GSG-linker-P2A ribosomal skipping sequence (pICSL00006; J. H. Kim et al., 2011). Downstream of the P2A sequence I included either the partial *Cai1* ORF present in the right homology arm (amplified using primers YD160-161), or the whole *Cai1* ORF as well as 600 bp of the native terminator (amplified using primers YD160 & YD162). The resulting L1s were cloned alongside the distal homology arm into two LMs. The first of these was a TIIGER cassette containing the GTSS selection elements (Fig. 4.6 pBarNeg). The second was a TIIGER cassette containing the GTSS selection elements plus a complete *Cai1* ORF and terminator downstream of it (Fig. 4.6 pBarPos); essentially an approximation of the what the *Cai1* locus should look like after a targeted replacement. To further improve stringency via NMD, I later amplified the *Pontellina plumata* turbo GFP ORF (it had the *St-LS1* intron 2 added ca. 380 bp downstream of the ATG start codon) from pICSL50016 using primers YD232-233. These primers added BsaI recognition sites and "CAGA"-"GAGC"

overhangs, which allowed me to treat the PCR product as a LM. These LMs were cloned into the pRKan290 LP.

To try to introduce the GTSS cassette via targeted substitution, I cloned the GTSS cassette followed by the GFP alongside the YDJJ162 Level M-L from earlier.

Cloning Targeted Replacement + Cre

This cloning was done with the help of visiting student Ainhoa Valero-Abad. To conduct RMCE, a LoxP acceptor locus was cloned by introducing heterospecific LoxP sites on either side of the 35S promoter in the TIIGER cassette. Heterospecificity was provided by the LoxP 551 (Hoess et al., 1986) upstream of the 35S, while the unidirectionality of the RMCE reaction was ensured by the Lox66 site downstream of the 35S – this variant recombines with Lox71 practically irreversibly (H. Albert et al., 1995). To ensure no uORF was introduced, the LoxP sites were cloned in a reverse orientation. LoxP 551 was cloned into a TIIGER cassette as a pair of oligonucleotides bearing BpiI recognition sites (YD181-182; treated as a L1). Lox66 was cloned downstream of the 35S promoter as a pair of oligonucleotides bearing BsaI recognition sites (YD186-187; treated as a L0). The Lox66 element had overhangs “TACT”-“AGTA”, which necessitated the reamplification of the 35S promoter with primers LG14 & YD185 to change its downstream overhang (“AGTA”→”TACT”). This PCR product was then cloned into a L1 construct alongside the Lox66 oligonucleotides and either the right homology arm, or the whole *Cai1* ORF and native terminator. These were then cloned into LMs alongside the distal left homology arm to approximate a TIIGER cassette prior to and after targeted substitution.

The former was later cloned into the LP_pRKan290 binary vector acceptor alongside the YDJJ162 LM to obtain a TIIGER-capable construct.

2.2.6 Transformation

Transient expression

Agrobacterium strains were grown up (28°C, 40 h, shaking), pelleted (3,500 g, 5 min), and resuspended in Infiltration buffer to an OD_{600} = 0.1, unless otherwise stated. This was then either

manually syringe-infiltrated into the adaxial side of leaves, or vacuum-infiltrated using a vacuum chamber (3 × 30 s bursts, -450 mbar).

To confirm *Cai1* functionality, 1 week-old *N. tabacum* seedlings were vacuum infiltrated with a combination of *Agrobacterium* strains carrying the appropriate plasmids at a total OD_{600}=0.2. Images were taken 8 dpi. *Agrobacterium* strains carrying plasmids with the genes *Delila* (pJAM2421) and *Ros1* (pJAM2424), as published by Albert et al. (2020), were kindly provided by Eugenio Butelli.

Leaf disk lux assays

The firefly luciferase cloned under the 35SΩ promoter into one of three plasmid backbones (pAGM31195, LP_pRKan290, and LP_pBIN19) was transiently expressed in the leaves of 5-weeks-old *N. benthamiana* plants at OD_{600}=0.1. Two days after infiltration, sodium citrate (100 mM; 40 µL per disk) and Triton-X 100 (6.7 nL per disk; Sigma-Aldrich; X100) were mixed and pipetted (40 µL) into a dark, fluorescence-compatible 96-well plate. Leaf disks were harvested with a 6 mm cork-borer and floated on top of the solution in the wells, adaxial side up. Luciferin (20 mM; 5 µL per disk; ThermoFisher Scientific; 336780100) was diluted in sodium citrate (100 mM; 35 µL per disk), and them pipetted (40 µL) into each well with a leaf disk. The leaf disks were then vacuum-infiltrated (3 × 30s; -450 mbar) and imaged with a NightOwl (Berthold Technologies Bioanalytics) imaging system.

Stable transformation

Plant transformations were done according to the following protocol, kindly provided by Matthew Smoker. *Agrobacterium* cultures were set up with the appropriate antibiotic selection regime (Table 2.1) and grown for 24 hours. The following day internode stem sections were harvested without nodes from 4–6-week-old potato plants grown in aseptic culture on MS medium. The internodes are sliced into 1 cm sections and placed into 20 mL of LSR broth. Overnight *Agrobacterium* culture (100 µL, OD_{600} = 1.0) is added to stem sections and incubated for 20 minutes at 40 rpm in the dark at 24°C. The stem sections are removed from the *Agrobacterium* suspension, blotted dry and plated at around 15-20 explants per dish on LSR1 media plus Acetosyringone (500 µM). The

plates are incubated under low light conditions (covered by sheet of paper) at 18C for 3 days. Afterwards, the explants are moved to LSR1 medium with Timentin and selection antibiotics and grown at high light conditions and 22C for 8 weeks, subculturing every 2 weeks. For shoot induction, plants were moved to LSR2 media with Timentin and selection antibiotics. Rooting of shoots was done in MS media with Timentin and selection antibiotics. Independent T_0 lines were recovered from different explants, thus ensuring they were the result of separate transformation and KI events. All transformed plasmids in this thesis are listed in Table 2.8.

All stable transformations were done by members of the TSL Tissue Culture Team. They also carried out regular callus subculturing and shoot recovery, except during KI experiments. During such experiments, the TSL Tissue Culture Team stopped caring for the calli after the end of the KI sector scoring period (usually around 28 days after co-cultivation) – thereafter I took over the regular subculturing of calli and recovery of shoots.

Antibiotic	Concentration (mg/L)
Kanamycin	100
Phosphinothricin	5
Timentin	320

Table 2.7. Selection agents used in Tissue Culture.

2.2.7 Breeding lines YD1-3

F_1 vigour screen

29 F_1 hybrids from a B26 × B100 cross (SP10016) were grown in a glasshouse until flowering, at which point inflorescences were bagged and tapped to encourage selfing. Nine lines which grew vigorously, flowered well, and set fruit readily were taken further.

RUBY transformability screen

The RUBY (He et al., 2020) L0 ORF (pICSL80094) was cloned into a L1 construct (YDJJ097) alongside the 35SΩ promoter (pICH51277) and the AtHSP18.2 terminator (pICSL60028). This was then cloned into a L2 acceptor (pICSL32281-LacZ) alongside a L1 *NptII* expression cassette (pICSL31841, position 1) creating the L2 vector YDJJ100. This was then stably transformed into B26, B100, and their F_1 progeny. The plates were photographed at 42 days post co-cultivation (dpcc) and from these the ratio of pink sectors to explants for each line was quantified. The F_1 progeny with the top three ratios were taken forward.

PLASMID NAME	ALTERNATIVE NAME
JAJJ101	pCas12a
JAJJ102	piCas12a
JAJJ103	pUL12-icas12a
JAJJ104	pU6:tRNA
JAJJ105	pFMV:tRNA
JAJJ108	pLP_pRKan290
JAJJ109	pLP_pBIN19
YDJJ007	DWF4:Cai1 L1
YDJJ008	DWF4:ANT1 L1
YDJJ009	35SΩ:Cai1 L1
YDJJ010	35SΩ:ANT1 L1
YDJJ015	DWF4:Cai1 L2
YDJJ016	35SΩ:Cai1 L2
YDJJ017	DWF4:ANT1 L2
YDJJ018	35SΩ:ANT1 L2
YDJJ021	pKI1
YDJJ022	pPre-crRNA
YDJJ025	pRAD54
YDJJ032	pYDJJ032
YDJJ041	pExtNpt
YDJJ042B	pMMEJ
YDJJ043	pNHEJ
YDJJ059	pΔΩ
YDJJ060	p35S(-106)
YDJJ100	NptII-35S:RUBY
YDJJ143	p2cuts
YDJJ144	p2cuts+int
YDJJ145	p1cut
YDJJ146	p1cut+int
YDJJ154	pBarNeg
YDJJ155	pBarPos
YDJJ170	PBarGFP
YDJJ180	pTIIGER-GTSS
YDJJ181	pGTSSneg
YDJJ185	35S:lux in pAGM31195
YDJJ186	35S:lux in LP_pRKan290
YDJJ187	35S:lux in LP_pBIN19

Table 2.8. Transformed constructs.

2.2.8 KI tracking

KI sector excision and scoring

Over the course of this thesis, two distinct methods were used to score KI sector frequency: the first method counted the sectors at the end of the 8-week culturing period, while the later method tracked their emergence over time.

Following the first method, after an 8-week culturing period transformed potato explants were visually inspected, and any suspected purple sectors were excised and subcultured on LSR1 media with Kanamycin and Timentin. Sometimes one explant gave rise to multiple, seemingly separate purple sectors. Regardless, these were always treated as one; sectors from the same explant were never subcultured or scored separately. The sector plates were photographed on the day of subculturing and again 7 days later. A purple sector was only counted as such if purple callus growth was visible after the 7-day period; sectors which disappeared or failed to grow were not counted (Fig. 2.6). The final frequency of purple sectors was calculated by dividing the number of confirmed sectors by the number of transformed explants. After growing to a moderate size, purple calli were moved to plates with LSR2 with Kanamycin and Timentin for shoot induction.

A B

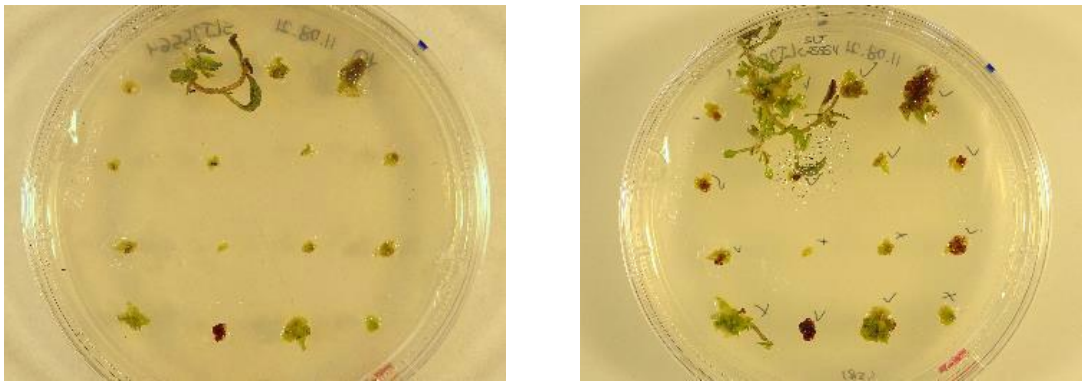


Figure 2.6. KI sectors after excision (A) and one week later (B).

Later, KI sector emergence was tracked over time. Shortly after co-cultivation, the plates were assigned a pseudonym and the name of the variable being tested (usually the transformed construct) was occluded. Starting around 14 days post co-cultivation (dpcc), calli were checked for purple sector emergence. Care was taken not

to allow the plates to open and to thoroughly check both the top and bottom side of calli for sectors. Sectors were generally scored every 2-3 days between 14 and 28 dpcc. Sectors appeared as early as 7 dpcc but few emerged before 14 dpcc. Sectors also appeared later than 28 dpcc but when an experiment was scored 14-49 dpcc only 26% of total scored KI sectors appeared after that point (SFig. 10). Crucially, when scoring using this second method, the calli forming on opposite ends of each explant were considered independent transformation events. Thus, I ultimately calculated the ratio between KI sectors and explant termini (double the number of explants). Only one sector was counted per terminus. After scoring was done, the plates were de-anonymized and the data were analysed using Time-to-Event analysis (details below).

2.2.9 Protein extraction and Western Blots

I completed all protein extraction following a protocol kindly provided by Camille-Madeleine Szymansky. Proteins for analysis were transiently expressed in *N. benthamiana* leaves and harvested at 3 dpi, 6 leaf disks (1 cm diameter) per construct, in a 2 mL Eppendorf tube, weighed, and flash-frozen in liquid nitrogen alongside 2 tungsten carbide beads/tube. The samples were ground using a Geno-Grinder® (1,200 rpm; 2 × 30 sec). Protein extraction buffer (100 mM Tris-HCl pH 8.0, 0.2% SDS, 10 mM DDT, and a protease inhibitor cocktail) was added in a 1:1 ratio of µL buffer per mg sample (e.g. 100 µL for 100 mg sample), shaken and incubated (5 min, RT). The samples were centrifuged (13, 000 g, 2 min, RT) and the supernatant was recovered. Blue Protein Loading Dye (NEB, B7703S) was added to the samples (0.33 × volume), after which they were vortexed and heated (95ºC, 5 min). 5 µL of each sample was initially we run alongside the PageRuler™ Prestained Protein Ladder (Thermo Scientific™) on an 8% SDS-PAGE gel, but the amount loaded was changed if uneven loading was observed. The gels were run at 80 V until samples entered the Resolving gel, and at 120 V thereafter. The proteins were transferred to a methanol-activated nitrocellulose membrane using the Trans-Blot Turbo Transfer System (Bio-Rad®) according to the manufacturer’s instructions. Following transfer, the membrane was incubated (30

min, RT, shaking) with 20 mL 5% powdered milk dissolved in washing buffer (20 mM Tris/HCl, 150 mM NaCl, and 0.1% Tween-20). Afterwards the membrane was rinsed with water and incubated (4°C, overnight) with 15 mL 5% powdered milk containing the primary anti-LbCpf1 mouse antibody, clone 2D5-6G11 (Sigma-Aldrich; MABE1824) diluted 1:10,000. Afterwards, I discarded the antibody solution and washed the membrane with washing buffer (3×10 min). I incubated the membrane (1 h, RT, shaking) with 15 mL 5% powdered milk containing the secondary HRP-conjugated anti-mouse secondary antibody, before washing as before and imaging on an Amersham ImageQuant 800 using the SuperSignal™ West Pico Plus kit (ThermoFisher Scientific; 34580) as directed by the manufacturer.

2.2.10 Editing the potato *GAI* gene

Cloning Cas9 constructs for *StGAI* editing

A L2 Cas9-expression gene editing construct was cloned into pICSL002220 – a version of the previously used pICSL4723_P1 L2 acceptor with the *NptII* gene and intronized Cas9 nuclease (Stuttmann et al., 2021) preassembled by Mark Youles. Into this acceptor I cloned a Cas9 guide acceptor with RFP counter-selection and internal BsaI recognition sites, under the control of the *A. thaliana U6-26* promoter and terminator. Next to this I also cloned the FASTred seed coat reporter (pICSL11123)(Shimada et al., 2010). The resulting L2 plasmid (YDJJ029) induced red colouration in the DH10B colonies which carried it.

YDJJ029 was then used in a diglig with BsaI where oligonucleotides (YD24-25) replaced the RFP module with the protospacer previously used by Tomlinson et al. (2019) to target the tomato *GAI* gene. The DH10B colonies which lost the RFP module also lost their red colour.

2.2.11 DELLA dwarf screening

GA3 spray screening

The T_0 plants were subject to two foliar sprays of GA_3 (100 mg/L) one week apart. The spray was applied until drippage from the leaves was observed, as done by Sharma et al. (1998). The F_1 lines were sprayed with GA_3 (200 mg/L) once per week for a total of four

times, again until drippage from the leaves was observed. The mock treatment consistent of spraying pure water.

F_1 genotyping via NGS amplicon sequencing

Crude genomic DNA was prepared from the F_1 progeny of line 457 × B100 using the following protocol, kindly provided by Xiao Lin. Leaf discs (diameter of ca. 1 cm) were collected in deep 96-well plate along with 2 metal beads per tube. NaOH (0.5 M; 40 µL) was added and the plate was spun down (4,000 g, 1 min). The samples were homogenized in a Geno Grinder™ (1,200 rpm; 2 min; Fisher Scientific) and spun down (8,000 rpm, 1 min). Tris-HCl (pH 7.5; 40 µL; 100 mM) was added and the tubes were shaken manually. The samples were spun down (8,000 rpm, 1 min) and 5 µL of the supernatant was transferred to a set of PCR tubes with Tris-HCl (pH 7.5; 180 µL; 100 mM). Crude DNA is of sufficient quality to support PCR amplifications up to 500 bp in size.

The crude DNA extracts were amplified with adapter primers (YD204-205) using Phusion DNA polymerase (DNA concentrations in the crude extracts were not measured, instead 1 µL of crude extract was added per 25 µL reaction; annealing temperature was 61°C for the first 10 cycles and 72°C for the following 25 cycles). Half of each reaction was run on a gel to confirm successful amplification, while the other half was amplified with barcoding primers. Eight barcoding forward primers (YD212-219) and four barcoding reverse primers (YD220-223) allowed for up to 32 samples to be mixed and submitted for NGS amplicon sequencing (AmpliconEZ service; Genewiz). Because of this, the samples were submitted in two batches. I handled the NGS amplicon sequencing results in Miniconda (conda version 24.3.0) running Jupyter Notebooks. This included read demultiplexing, mapping the demultiplexed reads to the *StGAI* locus, and querying the resulting alignments to calculate the proportions with or without the Valine deletion.

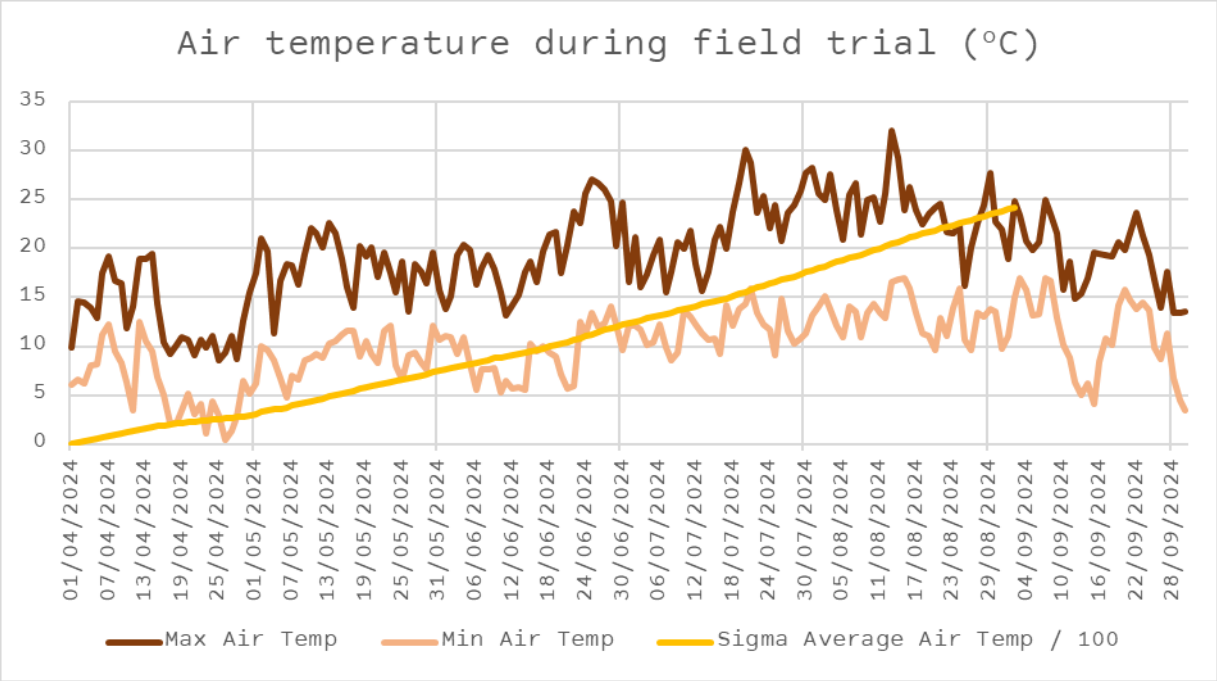
2.2.12 Field trial

Field trial design and conditions

Seed tubers, which were stored at 4°C, were subject to a chitting treatment before planting. As done by Marinus & Bodlaender (1978), the tubers were sprayed with GA_3 (10 mg/L) until completely wet, then left on a south-facing windowsill (1 week, RT).

Seed tubers were planted in rows 50 cm apart, with 25 cm of intra-row spacing (final planting density of 80,000 plants/ha). Considering the weight of dwarf and non-dwarf tubers, this is a seed rate of ca. 1 t/ha. This density is roughly in line with established practices in potato growing (Allen, 1978; van Dijk et al., 2022), while accounting for the seed potatoes' unusually small size (Ifenkwe & Allen, 1978) and the need to discriminate between tubers from different lines growing in the same row. One row of the cultivar Swift was planted around every plot to serve as guard crops. On 11.06.2024 the guard crops were cut back to ca. 30 cm above the ground as, due to the substantially higher quality of the commercially sourced seed tubers, they were outgrowing and excessively shading the trial plants.

The field trial was done on the Football Field, Church Farm, John Innes Experimental Field Station, Bawburgh, Norfolk, UK. Soil type was sandy clay loam. I planted the seed tubers on 22.04.2024 at a depth of ca. 5 cm. Later, field station staff hilled the potatoes, i.e. pilled soil around the base of the plants to promoter tuber growth. The plants were grown for 132 days before plant tops were cut off on 02.09.2024, and tubers harvested on 20.09.2024 (Fig. 2.7). I weeded the plots by hand roughly every 2 weeks to ensure the lines of interest were the only significant source of biomass in the field. Plants were treated with pesticides and fertilizers by field station staff as needed, including a total of 174 kg/ha of Nitrogen (Extran).



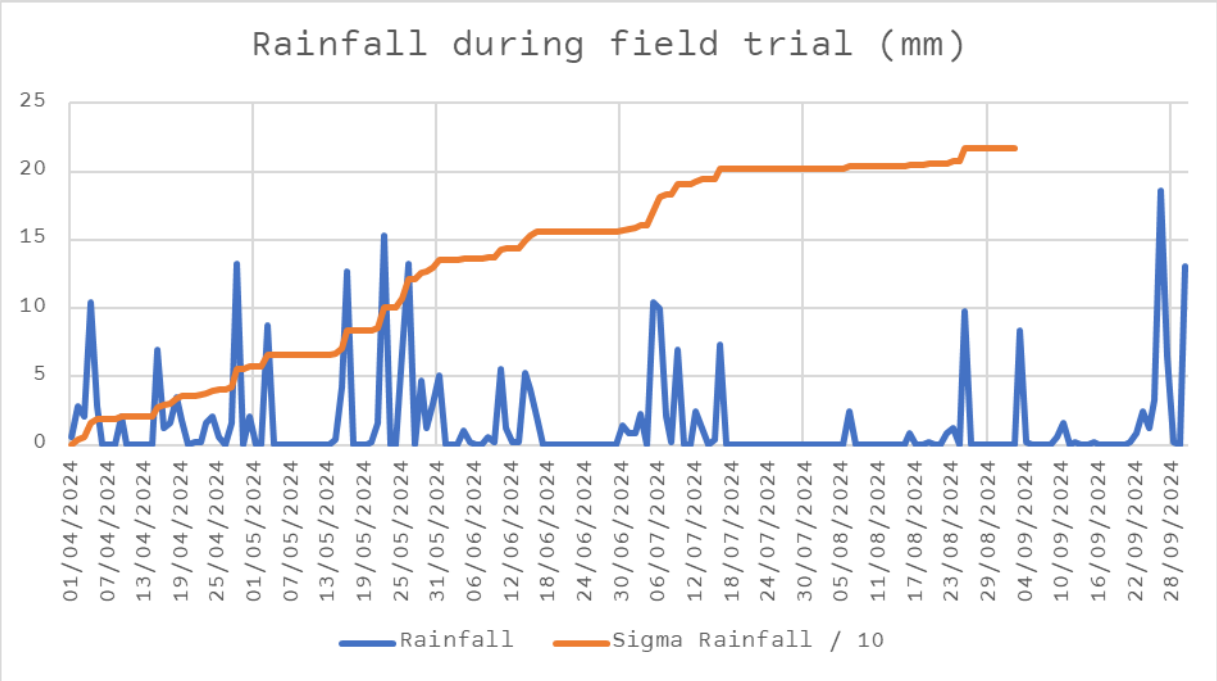


Figure 2.7. Field trial climactic conditions.
Data from Church Farm Weather Station. Location: 52°37’52.29”, 1°10’23.57”. Sigma curves terminate at 02.09.2024 when plant tops were cut off.

Field phenotyping via drone

Drone images were collected weekly by Phil Robinson and Martin Vickers using a DJI Mavic 3M drone carrying a 4/3 CMOS 20 MP RGB camera and a 5MP Multispectral camera (NIR/RE/R/G). Data were collected at a height of 12 m. To make drone measurements more accurate, real-time kinematics (RTK) corrections were obtained

from an NTRIP network using a mobile phone hotspot. This, combined with the drone's GPS data, delivers cm-level accuracy. Raw drone images were processed into orthomosaics by Phil Robinson using the Pix4D software. I further analysed these in QGIS Maidenhead (v3.36).

Despite the use of RTK corrections, the RGB (red-green-blue) and DSM (digital surface model) data from different dates were not perfectly aligned. To remedy this, five ground control points (GCPs) were used to manually align the images further. The positions of GCPs along the X and Y axes differed only slightly between dates and were manually corrected (avg. adjustment: 3.6 cm; max adjustment: 6.5 cm). Similarly, the DSMs from different dates were not perfectly aligned on the Z axis. To account for this, I measured the height values above 3 GCPs, where no growth had occurred, on the DSM layer of each date. Here, the differences between dates were more considerable (avg. adjustment: 31 cm; max adjustment: 65 cm) but the height of GCPs from the same date remained highly consistent (avg. $\text{variance}_{GCP1\text{-}3}$ = 7 cm; Table 2.7). The difference in mean DSM height between dates, relative to the first measurement (20th May) was used as a normalisation factor and subtracted from each pixel of that date's DSM.

Date	GCP1	GCP2	GCP3	delta1	delta2	delta3	delta average (norm. factor)	variance between dsm measurements
8-29	65.02	64.88	65.51	0.39	0.40	0.38	0.39	0.07
8-21	64.84	64.70	65.34	0.22	0.22	0.21	0.22	0.08
8-08	65.28	65.14	65.76	0.65	0.66	0.63	0.65	0.07
8-01	65.22	65.08	65.72	0.59	0.60	0.59	0.59	0.08
7-24	65.01	64.87	65.51	0.38	0.40	0.37	0.38	0.07
7-19	65.15	65.01	65.62	0.52	0.53	0.49	0.51	0.07
7-09	65.14	64.99	65.61	0.51	0.52	0.47	0.50	0.07
7-01	64.93	64.77	65.42	0.30	0.30	0.29	0.29	0.08
6-25	65.02	64.87	65.50	0.40	0.40	0.37	0.39	0.07
6-19	65.02	64.87	65.48	0.39	0.39	0.35	0.38	0.07
6-13	64.77	64.57	65.26	0.14	0.10	0.12	0.12	0.08
6-07	64.80	64.65	65.28	0.17	0.17	0.14	0.16	0.07
5-31	64.53	64.39	65.02	-0.09	-0.08	-0.11	-0.10	0.07
5-24	64.86	64.71	65.37	0.23	0.24	0.24	0.23	0.08
5-20	64.63	64.48	65.13	0.00	0.00	0.00	0.00	0.08

Table 2.9. DSM correction between dates.
All height measurements are in meters above sea level.

2.3 Statistical methods

Throughout this thesis, a variety of types of data was collected, which necessitated equally varied statistical approaches. To learn how to use these, I attended several statistical training courses provided by the Norwich Research Park Graduate Students Office, including: *Introduction to R for Statistics*, *Basic Statistics & Design Principles with R*, *Experimental Design and Analysis with R*, *Linear Mixed Models with R*. Additionally, I made use of the BSc course materials from the Quantitative Statistical Methods module (BL3320) from the University of St Andrews.

I conducted the statistical analysis in RStudio (version 2024.09.0+375) running R (4.4.1). I plotted the data either with base R plotting functions, or with the “ggplot2” package (Wickham, 2009). My colour choices were guided by the colourblind-friendly *Turbo* palette (Mikhailov, 2019), accessed through the package “viridis”.

2.3.1 Statistics used in KIs and TIIGER

The significance of the difference between observed and expected segregation ratios was tested using Fisher’s Exact Test, a

nonparametric method similar to a Chi-squared test but with superior performance when sample sizes are small (Bower, 2003). This was employed in R using the base function “fisher.test”.

Time-to-Event analysis

The emergence of KI sectors over time was compared using Time-to-Event analysis, also called Survival analysis. This is a statistical approach used to track and compare the probability of an event (e.g. individual death, machine failure, KI sector appearance) occurring in a population (e.g. patients, manufacturing batch, explant termini) over time (Schober & Vetter, 2018).

The event occurrence curves (incidence curves) of two groups were compared using a log-rank test – a non-parametric statistical test for null hypothesis testing. Log-rank tests are broadly analogous to conducting a Chi Squared test at each time point, followed by a summation of the resulting parameters and the calculation of a final P-value. A Bonferroni correction was applied to the P-values when conducting multiple Log-rank tests on the results of one experiment.

The Time-to-Event analysis was done in R using packages “survival” and “survminer”. As an example, the raw code conducting the analysis of one experiment is supplied in the Appendix (App. 2).

2.3.2 Field trial statistics

Below I describe the modelling approach used for every dataset, the final model, and a diagnostic plot to facilitate visual model evaluation (Residuals vs Fitted). This plots each predicted (Fitted) value (X axis) against its residual value (Fitted value minus the actual measured value; Y axis). Both axes retain the units of whatever the model predicts (kg, cm, etc.). Importantly, if the model fits the data well, then the residual values vary randomly, and the Residuals vs Fitted plot should not show any clustering.

Comparing seed tuber weights

To capture the size variability of the seed tubers used in the field trial, and to check if the seed tubers used for the dwarfs were significantly bigger than those of the non-dwarfs, I picked

12 random tubers from each group and weighed them. The data were analysed with a linear model (one-way ANOVA):

lm(formula = weight ~ group, data = df)

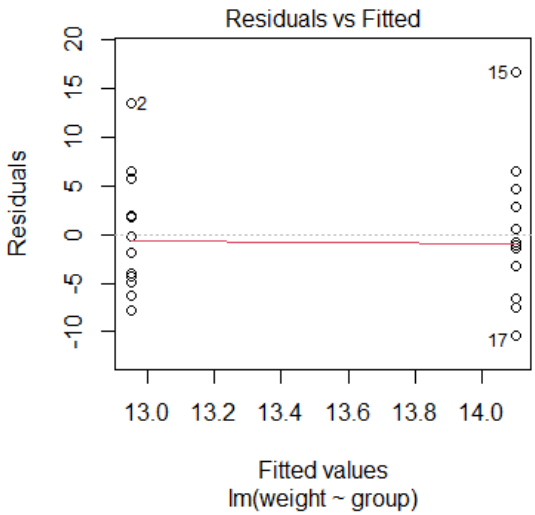


Figure 2.8. Diagnostic plot for weight comparison of tubers used as seed in the field trial.
The residual errors are randomly distributed above and below the two predicted means. Axes in grams.

Field trial manual plant height measurements

On 26.07.2024 (94 days after tuber planting in the field), plant height was measured with a vertical ruler, from the top of the hill to the tip of each plant's tallest stem. Four measurements were made per line, each from a different plant, except for line 1-1 where only three of four tubers sprouted. The plant height data was analysed with a Linear Mixed Effects model which fit the data very well:

lmer(height~geno+plot+(1|var), data=height)

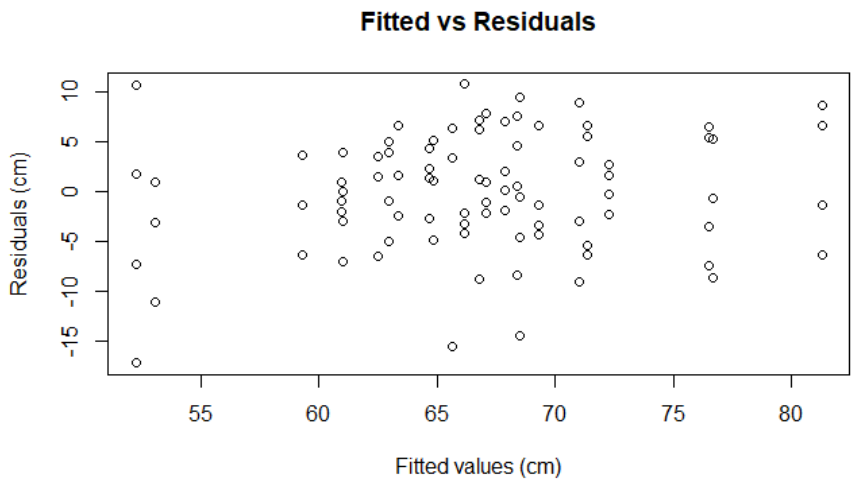


Figure 2.9. Diagnostic plot for manual plant height measurements. The "starry night" appearance of the plot indicates the residual errors are randomly distributed. Axes in cm.

Comparing field trial plot plant height data

The mean plot height, plotted against time, followed a Gaussian ("bell") curve. To model this, I adapted the base form of the Gaussian curve:

$$F(x) = Bell\ max\ height * e^{-\frac{1}{2}\left[\frac{X - Bell\ centre}{Bell\ width}\right]^2}$$

into:

$$Height(day\ X) = (Max\ canopy\ height + Genotype\ effect) * e^{-\frac{1}{2}\left[\frac{day\ X - Peak\ day}{Curve\ width}\right]^2}$$

I employed Non-linear Mixed Effects Modelling using the R package "nlme", with the data grouped by plot (due to repeated measurements of the same plot), and a random component affecting the baseline maximum height:

```
nlme(mean~max*exp(-0.5*((days-mid)/b)^2), data = deriv_dsm,
fixed = c(max~geno, mid+b~1), random = max~1,
start=c(max=0.54893, max.genodwarf=0.54893, mid=95.21822,
b=20.47498), groups=~plot)
```

The starting fixed effects estimates were based on a simpler nonlinear model (which did not account for the repeated measurements or the effect of genotype), which itself had starting values visually estimated.

The Gaussian Non-linear Mixed Effects model fit the data very well. The slight clustering of residuals at the lower fitted

values is likely due to guard crop overgrowth in the first few weeks, which encroached into the measurement area and resulted in "more plant height" than the model would expect at that time. The guard's excessive growth was addressed on the 11.06.2024 by cutting them back. The raw code conducting the Non-linear Mixed Effects model is provided in the Appendix (App. 3).

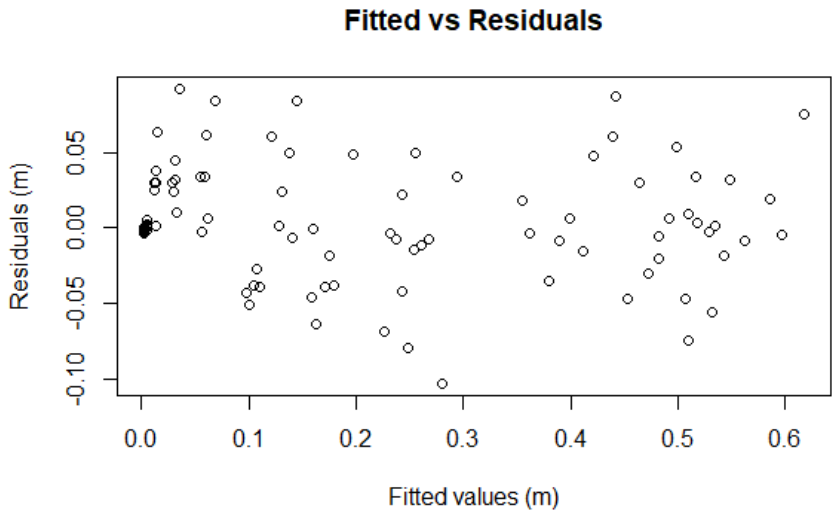


Figure 2.10. Diagnostic plots of the Non-linear Mixed Effects Model.
The "starry night" appearance of the plot indicates the residual errors are randomly distributed. Clustering in lower end likely due to guard plant overgrowth. Axes are in meters.

Comparing field trial yields

After the tubers were harvested, I weighed the fresh tuber yield of each line with an electronic scale. After weighting, some tubers which were affected with what appeared to be *Fusarium* dry rot, a hemibiotrophic disease, were discarded. Yield per line was converted into yield per plant by dividing the total yield by 4, except in line 1-1 where it was divided by 3 as only 3 of 4 tubers sprouted.

I investigated whether the lines' genotype affected their yield using the Student's t-test. This makes several assumptions which I tested first.

1) I confirmed the yields per plant were normally distributed in both dwarfs (Shapiro-Wilk normality test, W = 0.98, P-value = 1.0) and non-dwarfs (Shapiro-Wilk normality test, W = 0.96, P-value =

0.8) (Fig. 2.11 A). Code used:

```
shapiro.test(subset(tuber_yield, geno=="non")$yieldppl)
shapiro.test(subset(tuber_yield, geno=="dwarf")$yieldppl)
```

2) I confirmed both samples had essentially the same variance (F-test, $F_{(11,11)}$ = 1.08, P-value = 0.9). Code used for F-test:

```
var.test(yieldppl ~ geno, data = tuber_yield)
```

3) The t-test assumes the data points in one sample are independent of the other. This was true as the yields originated from different F_1 lines grown in separate plots.

4) A major potential issue was the fact that the factors of genotype and plot were perfectly confounded. Dwarf lines were grown in 3 plots, and non-dwarfs in another 3 plots, but as they never shared a plot it was impossible to completely separate the effect of genotype and plot. However, it was possible to test if plot affected yield per plant by comparing the variance between and within plots. A one-way ANOVA showed the variance between plots (Mean Square = 20801) was 3X less than the variance within plots (Mean Square = 63093), indicating that plot likely had an insignificant effect on yield ($F_{5,18}$ = 0.33, P-value = 0.9). This could also be seen by plotting the yield per plant against plot (Fig. 2.11 B). Code used for one-way ANOVA:

```
summary(aov(yieldppl ~ plot, data = tuber_yield))
```

Having confirmed the validity of the statistical approach, I tested the null hypothesis: that the average yield per plant did not differ between dwarfs and non-dwarfs. The average yield per dwarf plant (504 ± 241 g [SD]) and non-dwarf plant (470 ± 232 g [SD]) did not differ significantly (two-sample Student's t-test, $t_{(22)}$ = 0.35, P-value = 0.73). Code used for t-test:

```
t.test(tuber_yield$yieldppl~tuber_yield$geno, var.equal=TRUE)
```

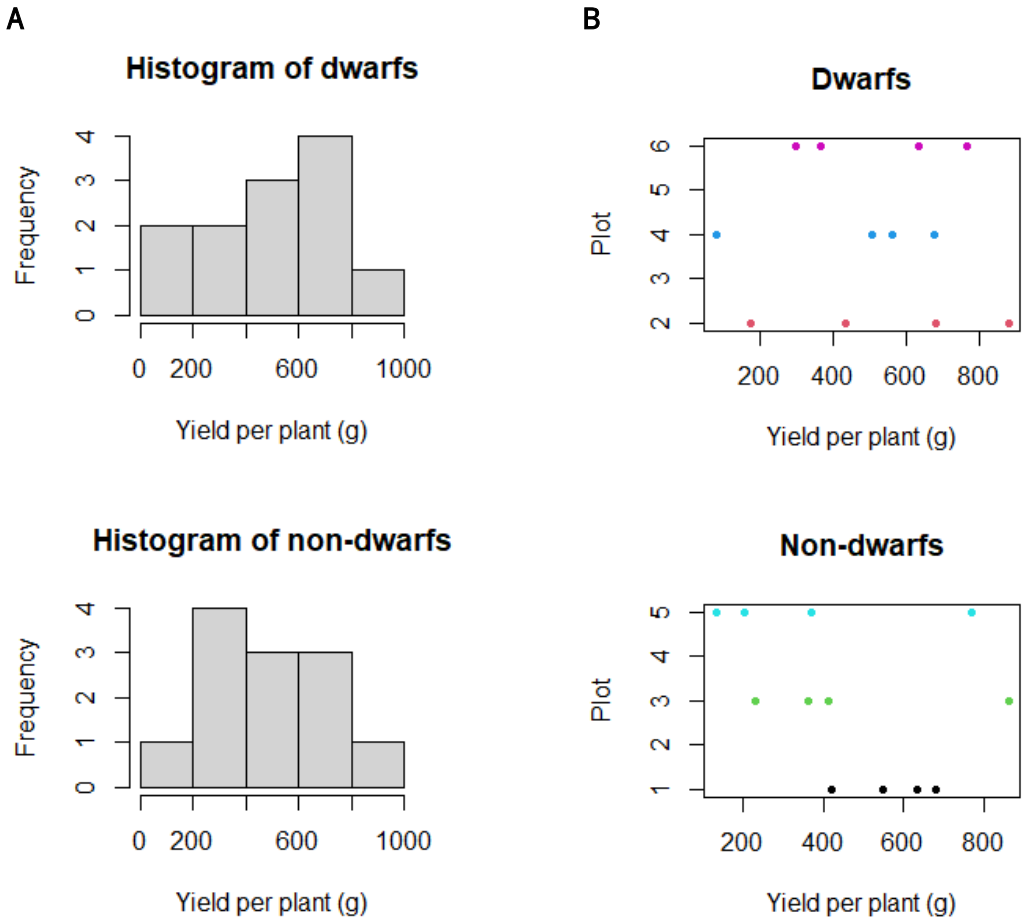


Figure 2.11. Descriptive plots of field trial yield data.
A Histogram of the yield per plant of dwarfs and non-dwarfs showing a normal distribution. **B** The yield per plant of individual dwarf and non-dwarf lines grouped by plot. The variance between plots is 3X less than that within plots, indicating plot had no insignificant effect on yield.

Comparing field trial tuber size distributions

I completed the tuber size distribution analysis in ImageJ (1.54d) running Java (1.8.0_345).

The histogram of the tuber size distributions showed a shift towards larger tubers in the non-dwarf harvest. However, most statistical tests compare the means or medians of distributions and would not adequately capture the difference observed. Instead, I employed the Two-sample Cramér-von Mises Test - a non-parametric test, sensitive to changes in distribution shape (Anderson, 1962). This was used to conduct null hypothesis testing, this being that the two sample distributions were sampled from the same underlying

distribution. The test was run using the “twosamples” statistical package and the following code:

```
test1=cvm_test(
            subset(size_data, size_data$geno=="dwarf")$area,
            subset(size_data, size_data$geno=="non")$area,
            nboots = 20000
            )
summary(test1)
```

Comparing field trial tuber shape

I completed the tuber shape analysis in ImageJ (1.54d) running Java (1.8.0_345).

The tubers’ aspect ratio (AR) data were not normally distributed, as a small minority of tubers were exceptionally long and had an AR > 2 (Shapiro-Wilk normality test, P-value << 0.001; Fig. 2.12 Left). To simplify statistical analysis, the AR data was BoxCox-transformed (Box & Cox, 1964) to achieve normality (λ = -1.111111; Shapiro-Wilk normality test, P-value = 0.18; Fig. 2.12 Middle). The transformed AR was then modelled with a linear model which accounted for the effect of the genotype (dwarf/non-dwarf) and the individual potato line:

```
lm(new_ar ~ geno + line, data=shape)
```

This fit the data well (Fig. 2.12 Right). The R code conducting the data transformation and subsequent analysis is provided in the Appendix (App. 4).

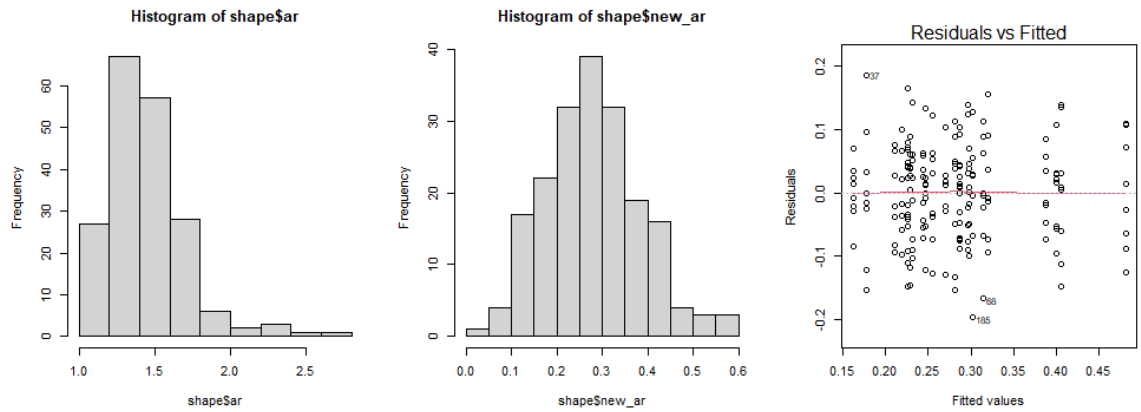


Figure 2.12. Plots of tuber shape data and linear model residuals. A Histogram of the tuber Aspect Ratio, showing a non-normal distribution. **B** Histogram of the transformed tuber Aspect Ratio, showing a normal distribution. **C** Residual vs Fitted plot for the linear model. The “starry night” appearance of the plot indicates the residual errors are randomly distributed.

CHAPTER 3. TARGETED GENE INSERTIONS

3.1 Introduction

The early work by Puchta et al. (1996) highlighted the limitations posed by available nucleases and the impact of competing repair pathways. Addressing the latter by shifting the balance towards HDR in plants has been a longstanding research goal (more extensively reviewed by Liu et al. (2019)). A popular genetic approach has sought the suppression of genes involved in the competing NHEJ pathway, like *KU70*, *KU80*, and *DNA Ligase 4*, and this has resulted in moderate increases in the frequency of GT and homologous recombination (Nishizawa-Yokoi et al., 2012; Qi et al., 2013; Vu et al., 2024). Conversely, attempts have been made to directly stimulate homologous recombination by the expression of participating genes, like *AtMIM* (Hanin et al., 2001), *ScRAD54* (Shaked et al., 2005), and *EcRuvC* (Shalev et al., 1999), again with some success. In parallel, small molecule effectors have been pursued an attractive, non-genetic means to modulate DSB repair pathways. 3-methoxybenzamide (3-MB), an inhibitor of poly(ADP)ribose polymerases, has shown promise. Early studies in tobacco revealed that 3-MB treatment increased HR rates severalfold (Puchta et al., 1995). More recently, Chetty et al. (2020) demonstrated a 2.7-fold increase in transformation frequency and improved biomass in potato explants treated with 3-MB. Despite these advances, however, none of the described methods have consistently elevated GT rates to levels practical for routine application.

An interesting approach to increasing HDR rates was explored by Baltes et al. (2014) using Geminivirus replicons in tobacco. Geminiviruses have circular single-stranded DNA genomes which are converted into double-stranded DNA intermediates by the host’s DNA polymerases, thus allowing for the expression of viral genes and viral genome replication (Fig. 3.1 A). Baltes et al. (2014) included the minimal viral elements required for genome circularization and replication on a T-DNA featuring GE reagents. This resulted in a proliferation of the vector *in planta*, and the

multiplication of the available repair template and facilitation of HDR (Fig. 3.1 B). There have since been mixed reports regarding the effectiveness of using viral replicons to achieve KIs across various systems. While some studies have reported positive outcomes (Butler et al., 2016; Gil-Humanes et al., 2017; M. Wang et al., 2017), others have found them to be counterproductive (Grandgeorge, 2021; Lawrenson et al., 2021).

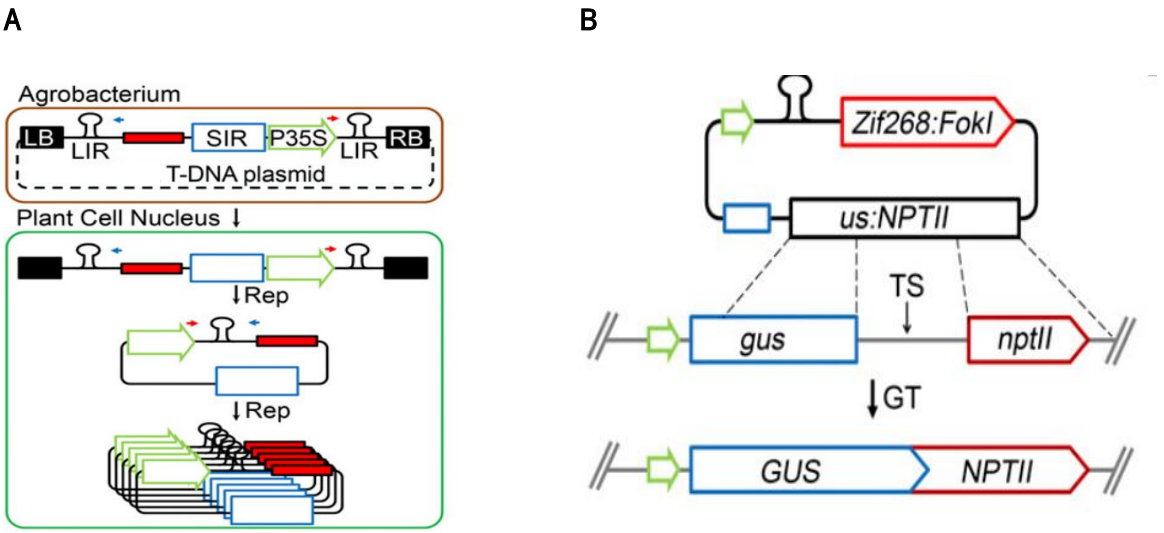


Figure 3.1. Use of Geminivirus replicons for improved KI rates. **A** Viral replication requires 3 elements: the large intergenic region (LIR), the short intergenic region (SIR), and the trans-acting replication- initiation protein (Rep). A T-DNA bearing those elements can circularise and proliferate *in planta*, thus potentially increasing the copy number of whatever other reagents are cloned into the T-DNA (red rectangle). **B** This tactic was also used to increase the amount of repair template available for KIs. Adapted from: Baltes *et al.* (2014).

Čermák et al. (2015) used viral replicons to KI genes in tomato but instead of detecting these through a disrupted selectable marker, the authors targeted an endogenous gene and developed an anthocyanin-based KI reporter (Fig. 3.2). Anthocyanin biosynthesis is regulated by R2R3-MYB transcription factors, bHLH transcription factors, and WD repeat proteins (Y. Zhang et al., 2014). Among these, MYB transcription factors primarily dictate anthocyanin production levels. Čermák et al. (2015) targeted the *SlANT1* MYB transcription factor gene, which induces robust and widespread anthocyanin accumulation when overexpressed (Mathews et al.,

2003). The authors induced a DSB just upstream of the *ANT1* start codon and provided a repair template, via T-DNA, containing the *NptII* selectable marker and the Cauliflower Mosaic Virus 35S promoter (35S) (Fig. 3.2 A). A KI resulted in the 35S promoter overexpressing *ANT1*, which was visually identifiable by the tissue's purple coloration (Fig. 3.2 B-H). This system enabled clear discrimination of KIs but also caused a moderately severe developmental phenotype in purple tissue, which complicated the recovery of KI events. Still, it offered several advantages: firstly, the frequency of purple sectors provided a convenient proxy for KI frequency, facilitating construct optimization. Secondly, the reporter's visual nature allowed it to capture KI events that would occur throughout the tissue culture stage, rather than just the events that occurred before or shortly after the introduction of some chemical selection. Lastly, the system's KI rates were more relevant to potential future applications, such as the insertion of R-genes, where the inclusion of a selectable marker in the KI cassette or target locus might be undesirable. Using this reporter system, Čermák et al. (2015) ultimately recovered two independent KI events from five purple calli that produced shoots. Although the exact efficiency was low and not clearly reported, it was a noteworthy achievement, given the absence of direct chemical selection for the insertion.

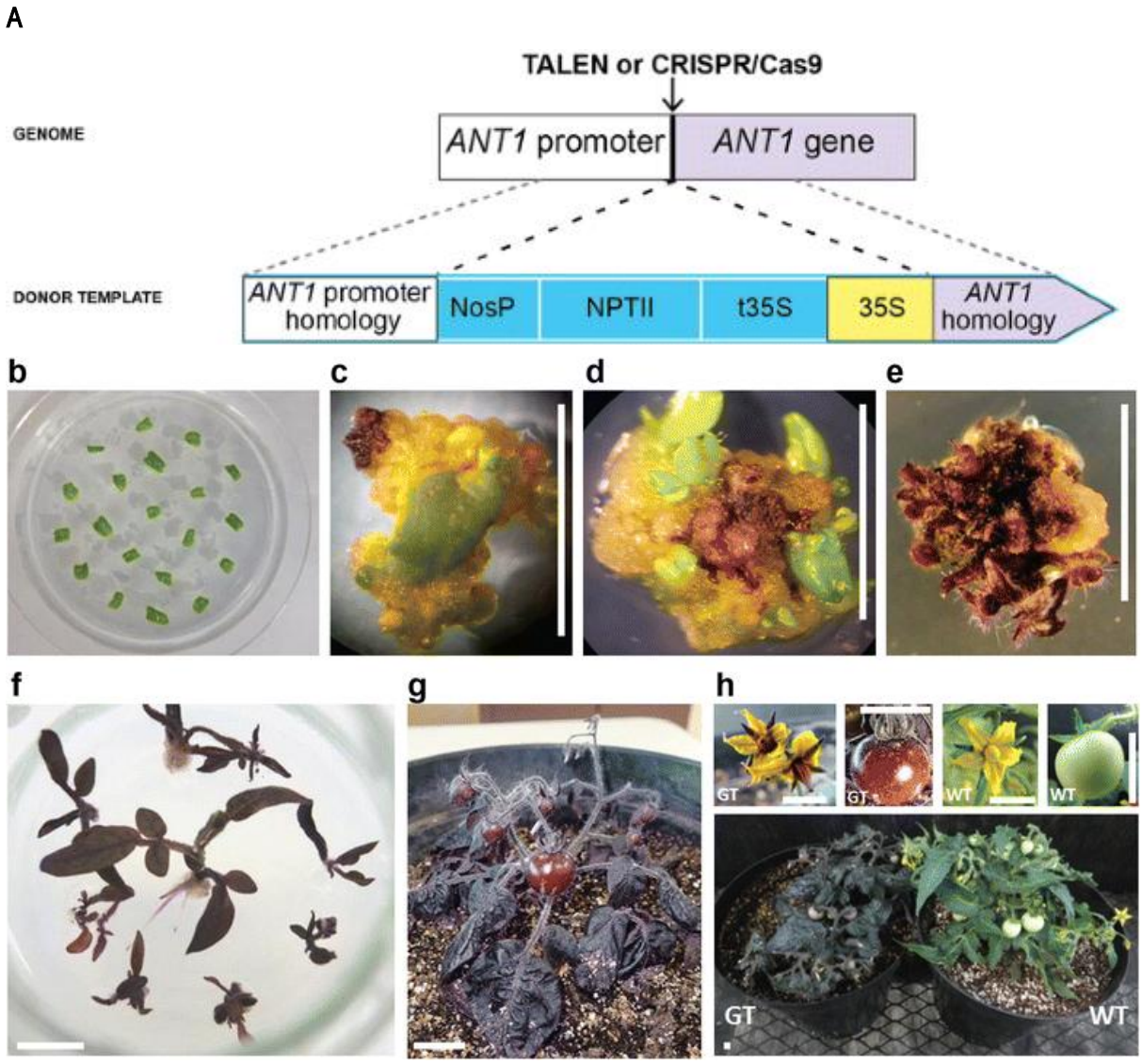


Figure 3.2. Anthocyanin KI reporter system.
A Diagram of a KI occurring upstream of the *ANT1* gene. Dotted lines between the genome and donor template indicate regions of homology intended to facilitate HDR. **B-H** Purple tomatoes recovery process. *ANT1* overexpression has a moderately severe developmental phenotype. HDR – homology-directed repair.
Adapted from: Čermák et al. (2015).

Considerable progress was also made with the introduction of better programmable nucleases. ZFNs and TALENs were a major development, as they enabled the targeting of endogenous loci with relative ease (Shukla et al., 2009; Townsend et al., 2009), but both were superseded with the introduction of the CRISPR/Cas9 gene editing system (Jinek et al., 2012). Once it was established that this could effectively cleave plant genomic DNA (Nekrasov et al., 2013), numerous groups attempted to conduct GT using this

exceptionally versatile and approachable nuclease. These attempts, which were thoroughly reviewed by Huang & Puchta (2019), took place in a wider variety of plant species than before, including some important crops, but the approach was largely limited to targeting a gene that results in chemical resistance if successfully modified, usually the *ALS1* gene or an orthologue thereof. One notable exception was the work by S. Li et al. (2018) who conducted the targeted substitution of several nucleotides in the endogenous rice gene *OsNRT1.1B*, obtaining an allele associated with higher nitrogen use efficiency (and no other obvious phenotype) in 6.72% of recovered plants. Usually, however, the frequency of recovering even selectable GTs remained low, and the prospect of efficient KIs of whole genes – remote.

The next significant advancement came with the introduction of the CRISPR/Cas12a system (also known as Cpf1) by Zetsche et al. (2015). This proved to be functional in plants and suitable for GT approaches (Begemann et al., 2017). The system was further improved by Schindele & Puchta (2020), who engineered a temperature-tolerant version of the Cas12a derived from *Lachnospiraceae bacterium ND2006* (ttLbCas12a), which cleaves more efficiently at the relatively cool temperatures conducive to plant tissue culture. Huang et al. (2021) used the improved enzyme to efficiently generate tobacco lines with targeted substitutions in their *ALS1* orthologue, 1/5th of which were T-DNA-free. The authors did this using ttLbCas12a, LbCas12a, and SaCas9 and found the first of these to be significantly more efficient than the rest, while work by L. Grandgeorge (2021) indicated ttLbCas12a is also superior to SpCas9 at induction of gene KIs in tomato. The expression level of the ttLbCas12a was later improved by the addition of numerous introns into its codon-optimised CDS (ittLbCas12a), which further improved GT rates (Lawrenson et al., 2024; Schindele et al., 2023). These iterative developments have made ttLbCas12a into the current nuclease of choice for GT in plants.

Schreiber et al. (2024) recently sought to enhance the performance of Cas9 and Cas12a at KIs by modulating the DNA repair pathways that address the DSBs induced by these enzymes. To achieve this,

they fused the nucleases to various 5' and 3' exonucleases, hypothesizing that strand resection following cleavage could promote HDR. The authors explored a broad range of parameters and evaluated the constructs' efficiency in a transient assay. Notably, fusing the UL12 exonuclease from *Herpes simplex virus 1* to ittLbCas12a increased KI rates approximately tenfold. However, most of their work focused on Cas9-based constructs and those are the ones they selected for stable transformation. In an *Arabidopsis* floral dipping experiment, the fusion of a UL12 orthologue to an intron-containing and codon-optimised Cas9 increased KI efficiency for a 1.4 kb cassette tenfold, with 1/6th of the recovered KI events being T-DNA-free. In wheat, which is much more challenging to transform, the same fusion achieved the successful KI of a 552 bp-long selectable marker CDS in a single plant out of 156 screened, while the unfused Cas9 yielded none among 251 screened plants. Thus, even a moderately efficient KI approach is dependent on efficient delivery of gene editing reagents to plant cells.

In closing, the field of GT has advanced significantly, with improvements in both cleavage efficiency and HDR frequency in plants. It seems likely that, by combining already published improvements into more complex approaches, further gains might be achieved. Much of the recent progress has been achieved in model systems like tobacco and tomato, offering great promise for other Solanaceous species like potato. While the recovery of stable KI lines has been reported at modest KI frequencies, these successes typically involve the insertion of relatively small cassettes into loci with easy to screen phenotypes. The delivery of large cargo such as R-genes with their own promoters and terminators— which are usually several kilobases long and lack an immediately detectable phenotype — remains a substantial challenge.

Aims

This chapter focuses on establishing an efficient method for gene KI in diploid potato, with the goal of KI breeding R-genes from wild potato relatives against *P. infestans*. This project had the following concrete aims:

- Transfer the anthocyanin KI reporter system demonstrated by Čermák et al. (2015) from tomato to diploid potato
- Use the above to recover a genuine gene KI
- Vary elements of the KI approach to establish an efficient method for KI breeding and investigate what factors constrain KI efficiency
- KI R-genes against *P. infestans* into diploid potato germplasm
- Demonstrate gain of resistance against *P. infestans* resulting from the R-gene KIs

3.2 Results

3.2.1 *Cai1*, a functional anthocyanin KI reporter in potato

Designating a target locus

The Solyntus diploid potato genome (van Lieshout et al., 2020) holds numerous MYB TFs homologous to *SlANT1* but the orthologue is chlorogenic acid inducer 1 (*Cai1*; Gene ID: 102590860; Soltu.DM.10G020820.1), also sometimes called *Mtf1* (Rommens et al., 2008). Details on how this locus was identified as the orthologue can be found in Section 2.2.7 but the same conclusion was also reached by Powell et al. (2022). I characterised *Cai1* within five elite diploid potato lines obtained from Solynta: B26, B98, B99, B100, and Solyntus. Lines B26 and B100 were of particular interest, due to their relatively high transformability or fertility, respectively. Anticipating potential differences in the *Cai1* allele sequences between lines, I designed several PCR primers (YDprF1-3 & YDprR4-7) to amplify across the locus (Fig. 3.3 A). The first tested primer combination (YDprF1/YDprR7) successfully amplified *Cai1* from Solyntus, B99, and B100, but not B26 and B98 (Fig. 3.3 B). I then tested all possible primer combinations on B26 and B98 (Fig. 3.3 C). Some combinations failed to amplify, indicating sequence divergence, and all amplicons were about 0.7 kb larger than expected, signifying that the *Cai1* alleles detected carried an insertion.

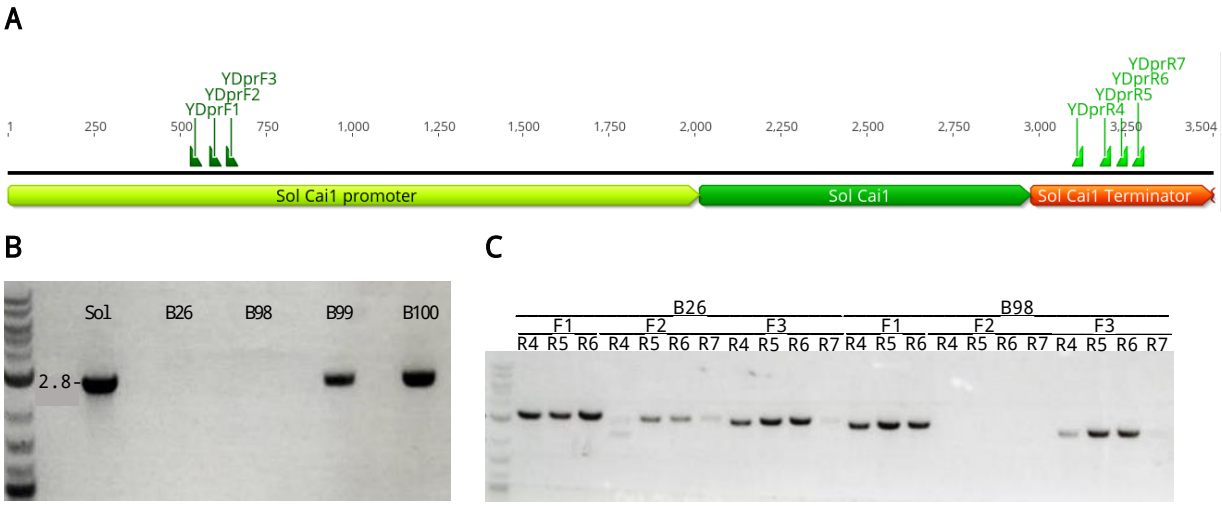


Figure 3.3. Cloning *Cai1* alleles.
A The Solyntus *Cai1* allele sequence with ORF (green), promoter (light green), terminator (orange) and flanking primers annotated. **B** Primers YDprF1 and YDprR7 produce an amplicon with gDNA from Solyntus, B99, and B100 but not B26 and B98. **C** The *Cai1* allele being amplified from B26 and B98 gDNA with all possible primer combinations except YDprF1/YdprR7. Ladder in all images is 1 kb plus (NEB).

The amplicons from all five lines were sequenced. The reads from lines B99, B100, and Solyntus all mapped perfectly to the *Cai1* allele from the Solyntus genome (SFig. 2). In contrast, mapping the reads from the larger amplicons from B26 and B98 revealed numerous SNPs and indels, and both had a 0.5 kb insertion just 60 bp upstream of the *Cai1* start codon. B26, which had its *Cai1* locus more extensively sequenced, also had an additional 0.2 kb insertion further upstream (Fig. 3.4 A). All sequenced alleles had complete ORFs which seemed to encode functional proteins. The B26 allele was further analysed and annotated (Fig. 3.4 B). The two insertions resembled transposable elements (TEs), with terminal inverted repeats and flanking direct repeats. The upstream TE (Fig. 3.4 B TE1.1) appears to have copied itself, indicating that it is likely a Type I TE element. The downstream TE (Fig. 3.4 B TE2) appears to be a non-autonomous Type II TE with terminal inverted repeats (TIRs). This was confirmed by submitting both sequences to the TE classifier tool *TEclass2* (Bickmann et al., 2023), which assigned these to the Class I-LTR Copia and Class II-TIR hAT classes. The TE-containing promoter found in B26's *Cai1*

allele is likely functional, as the whole allele shares 97.2% identity with that of the doubled monohaploid DM 1-3 516 R44 genome (v6.1; Pham et al., 2020) and that gene is expressed (SRR12168747).

Most of the cloning involved in the KI work going forward was based on the B26 *Cai1* allele established here and, despite doing tens of PCRs amplifying various parts of this locus over the course of this project, no evidence emerged to suggest any heterozygosity at *Cai1* in line B26.

A

Figure 3.4. The *Cai1* allele sequences.
A Alignment of the Solyntus *Cai1* allele, the B26 *Cai1* allele, and the B98 *Cai1* allele (partially sequenced). Differences between the alleles include SNPs, small indels, and insertions in the promoter region. All three alleles contain complete ORFs. **B** Alignment of the *Cai1* alleles from Solyntus (top) and B26 (bottom). TEs are annotated in red, Terminal inverted repeats (TIRs) in purple, Direct flanking repeats in blue. A potentially disabled TE (TE 1) is annotated in pale red, its mutated right-side TIR in grey. Pairwise Identity = 72.5%.

Conversely, several results hinted at heterozygosity for *Cai1* in B100. Firstly, PCRs amplifying Solyntus and B100 gDNA with primers YDprF1/YDprR7 consistently resulted in weaker bands with the latter, despite adding an equal amount of template (Fig. 3.3 B). Furthermore, PCR amplifying B100 gDNA across the *Cai1* locus with primers YD35/YD36 resulted in 2 bands, ca. 1.26 and 1.81 kb in length (Fig. 3.5). These are the predicted amplicon sizes based on

the Solyntus and B26 *Cai1* alleles, respectively, and sequencing reads from the TA-cloned PCR-products mapped perfectly (SFig. 3). Thus, the heterozygosity of B100 was confirmed and the sequence of both alleles established. Going forward, the Solyntus and B26 *Cai1* alleles are referred to as *Cai1-short* and *Cai1-long*, respectively.

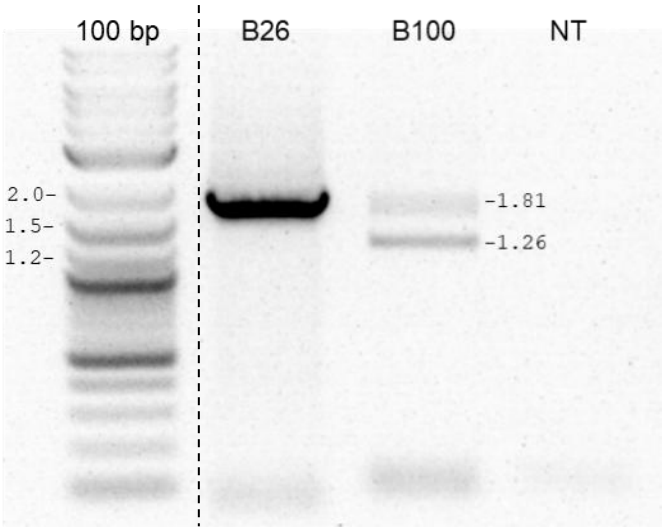


Figure 3.5. B26 and B100 are homo- and heterozygous for *Cai1*
B26 and B100 gDNA amplified with primers YD35/YD36 using the modified PCR protocol described in Section 2.2.3. The homozygous B26 yields a single 1.81 kb band, while the heterozygous B100 produces that, plus a 1.26 kb band. Ladder used was 1 kb+ DNA Ladder (NEB). Dotted line indicates where a section of gel was digitally removed. The whole gel image can be seen in Figure 3.21.

crRNA and homology arm design

Efficient KIs at *Cai1* require that the locus be targeted with the LbCas12a nuclease. A screen for suitable crRNA target sites was done on the *Cai1-long* allele, due to B26's superior transformability. According to publicly available RNA-Seq data (accession SAMN20673797), the *Cai1* ORF is 983 bp long, and consists of three exons, two short introns, and a 5'UTR just 11 bp long. Consistent with that, a canonical TATA box motif (…TG**TATA**TAAGTA…) was identified with high-confidence just 33 bp upstream of the ATG start codon. I analysed the 60 bp region upstream of the start codon and identified a single suitable crRNA which sits on top of the TATA box (crRNA1; for crRNA selection methodology, see Section 2.2.7). Regions roughly 800 bp upstream and downstream of crRNA1's cut site were designated as homology

arm left (HA-L; spans positions -862 to -34 relative to start of translation) and homology arm right (HA-R; spans positions -28 to 774 relative to start of translation), respectively (Fig. 3.6).

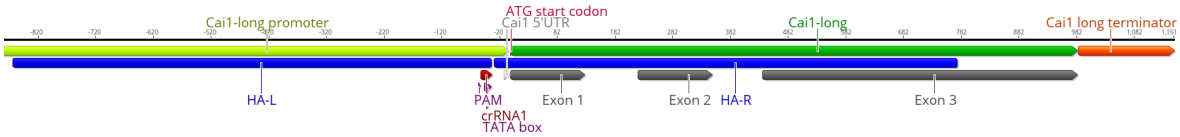


Figure 3.6. Targeting the *Cai1-long* allele.
The *Cai1-long* ORF (green) contains three exons (grey). The 5'UTR is just 11 bp (white) and a canonical TATA box (purple) sits just 33 bp upstream of the ATG. The crRNA1 protospacer sits on top of the TATA box. Homology arms (HAs, blue) span positions [-862; -34] & [-28; 774] relative to the ATG.

Confirming *Cai1* function

Rommens et al. (2008) reported that *Cai1* overexpression results in visible anthocyanin accumulation. To confirm that was also the case for the alleles present in B26, I amplified the part of the *Cai1-long* allele that would be downstream of any potential insertion – from the crRNA1 cut site until ca. 600 bp downstream of the STOP codon – and cloned it under the 35S promoter + TMV Ω leader sequence into the Level 1 binary vector pICH47751 (35S:Cai1). To try to avoid the developmental phenotype of *SlANT1* overexpression (Čermák et al., 2015; Grandgeorge, 2021), I cloned an analogous construct with the promoter of the *DWARF4* gene from *Arabidopsis*, which has been reported to not affect growth when driving *SlANT1* expression while still inducing colouration (DWF4:Cai1) (Jin et al., 2012). As positive controls, I created constructs with either promoter expressing *SlANT1*'s ORF and terminator (35S:ANT1 and DWF4:ANT1). This, and all other cloning described in this thesis, was done using the Golden Gate system unless otherwise stated (Engler & Marillonnet, 2014) (for more details on the cloning procedures used, see Section 2.2.5). The resulting plasmids were then expressed *in planta* via *Agrobacterium*-mediated gene transfer, which was also used throughout the thesis.

First, a transient leaf expression approach was pursued. This was complicated by the fact that obtaining visible anthocyanin

accumulation in leaves is difficult when overexpressing MYB TFs alone (X. Liu et al., 2015). To solve this, I co-expressed the *Cai1* and *ANT1* constructs with the *Delila* bHLH gene from *Antirrhinum majus*, which Albert et al. (2020) showed induces strong anthocyanin accumulation in *N. benthamiana* leaves in combination with its natural MYB *Ros1*. Syringe-infiltration into *N. tabacum* leaves gave inconsistent results (data not shown) so I repeated the experiment using vacuum-infiltration into *N. tabacum* seedling cotyledons (Fig. 3.7). Both *Cai1* and *ANT1* induced anthocyanin overaccumulation when expressed by the 35S promoter, the latter having a stronger effect than the former. The co-expression of *Delila* enhanced the pigmentation, but it was also observed without it. This suggested that the *Cai1* locus is a suitable target for the anthocyanin KI reporter system and indicated the resulting KI phenotype might be milder than that observed in tomato. In contrast, neither of the DWF4-driven constructs showed any anthocyanin accumulation, regardless of *Delila* co-expression. It was unclear if this was because the promoter is unsuitable for the task, or because it is simply not expressed well in tobacco cotyledons.

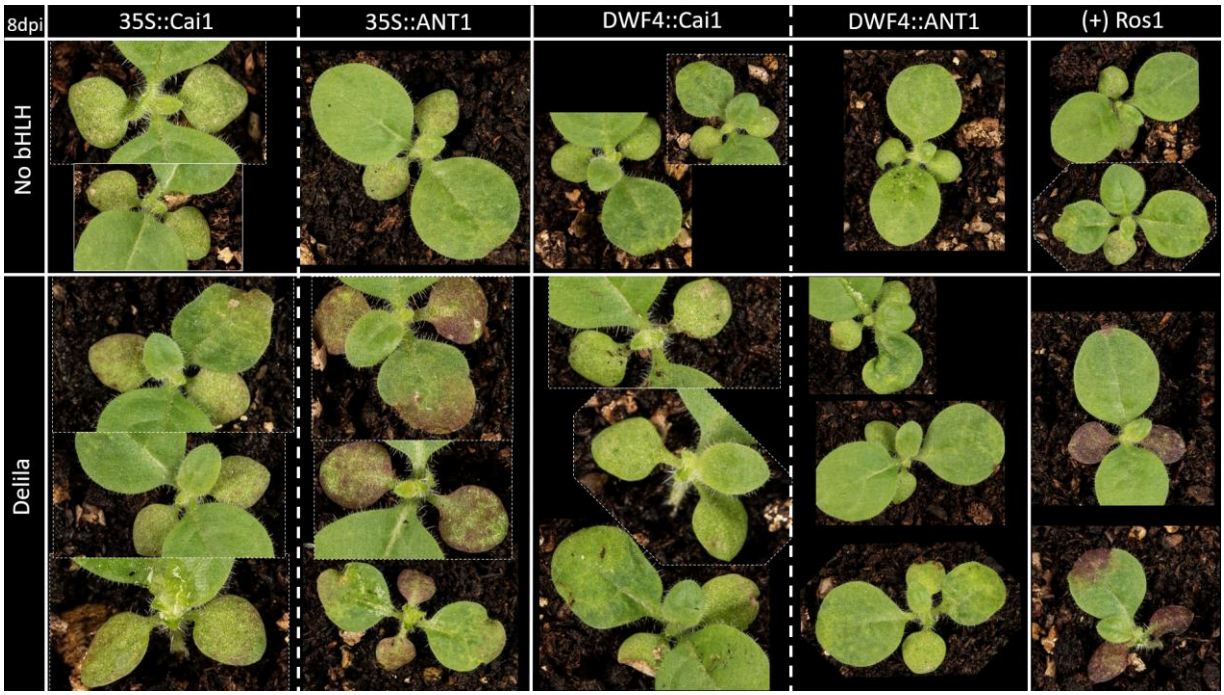


Figure 3.7. Transient expression in tobacco.
Several MYB transcription factor - promoter combinations (columns) were infiltrated into tobacco cotyledons with or without the bHLH gene *Delila* (rows). Both *ANT1* and *Cai1* induced visible anthocyanin

accumulation when expressed by 35S but *Cai1*'s phenotype was mild. The DWF4 promoter proved unsuitable. The *Ros1* control infiltrations produced robust pigmentation, as expected.

Next, I investigated the constructs' performance in a stable transformation in the potato line B26. To do this, I cloned the four Level 1 constructs (35S:Cai1, DWF4:Cai1, 35S:ANT1 and DWF4:ANT1) into multi-gene Level 2 constructs, featuring the *NptII* Kanamycin-resistance gene (YDJJ015-YDJJ018) (Beck et al., 1982). Stable transformation into B26 proved very efficient, as practically all explants produced vigorous callus. Calli transformed with the *Cai1* or *ANT1* gene driven by the 35SΩ promoter developed pronounced pigmentation. In contrast, no pigmentation was observed when the *DWF4* promoter was used (Fig. 3.8). This ultimately confirmed that the *Cai1* locus is an appropriate target for the anthocyanin KI reporter system and that the 35SΩ promoter is the better promoter choice. Here again, the overexpression phenotype seemed less severe than that in tomato, which made stable line recovery easier.

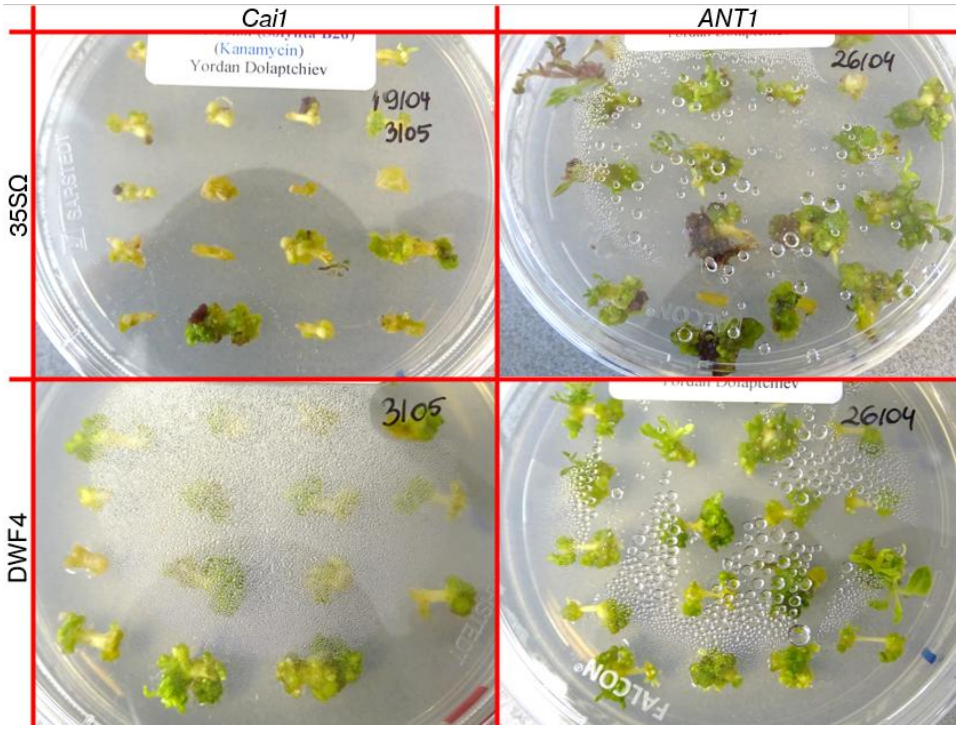


Figure 3.8. Stable transformation of *Cai1* and *Ant1*.
Both genes produced anthocyanin accumulation in callus when cloned next to the 35S promoter (top row). Analogous constructs with the DWF4 promoter did not produce pigmentation (bottom row). Thus,

both genes can result in anthocyanin bioaccumulation upon overexpression by 35S but not DWF4.

3.2.2 Gene Knock-Ins in potato are possible

KI construct cloning and transformation

The next objective was to replicate, in potatoes, the successful generation and recovery of KIs. To achieve this, I based the first KI construct on the best practices established by L. Grandgeorge (2021), which were themselves grounded in an extensive body of prior research. Below, I provide a brief overview of the plasmid and the rationale behind the design.

The first construct consisted of a programmable nuclease, a crRNA expression cassette, and a KI cassette (Fig. 3.9 pKI1). The programmable nuclease was the temperature-tolerant LbCas12a (ttLbCas12a) due to its superior performance at the cool temperatures involved in plant tissue culture (Schindele & Puchta, 2020). The crRNA used (crRNA1) was flanked by ribozymes to ensure "clean" expression of the RNA element without any additional nucleotides on either side (Gao & Zhao, 2014). The ttLbCas12a and crRNA were cloned in a head-to-head orientation, as that has been shown to result in higher editing rates (Castel et al., 2019). The KI cassette was flanked by the homology arms designated earlier, which were in turn flanked by copies of the crRNA1 target site (TS) as this has been shown to boost KI rates (Fauser et al., 2012; L. Grandgeorge, 2021; J. P. Zhang et al., 2017). Between the homology arms sat a *NptII* selectable marker and the 35S promoter + TMV Ω leader sequence which faced towards the partial *Cai1* ORF in the HA-R (Čermák et al., 2015; Grandgeorge, 2021).

Alongside pKI1, I created two additional plasmids to test potential ways to improve KI rates (Fig 3.9). Plasmid pPre-crRNA differed from pKI1 in that its crRNA was flanked not by ribozymes, but by Cas12a CRISPR repeats, thus resembling an unprocessed CRISPR array transcript. Xu et al. (2017) reported this alternative scaffold achieves higher editing frequencies with FnCas12a, and I attempted to employ it with LbCas12a. Plasmid pRAD54 differed from pKI1 by including the *RAD54* gene from *Saccharomyces cerevisiae* (Game & Mortimer, 1974). *Sc*RAD54 is a

protein involved in strand invasion, one of the steps of HDR, and ectopic expression has been reported to boost HDR rates in *Arabidopsis* (Shaked et al., 2005).

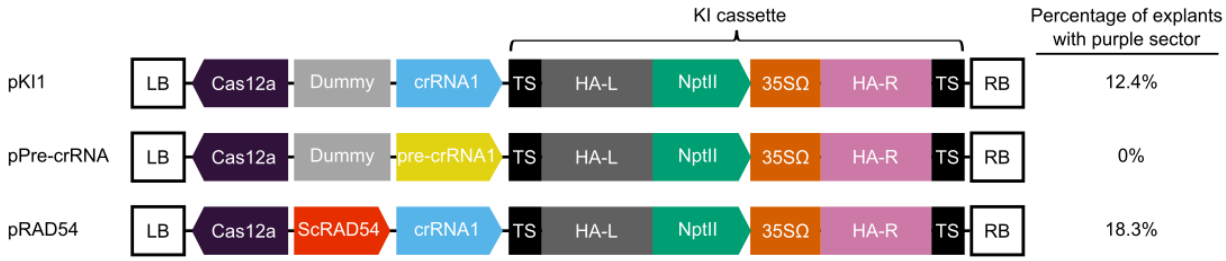


Figure 3.9. Initial set of KI constructs.
A comparison of the construct architectures used in the first KI experiments. pKI1 was closely modelled on the best practices established by L. Grandgeorge (2021) and yielded purple sectors with a moderate efficiency. pPre-crRNA employed an alternative crRNA architecture (R. Xu et al., 2017), which ultimately proved non-functional and resulted in no purple sectors. pRAD54 included the *ScRAD54* gene, which appeared to moderately boost the rate of purple sector formation (Shaked et al., 2005). LB – T-DNA Left Border, RB – T-DNA Right Border, Cas12a – ttLbCas12a, TS – crRNA1 Target Site. Purple sectors were harvested and scored at 56 days post co-cultivation. A more detailed diagram can be found in Section 2.2.5.

The plasmids were transformed into B26 and, starting from ca. 10 days post co-cultivation (dpcc), small purple sectors could be seen on some calli (Fig. 3.10 A). These grew but slower than the surrounding green tissue (for microscopic KI sector images see SFig. 4). At 56 dpcc the calli produced shoots but the sectors were largely overgrown, and no purple shoots were observed. Because of this, I excised the purple sectors and subcultured them separately. Ultimately, the pKI1 transformation produced viable sectors on 12.4% of explants (22/177), the pRAD54 transformation produced viable sectors on 18.3% of explants (30/164), while the pPre-crRNA transformation yielded none (Fig. 3.9). These encouraging results suggested that KIs in potato are achievable at moderate rates, and that *ScRAD54* might lead to a slight increase in frequency. The pre-crRNAs proved ineffective but provided a negative control for the experiment. Following sector subculture,

purple shoots were obtained, although green growth had to be regularly cut back (Fig. 3.10 B).

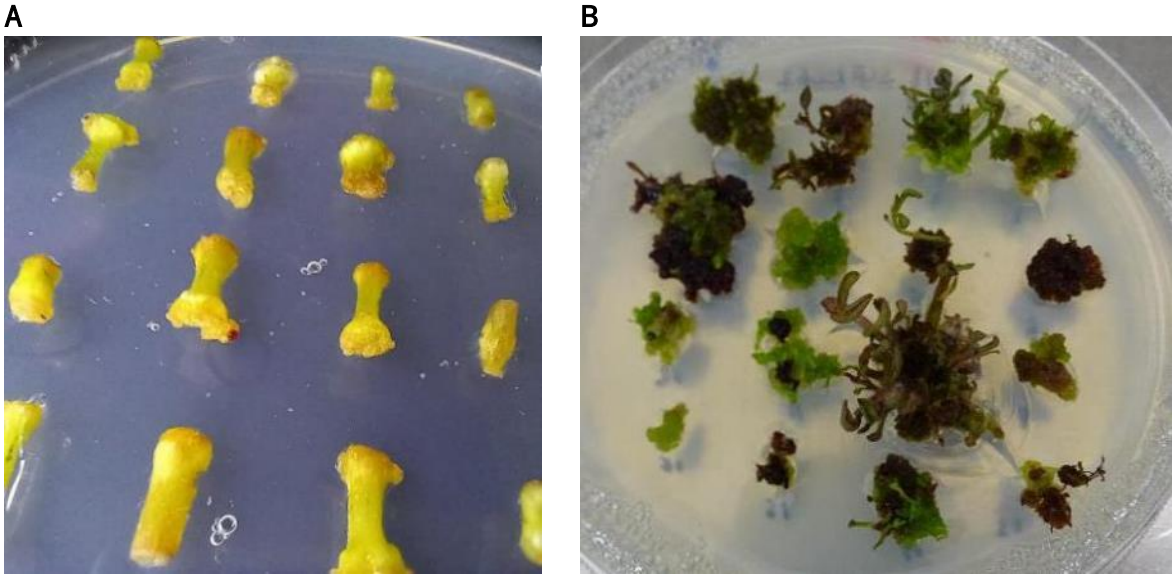


Figure 3.10. Purple KI sectors.
A B26 internode explants on callogenic media developing purple KI sectors ca. 28 dpcc. **B** Subcultured KI sectors on shooting media. Despite regular pruning, green callus and shoots repeatedly appear and threaten to overgrow the KI sector unless cut back.

Characterisation of T_0 KI lines

I recovered three independent purple putative KI lines from pKI1, and another three from pRAD54. These were moved to soil, where they grew well, set tubers, and some set berries. The purple pigmentation persisted in the plants' stems, flowers, berries, tuber skin, and in the embryos of some recovered seeds (Fig. 3.11 A). Notably, some T_0 lines exhibited mosaicism or variegated pigmentation, likely due to the KI selection being purely visual (SFig. 5). This was only observed in T_0s, and the plants regrown from tubers were phenotypically homogenous.

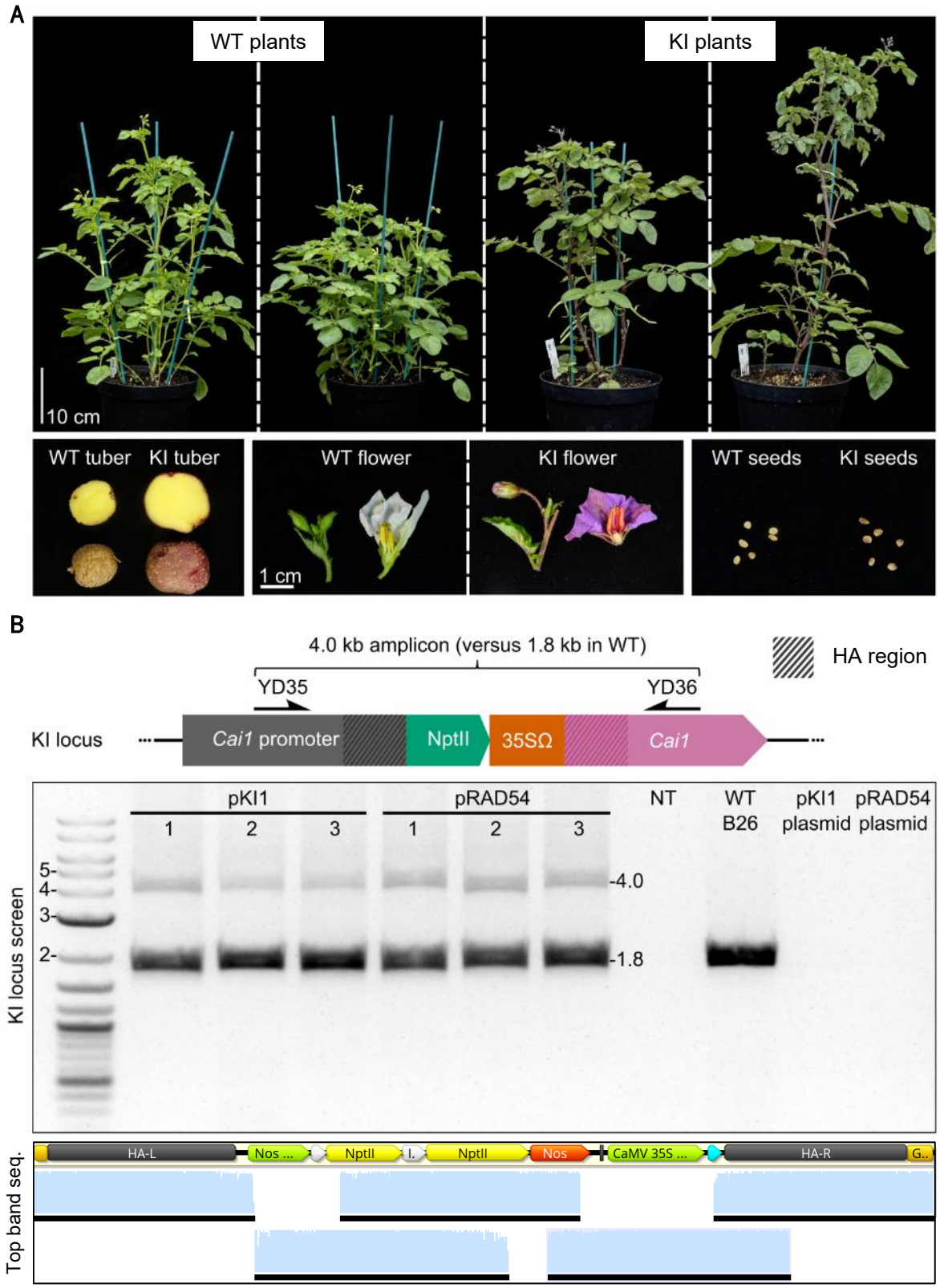


Figure 3.11. T_0 KI line phenotype and genotype.

A T_0 KI lines had a distinct phenotype compared to WT B26. Top row: WT B26 plants (left two panels) and T_0 KI plants (right two panels). Scale bar is 10 cm. Bottom row: comparison of WT and KI tubers (left), flowers (centre), and seeds (right). Scale bar is 1 cm. **B** The genotyping results of the first six T_0 KI lines. Top: a diagram of the KI locus. Amplifying it with primers YD35/YD36, which bind outside of the regions of homology, should yield a 4.0 kb product, while a WT locus should yield a 1.8 kb product. Middle: the PCR screen described above results in each T_0 line producing a 4.0 kb and a 1.8 kb band, indicating they are all heterozygous for an insertion at the *Cai1* locus. The reactions supplied with no template (NT) or one of the plasmids yielded nothing, as expected. Ladder is 1 kb plus (NEB). Bottom: Sanger sequencing reads from line pKI1-3 mapping perfectly to the expected KI locus. The yellow annotations flanking the homology arms are genomic DNA sequence from outside the KI cassette.

I genotyped the regrown T_0 lines by PCR using primers designed to anneal outside the homology arms flanking the targeted locus (Fig. 3.11 B top). While the wild-type (WT) locus would generate a 1.8 kb amplicon, integration of the KI cassette was expected to result in a 4.0 kb amplicon. In contrast, a T-DNA locus would not yield any PCR product. All six T_0 lines screened produced both a 4.0 kb and a 1.8 kb amplicon, indicating that they are heterozygous for an insertion at the *Cai1* locus (Fig. 3.11 B middle). The top band was consistently fainter than the bottom band, likely due to the shorter WT amplicon outcompeting the larger KI amplicon in the PCR. The WT B26 control produced only a 1.8 kb amplicon, confirming that the primers specifically amplify only that region in the B26 genetic background.

I excised, extracted, cloned and sequenced the obtained amplicons. The top bands from four lines were readily cloned (pKI1-1, pKI1-3, pRAD54-2, and pRAD54-3). Three of these (pKI1-1, pKI1-3, and pRAD54-3) had sequences that mapped perfectly to what a genuine KI should look like (Fig. 3.11 B bottom; SFig. 6). These sequences seamlessly transitioned from genomic DNA through HA-L, across the KI cassette, and back out through HA-R into genomic DNA. In contrast, pRAD54-2 sequencing revealed a 201 bp deletion in the

first exon of the *NptII* gene, also reflected in the slightly smaller size of its top band compared to others (Fig. 3.11 B, middle). Microhomologies at the deletion ends suggest it resulted from a spontaneous rearrangement (SFig. 7). Despite this anomaly, all sequenced top bands confirmed the presence of genuine KIs in the purple T0 lines, demonstrating that targeted gene insertions are both achievable and readily obtainable in diploid potato. Additionally, the purple phenotype proved to be a credible indicator of KI status, with no false positives identified.

I also cloned and sequenced three of the bottom bands. These revealed indels at the target site which disrupted the crRNA1 protospacer, thus “spoiling” the target site and prohibiting further cleavage (SFig. 8). This is likely a reflection of the competition between HDR and alternative DSB repair pathways that is thought to occur in somatic cells during KIs. Thus, this experiment indicates that KI lines can generally be expected to be heterozygous for a KI, as well as an indel at the targeted locus.

I further screened the T_0 lines for the presence of the T-DNA using primers that bind within the ttLbCas12a expression cassette. I observed the expected amplicon in all T_0 lines, except pKI1-2 (Fig. 3.12). The recovery of a T-DNA-free line was likely enabled by the inclusion of the *NptII* selection marker within the KI cassette. Because of this, following a successful KI cassette integration, the T-DNA is no longer selected for. Crucially, this indicated that KIs can occur independent of T-DNA integration with some frequency, as had been reported by Huang et al. (2021). The proportion of T-DNA-free lines was also similar (1:5 versus 2:8), although the exact frequency of such transient KIs is unclear, as the T-DNAs detected could have integrated independently of the KI process. Still, the fact a T-DNA-free line was identified from among just six T_0 lines suggests that it might be a moderately frequent phenomenon. Achieving such T-DNA-free KIs is an exciting prospect for crops where genetic segregation to discard such undesirable insertions is not possible. Lastly, the six T_0 lines were also PCR-screened for integrations of the plasmid backbone but none were detected (SFig. 9).

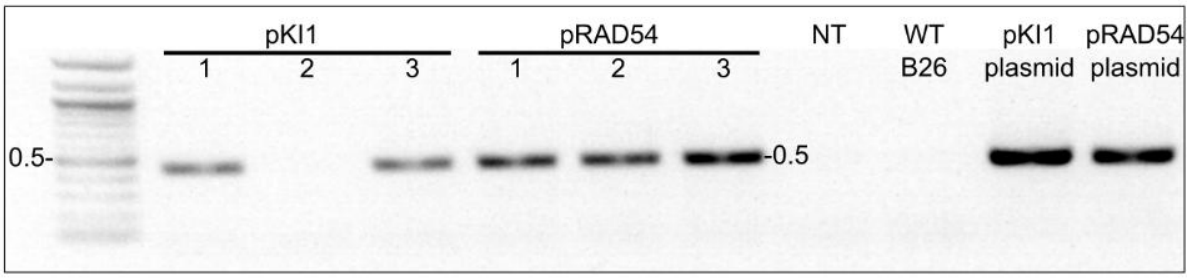


Figure 3.12. T_0 T-DNA screen.
The six T_0 KI lines and relevant controls were PCR-screened for the presence of ttLbCas12a, and thus the T-DNA, using primers AN73/AN79. The expected amplicon is 451 bp long and it was observed in all T_0 lines, except pKI1-2. The NT and WT controls yielded nothing, while the plasmid controls produced the expected amplicon. Ladder is 100 bp (NEB).

KIs are inherited in a Mendelian fashion

I investigated the inheritance of the KI locus in the pKI1-2 T-DNA-free line, as the absence of ttLbCas12a meant that no novel KIs could occur and skew the KI locus' segregation. I outcrossed the pKI1-2 line to B100, rather than self it, to avoid issues with the B26 genetic background's low fertility (further discussed in Section 3.2.4).

Attempts to track the segregation of the KI in the F_1 generation met with difficulty. I obtained two berries, yielding 130 and 124 F_1 seeds (SP10724). The seeds segregated for purple: WT seed colour in odd 30:100 and 26:98 ratios, significantly different from the 1:1 ratio predicted by Mendelian genetics. Later, some seemingly WT seeds sprouted into purple plants, indicating that seed colour is a poor proxy for the seedling's phenotype. Germination rates were low, something typical for seeds from B26, and some of the plants that did germinate were lost to a pest infestation in the glasshouse, so no segregation ratio could be established. Regardless, several F_1 plants carrying a KI and WT phenotype were obtained. The purple F_1s were homogenously purple but they did vary in the intensity of pigmentation (Fig. 3.13). A likely explanation for this is that the F_1 population is segregating for other factors which affect anthocyanin biosynthesis. Beyond this, all F_1s varied in morphology and flowering rate, indicating that there is substantial background

segregation occurring due to B26 and B100's high degree of heterozygosity. Ultimately, the KI allele proved heritable, but it could not be established if it segregates in a Mendelian fashion in the F_1 generation.

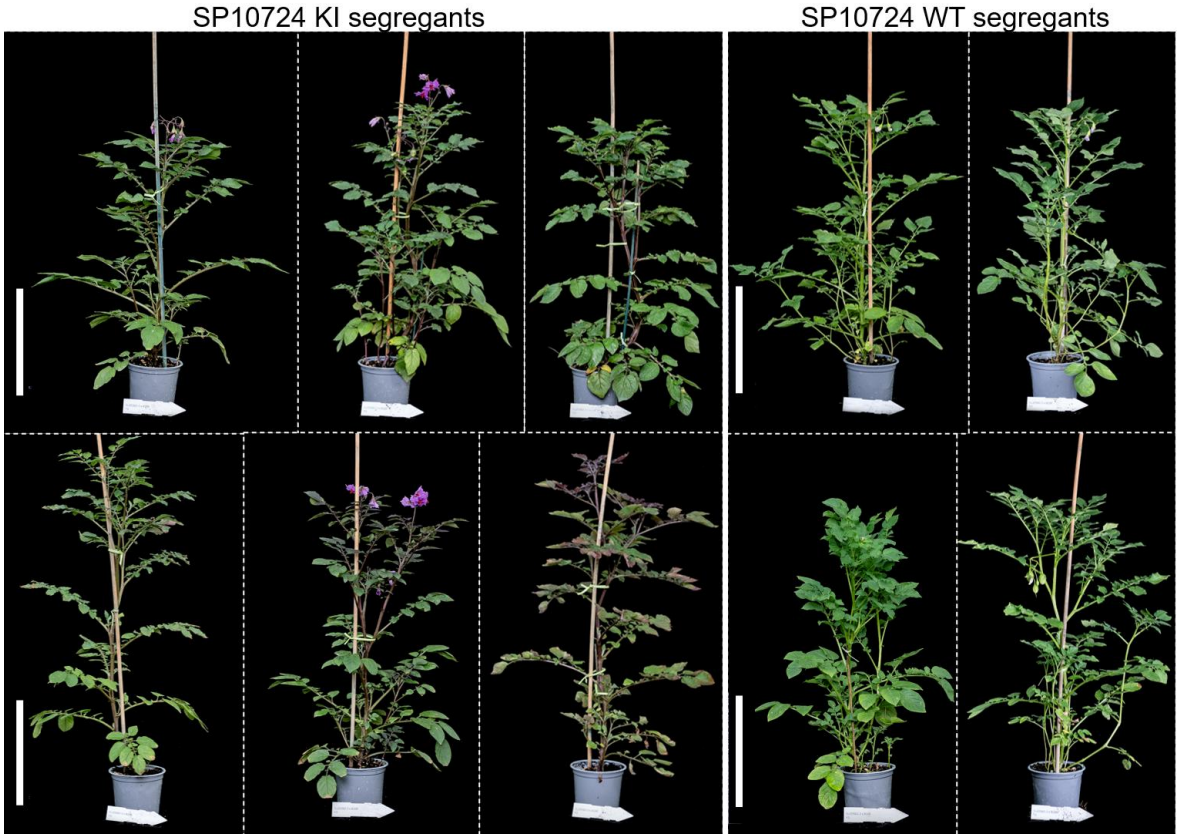
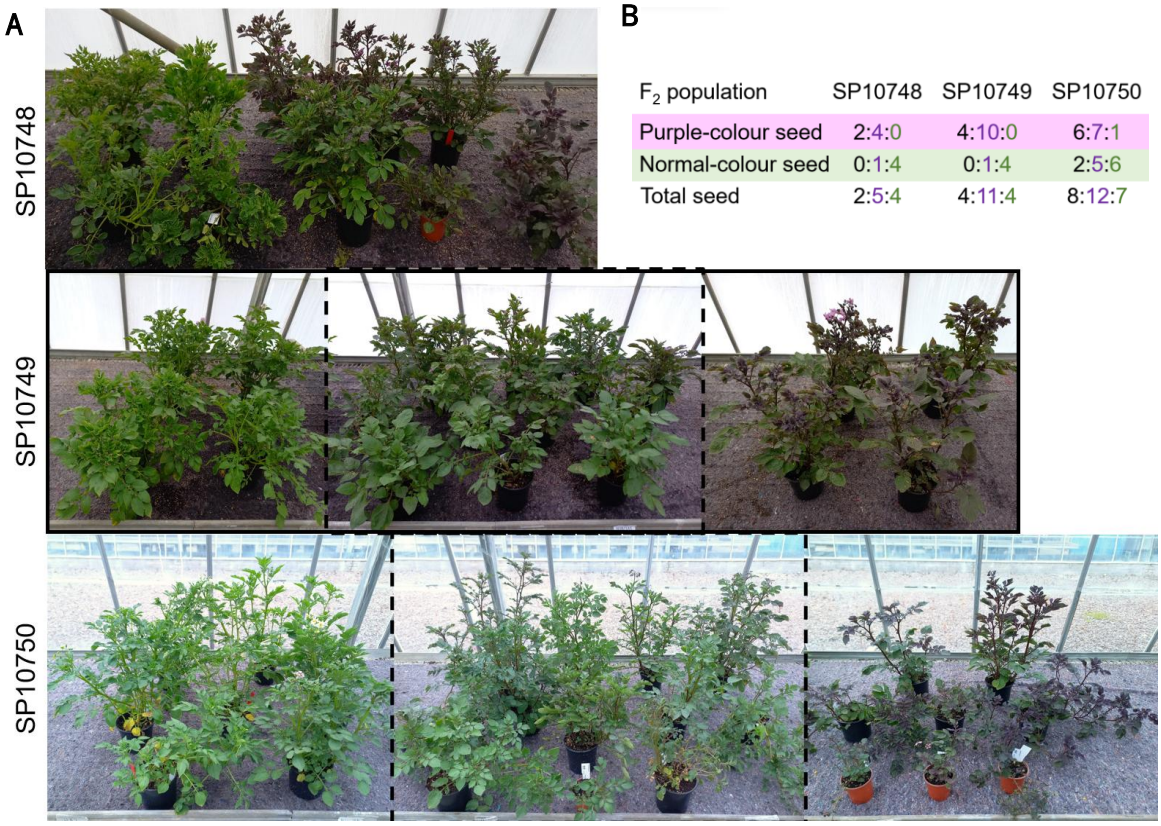


Figure 3.13. F_1 KI segregation.
The pKI1-2 × B100 (SP10724) F_1 population segregating for the anthocyanin KI phenotype. The KI plants' pigmentation was homogenous but differed in its intensity between segregants. Beyond that, the plants differed in morphology and flowering rate. Scale bar is 20 cm. Dotted lines indicate where images have been digitally merged to facilitate comparison.

To further investigate KI segregation, I proceeded to the F_2 generation. I collected selfed berries from three vigorous purple F_1 plants (SP10748-10750). As these plants were outcross-derived F_1s, they should be heterozygous for the KI, and Mendelian genetics predicts that the progeny from a selfed berry should segregate in a 1:2:1 ratio. Purple and normal-coloured F_2 seeds were sown separately, all germinating and growing well. The F_2 progeny's phenotype ranged from deeply pigmented to WT-like (Fig. 3.14 A). I grouped these plants into three phenotypic classes:

green, light purple, and deep purple, although there was some variability between plants within the same class. The observed segregation ratio in SP10748 (2:5:4), SP10749 (4:11:4), and SP10750 (8:12:7) did not significantly differ from the predicted 1:2:1 (Fisher's Exact Test, P-value = 1.0 for all) (Fig. 3.14 B). Thus, based on the phenotypic data, the KI locus seems to be inherited in a Mendelian fashion. Also, purple-coloured seed was highly enriched for plants with a light purple and deep purple phenotype, while WT-coloured seed was enriched for plants with the green phenotype.



F_2 population	SP10748	SP10749	SP10750
Purple-colour seed	2:4:0	4:10:0	6:7:1
Normal-colour seed	0:1:4	0:1:4	2:5:6
Total seed	2:5:4	4:11:4	8:12:7

Figure 3.14. KI F_2 phenotypic segregation.
A The F_2 progeny from three selfed F_1 KI plants (SP10748-10750) segregated for anthocyanin pigmentation and were manually assigned into three phenotypic classes: green (left), light purple (middle), and deep purple (right). Some phenotypic variability was still observed within classes. **B** Before being sown, the F_2 seeds were separated by colour. Purple seed was enriched for plants with a purple phenotype and *vice versa*. The segregation ratios of total seed were not significantly different from the 1:2:1 ratio predicted by Mendelian genetics.

I hypothesised that the deep purple, light purple, and green phenotypic classes consisted of homozygous KIs, heterozygous KIs, and null segregants, respectively. To test this, I genotyped nine plants from SP10749 and SP10750 each, three per phenotypic class, by PCR amplifying across the KI locus as previously described (Fig. 3.11 B top). Here homozygous KIs, null segregants, and heterozygous KIs are respectively expected to generate high bands (4.0 kb), low bands (1.26 or 1.8 kb, depending on the *Cai1* locus inherited from B100), or both. The band patterns produced by most plants (14/18) matched the prediction (Fig. 3.15). Where the pattern differed from the prediction, it fit with the adjacent phenotypic group. This is likely caused by background segregation for factors other than the *Cai1* MYB which affect anthocyanin production such as bHLH genes (Y. Liu et al., 2016). In potato, there are numerous genes involved, and none of these have been characterised in the heterozygous lines these populations stem from. Indeed, the variability in pigmentation intensity between individual lines within the same assigned phenotypic class, as noted earlier, is indicative of such segregation taking place. This would explain, for example, why line Green 1 from SP10749 was heterozygous for the KI yet was assigned to the Green phenotypic group. Still, the genotyping data largely agrees with the phenotypic data and is consistent with the KI locus being inherited in a Mendelian fashion. Together, these two datasets demonstrate that it is possible to breed for homozygous KIs, and that having two copies of the KI allele results in a more intense phenotype (Fig. 3.14 A).

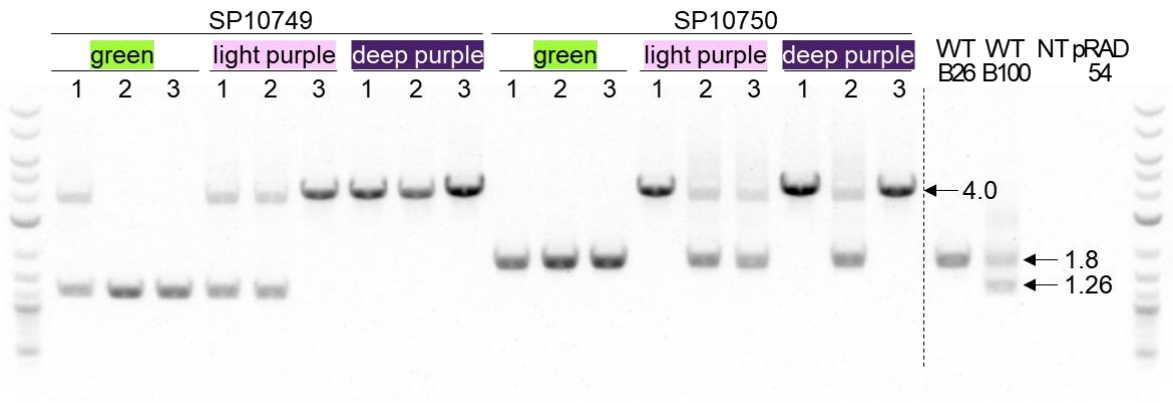


Figure 3.15. KI F_2 genotypic segregation.

Plants from two F_2 populations (SP10749 & SP10750) were genotyped at the KI locus by PCR amplification with primers YD35/YD36, which bind outside of the regions of homology. Plants from the three phenotypic classes (green, light purple, and deep purple) are predicted to be homozygous KIs, heterozygous KIs, and null segregants, respectively. Homozygotes should yield only a high band (4.0 kb), null segregants should produce only a low band (1.26 or 1.8 kb, depending on the *Cai1* locus inherited from B100), while heterozygotes should yield both bands. The genotyping largely agreed with the assigned phenotypic classes (14/18). The WT controls produced either one 1.8 kb band (B26) or two 1.8 kb and 1.26 kb bands (B100), as previously established in Section 3.2.1. The reactions supplied with no template (NT) or the pRAD54 plasmid yielded nothing, as expected. Ladder is 1 kb plus (NEB).

T-DNAs can be segregated away

Having shown that KIs are heritable, I attempted to breed a T-DNA-free line from a T-DNA-positive T_0 plant. Simultaneously, I wanted to establish whether the T-DNA and KI loci are linked, as reported by Lawrenson et al. (2021). To do this, I outcrossed the pRAD54-3 line to B100, yielding a berry with 23 purple seeds, and 48 WT-coloured seeds. I sowed 32 seeds, which yielded 5 plants with a WT phenotype, and 8 plants with a KI phenotype. I then genotyped these F_1s with primers which amplify a part of the RAD54 coding sequence located on the T-DNA. No amplicons were observed with lines 5 and 12, the latter of which had a KI phenotype, while the remaining 11 lines produced the expected band (Fig. 3.16). Thus, in line 12 the T-DNA had successfully been segregated away from the KI. The otherwise widespread T-DNA signal can be explained by pRAD54-3 harbouring three T-DNA insertions, which segregate independently of the KI locus, and of each other. This hypothesis predicts a 1:7 segregation of T-DNA negative: positive lines, which is not significantly different from the observed 2:11 ratio (Fisher’s Exact Test, P-value = 1.0), although the sample size is very small. Ultimately, it seems KI loci are not necessarily linked to T-DNA loci as previously reported, and in my experiments the T-DNAs were readily segregated away.

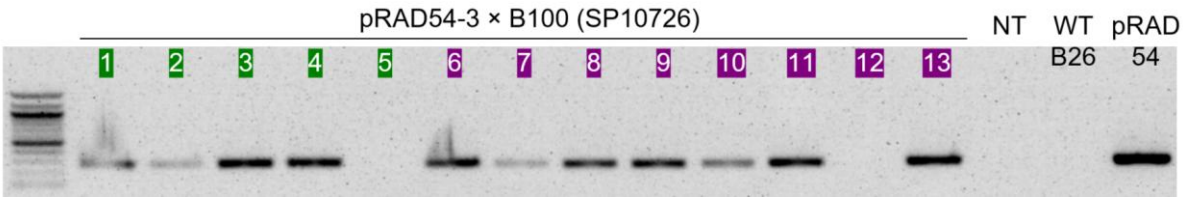


Figure 3.16. T-DNAs and KIs are not genetically linked.
T-DNA screen of an F_1 population segregating for a KI and the T-DNA. The PCR was conducted with primers YD131/YD246, which amplify a section of the RAD54 coding sequence 261 bp long. Five of the screened plants had a WT phenotype (green numbers 1-5), while eight had a KI phenotype (purple numbers 6-13). All plants were T-DNA-positive, except for lines 5 and 12, the latter of which had a KI phenotype. Thus, in line 12 the T-DNA had successfully been segregated away from the KI. The presence of a RAD54 coding sequence in most F_1 segregants indicates the T_0 parental like pRAD54-3 carried multiple T-DNA insertions. The pRAD54 plasmid produced the predicted band. The reactions supplied with no template (NT) or WT B26 gDNA yielded nothing, as expected. Ladder is 100 bp (NEB).

3.2.3 Exploring alternative KI approaches

In the preceding section I established a method for HDR-mediated gene KIs in diploid potatoes, achieving moderate efficiency and utilizing a convenient anthocyanin-based reporter system (Fig. 3.17). Here long homology arms were used to guide the insertion of a KI cassette containing a selection marker and the 35SΩ promoter, building on previous successes reported by Čermák et al. (2015) and L. Grandgeorge (2021). In this section, I explore alternative approaches to make this technique more practical and versatile.

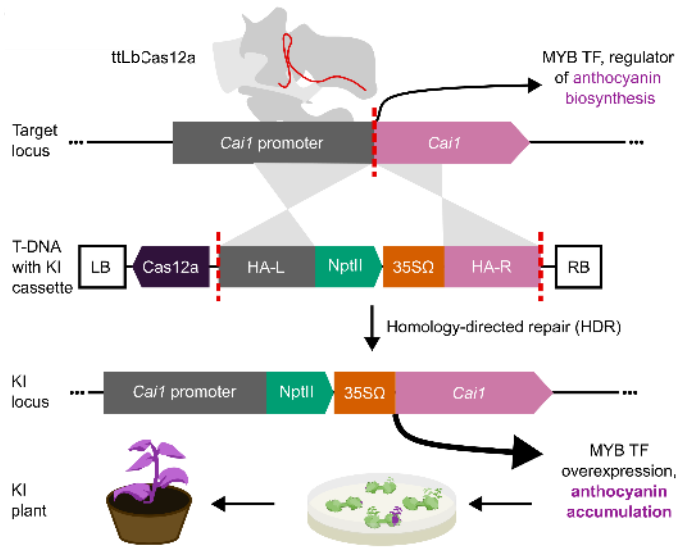


Figure 3.17. KI method diagram.
The targeted insertion locus and the repair template are cleaved by ttLbCas12a (red dotted lines). HDR between the locus and the template's HA-L & HA-R (grey cones) results in targeted cassette integration, MYB over-expression, and purple pigmentation.

One modification was motivated by the fact that T-DNAs are present in most recovered KI lines and that these can later be segregated away in diploid potatoes. Because of this, the selection marker needn't be included in the KI cassette; instead, it can be placed elsewhere on the T-DNA so it can later be removed by breeding. To this end, I created a version of the pKI1 plasmid, where the *NptII* gene was cloned at the far-left of the T-DNA, while a short dummy sequence was left to occupy its position in the KI cassette (Fig. 3.18 pExtNpt). This construct yielded KI sectors on 27.5% of kanamycin-resistant explants, indicating the position of the selectable marker is not crucial to the KI process. The increase in KI frequency was likely due to the decrease in KI cassette size from 2.2 kb in pKI1, to ca. 550 bp in pExtNpt. Going forward, KI plasmids would feature a selectable marker outside of the KI cassette. As Kanamycin would henceforth select against desirable T-DNA-free KI lines, its use in rooting media was discontinued.

I also tried to achieve KIs using alternative repair pathways, as HDR rarely occurs in somatic plant cells. DSBs are generally repaired by NHEJ and it has previously been employed in gene targeting in plants (J. Li et al., 2016, 2018; S. Li et al., 2018; Weinthal et al., 2013). MMEJ has not been utilised in this way, but it is known to be active in plants and can use short microhomologies to facilitate DNA integration (van Kregten et al., 2016). Furthermore, preliminary work on MMEJ-mediated KIs in

tomato suggested that this pathway might be a suitable alternative to HDR (L. Grandgeorge, unpublished data). To try and employ these DSB repair mechanisms, I created versions of pKI1 with the KI cassette flanked by either no homology arms, or 25 bp-long microhomology arms (Fig. 3.18 pNHEJ & pMMEJ). Surprisingly, pMMEJ performed worse, with the explants yielding just one KI sector, while pNHEJ generated KI sectors on 5.9% of explants. Ultimately, both NHEJ and MMEJ seemed to be less efficient pathways for achieving KIs and were not pursued further.

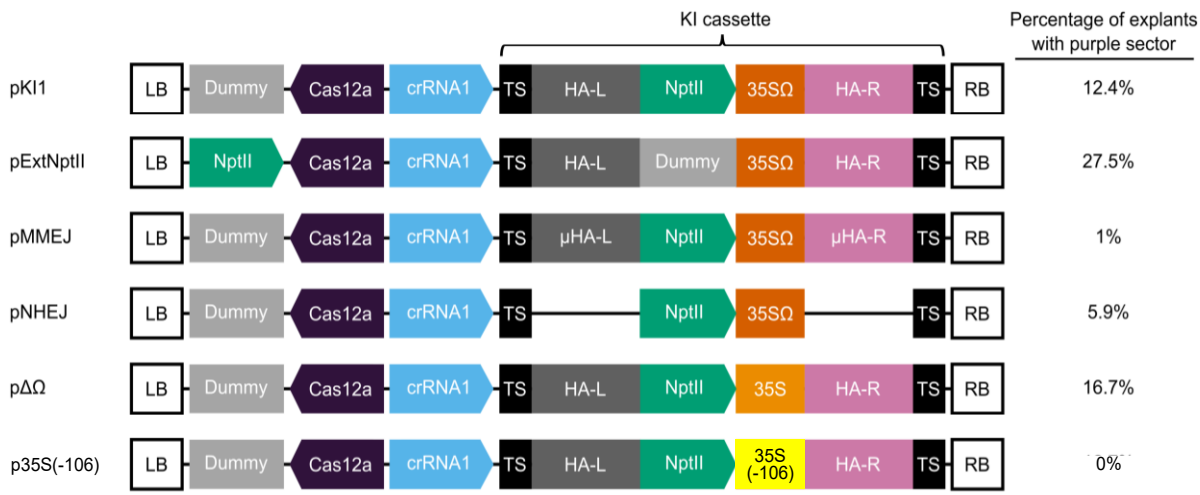


Figure 3.18. Alternative approaches to KIs.
Several plasmids with variations on the design of pKI1 (also illustrated for reference). LB – T-DNA Left Border, RB – T-DNA Right Border, Cas12a – ttLbCas12a, TS – crRNA1 Target Site, μHA - microhomology arm, 35S(-106) – 35S promoter truncated at the -106 position relative to the transcription initiation site. Purple sectors were harvested and scored at 56 days post co-cultivation.

Lastly, I tried to KI a milder promoter to minimise the developmental phenotype while retaining the clear pigmentation of KI sectors. To do this, I cloned a version of pKI1 without the TMV Ω leader sequence downstream of the 35S promoter (Fig. 3.18 pΔΩ), and another version with the 35S promoter truncated after position -106 (Fig. 3.18 p35S(-106)). pΔΩ transformation resulted in KI sectors on 16.7% of explants. The KI sectors and plants were not noticeably less pigmented that those obtained from pKI1, indicating that the TMV Ω leader sequence is not necessary for the anthocyanin reporter system. The callus and plant growth also

seemed marginally better, although there was no straightforward way to quantify this. In contrast, the p35S(-106) transformation did not produce any KI sectors. This is likely due to the truncated 35S promoter being nonfunctional, although previous research has claimed this fragment to be transcriptionally active in tobacco cells (Odell et al., 1985). Based on these results, I used the full 35S promoter without the TMV Ω leader sequence in future plasmids.

3.2.4 Improving KI efficiency

Having established a KI strategy designed to produce more vigorous events and facilitate the removal of the selectable marker through breeding, I shifted my focus to enhancing the rate of KIs. Achieving higher KI efficiency is not only valuable for increasing the number of successful events per transformation but also for expanding the range of potential applications. If efficiency can be sufficiently elevated, it may become feasible to target genes without relying on an anthocyanin reporter, enabling a genotyping-based screening approach instead.

Small molecule effectors

As mentioned in the introduction, small molecule effectors have previously been reported to increase the frequency of KIs by modulating the DSB repair pathways. Here, I attempted to do the same. 3-methoxybenzamide (3-MB) is an inhibitor of poly(ADP)ribose polymerases, which play a complex role in stress and DNA repair, and has been reported to increase potato transformation efficiency (0.2 mM; Chetty et al., 2020) and the frequency of homologous recombination in tobacco seedlings (2.5 mM; Puchta et al., 1995). Another chemical, SCR7, is a proprietary inhibitor of DNA ligase IV, which plays a role in NHEJ, and has been used to increase gene editing rates in mammalian cells (1 μM; Maruyama et al., 2015) and *Magnaporthe oryzae* (Andy Foster, private correspondence). To assess the chemicals' effect on KIs, transformations of pKI1 had either one added to the media and were then compared to a mock treatment (Table 3.1). The mock treatment produced sectors on 14.3% of explants, which was consistent with pKI1's previous performance of 12.4%. When treated with 3-MB (0.2 mM), explants grew normally and produced purple sectors slightly more often

(15.7%). In contrast, treatment with SCR7 (1 μM) led to poor transformation, slow callus development, and a reduced frequency of purple sectors at 9.2%. These detrimental effects made sense in lieu of a recent paper by Kralemann *et al.* (2022), which implicates cNHEJ in T-DNA integration. As 3-MB was well-tolerated, the effects of higher concentrations (1 mM and 2.5 mM) were assessed. At 1.0 mM and 2.5 mM, 3-MB resulted in KI sectors on 11.4% and 0.7% of explants, respectively. The two concentrations had strikingly different effects on the efficiency of the transformation: at 1.0 mM the callus grew well and had an elevated propensity for shoot formation, while at 2.5 mM callus growth was strongly inhibited (Fig. 3.19). Notably, the experiment with 3-MB at 1.0 mM provided the only example where an entirely purple shoot was recovered prior to selective KI sector subculture. Ultimately, no small molecule effector was identified which made a substantial positive difference in KI frequency.

Treatment	Concentration (mM)	Sectors recovered	Transformed explants	Explants with sectors (%)
Mock	-	12	84	14.3%
3-MB	0.2	17	108	15.7%
	1.0	17	149	11.4%
	2.5	1	145	0.7%
SCR7	0.001	14	152	9.2%

Table 3.1. KI rates when treated with 3-MB & SCR7.

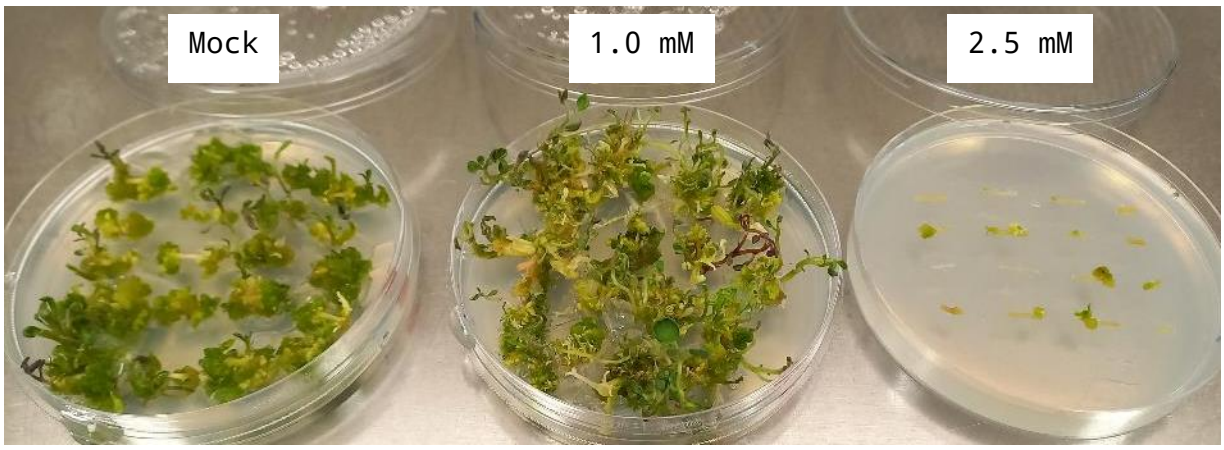


Figure 3.19. Effects of 3-MB on shoot growth.
B26 explants transformed with pKI1 on media supplemented with 3-MB at 0 mM (mock), 1.0 mM, or 2.5 mM. At 1.0 mM, callus growth

proceeded well while the rate of shoot formation was notably increased. Notable was the entirely purple shoot produced prior to selective KI sector subculture. At 2.5 mM, 3-MB effectively arrested callus formation. Picture was taken 56 days post co-cultivation.

Breeding a potato for KIs

An indirect way of improving KI efficiency would be to breed a better diploid potato line to work with. Thus far, KIs were done in line B26 due to its superior transformability. However, breeding the resulting T_0 plants was difficult because the line suffered from a flower bud abortion phenotype (Fig. 3.20 A). Additionally, B26 has poor pollen germination (Fig. 3.20 B), meaning that it cannot be used as a male parent or produce selfed berries reliably. Another diploid potato line called B100 flowers well, produces viable pollen (Fig 3.20 C), and readily selfs, but transforms relatively poorly. Efficient GE work requires a line which performs well in both aspects.

A

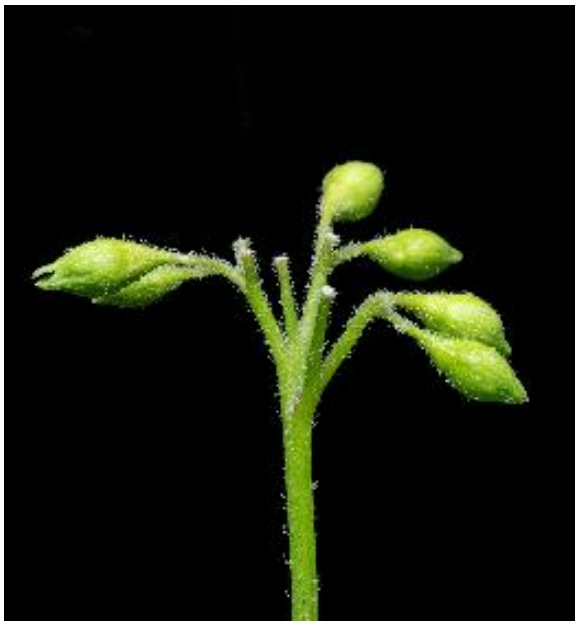
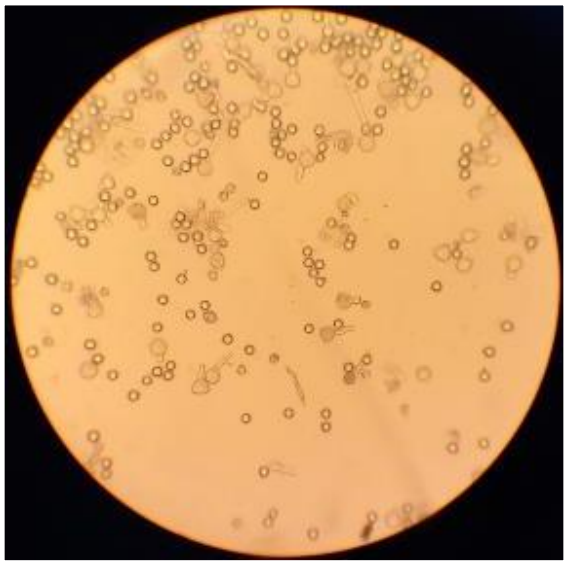


Figure 3.20. B26 and B100 fertility phenotypes
A B26 inflorescence displaying its flower bud abortion phenotype. **B** B26 pollen suffers from a low rate of germination *in vitro*, indicating low male fertility. **C** B100 pollen has a high rate of germination.

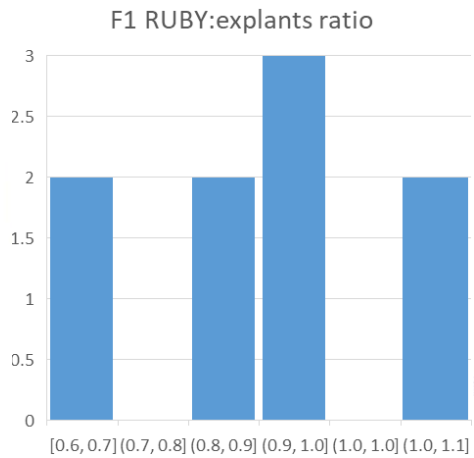
B

C



To generate such a line, I crossed B26 and B100 (SP10016). Berries were obtained and seed sown, yielding 29 F_1 hybrids. The plants were grown in a glasshouse until flowering, and I selected nine lines which grew vigorously, flowered well, and set fruit readily. Next, the lines were screened for their transformability. B26, B100 and their nine selected progeny were transformed with a plasmid bearing *NptII* and the RUBY visual reporter (He et al., 2020) under the 35S promoter. For each line, I calculated the ratio of red RUBY sectors to explants. The F_1s segregated for transformability in what appeared to be a 1:2:1 ratio, with some even surpassing B26 (Fig. 3.21 A&B). I named the top three lines YD1, YD2, and YD3 (Fig. 3.21 B&C).

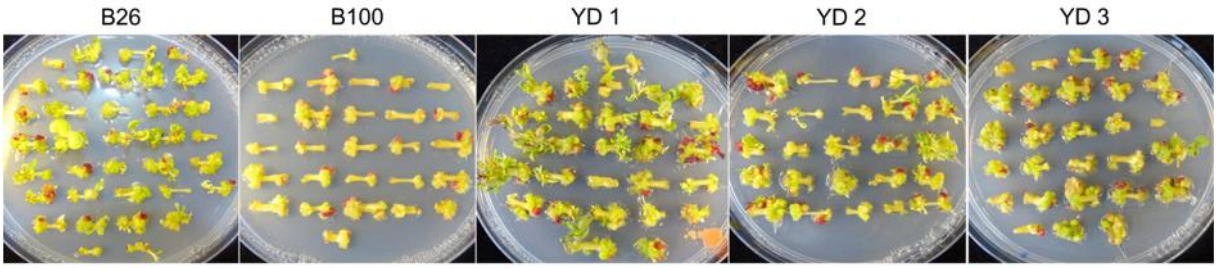
A



B

line	sectors	explants	ratio
1-11	37	59	0.63
1-2	49	55	0.89
1-3	40	58	0.69
1-5 (YD1)	49	55	0.89
1-8	40	50	0.80
2-16	39	49	0.80
2-2	44	50	0.88
2-6 (YD2)	54	50	1.08
2-7 (YD3)	58	52	1.12
B100	36	52	0.69
B26	65	68	0.96

C



D

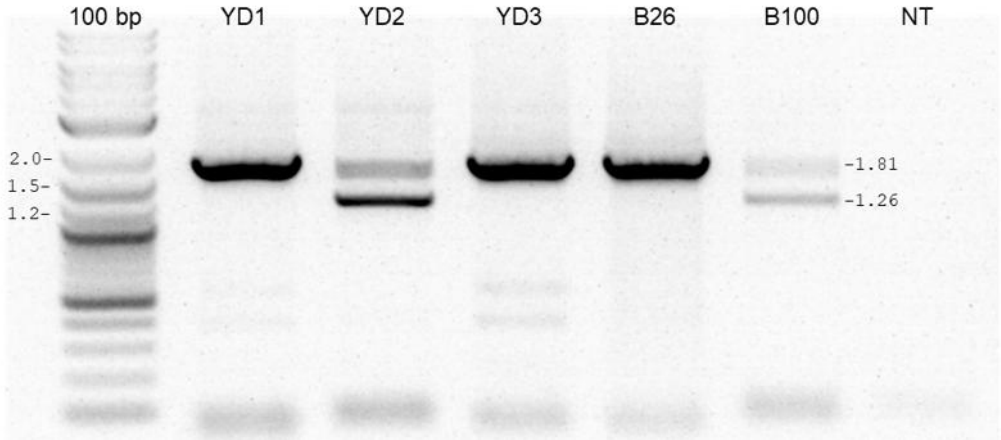


Figure 3.21. Transformability and genotyping of B26 × B100 F_1 hybrids.

A Histogram of RUBY sector: explant ratio of the 9 B26xB100 F_1 hybrids. The line's propensity for RUBY sectors appears to follow a 1:2:1 ratio. **B** Tabulated data of transformation efficiencies of B26, B100, and the nine hybrids. Top-performing hybrids were titled YD1-3 and highlighted in yellow. **C** Representative plate images of B26, B100, and YD1-3. **D** Genotyping data for B26, B100, and YD1-3. PCR-amplified across the *Cai1* locus using primers YD35/YD36. YD1, YD3, and B26 are homozygous at *Cai1*, while YD2 and B100 are heterozygous.

Lastly, I wanted to compare the propensity of the different lines to yield KI sectors. To do this, I first genotyped them at the *Cai1* locus as done previously. As already established, the homozygous B26 yields a single 1.81 kb band, while the heterozygous B100 produces two bands at roughly 1.81 kb and 1.26 kb. YD1 and 3 are homozygous for the *Cai1-long* allele like B26, while YD2 is heterozygous like B100 (Fig 3.21 D). Based on the substantial differences between the *Cai1-long* and *Cai1-short* alleles (Fig. 3.4 B), it seemed likely that the homology arms targeting one allele might be too different to target the other (Percent Identity in HA region = 62.4%). Furthermore, there is a SNP in the Cas12a target site being used (crRNA1), which should stop the nuclease from efficiently cutting the *Cai1-short* allele. Hypothetically, this means that it is only possible to target one allele at a time, and KIs in heterozygous lines should be less efficient due to a 50% reduction in viable targets. To test this prediction, the progeny and both parental lines were transformed with the same KI construct, targeting the *Cai1-long* allele as done until now.

To gain a better understanding of how KI sector formation occurs over time, I employed Time-to-Event analysis and counted the KI sectors multiple times between 7 and 28 dpcc (for detail see Section 2.2.6). Furthermore, transformation efficiency was quantified as the ratio of sectors to explant termini, as the two ends of an internode explant can generally by expected to harbour two distinct transformation events. This new approach to purple sector scoring was used in this and all other experiments going forward.

KI sectors started to appear ca. 10 dpcc and seemed to slow down at 28 dpcc. As predicted, the homozygous lines YD1, YD3 and B26 did best, but they performed very different from one another (Fig. 3.22). Throughout the tracked period, both YD1 (P-value = 0) and YD3 (P-value = 6e-04) significantly outperformed B26, yielding more than twice as many KI sectors relative to explant termini. In contrast, the heterozygous lines YD2 and B100 produced essentially no KI sectors, performing equally poorly despite YD2 transforming essentially as well as YD1 and YD3. This suggested lines YD1-3

were segregating for some genetic factor that was key to the KI process. This factor seemed to affect the line's propensity for KIs independent of the line's general transformability. Lastly, the segregation for this factor appeared to coincide with the state of zygosity at the targeted locus. This topic is further discussed in Section 3.3.8. Going forward, KI work could proceed in either YD1 or YD3, but I opted for the former as it produces shoots in tissue culture more readily.

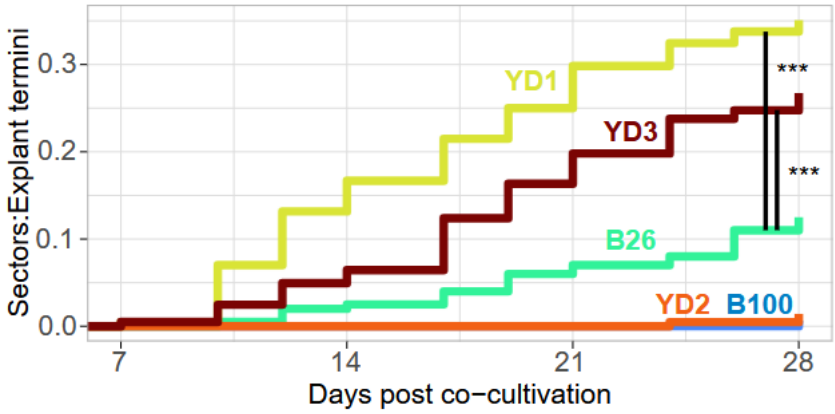


Figure 3.22. KI rates among potato lines.
Incidence curves of KI sector emergence in diploid potato lines B26, B100, YD1, YD2, and YD3 over time. YD1 & YD3 yielded the most KI sectors, significantly more than B26. IN contrast, YD2 and B100 yielded practically none. The construct transformed in all lines was piCas12a (more details below). Experiment done in two biological repeats with similar results.

Choosing a better programmable nuclease

To further increase KI efficiency, I attempted to optimize the elements cloned into the KI plasmids. The following constructs incorporated the best practices established thus far, as well as some additional improvements. Firstly, all future plasmids included the FAST-Red fluorescent seed selection marker outside of the KI cassette, where it could help with rapidly breeding away T-DNAs (Shimada et al., 2010). The *IPT* gene, which disrupts shoot morphology and rooting (Smigocki & Owens, 1989), was included just outside of the T-DNA left border to select against readthrough insertion events beyond the left border. To measure how constructs perform when delivering authentic R-genes, rather than just *NptII* or the 35S promoter, the *Rpi-vnt1* gene (ca. 4.0 kb) was included in

the KI cassette.

The first set of constructs were assembled to test the effect of two collaborator-provided constructs. One was an intron-containing HA-tagged version of ttLbCas12a (piCas12a), with reportedly better expression in plants (Schindele et al., 2023). The other was an intron-containing ttLbCas12a fused to a viral 5'-3' exonuclease (pUL12-iCas12a), which increased HDR rates in *Arabidopsis* and wheat by facilitating end resection at DSB sites (Schreiber et al., 2024).

In this experiment, KI sector counting was extended to 49 dpcc but emergence rates decreased with time (SFig. 10). 74% of the final sector count was already present by 28 dpcc so future scoring was limited to that timepoint. The KI sector formation rate with piCas12a was significantly higher than that of the control plasmid (pCas12a) featuring the classic ttLbCas12a (P-value = 0) (Fig. 3.23). Conversely, the KI rates observed with pUL12-iCas12a did not significantly differ from those with piCas12a (P-value = 1). These results suggested that in this experimental context, the amount of nuclease, and thus the rate of DSB induction, is rate-limiting for KIs, while the end resection step of HDR either is not, or the UL12 exonuclease tag does not facilitate it.

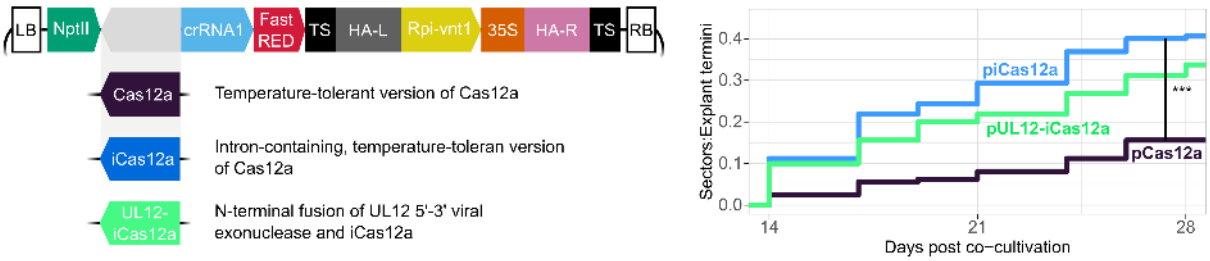


Figure 3.23. Comparing different nucleases.
Left: diagram of three KI plasmids featuring the ttLbCas12a, a version of ttLbCas12a with introns, or a version with introns and a viral exonuclease fused at the N'-terminus. Right: the plasmid with an intronized Cas12a yielded significantly more KI sectors that the plasmid without introns in the Cas12a. Adding the viral exonuclease made no significant difference. LB – T-DNA Left Border, RB – T-DNA Right Border, TS – crRNA1 Target Site. Experiment done in two biological repeats with similar results.

The conclusions drawn from the KI data predict that piCas12a should express the nuclease at a higher level than pCas12a, while

assuming the addition of the exonuclease tag would not change the level of expression. To test this, I transiently expressed pCas12a, piCas12a, and pUL12-iCas12a in *N. benthamiana* leaves, followed by a protein extraction and a Western Blot to detect the nuclease's protein levels. Across two biological repeats, Cas12a and iCas12a seemed to have equal protein levels, despite producing KI sectors at very different rates, while UL12-iCas12a was barely detectable, despite producing KI sectors at a similar frequency to iCas12a (Fig. 3.24). This inconsistency was likely an artifact of transient expression, which occurs under very different conditions to the predominantly stable expression seen in the KI experiment. A comparison of the Cas12a protein levels in stable lines transformed with pCas12a, piCas12a, and pUL12-iCas12a has been planned but is still outstanding.

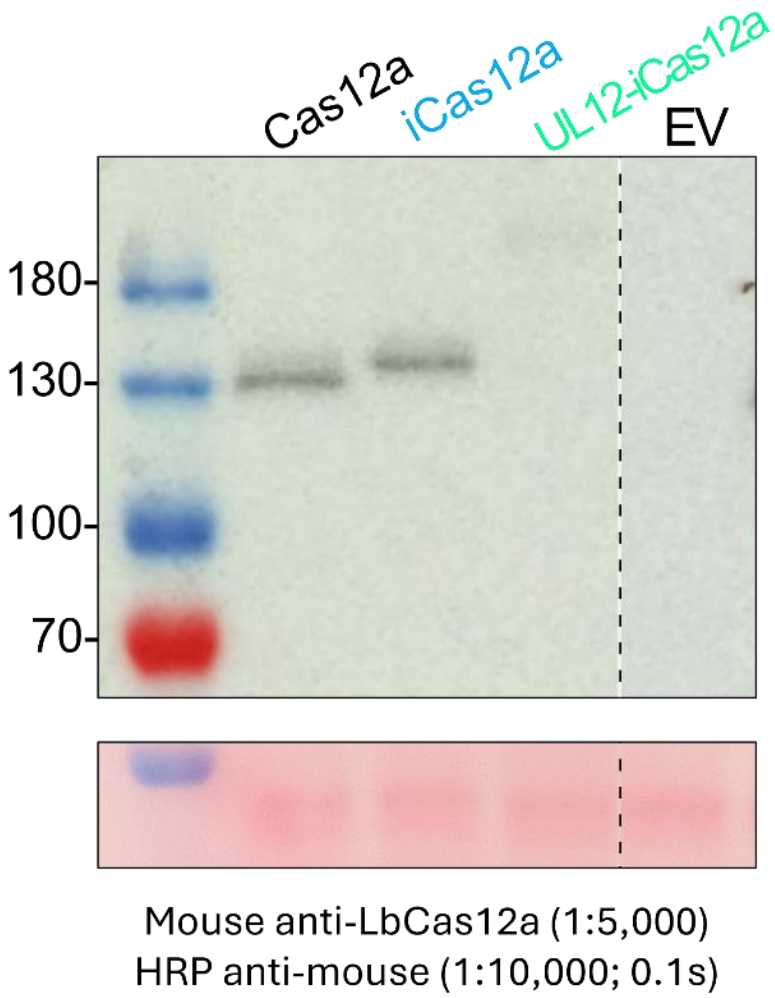


Figure 3.24. Cas12a protein levels with transient expression. The nucleases detected were ttLbCas12a (144 kDa), 3xHA-ittLbCas12a (151 kDa), and UL12-ittLbCas12a (220 kDa). Constructs were

transiently expressed in *N. benthamiana* leaves. Surprisingly, the presence of introns did not affect protein levels during transient expression. In contrast, the exonuclease fusion accumulated at a very low level. Two biological repeats done with similar results.

Testing crRNA expression cassettes

Thus far, my crRNAs have been flanked by the Hammerhead and Herpes delta virus ribozymes to ensure clean expression (Gao & Zhao, 2014). However, a tRNA-flanked architecture is also widely used and is particularly well-suited to multiplexing (Port & Bullock, 2016). The relative efficiency of crRNAs produced by these two expression systems has been investigated, but remains without a clear answer (J. Li et al., 2020; Milner et al., 2024; G. Yuan et al., 2022). To address this, a version of pUL12-iCas12a with a tRNA-flanked crRNA was cloned (pU6:tRNA). Additionally, I wanted to know if the choice of promoter and terminator for the crRNA makes a difference. All crRNAs until now were expressed using the U6-26 Pol III promoter and terminator from *Arabidopsis*. This is a relatively mild promoter-terminator pair, and there have been recent reports that stronger ones result in higher rates of GE (B. Li et al., 2022). To test this, a version of pU6:tRNA was cloned with the crRNA expressed by the strong Pol II Figwort Mosaic Virus (FMV) promoter (Mait et al., 1997) and a synthetic double terminator EU-*Nb*ACT (Diamos & Mason, 2018) (pFMV:tRNA). Surprisingly, neither the change of the expression cassette, nor the promoter and terminator made a significant difference (Fig. 3.25). The latter was particularly unexpected, as the two promoter-terminator pairs are likely to result in drastically different expression levels. A likely explanation is that crRNAs are already transcribed at a level far beyond what is required by the available Cas12a nucleases. If the enzymes are saturated with guides, then cloning a stronger promoter and terminator around the crRNA cassette would not impact cleavage rate, nor the KI rate. A comparison of the crRNA levels in stable lines transformed with pUL12-iCas12a, pU6:tRNA, and pFMV:tRNA via RT-qPCR has been planned but is still outstanding.

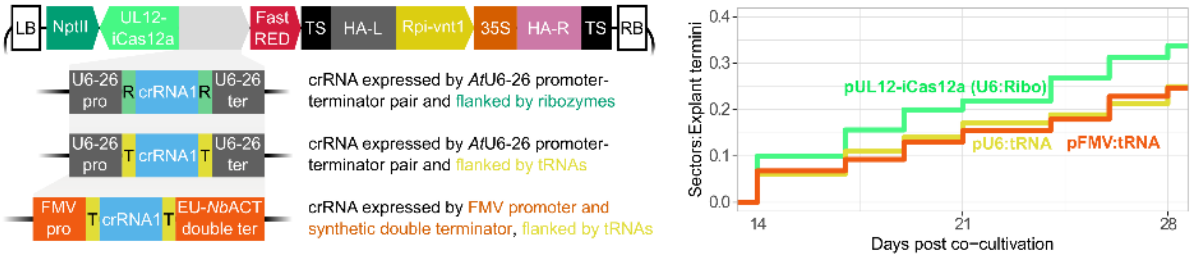


Figure 3.25. Comparing crRNA expression cassettes.
Left: diagram of three KI plasmids featuring three different crRNA expression cassettes. Right: all three plasmids performed similarly. LB – T-DNA Left Border, RB – T-DNA Right Border, TS – crRNA1 Target Site, R – ribozyme element, T – tRNA element. Experiment done in two biological repeats with similar results.

Choosing a better plasmid backbone

The last element to be optimised was the plasmid backbone. This oft-overlooked part of the binary vector is crucial to achieving efficient transformation, which is a prerequisite for efficient KIs. The previous constructs featuring the *IPT* gene were cloned into the pAGM31195 plasmid (X. Lin et al., 2022), which is a broadly standard binary vector featuring the pRiA4b ori for replication in *Agrobacterium* (plasmid maps can be found in SFig. 11). An alternative plasmid backbone is LP_pRKan290, which is a Golden Gate-compatible, Kanamycin-resistant binary derived from the pRK290 vector (Ditta et al., 1980). This is a large (ca. 20 kb) plasmid with numerous trans-acting replication and maintenance regions (Kornacki et al., 1984), and it has proven superior at transforming even recalcitrant tetraploid potato varieties (Youles et al., 2024). Another potential plasmid backbone is LP_pBIN19, which is a smaller (ca. 9.5 kb), Golden Gate-compatible derivative of the classic pBIN19 binary vector (Bevan, 1984). To compare these constructs' performance in inducing KIs, I cloned the same T-DNA cassette as in pUL12-iCas12a into LP_pRKan290 (pLP_pRKan290) and LP_pBIN19 (pLP_pBIN19). When transformed (Fig. 3.26), pLP_pRKan290 yielded significantly more KI sectors than pUL12-iCas12a (P-value = 0.002), while pLP_pBIN19 produced significantly fewer (P-value = 0.002). The scale of the difference in performance between the LP_pRKan290 and LP_pBIN19 plasmid backbones was unexpected, as both are ultimately derivatives of the same pRK2 plasmid (Bevan, 1984; Ditta et al., 1980) and share

the same origin of replication (ori), although pBIN19 has undergone more extensive truncations.

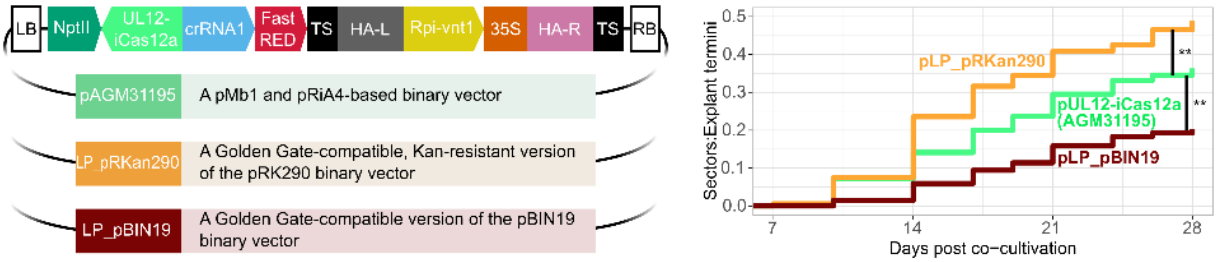


Figure 3.26. Comparing plasmid backbones.
Left: diagram of three KI plasmids cloned into one of three plasmid backbones. Right: the plasmid cloned into pRKan290 yielded significantly more KI sectors than that cloned into pAGM31195. This in turn proved significantly more efficient than the plasmid cloned into pBIN19. LB – T-DNA Left Border, RB – T-DNA Right Border, TS – crRNA1 Target Site. Experiment done in two biological repeats with similar results.

I also evaluated the performance of the plasmid backbones in transient expression assays. To do this, I cloned the firefly luciferase gene, under the control of a constitutive promoter (35SΩ), into each of the three vectors and transiently expressed them in *N. benthamiana* leaves. I then harvested the leaf disks and measured the luciferase activity (Fig. 3.27). Among the tested vectors, LP_pBIN19 consistently produced the highest luciferase activity, followed by LP_pRKan290, while pAGM31195 exhibited the lowest performance. Thus, it appears LP_pBIN19 is simultaneously the best plasmid backbone to use for transient gene expression, and the worst backbone for KIs. This discrepancy between the vectors' KI performance and their luciferase activity, along with the earlier observed inconsistency between KI rates and nuclease protein levels, underscores a key limitation of using transient assays to infer performance in gene targeting.

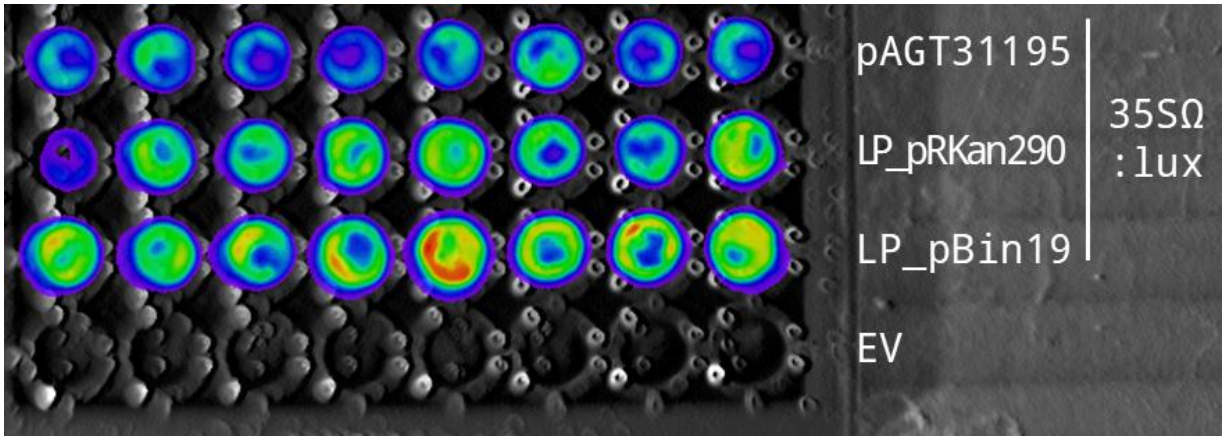


Figure 3.27. Transient expression performance of plasmid backbones.
The firefly luciferase gene was cloned under a constitutive promoter and into one of three plasmid backbones: pAGM31195, LP_pRKan290, and LP_pBIN19. The strength of signal consistently followed the order: LP_pBIN19 > LP_pRKan290 > pAGM31195. Imaging done at 2 dpi. The quantified signal strength can be found in Supplementary Figure 12. Experiment was done in three biological repeats with similar results. EV - empty vector negative control.

3.2.5 KI of *Rpi-vnt1*, an R-gene against *P. infestans*

Purple YD1 lines were recovered from the transformations of constructs bearing a KI cassette with *Rpi-vnt1* and a Cas12a with introns (piCas12a), introns and an exonuclease (pUL12-iCas12a), and the two alternative crRNA expression cassettes (pU6:tRNA & pFMV:tRNA) (Fig. 3.28 A). A single purple sector from YD2 was also recovered from its transformation with piCas12a (Fig. 3.22). The shoots it produced had an abnormal phenotype, and the purple pigmentation was not particularly intense.

The first five recovered YD1 lines and the single YD2 line were genotyped across the KI locus (Fig. 3.28 B). All lines yielded an intense band of 1.8 kb or smaller, indicating the presence of a WT allele or one bearing a small indel (Fig. 3.28 C). Three lines yielded a second, higher band but only that of line piCas12a-II had the expected size of 6.4 kb. The top bands of pUL12-iCas12a-I and pFMV:tRNA-I were roughly 3.5 and 4.0 kb, which might reflect incomplete cassette insertions.

The first five recovered YD1 lines and the single YD2 line were screened for Cas12a and confirmed to all carry a T-DNA (Fig. 3.28 A). This was expected as the *NptII* gene sat on the T-DNA outside the KI cassette, and the calli were selected on Kanamycin. The lines were then screened for KIs using whole-insert amplification (Fig. 3.28 B). This is process is ongoing, as the increased size difference between the KI band (6.4 kb) and WT band (1.8 kb) necessitates extensive PCR optimization. However, at least one line (piCas12a-II) produced the expected double band pattern (Fig. 3.28 C). This line has an intense purple phenotype, and efforts are underway to further characterise it and segregate away its T-DNA (Fig. 3.28 D).

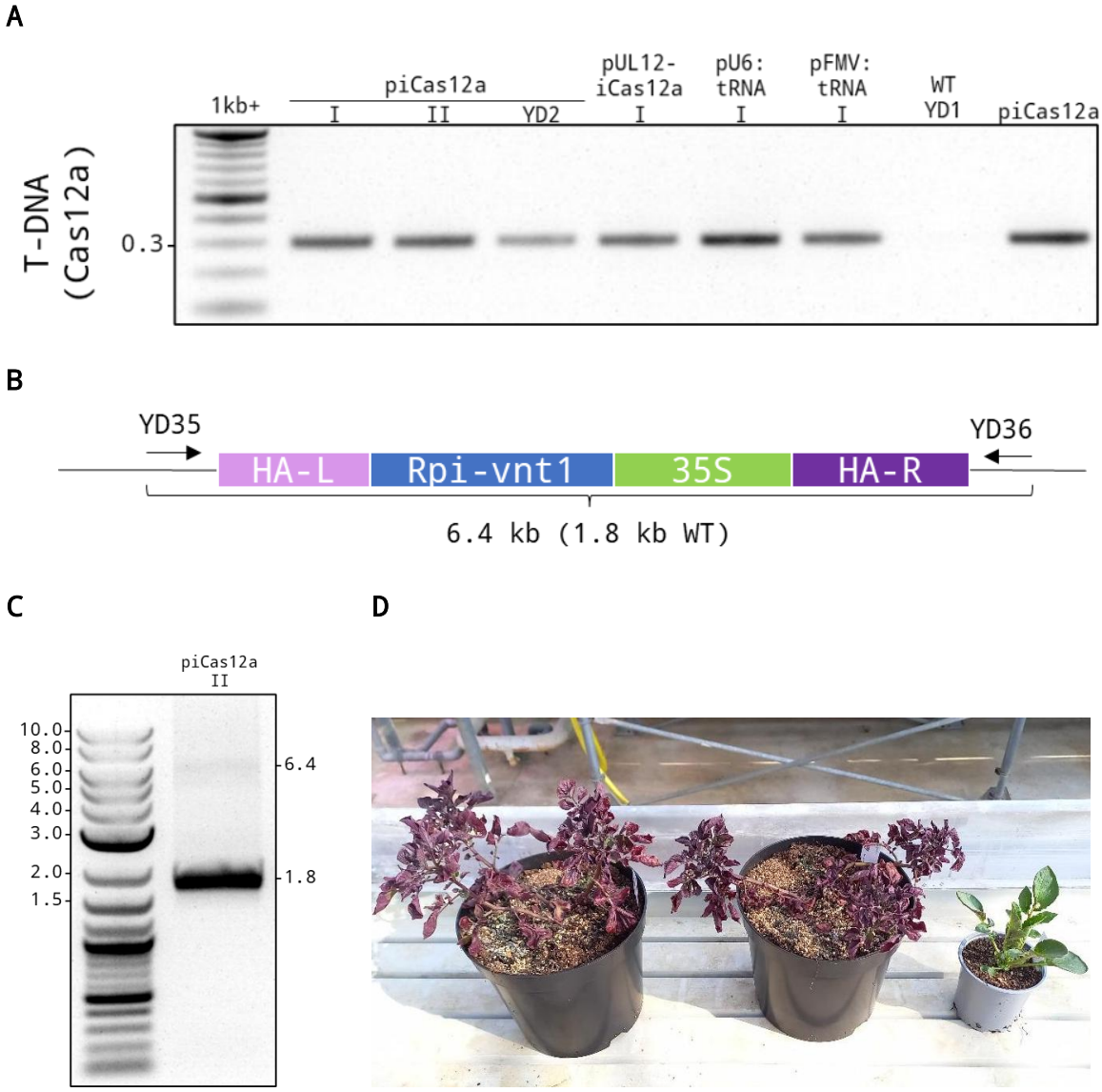


Figure 3.28. Phenotype & genotype of T_0 plants carrying *Rpi-vnt1*. **A** T-DNA screed (primers YD227-228). **B** Diagram of the predicted KI locus and the whole-insert amplification being attempted. **C** Whole-

insert amplification of line piCas12a-II on a 0.7% TAE gel. **D** Phenotype of piCas12a (left and middle) and a WT YD1 plant (right). The WT plant is smaller because the tuber sprouted later than expected. Ladder in all gel images was 1kb+ (NEB).

3.3 Discussion

In this chapter, I aimed to develop an efficient method for gene KIs in diploid potato. This involved implementing a KI reporter system and optimizing several aspects of the KI process. Additionally, I sought to apply this approach to integrate strong R-genes against *P. infestans* from wild potato relatives into the diploid germplasm. Significant progress was achieved, and valuable insights were gained into the factors influencing KI success. However, some conclusions are based on assumptions that require experimental validation, and a few critical questions remain unanswered. Despite these challenges, the advancements reported in this chapter, alongside progress in the broader field of GT, suggest a promising future where KIs may soon become a reliable and routine technique in diploid potato and other plant species.

3.3.1 A functional anthocyanin KI reporter in potato

The anthocyanin KI reporter system demonstrated by Čermák et al. (2015) proved highly effective in potato. *Cai1* overexpression produced a milder phenotype than *SlANT1*, resulting in callus sectors and plants which were immediately distinguishable from the WT while still reasonably vigorous.

KI plants recovered were moderately vigorous but their performance once outside of tissue culture varied throughout the year. On average, KI plants grew slower and were more elongated that their WT counterparts. This difference was particularly noticeable in plants grown in the darker months of the year and largely disappeared in those grown in the summer. Thus, the lower vigour in KI lines is at least partially caused by the blockage of light by the high anthocyanin levels.

The KI reporter gene’s developmental phenotype in the callus stage was more problematic. KI sectors grew slower than the WT callus

around them, which necessitated the excision of KI sectors from the calli and the periodic cutting back of green callus tissue during subculturing. This was laborious and resulted in small KI sectors occasionally being overgrown and lost.

Initially the KI cassette delivered the 35S promoter and the TMV Ω leader sequence. The latter was later dropped as the promoter alone still produced purple callus sectors and plants. This change did not result in an obvious difference in vigour against the background variability between lines obtained through tissue culture. Looking to reduce the extent of overexpression and improve KI tissue growth, I cloned a KI cassette with a truncated 35S(-106) promoter but this proved to be non-functional. Conversely, in the following chapter I tried using a version truncated after position -200 (35S(-200)) which did result in KI sectors and plants with a deep purple colour (Section 4.2.1). Again, no obvious improvement in vigour was observed. Thus, it appears that if there is an ideal level of *Cai1* expression, it is difficult to arrive at by truncating the constitutive 35S promoter. Recently, a preliminary experiment with a nopaline synthase (Nos) promoter driving *Cai1* resulted in robust callus pigmentation but little-to-no shoot pigmentation. If future work bears this out, the Nos promoter might become the best choice for KIs.

One limitation of the anthocyanin reporter is that it functions solely as a visual marker, rather than a chemical selection tool. As a result, some of the T_0 lines recovered were mosaics, displaying both purple and green tissue. I resolved this by regrowing the plants from purple tubers with purple eyes, which consistently yielded homogenous plants. Another drawback is that, while the *Cai1* MYB has a major impact on anthocyanin biosynthesis, it is not the only gene involved (Y. Zhang et al., 2014). As a result, F_2 populations like SP10749 and SP10750, which originated from heterozygous founders B26 and B100, exhibited additional variation for anthocyanin pigmentation (Fig. 3.14). This background variation complicated attempts to confirm the KI loci's Mendelian inheritance.

3.3.2 Nature of recovered KI events

All recovered KI lines were heterozygous. This could partially be explained by the reporter system used, as KIs appeared to reduce plant vigour more severely in homozygous lines (Fig. 3.14). Consequently, by selecting the most vigorous shoots in tissue culture, I might have unintentionally selected against homozygous KIs. The inability to readily distinguish between heterozygous and homozygous events due to the dominance of the purple pigmentation phenotype is an inherent limitation of the anthocyanin KI reporter. As a result, biallelic KIs may have occurred but gone undetected. This, combined with the practice of counting only one sector per explant (and later per explant terminus), means that while the KI frequencies reported are suitable for relative comparisons between constructs within the same experiment, they should not be interpreted as absolute frequencies of KI occurrence.

Of the four pKI1 events genotyped, three yielded amplicon sequencing results which perfectly matched the expected KI locus (Fig. 3.11, SFig. 6). The exception, line pRAD54-2, contained a 201 bp deletion in the *NptII* gene located in the KI cassette. Notably, this deletion was flanked by a 7 bp “CGATTGG” microhomology. This appears to be the result of a spontaneous rearrangement, possibly via MMEJ. Rearrangements in the delivered KI cassette have been reported before. For example, Puchta et al. (1996) reported on a 14 bp deletion in the delivered cassette which the authors hypothesised was due to a secondary structure in the single-stranded T-DNA causing a DNA polymerase slippage. No such secondary structure is present in the deleted region in pRAD54-2, and the scale of the deletion is over ten times larger, so it is likely that more than one mechanism might lead to deletions in KI cassettes.

More than one mechanism also seems to account for the KI cassette’s integration into the target locus. This is indicated by the fact that one out of six T_0 KI lines recovered from the first KI attempt was T-DNA-free (Fig. 3.11). The process of transformation via *Agrobacterium* co-cultivation is complex and some of the details are still unresolved, particularly the steps

between T-DNA delivery to the host nucleus and T-DNA integration into the genome. T-DNA arrives in the plant nucleus in the form of a protein-coated single-stranded DNA, also known as a T-strand (Gelvin, 2010). Multiple T-strands can be delivered to same nucleus and this might frequently be the case (Liang & Tzfira, 2013; Pereman et al., 2019). In the nucleus, the T-strands are converted into double-stranded T-DNA intermediates and this has been found to coincide with the earliest detectable transgene expression (Pereman et al., 2019). These intermediates can persist in the host nucleus, they can combine together into larger circular or linear extrachromosomal forms (Singer et al., 2012), or they can integrate into the host genome (Kralemann et al., 2022, 2024; van Kregten et al., 2016). Transient expression in the absence of selection tends to peak around 3-4 days post inoculation and slowly decline thereafter, indicating that most of the T-DNA is silenced or degraded several days after inoculation (Janssen & Gardner, 1990). Pereman et al. (2019) used single molecule imaging to detect an average of 8 double-stranded T-DNAs per nucleus in *N. benthamiana* leaf cells at 23 hours after inoculation. The authors pointed out that this contrasted with the 1-2 T-DNA loci usually found in transgenic *N. benthamiana* lines (Does et al., 1991) and suggested most of the delivered T-DNA is ultimately lost through degradation or dilution during cell division.

The recovery of a T-DNA-free KI line hints that KIs can occur during transient expression; we refer to this as the one-step process (Fig. 3.29). In the first KI experiment, the selectable marker was located within the KI cassette and any KIs that occurred prior to T-DNA integration would disable the selection for the T-DNA. If a one-step KI is combined with the subsequent loss of all forms of T-DNA from the cell lineage, a T-DNA-free KI line results. Alternatively, the KI can occur after T-DNA integration, which we refer to as the two-step process. Notably, T-DNA integration can occur independent of the KI which could result in one-step events that carry a T-DNA. Thus, it is not possible to precisely compare the relative frequencies of these mechanisms based on the results obtained thus far. However, the fact that most KI sectors appear after 10 days or more, when

stable expression predominates (Janssen & Gardner, 1990), suggests most KIs are the product of the two-step process.

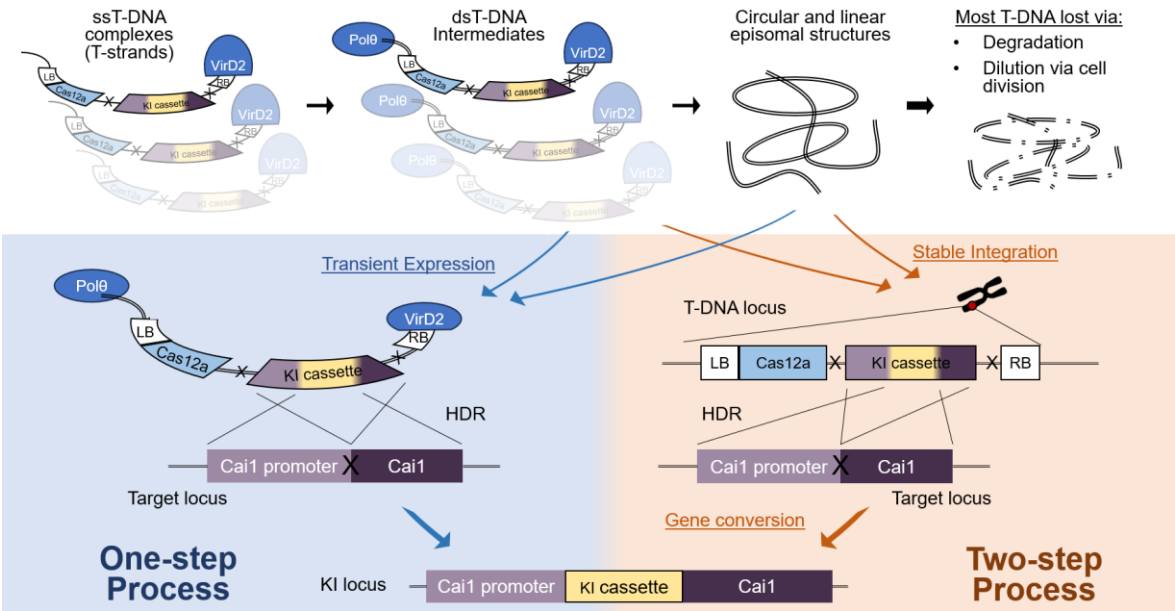


Figure 3.29. One-step and Two-step KI process.
Along the top: T-DNA arrives in the nucleus and a protein-coated single-stranded DNA called a T-strand. Multiple T-stands can be delivered to the same nucleus. These T-strands are processed into double-stranded T-DNA intermediates by the host DNA repair machinery. These intermediates can then associate into larger circular or linear episomal structures. Both these larger structures and the double-stranded intermediates can integrate into the host genome, but the majority is lost to either degradation or diluted during the process of cell division. For more information, see Kralemann et al. (2022). **Bottom left:** In the one-step KI process, the transiently expressed Cas12a cleaves the target locus and some form of double-stranded T-DNA, followed by HDR. T-DNA integration then may or may not occur, the latter resulting in a T-DNA-free line. **Bottom right:** In the two-step process, T-DNA integration occurs first. The Cas12a is stably expressed and cleaves the target and T-DNA loci. IF HDR occurs, the T-DNA locus changes the target locus in a form of gene conversion. Pol θ – DNA polymerase θ; VirD2 – *A. tumefaciens* VirD2 protein; LB – Left Border; RB – Right Border; X – Cas12a target site; dsT-DNA - double-stranded transfer DNA; HDR – homology-directed repair.

The KI T_0 lines which had *Rpi-vnt1* introduced all carried a T-DNA, likely because of the *NptII* gene's position on the T-DNA external to the KI cassette and the selection of the calli with Kanamycin. Importantly, the copy of *Rpi-vnt1* on the T-DNA should confer resistance against *P. infestans*. Because of this, it is necessary to segregate away such T-DNA insertions before the gain of resistance through R-gene KI can be demonstrated. At least one of the lines (piCas12a-II) seems to carry a genuine R-gene KI and efforts are underway to backcross this with a WT YD1 to segregate away the T-DNA and test the functionality of the knocked-in R-gene.

It remains unclear what proportion of the purple sectors observed on the calli result from HDR, and what proportion may arise from alternative DNA repair mechanisms. Notably, the plasmid pNHEJ, which lacks homology arms, still produced sectors on 5.9% of explants – compared to 12.4% with pKI1 (Fig. 3.18). This suggests that a non-negligible fraction of purple events could represent targeted insertions mediated by NHEJ. Such insertions are typically expected to carry indels at both junctions due to the error-prone nature of NHEJ. The genotyping methods used in this thesis should be capable of distinguishing these from KIs generated via HDR. HDR-mediated KIs, when genotyped across the Cai1 locus, should yield amplicons of the expected size and sequence, as well as a shorter amplicon if the KI is heterozygous. In contrast, NHEJ-mediated insertions would likely produce amplicons of variable length and sequence at the junctions. Thus far, all PCR-screened putative KI lines have yielded KI bands of the expected size (Fig. 3.11; Fig. 3.15; Fig. 3.28) and sequence (Fig. 3.11; SFig. 6), except pRAD54-2 whose KI band was slightly shorter than expected. However, this was shown to be due to a 201 bp deletion inside the NptII gene, while the junctions of the insert remained free of indels (SFig. 7). Therefore, if purple sectors do form via NHEJ, this likely occurs at a frequency at least several times lower than that of sector formation via HDR. While this might seem to conflict with the moderate frequency of purple sector formation with pNHEJ, it is possible the sectors observed were not true NHEJ events. While pNHEJ lacked homology arms, its KI cassette was still flanked by two copies of the crRNA

target site found at the *Cai1* locus. Cas12a cuts in a staggered manner, which leaves behind overhangs 4-5 nts in length. This raises the possibility that pNHEJ insertions occurred via a mechanism such as MMEJ, which utilizes microhomologies of such length (van Kregten et al., 2016). Unfortunately, shoots from the pNHEJ transformation were not recovered for genotyping.

3.3.3 Competition between HDR and NHEJ

As previously discussed, NHEJ and other error-prone pathways are the dominant form of DSB repair in somatic plant cells. Consequently, cleavage of crRNA target sites frequently results in insertions or deletions of various sizes. In addition to not facilitating KIs, this often results in the disruption of the target site which precludes any further cleavage by the Cas12a. This process, sometimes referred to as target site “spoilage”, could be seen in the sequence of the lower bands when KI lines were genotyped across the targeted *Cai1* locus (Fig. 3.11). While the bands’ roughly 1.8 kb size suggested that they might be produced by a WT non-targeted allele, sequencing instead revealed they carried disrupted crRNA1 target sites with 8-19 bp deletions (SFig. 8). These results are consistent with target site spoilage playing a major role limiting KI rates. This process also offers a mechanistic explanation for Cas12a’s reported superior performance in HDR compared to Cas9 (Grandgeorge, 2021; Tóth et al., 2016). Cas12a cuts at the far end of its protospacer, cleaving only the last 1-2 bp. As a result, NHEJ can often occur without the target sequence being disrupted, thus allowing the protospacer to be targeted again and offering another opportunity for HDR to take place. In contrast, Cas9 cuts in the middle of its recognition sequence, where even single nucleotide indel can disrupt the protospacer.

If target site spoilage reduces KI frequency, it might be possible to mitigate this through informed construct design. While little can be done to reduce spoilage at the endogenous locus, the protospacer copies flanking the KI cassette could, for example, be multiplied to allow for additional chances to achieve HDR (Fig. 3.30). To date, such an experiment has not been attempted.

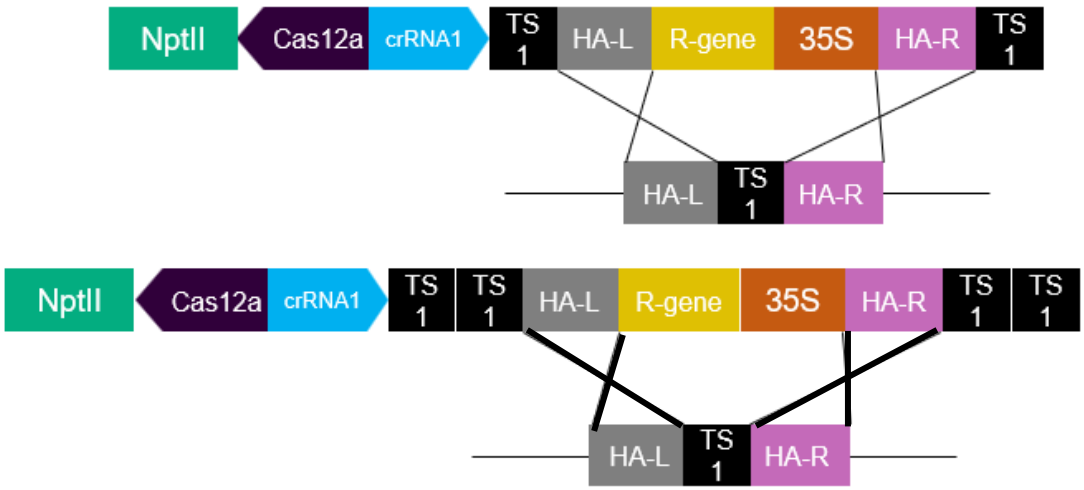


Figure 3.30. Example of KI constructs with 1 (top) or two (bottom) copies of the crRNA1 target site flanking the KI cassette.
The presence of an additional crRNA protospacers flanking the KI cassette might allow for HDR to take place even if another protospacer is "spoiled" by NHEJ. Such a construct is proposed but has not been cloned.

The classic approach to tackling the issue of NHEJ competition has been to attempt to either downregulate NHEJ or upregulate HDR. I also attempted this via several approaches. I ectopically expressed the *ScRAD54* gene under the constitutive 35S promoter, which was reported to boost the rates of HDR in *Arabidopsis* (Shaked et al., 2005). However, including the gene on the T-DNA made no significant difference to KI rates. Next, I included small molecule effectors to inhibit NHEJ and promote HDR. Neither 3-MB, a poly(ADP-ribose)polymerase inhibitor reported to increase homologous recombination (Puchta et al., 1995), nor SCR7, a proprietary inhibitor of DNA Ligase IV involved in NHEJ, increased KI rates (Table 3.1). However, 3-MB at a concentration of 1 mM significantly increased shoot formation in the callus, which might be useful in recovering KI plants. SCR7 application proved counterproductive, as it effectively inhibited the process of transforming of the explants, which NHEJ has since been found to be involved in (Kralemann et al., 2022). While this makes it unsuitable in the context of stable transformation, it might be useful if KIs were attempted via transient expression (discussed more in Section 3.3.5). Ultimately, none of my attempts to shift the balance between NHEJ and HDR increased KI rates.

Recently, Schreiber et al. (2024) reported that they boosted the rates of gene KIs in *Arabidopsis* and wheat using Cas12a-exonuclease fusions, which are believed to yield DNA brakes with 3'-overhangs more suitable as substrates for the process of HDR. Unfortunately, these Cas12a-exonuclease fusions failed to increase KI rates in my experiments.

3.3.4 Frequency of KI events

Increasing KI efficiency to a practical level is key. To do this, I experimented with modifying three aspects of KI construct design: the nuclease used, the crRNA expression cassette, and the plasmid backbone. The results from this were interpreted within a simplified conceptual model of the KI process (Fig. 3.31 A).

The nucleases I used were the ttLbCas12a (Huang et al., 2021), an HA-tagged intron-bearing version of ttLbCas12a (Schindele et al., 2023), and an intron-bearing version of ttLbCas12a tagged with the UL12 viral 5'-3' exonuclease (Schreiber et al., 2024). Ultimately, the addition of introns boosted KI rates significantly, while the further fusion of an exonuclease made no difference. Grützner et al. (2021) found the addition of introns to Cas9 increased editing efficiency in *Arabidopsis* due to a related increase in Cas9 protein levels. The increase in KI I observed was likely also due to an increase in Cas12a protein levels, which results in more cleavage, thus indicating that these aspects were rate-limiting (Fig. 3.31 B). However, detection of said protein levels in stably transformed lines remains to be done. It is unclear if further increases in Cas12a levels will result in even higher KI rates.

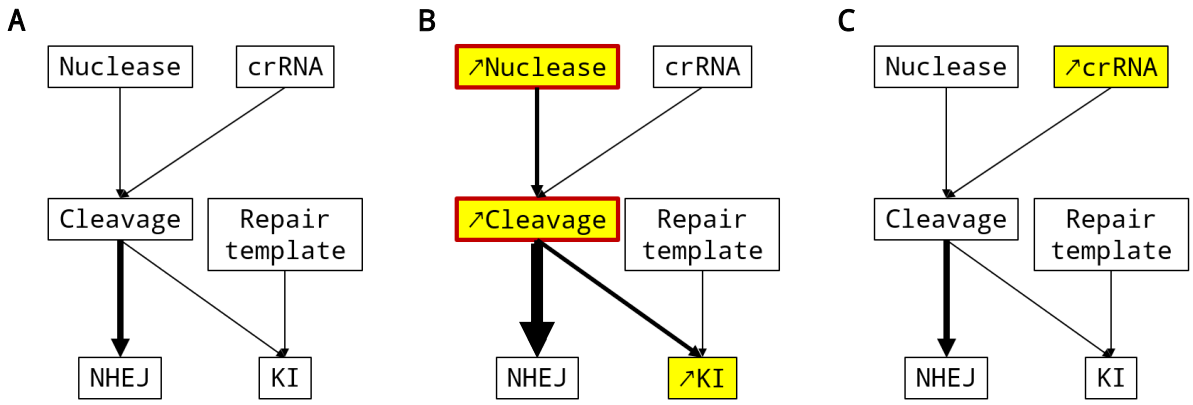


Figure 3.31. Model of the KI process.

When an increase (yellow) in a certain element of the KI process increases KI frequency, this indicates that this element is rate-limiting (red border). **A** Model of the KI process. Both a nuclease and a crRNA are necessary to induce target site cleavage. Cleavage, in turn, is required alongside repair template availability to induce KIs. However, most cleavage events are thought to be repaired by NHEJ instead. **B** The addition of introns to the nuclease led to higher KI rates, indicating nuclease expression and therefore target site cleavage are rate-limiting. **C** Changing the promoter and terminator on crRNA1 made no difference to the KI frequency, hinting that the level of crRNA produced is not rate-limiting. Note, that both the change in nuclease and crRNA levels is yet to be experimentally confirmed.

I compared the KI performance of crRNAs expressed by one of two promotor-terminator pairs (*At*U6-26 versus FMV promoter & synthetic double EU-*Nb*ACT terminator (Diamos & Mason, 2018; Mait et al., 1997)) and flanked by one of two excisable element systems (ribozymes versus tRNAs (Gao & Zhao, 2014; Xie et al., 2015)). The excisable elements are included to ensure “clean” transcription of crRNAs without any unwanted nucleotides on either side. Both ribozymes and tRNA are widely used in the GE field but attempts to compare their relative efficiency have thus far yielded inconclusive or conflicting results (J. Li et al., 2020; Milner et al., 2024; G. Yuan et al., 2022). The two promoter-terminator pairs are likely to yield different levels of transcripts, as they differ in origin and are transcribed by different RNA polymerases (Pol II & III). Ultimately, no significant difference was found between the KI rates with either promoter-terminator pair or excisable elements. There are two possible conclusions: either the crRNA expression systems yielded roughly similar levels of crRNA, or the levels were different but that did not affect the cleavage rate and KI frequency (Fig. 3.31 C). The latter appears more likely but still must be confirmed by directly measuring the levels of crRNA transcribed, potentially using the RT-qPCR method publish by (Varkonyi-Gasic et al., 2007).

Finally, I compared the effects of varying the plasmid backbone used to transform the same T-DNA (SFig. 11). The later KI

constructs, which featured the *IPT* gene next to the left T-DNA border, were cloned into the pAGM31195 plasmid. This is a broadly standard binary vector, 8.9 kb in length, replicated in *Agrobacteria* with the high copy number pRIa4b replication region (Nishiguchi et al., 1987; Oltmanns et al., 2009). The second plasmid backbone was LP_pRKan290 – a 20 kb Golden Gate-compatible, Kanamycin-resistant derivative of pRK290 (Ditta et al., 1980), which is itself a derivative of the 56 kb pRK2 (Ingram et al., 1973). pRK2 and its homologues have been described as "highly potent, self-transmissible, selfish DNA molecules with a complicated regulatory circuit" (Adamczyk & Jagura-Burdzy, 2003). Reflecting this, LP_pRKan290 contains multiple trans-acting replication and maintenance regions like *trfA* and *trfB* (Kornacki et al., 1984), as well as a low copy origin or replication *oriV* (Meyer & Helinski, 1977; Oltmanns et al., 2009). The third plasmid backbone was LP_pBIN19 – a 9.5 kb Golden Gate-compatible version of pBIN19, whose replicon is a highly truncated derivative of pRK2 (Bevan, 1984). While LP_pBIN19 shares the same *oriV*, it lacks many of the replication and maintenance genes found in LP_pRKan290 and pRK2.

The LP_pRKan290 backbone yielded significantly more KIs than pAGM31195, while LP_pBIN19 produced significantly fewer. This boost to KIs is likely a result of an improved stable transformation efficiency. This superior performance likely stems from the replication and maintenance genes which the plasmid backbone harbours. Such systems are lacking in pBIN19, which has gradually decreased in use for stable transformation due to a lacklustre efficiency, and this is likely the cause for LP_pBIN19's poor KI performance. Interestingly, there was no correlation between the plasmids' performance in stable and transient expression where LP_pBIN19 worked better than the other two backbones (Fig. 3.27).

Remarkably, despite extensive efforts to optimize GE in plants, essentially all have relied on plasmids designed primarily for efficiency in cloning in *E. coli* — favouring small size, high yield, and convenient restriction sites — rather than for optimal performance in transformation. When plasmid backbones have been

considered, optimization has largely focused on the ori, with little attention given to other elements outside of the T-DNA whose roles are not always well understood (Szarzanowicz et al., 2024; Zhi et al., 2015). The superior performance of LP_pRKan290 suggests a need to rethink this approach, as naturally occurring selfish DNA elements might hold the key to significant gains in both GM and GE efficiency.

The gains in KI efficiency already achieved in this chapter have been significant. However, two questions remain: quantitatively, just how much of an improvement has been achieved, and is the efficiency now good enough that a KI could be recovered without the need for an anthocyanin reporter? The answer to the first question can be arrived at by cloning a version of the very first KI construct (pKI1) featuring all the established improvements and then transforming them side-by-side. The second question could be answered by the KI of a cassette without the 35S promoter, followed by PCR screening. Both experiments are currently being planned.

3.3.5 Recovering T-DNA-free lines

As previously discussed, one out of six KI lines recovered from pKI1 and pRAD54 was T-DNA-free (Fig. 3.12), which is consistent with prior publications (Huang et al., 2021; Schreiber et al., 2024). These calli and plants were selected on Kanamycin, and a T-DNA-free line was obtained because the *NptII* gene was in the KI cassette. However, it is preferable to not have a selectable marker next to the delivered gene. Because of this, the *NptII* gene was moved outside of the cassette. This meant that employing Kanamycin selection will now select against any such T-DNA-free events, so Kanamycin use in the rooting media was discontinued. This would not be an issue in diploid potatoes, given that a T-DNA locus can be segregated away (Fig. 3.16), but it would be in a crop like tetraploid potatoes where backcrossing isn’t an option. In that case, it might be possible to conduct KIs without any antibiotic selection. Potatoes are generally very amenable to transformation, so much so that antibiotic-free transformations have been demonstrated before (Miroshnichenko et al., 2024). A preliminary experiment in YD1 has confirmed that KI sectors are

still produced in the absence of any antibiotic selection at a rate roughly three time lower than usual (data not shown). Without chemical selection, KIs can still be generated via either the One-Step or Two-Step processes described in Section 3.3.2 but the competition with WT untransformed tissue is more intense. In this context, NHEJ inhibitors like SCR7 might finally produce *in planta* the HDR-boosting effects reported in mammalian cells (Maruyama et al., 2015; Vartak & Raghavan, 2015). Alternatively, antibiotics in the media can be used to select not for the T-DNA, but for a chemical resistance arising from the KI, such as that reported by Butler et al. (2016) or the GT-specific selection system presented in the following chapter.

3.3.6 Segregating away T-DNAs

If a T-DNA does integrate into a KI line, it would generally be necessary to segregate it away by crossing. That was straightforward to do in diploid potatoes, but it could be made easier. The F_1 progeny derived from the pRAD54-3 × B100 segregated 11:2 for the presence: absence of a T-DNA (Fig. 3.16). This was compatible with the 7:1 ratio expected from an F_1 population deriving from a parent with three independent T-DNA insertions. The triple T-DNA insertion might be a result of the plasmid backbone initially used (pICSL4723-LacZ) which the high copy number pVS1 *ori* (Zhi et al., 2015). The low copy number LP_pRKan290 backbone adopted later might be less prone to producing such multi-insert events, although some reports have cast doubt on the link between the *ori*, the plasmid copy number, and the frequency of single copy T-DNA events (Zhi et al., 2015). The T-DNAs that do integrate can then be quickly segregated away if the FAST-Red seed coat fluorescence marker is included on the construct, as it works well in potato (Shimada et al., 2010).

3.3.7 Segregation for transformability

I crossed B26 and B100 to obtain a line with the former's high transformability, and the latter's vigorous flowering. F_1 lines selected for their strong growth and fertility were transformed with a 35S:RUBY construct, resulting in a segregation which approximated 1:2:1. The ratio suggested heterozygous parental

lines, which is not unexpected given B26 and B100's high degree of heterozygosity, but their differing performance in tissue culture suggested some form of epistatic interaction. One hypothesis that neatly explains the results involves two loci: a transformability locus (*T*) and a recessive epistatic locus (*I*) (Table 3.2). Both parents are heterozygous for *T* (*Tt*), but B26 is homozygous for the dominant allele at *I* (*II*), while B100 is homozygous for the recessive allele (*ii*). In this case, all F_1s will be heterozygous for *I* and its effects will not be visible, while *T* segregating will result in the transformability ratio of 1:2:1. While predicated on only 9 F_1s, this hypothesis does accurately predict that B100 and B26's phenotypes will belong to the *Iitt* and *IiTt* putative phenotypic classes. If the perceived segregation ratio is confirmed with a larger number of F_1 progeny, and a high-throughput phenotyping method for transformability were established, it could be possible to map the inferred loci.

TF eff	Low		Mid		High
P	B100 *iiTt*	×	B26 *IITt*		
F_1	*Iitt* 1	:	*IiTt* 2	:	*IiTT* 1

Table 3.2. Hypothetical inheritance mechanism for transformability.
The mechanism features a recessive epistatic locus (*I*) which masks the transformability locus (*T*) when homozygous recessive (*ii*).

Stable transformation via *Agrobacteria* co-cultivation is a complex process affect by numerous factors and plant genotype is one of them (Azizi-Dargahlou & Pouresmaeil, 2024). One limitation of the phenotyping performed is that it is unclear if the plants are segregating for transformability, or just regeneration capacity. For example, Taylor & Veilleux (1992) found 2 independent loci determined leaf disk regeneration rates in *Solanum phureja*¸ a diploid wild potato, with no *Agrobacteria* involved. Conversely, El-Kharbotly et al. (1995) identified a locus for transformability in diploid potato which appeared to segregate independently from

performance in regeneration. Furthermore, they found this putative transformability locus to be linked with the *R1* locus and a molecular marker on chromosome 5. As transformation with *Agrobacteria* is fundamentally an infection process, it is perhaps not surprising that a transformability locus might co-segregate with other immune receptors.

3.3.8 Segregation of KI amenability

The highly transformable lines YD1-3 and the parental lines B26 and B100 differed dramatically in their amenability for KIs. The lines homozygous for *Cai1* yielded KI sectors but at different rates, with both YD1 and YD3 significantly outperforming B26, while the heterozygous YD2 and B100 produced essentially no KIs. It appeared that a genetic factor for KIs was segregating, and it co-occurred with homozygosity at the *Cai1* locus with marginal statistical significance (Fisher's Exact Test, P-value = 0.1).

This could be explained by a hypothetical process of "template competition" (Fig. 3.32). When a targeted locus is cleaved and HDR occurs, the KI cassette must compete with any alternative templates. In plants, there will be at least one other allele for every gene, which is likely to have a more extensive homology to the cleaved locus than any provided template. If the plant is diploid and heterozygous, and the alternative allele is not targetable, then HDR based on it will preclude future KIs. If the plant is diploid and homozygous for the targeted allele, then HDR at the targeted locus will restore it and allow for it to be cleaved again, thus multiplying the chances of the otherwise uncompetitive KI cassette to become integrated. Notably, if a diploid and homozygous plant is targeted and NHEJ results in a spoiled site, the plant becomes heterozygous for the purpose of KIs. While it should still be possible for a heterozygote to conduct HDR using the KI cassette, the lack of KIs obtained from the well-transforming YD2 line indicates this is rare. Notably, template competition implies that plant species with high states of ploidy might be less suitable for KIs due to the presence of more competing endogenous alleles.

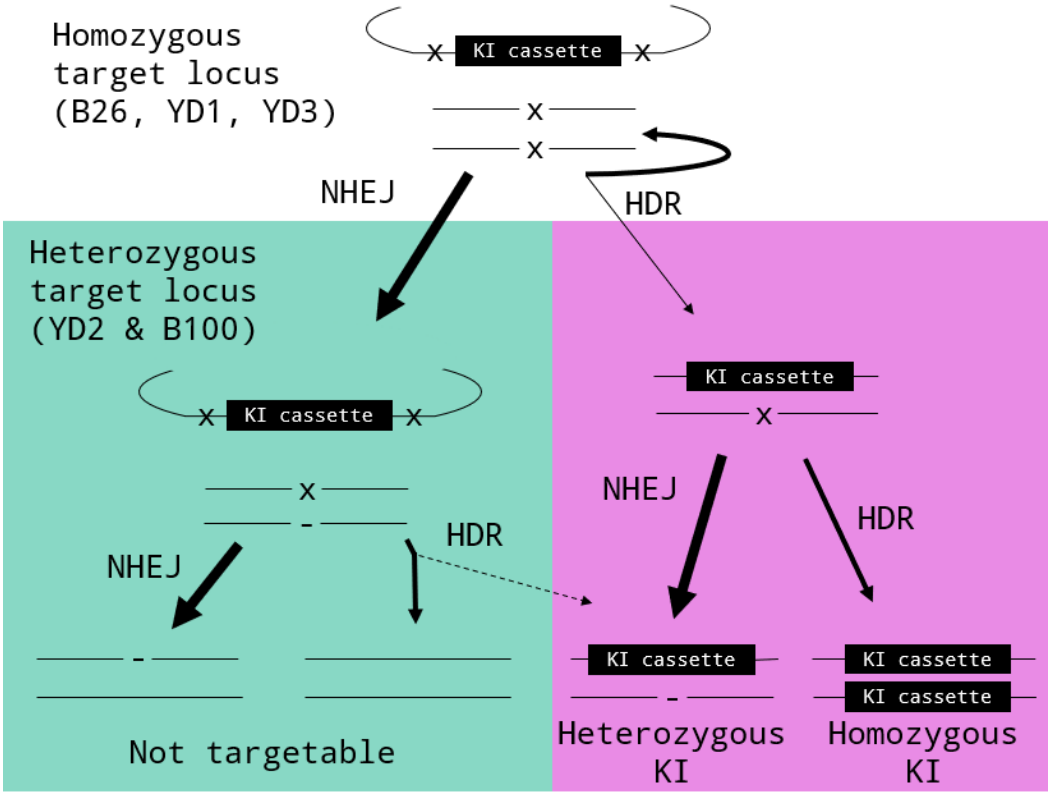


Figure 3.32. Template competition hypothesis.

Top: During a KI attempt, the cleavage of a protospacer at a homozygous *Cai1* locus (e.g. B26, YD1, or YD3) might often result in protospacer spoilage via NHEJ (left; green). If HDR occurs (right; magenta), then the targeted locus is either restored to its initial state or converted into a KI event, depending on whether the sister allele or the KI cassette are used as a repair template. If the former occurs, the locus can be cleaved again. Here, I refer to this as "template competition".

Left: If the KI experiment takes place in a line where only one allele is targetable (e.g. YD2, B100), or if the line was homozygous but one allele has already been spoiled by NHEJ, then cleavage of the targetable allele can result in protospacer spoilage if NHEJ **OR** HDR occur, provided the latter uses the non-targetable sister allele as a repair template. If HDR strongly prefers the sister allele to the T-DNA-provided KI cassette, this could explain why YD2 & B100 do not yield KIs (dotted line). Beyond this, the target locus becomes non-targetable and the likelihood of KIs is greatly diminished.

Right: If a KI occurs and the sister allele is still targetable, this might be spoiled by NHEJ or be the subject of a second KI. Presumably, the latter might be more likely than before due to the

availability of a knocked-in sister allele to serve as a template. However, the developmental penalty associated with homozygosity for a *Cai1* overexpression allele might have selected against this outcome here. **X** – viable protospacer; **–** - “spoiled” protospacer.

CHAPTER 4. TARGETED ITERATIVE INSERTION OF GENES VIA RECOMBINATION (TIIGER)

4.1 Introduction

In the preceding chapter, substantial progress was made in establishing efficient gene KIs in potato, but all recovered KI plants were intensely pigmented and reduced in vigour. As previously discussed, this could be resolved with the right choice of promoter, ideally one endogenous to potato, but the fundamental issue remains that the cassette lands in front of the *Cai1* ORF, replaces the native promoter and changes the phenotype.

Another limitation is that, if a KI is somehow selected for, there is no straightforward way to reuse the same selection again and add further genes to the same line. This can be a major constraint in the context of a breeding programme with dynamic goals and priorities. It is essential to be able to ensure that each hard-won elite breeding line has all the right genes added (and none of the wrong ones). These issues could be greatly alleviated if there was a method to efficiently deploy genes in an iterative manner, whereby the marker genes can be added and removed as needed.

Such a method requires the implementation of successive targeted gene replacements as well as insertions. There are some prior examples of such replacements in plants. Weinthal et al. (2013) used ZFNs to replace the *green fluorescent protein* (*GFP*) gene with a Hygromycin-resistance gene in tobacco and *Arabidopsis*. However, this was done using NHEJ and the efficiency was very low, with only 1/51 screened plants harbouring a clean replacement event, even after selecting with Hygromycin. Fauser et al. (2012) conducted the targeted replacement of an *NptII* gene with a cassette restoring a disrupted GUS reporter in *Arabidopsis*. This was done via HDR and GUS-positive lines were recovered at a low frequency (<1%). Ultimately, both attempts targeted a T-DNA locus with an easily detectable phenotype. Other papers that have mentioned gene or allele replacement have generally referred to targeted substitutions, targeted insertions, or some mixture of

the two (Begemann et al., 2017; J. Li et al., 2016; Y. Zhang et al., 2012). To date, no efficient targeted replacement of an endogenous locus has been reported in plants.

Aims

The goal of this chapter is to develop a method for the targeted iterative insertion of genes via recombination (TIIGER). The aim is to deliver R-genes against *P. infestans* into diploid potato at a defined location, in this case upstream of the *Cai1* gene promoter, without any residual unintended sequences (i.e. “scar sequence”) in the final product.

An implementation of TIIGER was conceived with two distinct phases. In phase 1 (Fig. 4.1 Top), a WT plant is transformed with a construct (T-DNA1) with a Cas12 that targets the proximal *Cai1* promoter either once in the middle (orange Cas12a), or once on both sides (green Cas12as). The HA-R is homologous to the *Cai1* ORF, as seen previously. In contrast, the HA-L varies; initially one with homology to the distal *Cai1* promoter is used (dHA-L). This should result in a targeted replacement of the proximal promoter region with the TIIGER cassette in the case of HDR. The cassette contains a gene of interest (R-gene 1), the 35S promoter, and the *cytosine deaminase* (*codA*) gene from *E. coli*. If integrated correctly, a phase 1 event results in a purple plant with a *codA*-conferred susceptibility to 5-fluorocytosine (5-FC) (Stougaard, 1993). At the end of phase 1, purple tissue on green callus is selected with Kanamycin (an *NptII* gene sits outside of the cassette) with the expectation that a T-DNA is integrated unlinked to the Cai1 locus.

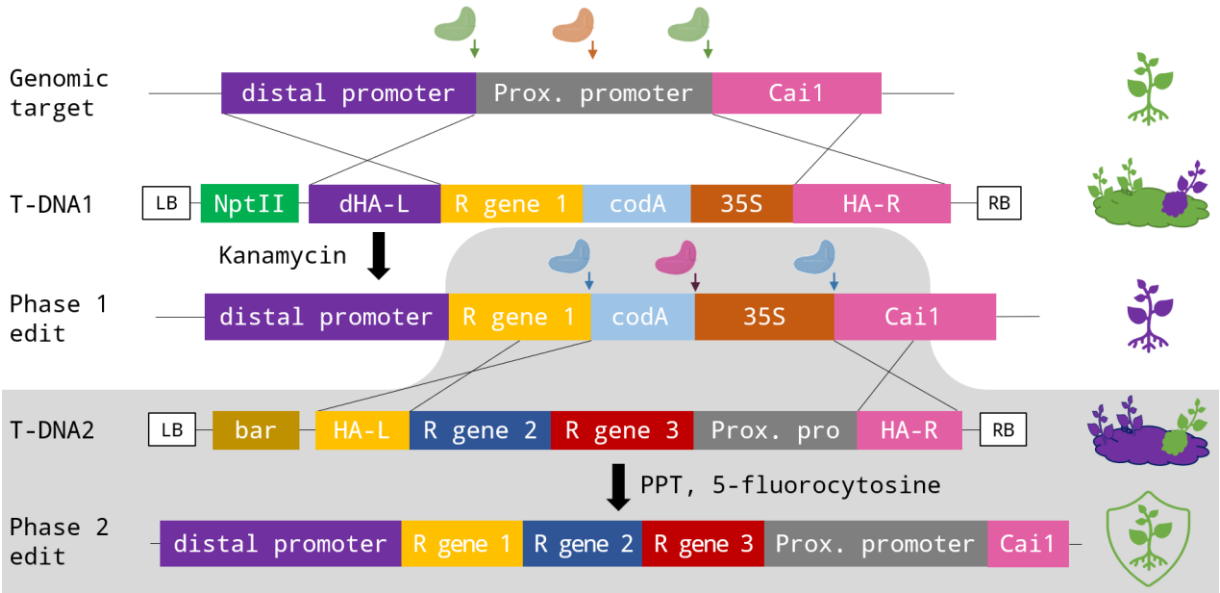


Figure 4.1. Biphasic TIIGER approach.
Top white space: Phase 1. The proximal promoter region is targeted once (orange Cas12a) or twice (green Cas12as) and replaced with an R-gene and the *codA*-35S selection cassette. Select for purple phase 1 events on green callus with Kanamycin.
Botton grey space: Phase 2. A purple phase 1 event is targeted once (pink Cas12a) or twice (blue Cas12a) to replace the *codA*-35S selection cassette with additional R-genes and the original proximal promoter. Select for green phase 2 events on purple callus with PPT and 5-fluorocytosine. The phase 2 event can again be targeted in a phase 1 mode, allowing for the iterative delivery of genes.
dHA-L – HA-L with homology to the distal *Cai1* promoter; Prox. promoter – proximal *Cai1* promoter; PPT – phosphinothricin; bar – bialaphos (PPT) resistance gene.

In phase 2 (Fig. 4.1 Bottom), a line carrying the phase 1 event is super-transformed with a construct targeting the *codA*-35S selection cassette (T-DNA2), again either once in the middle (pink Cas12a) or once on both sides (blue Cas12as). The TIIGER cassette contains additional genes of interest (R-genes 2&3), the native proximal *Cai1* promoter replaced in phase 1, and a HA-L homologous to the downstream end of R-gene 1. In case of HDR, the *codA*-35S selection cassette should be replaced with the additional genes of interest and proximal promoter. This phase 2 event would result in green, 5-FC-insensitive plants. Here, green tissue on purple

callus is selected with phosphinothricin (PPT, selects for the T-DNA which features the *bar* gene outside of the cassette) and 5-FC.

Ultimately, the TIIGER approach should result in the targeted integration of one or more genes of interest, without any scar sequence, next to a WT *Cai1* allele. Furthermore, provided both T-DNAs used are segregated away from the phase 2 event, or their integration was avoided as discussed in the previous chapter, the phase 2 event can again be targeted in a phase 1 mode, allowing for the iterative delivery of genes at the TIIGER locus.

Lastly, this method should benefit from the development of novel selection systems for GT events. The anthocyanin visual reporter used thus far requires the manual selection of purple callus tissue and shoots during subculturing. A reporter where GT results in chemical resistance would be more practical. This has previously been done by targeting either a pre-delivered disrupted selection marker (Weinthal et al., 2013), or the endogenous *ALS1* locus to induce chlorsulfuron resistance (Butler et al., 2016). However, in these approaches the targeted locus and the resulting resistance are genetically linked. A hypothetical GT-specific selection (GTSS) system where cassette integration at an arbitrary locus results in cassette-encoded chemical resistance would make recovering phase 1 events significantly easier. Furthermore, by alternating the type of chemical resistance delivered, it might be possible to conduct multiple TIIGER phases to rapidly deliver numerous genes before reconstituting the WT allele via the TIIGER phase 2.

In pursuit of the TIIGER method outlined above, this chapter aimed to answer the following concrete questions:

- Can targeted replacements be efficiently obtained by using a distal HA-L?
- Is it better to cleave the region being replaced once in the middle, or once on both sides?
- Does the TIIGER cassette size affect the efficiency of the targeted replacement?
- Can GTSS be established to facilitate additional steps in TIIGER?

- Can the TIIGER method ultimately deliver multiple R-genes with no scar sequence to a diploid potato line and thus make it resistant to *P. infestans*?

4.2 Results

4.2.1 Developing TIIGER

Efficient targeted replacements using two crRNAs at the *Cai1* locus

The 829 bp region of the *Cai1* promoter previously used as HA-L was designated as the Proximal promoter to be removed. For that purpose, the Distal promoter (800 bp region immediately upstream) was cloned as dHA-L (Fig. 4.1 A; spans residues –1,571 to -863 relative to the start of translation).

Two suitable crRNAs which cut just upstream and in the middle of the Proximal promoter were designated as crRNA5 and crRNA3, respectively. Phase 1 constructs were designed to either cut the *Cai1* locus twice with crRNAs 5 & 1 (Fig. 4.2 B p2cuts) or just once with crRNA3 (Fig. 4.2 B p1cut) using the ttLbCas12a nuclease. Versions with the intron-bearing ttLbCas12a were also cloned to test if improved nuclease expression affects targeted replacements as it did KIs (Fig. 4.2 B p2cuts+int and p1cut+int). Furthermore, the constructs incorporated the improvements from the previous chapter: the LP_pRKan290 plasmid backbone, the FAST-Red seed selection marker, and the 35S promoter. Lastly, all constructs carried a minimal TIIGER cassette, featuring just the *codA* gene and 35S promoter.

All constructs produced purple sectors (Fig. 4.2 C). Constructs which cut the target locus twice significantly outperformed the constructs which cut once (P-value>0.001). Intriguingly, the intron-containing ttLbCas12a performed significantly better when cutting twice (P-value>0.001) but made no difference when cutting just once (P-value=1.0).

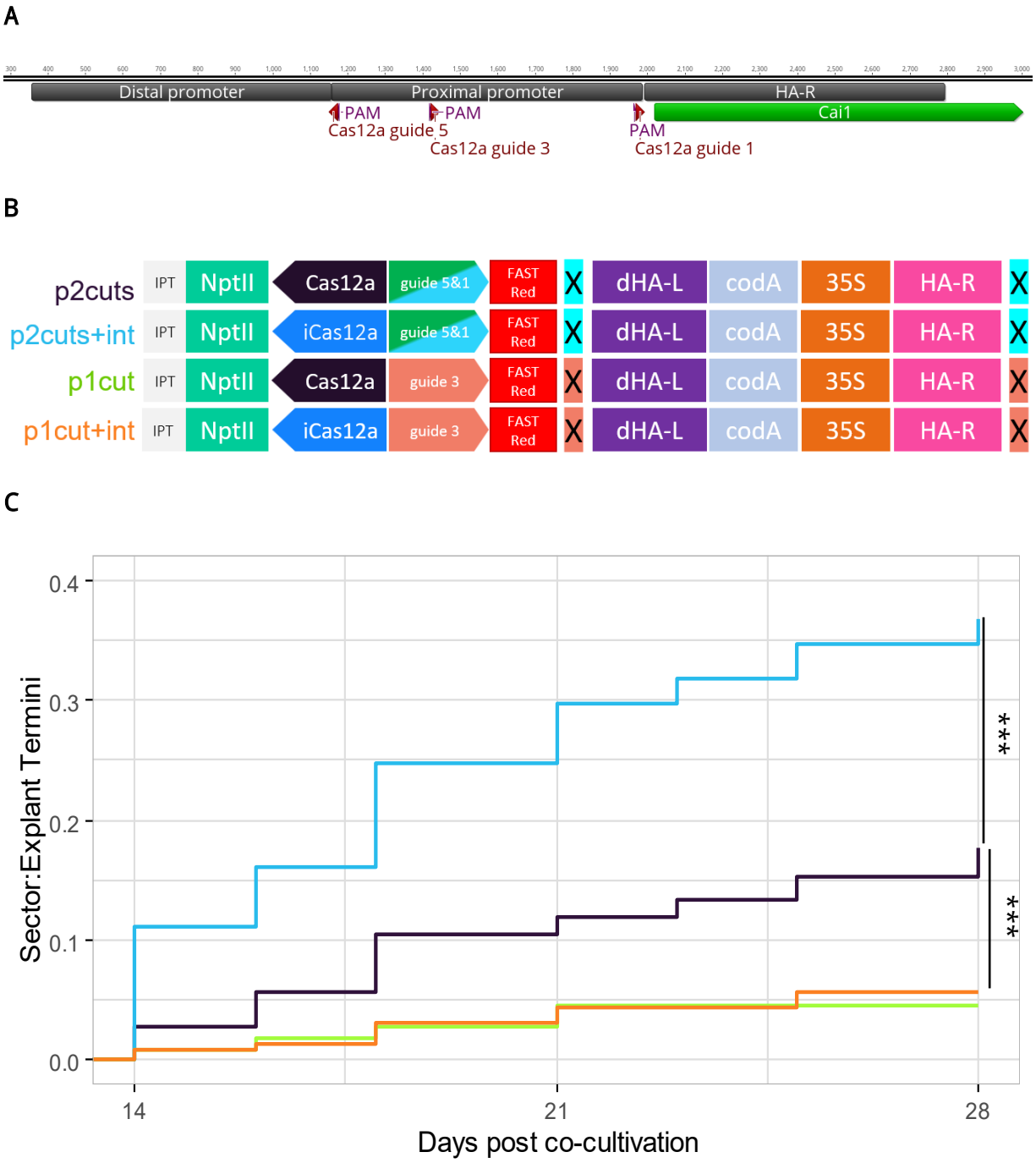


Figure 4.2. Efficient targeted replacements.
A The *Cai1-long* locus with the HA-R and promoter regions annotated. The crRNA1 & 5 protospacers flank the Proximal promoter, while the crRNA3 protospacer is between them, separated by 550 and 280 bp, respectively.
B Diagram of Phase 1 constructs, which differ in using the crRNA5&1 / crRNA3 or in using the ttLbCas12a / intron-containing ttLbCas12a. The protospacers flanking the TIIGER cassette were targetable by either crRNA1 (cyan) or crRNA3 (orange).
C Trace colour corresponds with construct name colour in **B**.

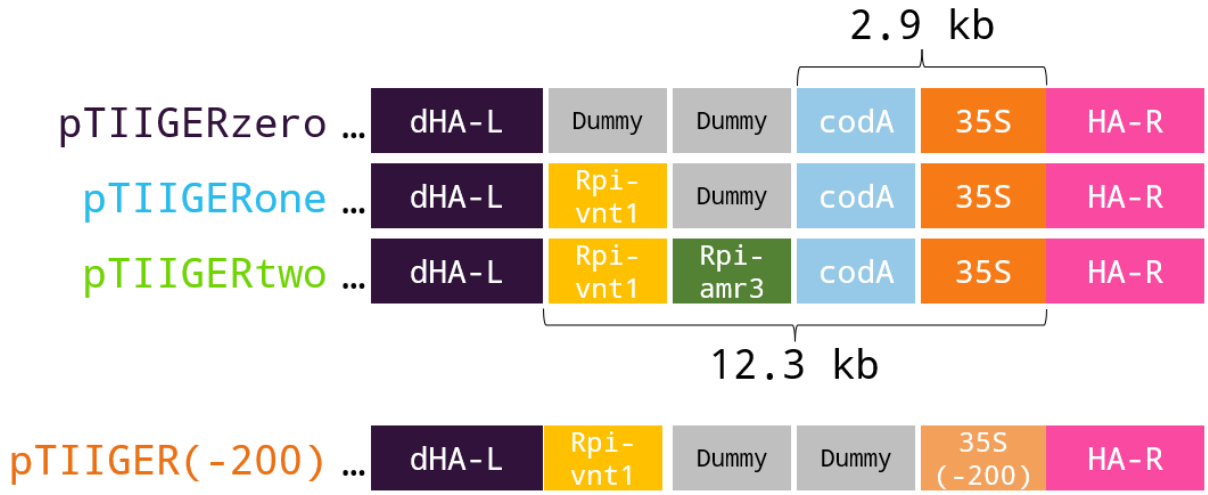
Indicated differences have Bonferroni-adjusted P-values < 0.001. Results are from a single biological repeat; second repeat is still outstanding.
iCas12a – intron-bearing version of ttLbCas12a.

These results indicate that if the proximal promoter is cut on both sides, cleavage rate is rate-limiting for targeted substitutions. In contrast, if a single cut is made in the middle, another factor is rate-limiting, likely a step of the HDR process disrupted by the remnants of the Proximal promoter, which constitute 280 to 550 bp regions of non-homology (Fig. 4.2 A). Going forward, I used two flanking crRNAs and an intron-containing ttLbCas12a to achieve targeted replacements.

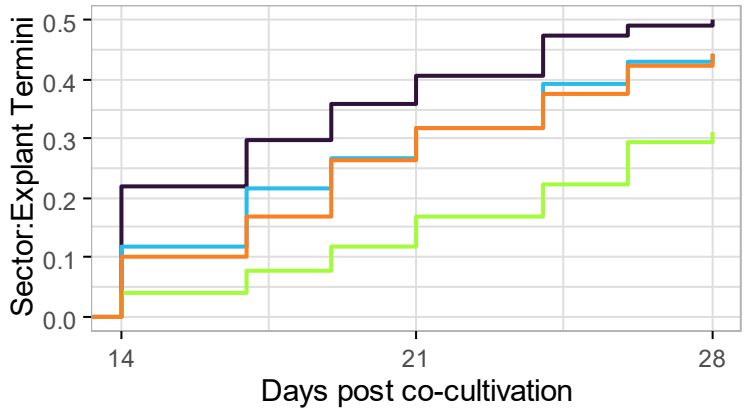
Targeted replacement of cassettes up to 12.3 kb

To test if TIIGER cassette length affects the frequency of targeted replacement, I cloned three different TIIGER cassettes into otherwise identical phase 1 constructs (Fig. 4.3 A). These included the *codA*-35S selection cassette, plus none, one, or two R-genes (named pTIIGERzero, pTIIGERone, and pTIIGERtwo). The R-genes in question were either *Rpi-vnt1* (S. J. Foster et al., 2009; Pel et al., 2009) or *Rpi-amr3* (Witek et al., 2016), both with their native promoter and terminator. These TIIGER cassettes differed in size between 2.9 and 12.3 kb. Additionally, I wanted to test if in phase 2 it is better to have a phase 1 event with the *codA* gene for negative selection or is it better to omit this and shorten the region for replacement as much as possible. To this end, I cloned a version of pTIIGERone with no *codA* gene and a 35S promoter truncated at the -200 position (pTIIGER(-200)).

A



B



C

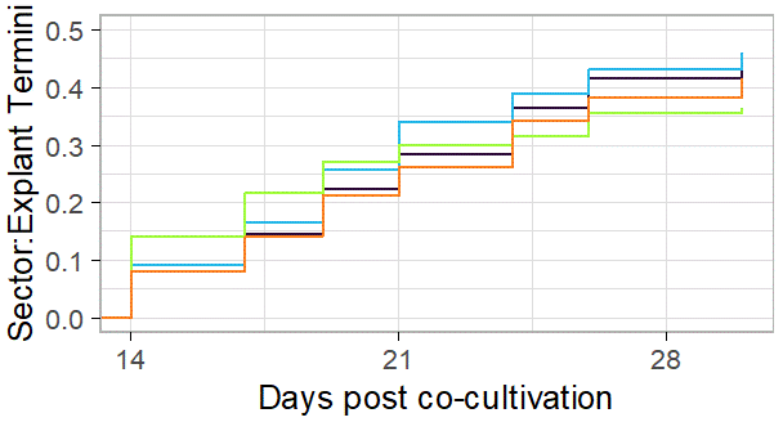


Figure 4.3. Effect of TIIGER cassettes size on efficiency.
A A diagram of the four constructs' TIIGER cassettes with different cassette sizes. The "…" alludes to the rest of the construct, which is not show but matches that of p2cuts+int in Fig. 4.2 B. The cassette size of pTIIGERone and pTIIGER(-200) are not shown but are 6.9 and 4.3 kb, respectively. **B** Trace colours match plasmid name colour. Data is from a single biological repeat

in YD1. **C** Trace colours match plasmid name colour. Data is from a single biological repeat in YD3.

All four constructs yielded purple sectors with good frequency. When transformed into the YD1 line, the constructs with larger payloads proved moderately less efficient than those with shorter cassettes, with pTIIGERtwo yielding roughly half as many sectors as pTIIGERnone (Fig. 4.3 B). However, when the four constructs were transformed into the YD3 line, all four constructs performed the same (Fig. 4.3 C). This inconsistency might be due to YD3's propensity to develop brown spots on its callus, which can interfere with purple sector scoring, or it might reflect a genuine difference between genotypes. Constructs pTIIGERzero, pTIIGERone, and pTIIGERtwo were re-transformed into both YD1 and YD3 in parallel, but the transformations worked poorly, and all vectors yielded very few sectors (data not shown). Thus, while more experiments are necessary, it seems that up to 12.3 kb the TIIGER cassette size reduces targeted replacement frequency moderately or not at all.

Phase 1 T_0 genotyping

To date, two independent T_0 lines have been recovered from pTIIGERzero, four from pTIIGERone, and three from pTIIGERtwo. These lines all carried a T-DNA due to the external NptII gene and the Kanamycin selection employed (Fig. 4.4 A). Genotyping the lines at the *Cai1* locus proved more challenging. I attempted to amplify across the targeted locus using the primer pair YD134 & YD36 which bind outside the homology arms (Fig. 4.4 B). However, the large size of the TIIGER amplicons (4.8-14.1 kb) and the presence of a much smaller WT allele (2.7 kb) necessitate extensive PCR optimisation, which is ongoing. Still, at least one line (pTIIGERzero-III, Fig. 4.4 C) produced a band pattern consistent with a targeted replacement (Fig. 4.4 D). Efforts are underway to further characterise this line and backcross it to WT YD1 to recover T-DNA-free TIIGER lines.

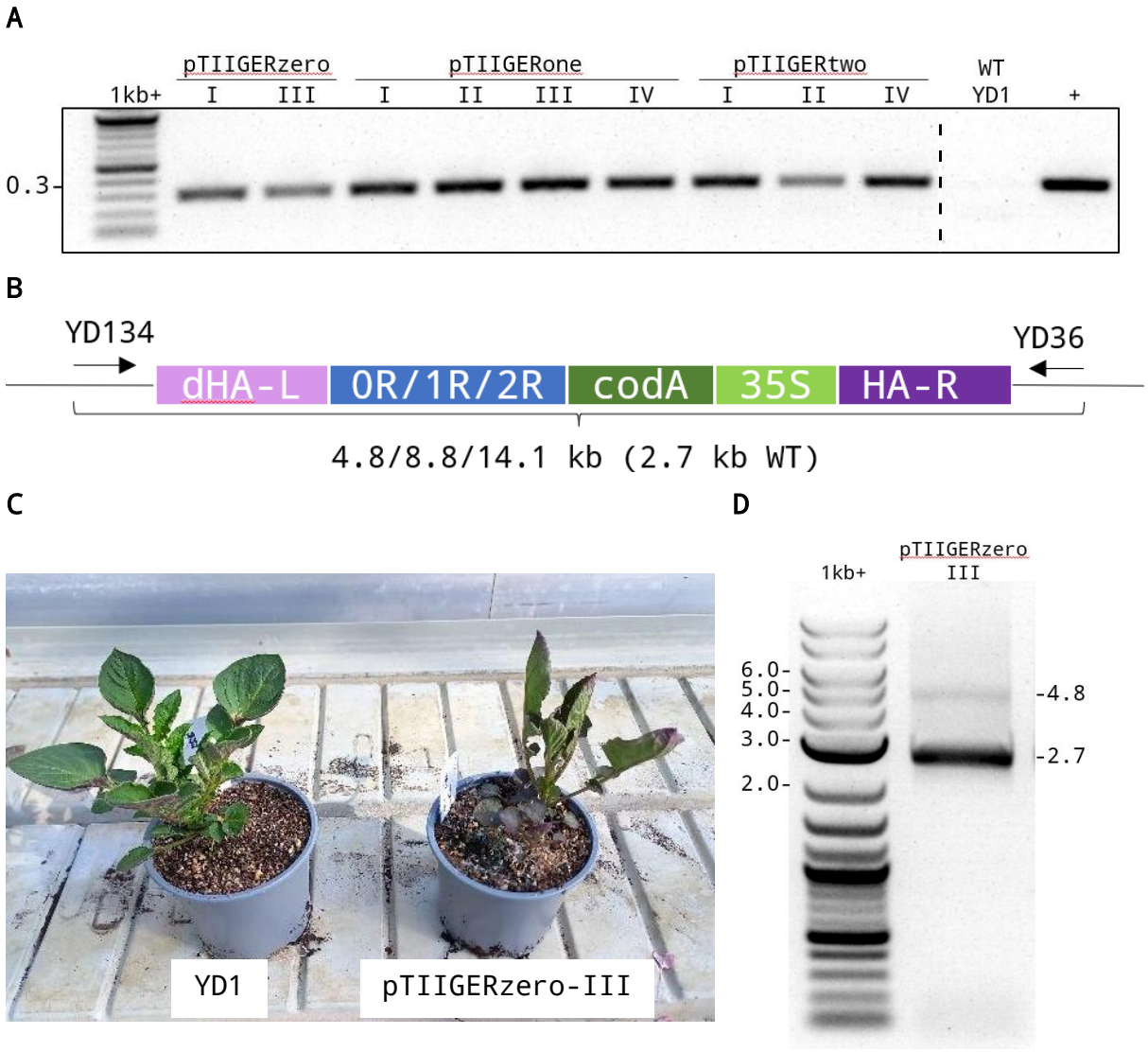


Figure 4.4. Genotyping TIIGER T_0 lines with multiple genes delivered.

A Primers YD227/228 amplify a 300 bp region of the LbCas12a ORF. The positive control (+) was the pTIIGERtwo plasmid.

B Diagram of expected TIIGER loci. The YD134 & YD36 primers should yield a longer PCR product when amplifying a TIIGER allele resulting from pTIIGERzero (4.7 kb), pTIIGERone (8.8 kb), or pTIIGERtwo (14.1 kb), than when amplifying the WT sequence (2.7 kb).

C pTIIGERzero-III's purple phenotype, compared to a WT YD1.

D Genotyping pTIIGERzero-III's yielded the double band pattern expected from a genuine TIIGER event.

4.2.2 GT-specific selection (GTSS)

How to select for a GT event at any locus but not a T-DNA insertion

A major hurdle to establishing GTSS is ensuring the selectable marker confers chemical resistance after targeted integration but not if the T-DNA has inserted at a random location. One approach is to clone a TIIGER cassette with a selection marker like the *bar* gene with a weak promoter (*Nos*) and no terminator. Instead, the *bar* CDS is fused to the partial *Cai1* ORF in the HA-R via the P2A ribosomal skipping sequence (J. H. Kim et al., 2011) (Fig. 4.5 A). As the resulting fusion gene lacks a terminator, it should express poorly (Pérez-González & Caro, 2018). If, however, the cassette is integrated at the *Cai1* locus, the complete *Cai1* ORF and native terminator should support efficient transcription and RNA processing. The P2A sequence will then allow for the translation of *bar* and *Cai1* as separate polypeptides, thus allowing targeted replacements to be selected for with PPT.

One potential issue was that, should the T-DNA integrate upstream of an endogenous terminator, the *bar* gene could be expressed. To avoid this, I cloned the 909 bp-long intron-bearing version of *TurboGFP* gene (Evdokimov et al., 2006) in-frame to the right of HA-R, again with no terminator (Fig. 4.5 B). Notably, the PAM of the crRNA1 protospacer flanking HA-R introduced a STOP codon ca. 384 bp upstream of the intron in *TurboGFP*. I hypothesised the distance between the STOP codon and the last intron (>50 bp) and the overall length of the resulting 3’UTR (>300 bp) will ensure nonsense-mediated decay precluding any T-DNA-based expression (Hori & Watanabe, 2007).

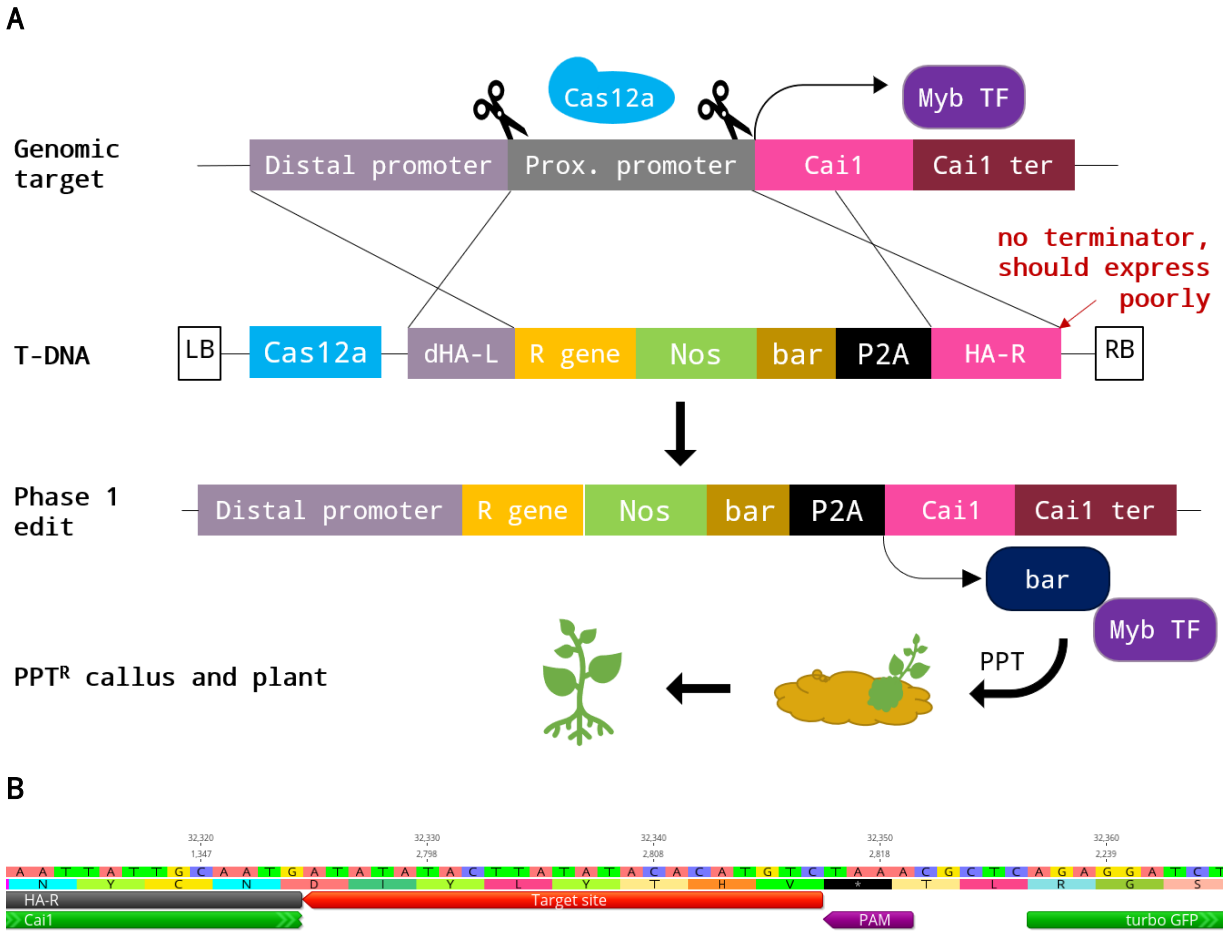


Figure 4.5. GT-specific selection.
A A TIIGER cassette containing a fusion of the *bar* gene and a partial *Cai1* ORF with no terminator which shouldn't express well. If HDR occurs, the cassette is integrated in-frame with the rest of the *Cai1* gene and terminator, which should allow for efficient transcription. The P2A ribosomal skipping sequence should then allow for the translation of *bar* and *Cai1* as independent proteins. *Bar* expression should result in PPT-resistance.
B A *TurboGFP* ORF fused in-frame with the partial *Cai1* ORF in the HA-R. The PAM of the flanking crRNA1 protospacer introduces a STOP codon (black asterisk).

To test if the GTSS selection would work, I cloned three plasmids (Fig. 4.6 A). Two of them followed the cassette design outlined above, either with (pBarGFP) or without (pBarNeg) the *TurboGFP*. The third construct was intended to resemble an integrated TIIGER cassette, featuring a complete *Cai1* ORF and terminator.

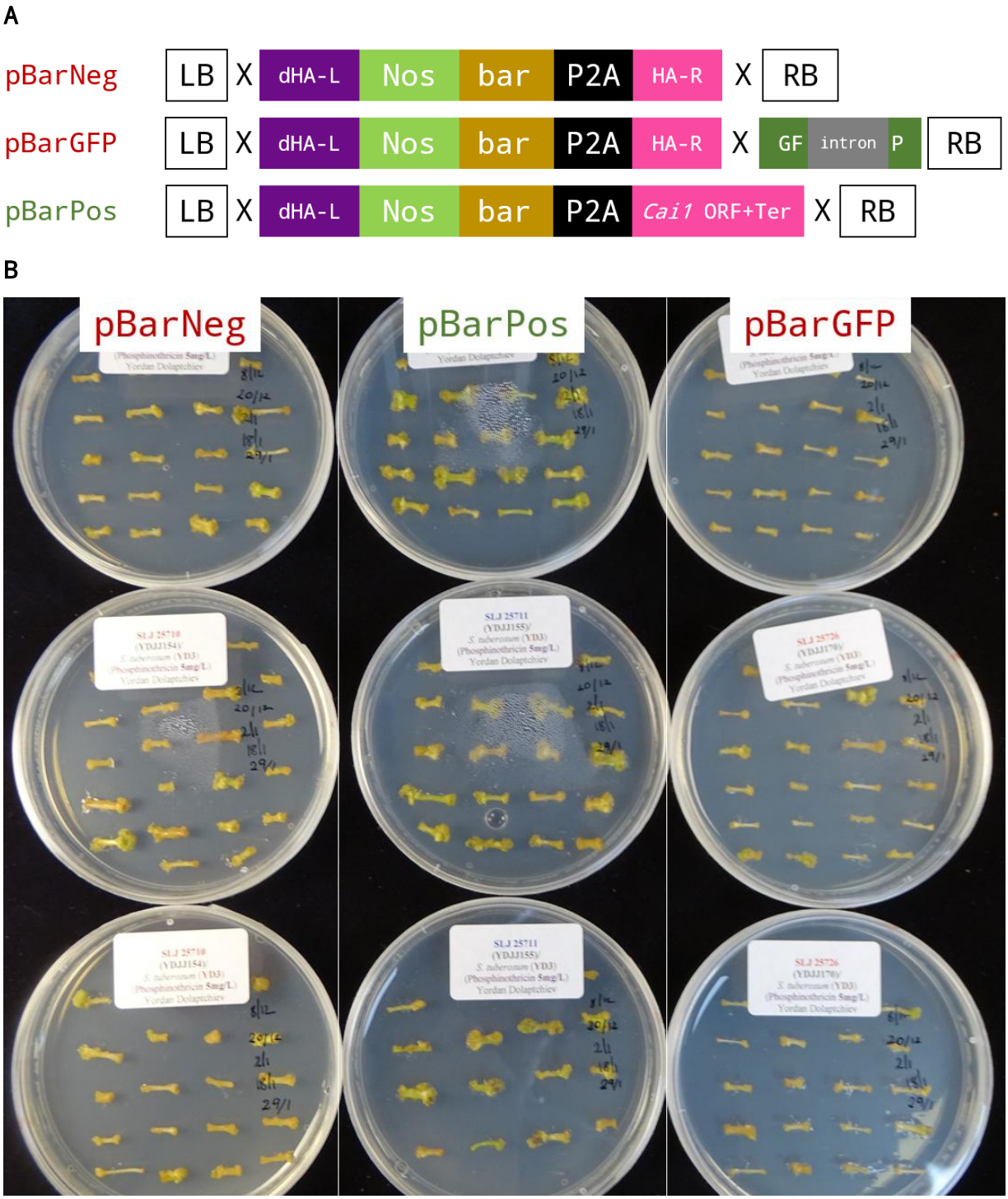


Figure 4.6. GTSS can be stringent.

A Diagram of the three constructs designed to resemble a non-integrated (pBarGFP & pBarNeg) and successfully integrated (pBarPos) GTSS cassettes.

B Image was taken 49 dpcc. pBarPos yielded vigorous callus, while the other two did not. pBarGFP performed particularly poorly. Thus, an integrated cassette should improve the vigour of any KIs on the callus, thus potentially allowing for their selection.

All three were transformed into potato and selected with PPT. They performed very differently, with pBarPos yielding vigorous callus, pBarNeg producing smaller, yellow calli, and pBarGFP yielding almost no calli at all (Fig. 4.6 B). Observed macroscopically under UV light, no GFP signal could be detected in the pBarGFP calli (data not shown). Thus, it seemed the complete *bar-Cai1* fusion could confer effective resistance to PPT, while the incomplete version in the T-DNA could not, especially if the *TurboGFP* gene cassette was included. These encouraging results indicated the GTSS approach is feasible and stringent.

Additionally, purple pigmentation could be seen in the pBarPos calli. It seemed the anthocyanin reporter functions even with a moderately strong promoter like Nos; such milder overexpression could result in purple tissue with superior vigour.

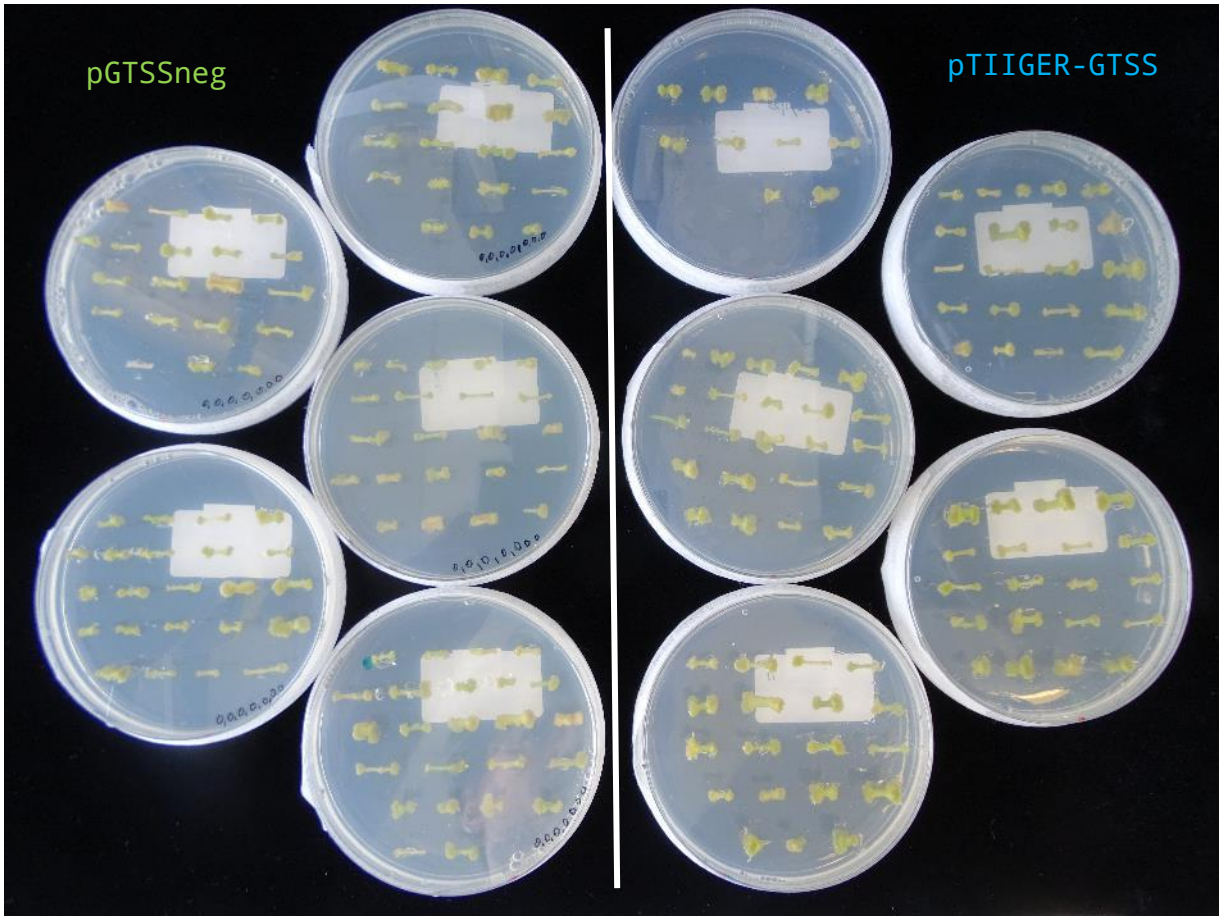
Using GTSS to recover TIIGER phase 1 events

Next, I tried to recover TIIGER events using GTSS (Fig. 4.7 A). To do this, I cloned an intronized ttLbCas12a nuclease and a pair of crRNAs flanking the proximal *Cai1* promoter alongside the cassette used in pBarGFP (pTIIGER-GTSS). As a negative control, I cloned a version of pTIIGER-GTSS with a pair of crRNAs without compatible protospacers in the T-DNA of the Solyntus genome (pGTSSneg).

These were transformed into YD1. The calli grew slowly, something typical for selection with PPT, but at the time of writing (28 dpcc), the calli transformed with pTIIGER-GTSS grew better than those transformed with pGTSSneg, although the difference was subtle (Fig. 4.7 B). The pTIIGER-GTSS calli also continued to produce purple sectors (Fig. 4.7 C) but these were less intensely pigmented and remained small. This indicates that further improvements to GTSS are necessary to improve stringency and reduce the level of *Cai1* overexpression.

A

B



C

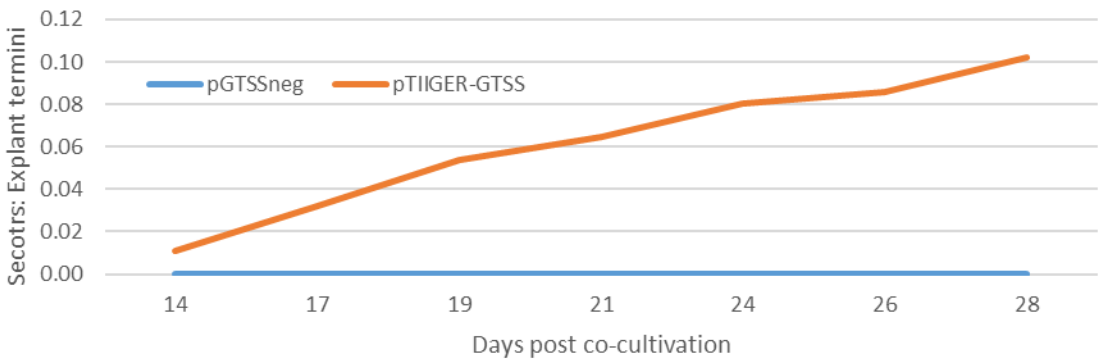


Figure 4.7. TIIGER events recovered with GTSS.
A Phase 1 TIIGER constructs with the bar-GTSS system. pTIIGER-GTSS carries a pair of crRNAs flanking the proximal *Cai1* promoter. pGTSSneg is a negative control with a pair of non-targeting

crRNAs. **B** The calli at 28 dpcc. **C** Calli transformed with pTIIGER-GTSS unexpectedly still yielded purple sectors.

4.2.3 Targeted Replacement + Cre

The efficient targeted replacements demonstrated above have applications beyond the TIIGER method. In this section, a project is presented in which targeted replacements are used to introduce recognition sites for the Cre-Lox system. This system consists of the Cre recombinase and 34 bp LoxP sites which are efficiently recombined, resulting in insertions, deletions, or inversions (Gilbertson, 2003). The targeted introduction of such short recognition sites, followed by recombinase-mediated DNA insertion, has previously been used to insert gene into plant genomes at pre-defined locations (C. Sun et al., 2023). While this approach invariably leaves scar sequence at the insertion site, it tends to be very efficient. Such recombinase-mediated gene insertion might provide an alternative to classic transgenesis, resulting in more consistent expression due to the insert's consistent genomic context. The high efficiency could also allow for the simultaneous introducing of very large cassettes containing numerous genes. To this end, I attempted to combine the Cre-Lox system, anthocyanin reporter, and targeted replacements into an efficient platform for gene insertion.

Introducing a LoxP acceptor locus

A classic Cre-Lox-based strategy for gene delivery is recombinase-mediated cassette exchange (RMCE), which relies on the use of heterospecific LoxP sites. These have been mutated to recombine with some loxP allelic variants but not others (e.g. LoxP511; Hoess et al., 1986). If two loci carry gene cassettes flanked by the same pair of incompatible heterospecific LoxP sites, Cre can exchange the cassettes (Lyznik et al., 2003). While this process is generally reversible, the reverse reaction can be inhibited by using the LoxP66 & LoxP71 pair of recognition sites, which recombine to form an inactive site (LoxP72) (H. Albert et al., 1995).

I designed an implementation of RMCE at the *Cai1* locus (Fig. 4.8 A). First, a targeted replacement of the proximal promoter would

introduce the 35S promoter flanked by the LoxP511 and LoxP66 sites. An event with this LoxP locus would then be super-transformed with a construct expressing the recombinase and bearing a LoxP cassette containing an R-gene and the proximal *Cai1* promoter, flanked by the LoxP511 and LoxP71 sites. If RMCE occurs, then the R-gene and proximal promoter land in front of the *Cai1* gene, removing the 35S promoter. As the 35S promoter would be lost, the RMCE products would be selected by loss of pigmentation and a gain in vigour.

Before I could attempt the RMCE process, I had to introduce the necessary LoxP sites upstream of Cai1. I cloned a cassette with the 35S promoter flanked by LoxP511 and LoxP66. The recognition sites were cloned in a left-facing orientation to avoid introducing premature START codons in front of *Cai1*. Another version was cloned with the full *Cai1* ORF and native terminator to simulate a successful integration. A transformation of both in parallel confirmed that the successful integration resulted in purple pigmentation, while the cassette alone did not (data not shown). The cassette was then included in a GT construct, featuring the intron-bearing Cas12a and crRNAs flanking the proximal promoter (pCre1) (Fig. 4.8 B). pCre1 was transformed into YD1, yielding purple sectors at a high frequency (Fig. 4.8 C).

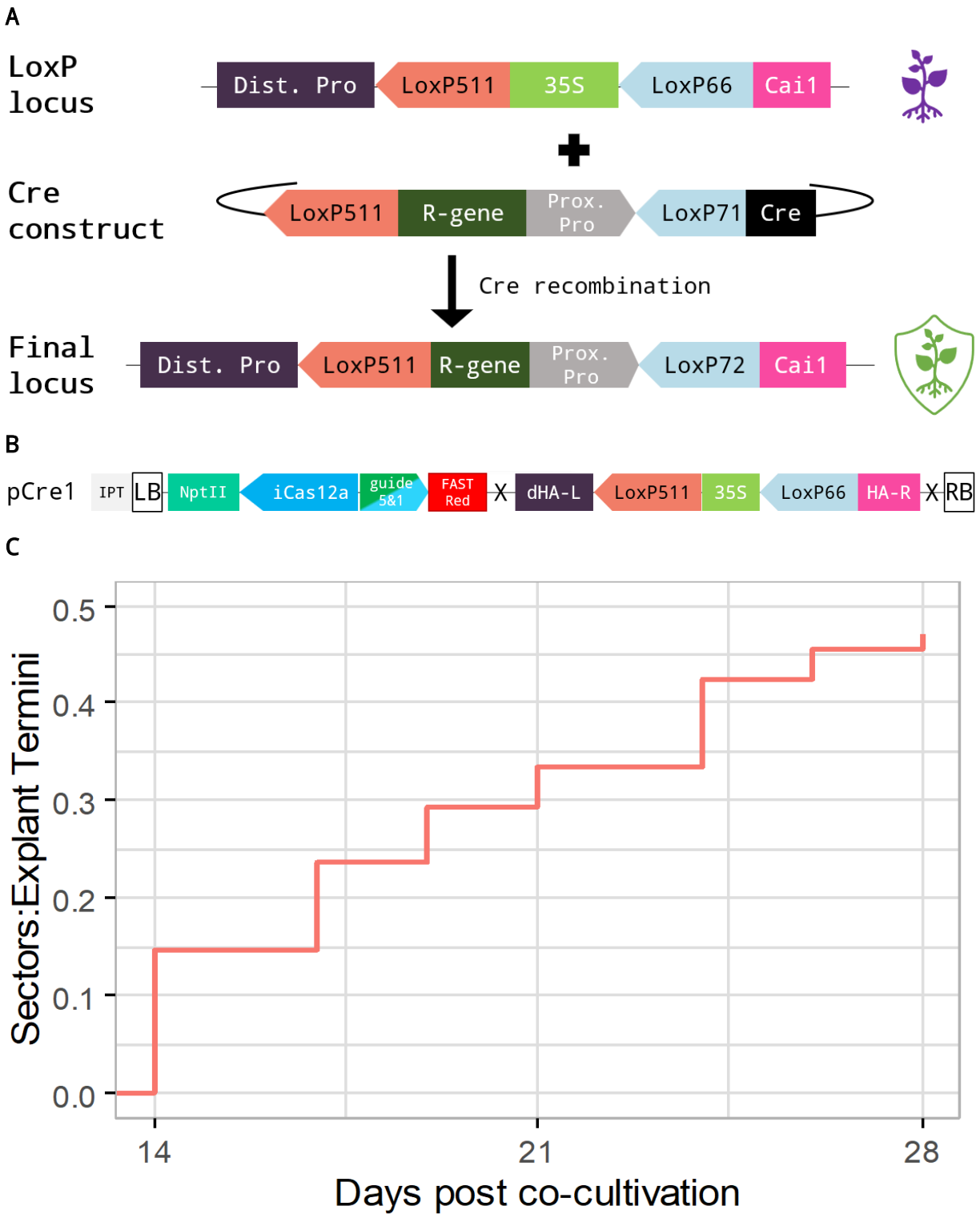


Figure 4.8. Introducing a LoxP acceptor locus.
A Combining of the recombinase-mediated cassette exchange (RMCE) and the anthocyanin reporter. If a LoxP551/LoxP66-flanked cassette with the 35S promoter were introduced in front of the *Cai1* gene via targeted gene replacement, subsequent RMCE events in that line can be selected for via loss of pigment and gain of vigour. The LoxP sites are in the reverse orientation to ensure no ATG start codons are erroneously introduced; both LoxP66 and the LoxP72

resulting from the RMCE process lack such codons. **B** A construct for the replacement of the *Cai1* proximal promoter with the LoxP cassette. **C** The emergence of purple sectors following transformation of YD1 with pCre1. The experiment was done in two biological repeats with similar results.

I recovered three T_0 lines which exhibited the pigmented and reduced vigour phenotype typical of the anthocyanin reporter. I genotyped these with the YD134 & YD36 primers, as done previously (Fig. 4.9 A). In contrast to previous KI and TIIGER events, here the whole-insert amplification of the TIIGER locus resulted in a smaller amplicon (2.5 kb) than the WT allele did (2.7 kb). Of the three lines recovered, only pCre1-III produced the expected 2.5 kb band (Fig. 4.9 B). This was amplified again at scale (Fig. 4.9 C), cloned and sent for sequencing, which confirmed the accurate integration of the 35S-LoxP acceptor locus (Fig. 4.9 D). The three lines were also confirmed to carry a T-DNA (Fig. 4.9 E); efforts are ongoing to backcross pCre1-III to WT YD1 and recover T-DNA-free lines.

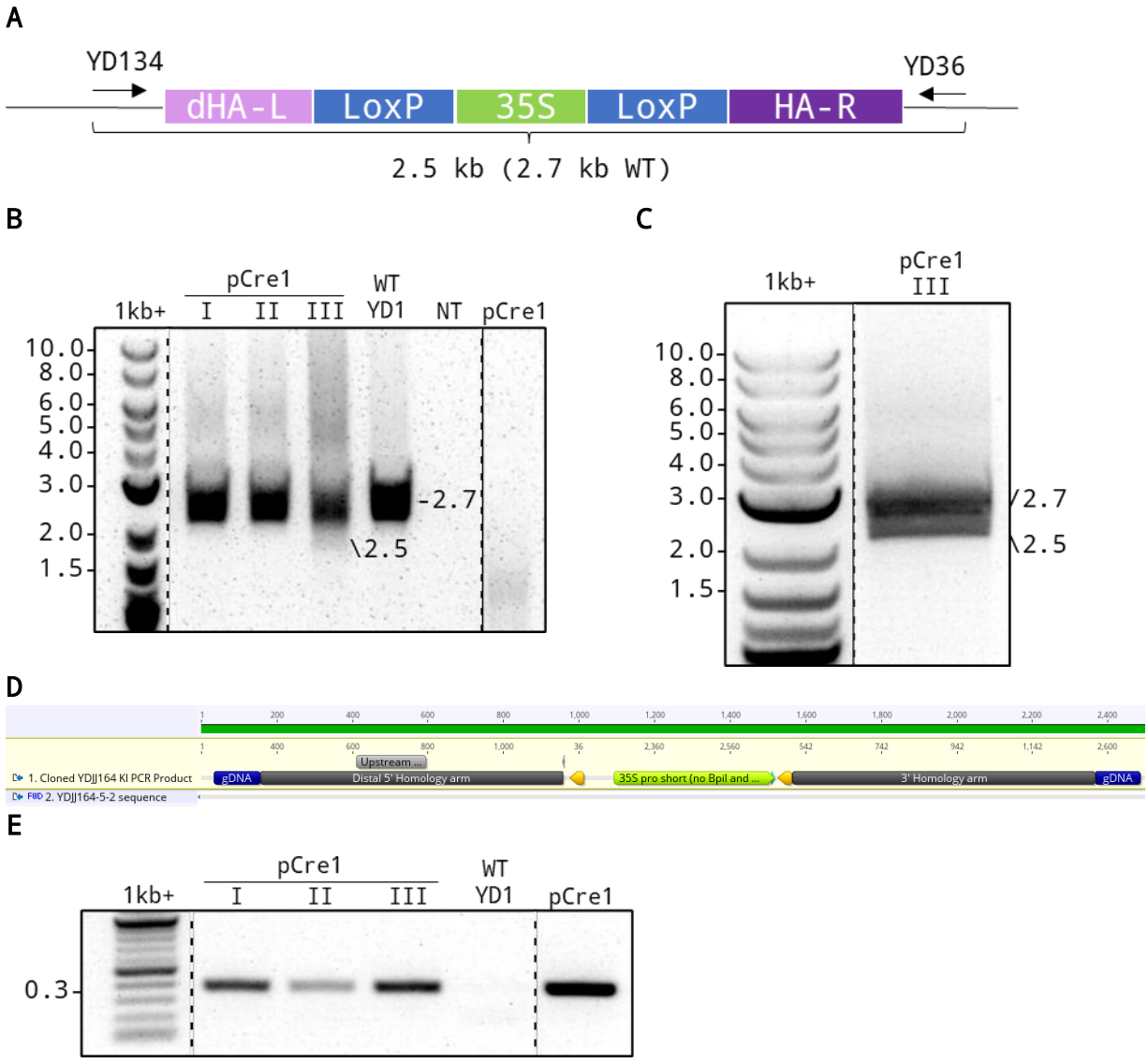


Figure 4.9. LoxP line genotyping.
A Diagram of expected LoxP TIIGER locus. The YD134 & YD36 primers should yield a shorter amplicon (2.5 kb) than the WT sequence (2.7 kb). **B** Lines pCre1-I and pCre1-II yielded only the WT amplicon, while pCre1-III produced two bands. Unfortunately, the gel was cast using a wide comb, which made distinguishing individual bands difficult. **C** A clearer gel showing the two amplicons produced by pCre1-III. **D** Sequencing results showing a perfect match between the cloned 2.5 kb band and the predicted TIIGER locus. **E** All three pCre1 T_0 lines carry a T-DNA.

Next, I designed a construct to conduct the RMCE process (Fig. 4.10 pCre2). This carries the Cre recombinase and an RMCE cassette containing R-genes and the original proximal promoter, flanked by the LoxP511 and LoxP71 recognition sites. Initially, I will attempt RMCE via stable transformation using the self-excising Cre

approach published by Mlynárová & Nap (2003). To this end, the recombinase is flanked by heterospecific LoxP2272 sites (Lee & Saito, 1998), which are in turn nestled within a disrupted *bar* gene. This arrangement should result in the Cre efficiently excising itself, as well as recombining LoxP sites *in trans*, and such events should be selected for with PPT. A FAST-Red cassette is included at the far-right on the T-DNA to facilitate the removal of integrated T-DNAs via breeding. The cloning of pCre2 is ongoing but, should RMCE prove achievable, the ultimately aim to accomplish RMCE through transient expression. This might be achieved with a version of pCre2 without the disrupted *bar* gene, LoxP2272 sites, and an IPT gene in place of the FAST-Red marker.

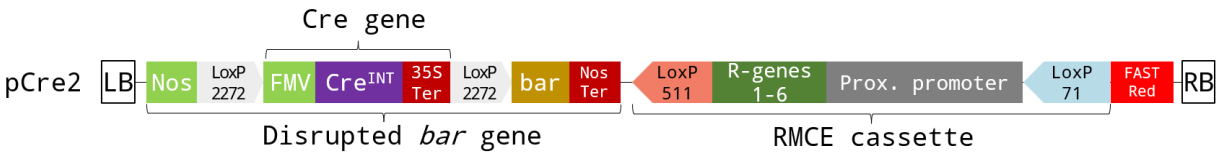


Figure 4.10. Cre-expressing construct for RMCE process. Alongside the RMCE cassette, pCre2 contains a *bar* gene disrupted by an intronized Cre recombinase gene flanked by heterospecific LoxP2272 sites. Mlynárová & Nap (2003) report that the Cre should self-excise, recombining other LoxP sites *in trans*, and that such events can be select for using the now reconstituted *bar* gene. The RMCE cassette contains an R-gene and the proximal *Cai1* promoter removed by pCre1.

4.3 Discussion

In this chapter, I aimed to develop the TIIGER method to deliver R-genes against *P. infestans* into diploid potato germplasm. The biphasic method was designed to ultimately yield targeted insertions without any selection markers or scar sequence present. Substantial progress was made towards establishing TIIGER, with optimisations allowing for recovery rates of extensive targeted replacements similar to KI rates seen in the previous chapter. Additionally, the GTSS system was developed, which could greatly facilitate TIIGER and GT in general. Lastly, a side-project combining targeted gene replacements and RMCE was proposed, and

significant progress was made towards it. However, outstanding questions remain for all these research avenues, as is discussed below.

4.3.1 TIIGER phase 1 works efficiently

The TIIGER approach is predicated on conducting efficient and successive targeted replacements at endogenous plant loci. To do this, I modified the gene KI method used in the previous chapter. I changed the left homology arm to correspond to a distal *Cai1* promoter region of the same length (ca. 800 bp) as the proximal. It is currently unclear how the length of the proximal region, which gets removed in case of HDR, might affect the efficiency of targeted replacements.

The efficiency of replacing the 829 bp proximal *Cai1* region was high, provided it was cleaved at both ends rather than once in the middle. An intuitive explanation for this is that a single cut leaves behind regions of non-homology between the targeted locus and template which disrupt the subsequent HDR process. Additionally, it might be that having more than one DSB at the locus is itself conducive to HDR. A major limitation of NHEJ and MMEJ is that, should the wrong ends be rejoined, chromosomal aberrations can occur. It could therefore be hypothesised that, if multiple DSBs cooccur, alternative pathways like HDR might be upregulated in some manner. Such a hypothetical effect might explain why targeting the T-DNA on either side of the KI cassette increases GT rates – a phenomenon which currently lacks a firm mechanistic explanation (Fauser et al., 2012). Furthermore, if such an effect did exist, then targeting the proximal promoter region at three points simultaneously (using crRNA1, crRNA3, and crRNA5) might result in a further increase in the rates of targeted replacement.

The effect of the TIIGER cassette's size on purple sector frequency was investigated and found to be surprisingly limited. The targeted replacement of a 2.9 and a 12.3 kb cassette seemed to occur with equal frequency in YD3, while in YD1 the latter was roughly half as efficient as the former. More experimentation is required to probe this discrepancy, but the TIIGER process might be capable of efficiently integrating even large, multi-gene

cassettes. Thus far, only one line has been confirmed to carry a pTIIGERzero event, likely due to the TIIGER cassette's small size and relative ease of genotyping. Efforts are ongoing to characterise lines that should carrying one or two R-genes at the TIIGER locus.

In chapter 3, I distinguished "one-step" and "two-step" KI events; for one step events, the T-DNA does not insert, and limited current data suggest 10-20% of events may be one-step provided the T-DNA is not selected for. If TIIGER phase 1 events can be recovered that arise from a one-step event, there will be no need to segregate away the T-DNA, and the line with the phase 1 event can be used directly for phase 2. This could greatly accelerate the process and guarantee a consistent level of plant material transformability between phases. Efforts to conduct phase 1 of TIIGER transiently, in the absence of Kanamycin selection, are ongoing.

Notably, phase 2 of TIIGER has yet to be demonstrated, as is the associated selection against the *codA*-35S cassette and restoration of the WT phenotype. As targeted replacements are achievable in principle, the focus shifts to whether such events can be efficiently selected for. The *codA* negative selection marker has a long history of use in plans (Stougaard, 1993), including potatoes (Kondrák et al., 2006). Furthermore, preliminary experiments in YD1 have also yielded encouraging results (data not shown). The loss of the 35S promoter is also certain to result in the loss of the purple phenotype. Crucially, both 5-FC tolerance and loss of pigmentation are recessive phenotypes, which necessitates the TIIGER phase 2 be conducted in a heterozygous line for the phase 1 event.

Another potential complication is that a simple deletion of the *codA*-35S cassette is likely to yield the same phenotype as a TIIGER phase 2 targeted replacement. This is likely to result in some background of deletions among recovered plants. If this is found to be a substantial issue, it might be necessary to target the *codA*-35S cassette just once at the end of the *codA*'s terminator, rather than once on each side. While this is likely to

reduce the frequency of targeted replacements, the single DSB is less likely to result in the deletion of the *codA*-35S cassette.

The potential issues with selecting phase 2 events also motivated the cloning of the pTIIGER(-200) construct, which lacks the *codA* gene and carries a truncated 35S. If the length of the region being removed does affect TIIGER efficiency, then substituting such a minimal cassette might be more efficient.

4.3.2. GTSS

I attempted to establish a selection system where cassette integration at a predefined locus results in cassette-encoded chemical resistance. The major challenge was to stringently select for the cassette after but not prior to integration. In GTSS, this was attempted by cloning a selectable marker lacking a terminator, which only becomes part of a complete gene cassette following GT. The selection was then further improved by the addition of an intron-bearing *TurboGFP* gene downstream of the HA-R.

The first implementation of GTSS featured the *bar* gene and was used to try and select for TIIGER phase 1 events. While the TIIGER construct produced more vigorous calli than the control, it still yielded purple sectors, and these were still small compared to the calli. It seems this iteration of GTSS might alleviate the need to continuously cut back green tissue from the selected purple calli, but it does not dispense with it. Ultimately, further work is necessary to obtain a GTSS marker with higher stringency and a lessened impact on vigour. Both issues could be partially addressed by substituting the *bar* gene for *NptII*. Using a milder promoter might also help improve callus vigour.

More comprehensive changes to the GTSS approach are also being considered. Two features of the GTSS approach were borne out of the context of the *Cai1* KI work. Firstly, cassettes have thus far been targeted upstream of *Cai1*, and this is reflected in the GTSS marker lacking a terminator. Alternatively, cassettes could be targeted downstream of *Cai1* and GTSS markers could instead lack a promoter. This might even result in more stringent selection, but it would make the system reliant on the endogenous gene’s promoter. Secondly, *Cai1* has been the targeted locus thus far, but GTSS can theoretically select for GT events in front of any

endogenous ORF and the TIIGER and GTSS approaches have no inherent link to it.

Conversely, the *StALS1* gene might be a better target (Fig. 4.11). In such an implementation, GTSS-bar could be used to select for an initial TIIGER phase 1 event, selecting with PPT. In a subsequent intermediate TIIGER phase, the bar-P2A cassette could be replaced with a TIIGER cassette containing additional genes and a GTSS-nptII selection module, selecting with Kanamycin. An arbitrary number of intermediate TIIGER phases can follow, alternating between GTSS-bar and GTSS-nptII modules and adding and removing genes as needed. Notably, such intermediate phases can be conducted in either homo- or heterozygous lines, as the GTSS phenotype is dominant. Whenever a “clean” version of the TIIGER line is desired, a final TIIGER phase 2 can be conducted. During this, the GTSS module is replaced with the original *StALS1* promoter, while introducing the P186A mutation on a partial ALS fragment into *StALS1* via the HA-R. This confers chlorsulfuron resistance and should be easily selected for. Thus, by targeting the *StALS1* locus, it might be possible to avoid the predicted issues with recovering the final TIIGER phase 2 event at *Cai1*. Such a system, if successfully implemented, would grant the operator the ability to deploy genes with great versatility and accuracy, while having the option to convert any iteration into a product compatible with the Precision Breeding Act (Genetic Technology (Precision Breeding) Act, 2023).

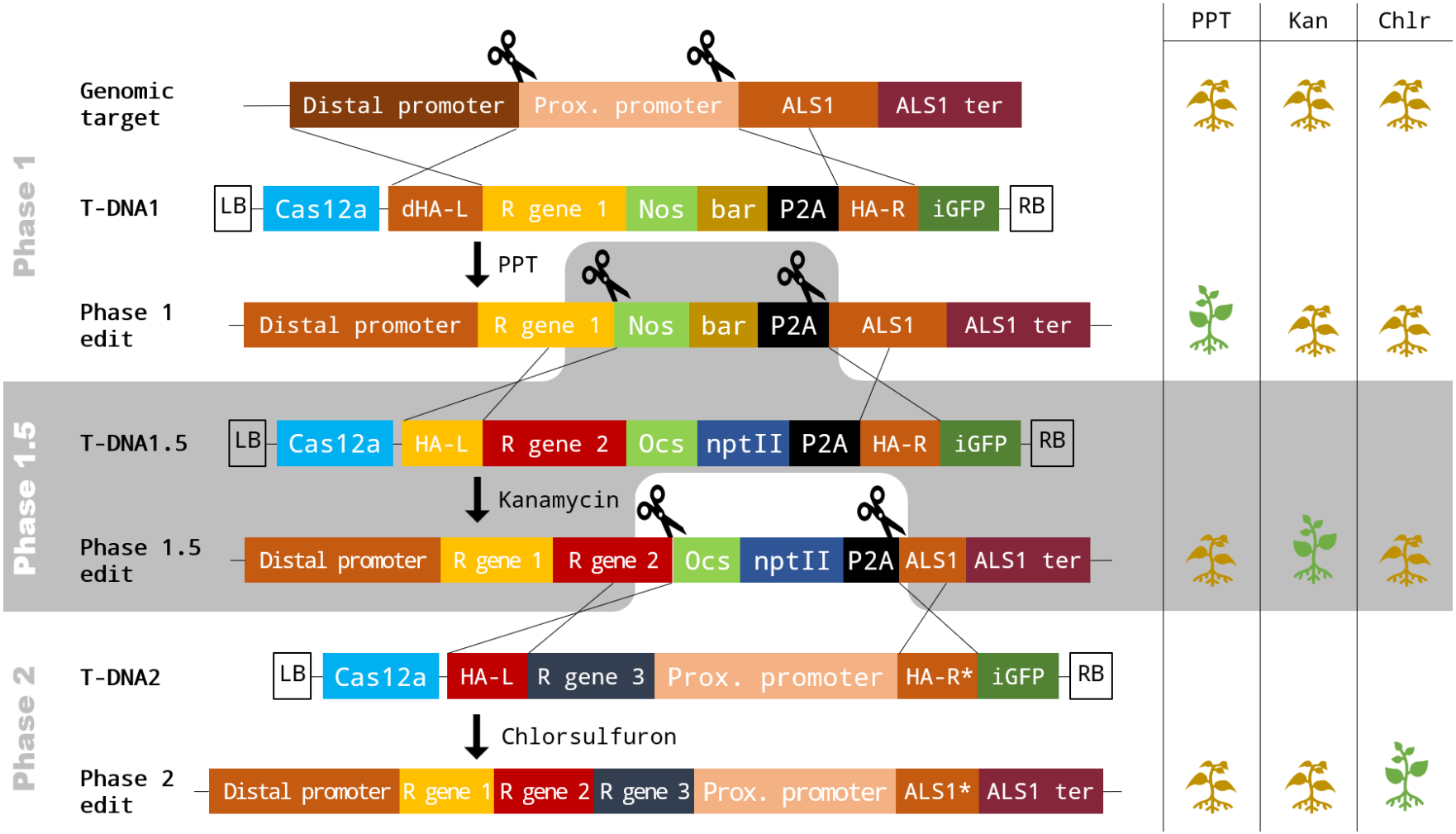


Figure 4.11. TIIGER-GTSS at the *StALS1* locus.

Chlr – chlorsulfuron. Asterisk behind ALS1 and HA-R signified they carry the P186A mutation.

4.3.3. Targeted replacement + Cre

This project aims to combine the anthocyanin reporter and the Cre-Lox recombinase system to efficiently deliver genes and select for them thereafter. In Section 4.2.3 I described the introduction of a LoxP acceptor cassette at the *Cai1* locus and the design of an RMCE construct to make use of it. The super-transformation of the pCre1-III line with pCre2 is still outstanding.

The self-excising Cre approach was chosen because it should allow for the efficient recovery of RMCE events. However, the use of chemical selection results in stable pCre2 integration somewhere in the genome. Ultimately, it would be preferable to conduct RMCE during transient expression, thus avoiding the need for later breeding. Such a construct wouldn't feature the *bar* gene and LoxP2272 sites flanking the recombinase, and FAST-Red cassette could be substituted for the *IPT* gene.

Mlynárová & Nap (2003) used a version of the recombinase featuring an intron for better expression. I am also including this, as well as hemagglutinin (HA) tags on the Cre and one of the R-genes delivered via RMCE. This could then be used to compare the variability of protein expression levels of RMCE-delivered genes versus genes delivered via transgenesis (Cre) across stably transformed lines.

CHAPTER 5. ENGINEERING COMPACT GROWTH IN POTATO

5.1 Introduction

As discussed in the General Introduction, the lack of genetic gain in potato varieties contrasts with the progress made in crops like maize over the last century. This improvement has mainly come from breeding for tolerance to high planting densities, which can reach up to 90,000 plants per hectare today (Djaman et al., 2022; Du et al., 2024; Duvick, 2001). To make maize plants tolerant, breeders have optimised the plants' architecture for growth at high densities by modulating aspects like leaf angle (X. Wang et al., 2022), the shade-avoidance response (Wei et al., 2018), and reduced plant height (Paciorek et al., 2022; Shang et al., 2020). Dwarfed varieties have proven to be particularly agronomically valuable in several crops and have seen wide adoption across since the Green Revolution. None of today's widely grown potato varieties are dwarfed but there is good reason to expect the crop would benefit from more compact growth.

Potato growth and productivity

The link between potatoes' canopy structure and productivity has been extensively studied. An inherent challenge is the complexity of the canopy – potatoes tend to have a bushy, sprawling growth habit, and measuring and quantifying this can be difficult. A widely used metric is the leaf area index (LAI) — the ratio of leaf area to the ground area it covers (Watson, 1952). Researchers consistently found that as canopies developed, so too did their ability to intercept photosynthetically active radiation (PAR, a measure closely correlated with plant total dry weight), until an LAI of approximately 3 (Fig. 5.1) (Burstall & Harris, 1983; Firman & Allen, 1989; Gordon et al., 1997; Khurana & McLaren, 1982). Beyond this, PAR interception quickly plateaued, and additional leaf growth did not contribute to yield. This is due to mutual shading of the leaves, which can ultimately transition from carbon sources into carbon sinks.

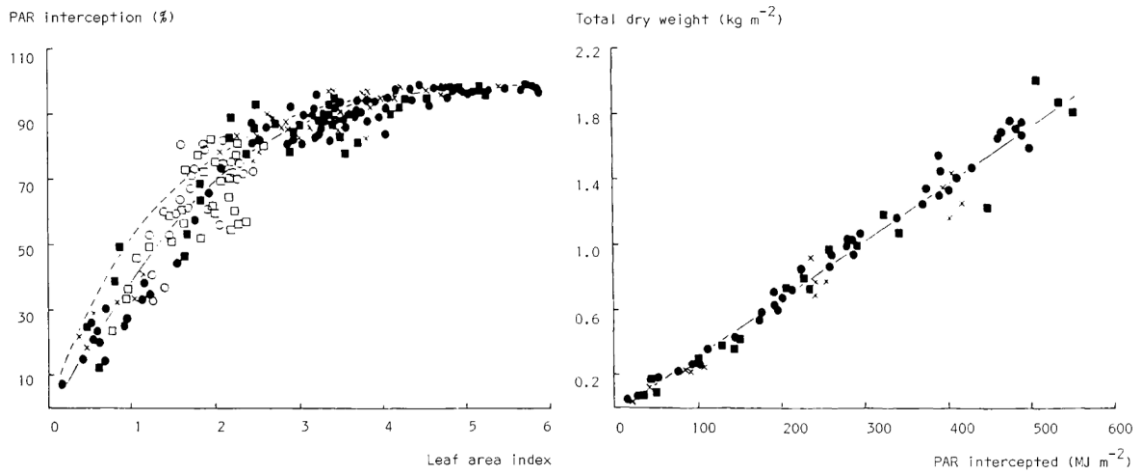


Figure 5.1. Relationship between PAR interception & LAI (left) and total dry weight & PAR interception (right) in potato.
An increase in leaf mass, as measured by the leaf area index (LAI), results in more photosynthetically active radiation (PAR) interception to a point (LAI ≈ 3), after which it plateaus. Potato total dry weight is directly proportional to the amount of PAR intercepted. Source: Khurana & McLaren (1982).

These findings also allowed some to postulate an ideal pattern of growth for potato: "one […] with the attainment of an LAI of […] 3 shortly after initiation, this leaf area being maintained (but not increased) for as long as possible" (Ivins & Bremner, 1965). This proposition differs substantially from the growth patterns of popular potato varieties, whose canopy typically peaks between at a LAI of 4.5-6.5 (Boyd et al., 2002; Burstall & Harris, 1983; Jefferies & MacKerron, 1993; Khurana & McLaren, 1982). This discrepancy implies that much of the foliage in the field might not contribute significantly to the final yield. Instead, it might hinder crop performance by reducing canopy ventilation and lowering the harvest index[1], thus constituting a potential major inefficiency in potato cultivation. Also, because higher planting densities lead to a higher peak LAI, potatoes' suboptimal canopy places a limit on planting density and the maximum possible yield per unit area (Allen, 1978). If a more compact growth habit was engineered into potato, this could conceivably lessen the effects

[1] Harvest index – the ratio of a crop's dry matter yield to its total dry biomass. A measure of productive efficiency.

of mutual shading, enable more efficient growth at higher planting densities, and ultimately increase yields per unit area.

Dwarfism alleles – nature and mechanism

In multiple crops, breeders have used dwarfism alleles to introduce short stature into elite lines. These generally act through the signalling axis of Gibberellins (GAs), GA receptors, and DELLA proteins, and this is the main mechanism plants use to control growth (for a thorough review of this topic, see T. P. Sun (2011)). GAs are a diverse class of phytohormones in higher plants and function as positive regulators of growth (Hedden & Sponsel, 2015). Bioactive GAs like GA_3 achieve this by inducing the proteasome-mediated degradation of DELLA proteins, which are negative growth regulators (Fu et al., 2002). This requires the formation of a tripartite complex involving the gibberellin, a GA receptor, and a DELLA protein (Murase et al., 2008). Crucial to this interaction is the DELLA protein's DELLA motif, consisting of Asp-Glu-Leu-Leu-Ala residues ("DELLA" in single letter code) and other nearby residues which form the binding site for the GA receptor. Importantly, if these residues are mutated or deleted, this can disrupt the DELLA protein's degradation and thus the disinhibition of growth by GAs. Several key dwarfism alleles are dominant mutant alleles of DELLA genes (Q. Liu et al., 2021). For example, the wheat *Reduced height 1* gene codes for a DELLA protein and in both widely-used dwarf alleles (*Rht1-B1a* and *Rht1-D1b*) the N-terminus containing the DELLA motif has been truncated, resulting in a non-degradable DELLA protein and a GA-insensitive dominant dwarfed mutant (Peng et al., 1999). The barley orthologue called *Slender1 (Sln1)* has a dominant dwarf allele (*Sln1d*) where a substitution near the DELLA motif (…DELLAAL**G**… → …DELLAAL**E**…) results in a plant with short stature and reduced GA sensitivity (Chandler et al., 2002). Notably, these alleles are gain-of-function mutants, while loss-of-function mutants like KOs of DELLA genes are usually recessive and result in constitutively disinhibited, highly elongated growth and male sterility (Chandler et al., 2002; Ikeda et al., 2001; Tomlinson et al., 2019).

An analogous dwarfism allele was engineered by Tomlinson et al. (2019) in tomato. To do this, the authors targeted the tomato

DELLA gene *GA insensitive* (*GAI*), which is an orthologue of *GAI* in *Arabidopsis* and of *Rht1* and *Sln1*, with CRISPR/Cas9 using a sgRNA which cuts immediately downstream of the DELLA motif (Fig. 5.2 A). They obtained several highly elongated, loss-of-function mutants, as well as a dominant gain-of-function mild dwarfism allele. This was called *PRO*^{D} and shown to be the result of the deletion of an Alanine from the targeted motif (…DELL**A**VLG… → …DELLVLG…) (Fig. 5.2 B). The plants carrying *PRO*^{D} were 14-30% shorter than the WT in the glasshouse, exhibited reduced GA sensitivity, and yielded slightly smaller fruit. This research demonstrated that the mechanism which confers dwarfism in monocot crops can also be exploited in a dicot Solanaceous species, thus providing a template for engineering dwarfism into potato.

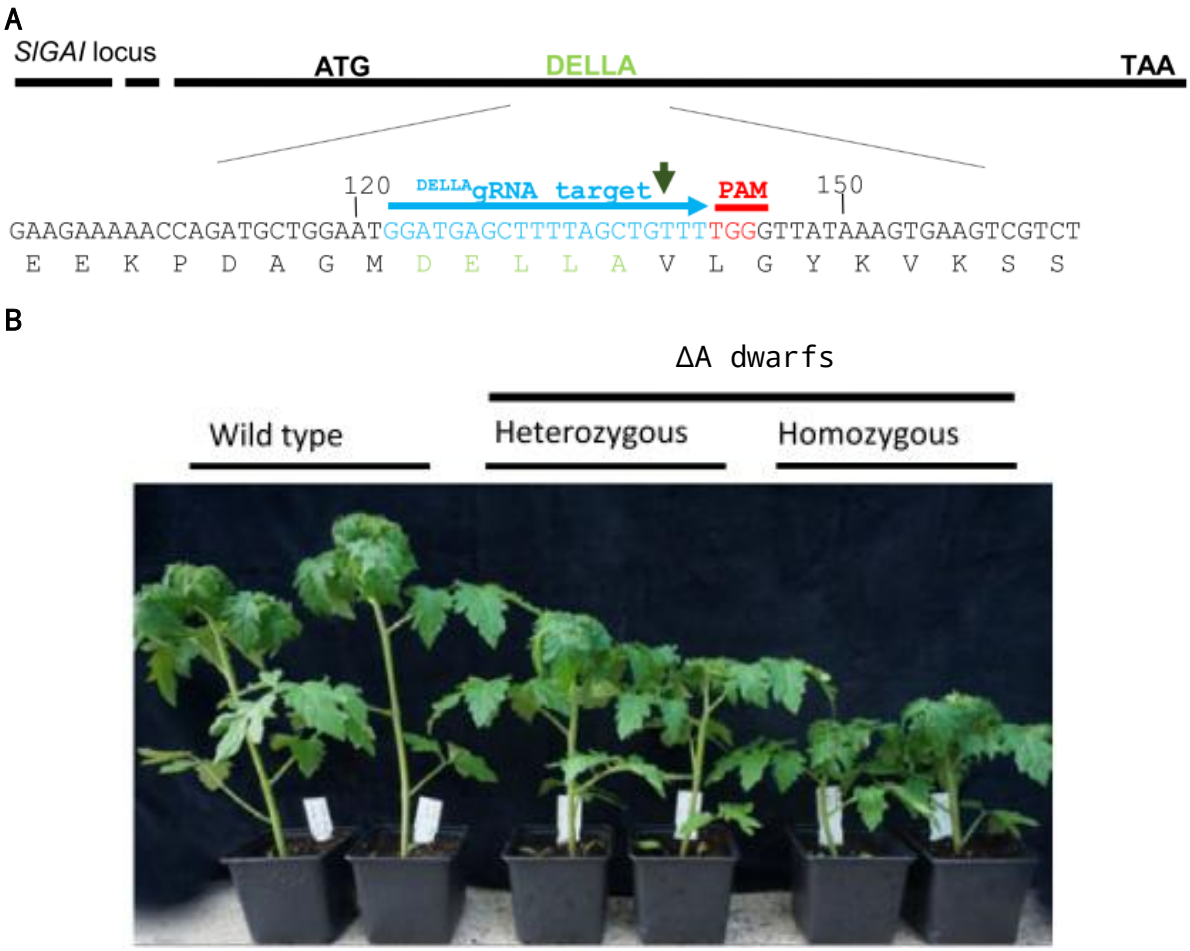


Figure 5.2. DELLA dwarf tomatoes.
A Targeting the tomato *GAI* locus with CRISPR/Cas9. **B** The glasshouse phenotype of plants homo- and heterozygous for the ΔA dwarfism allele. Modified from: Tomlinson et al. (2019).

Notably, dwarfed stature has previously been engineered into potato using a different mechanism. Carrera et al. (2000) knocked down the expression of the potato GA biosynthetic gene *StGA20ox1* in the wild landrace *Solanum tuberosum ssp. Andigena* by expressing an antisense version of the gene's coding sequence. This resulted in plants that had lower GA levels, were 37–58% shorter than WT, set tubers faster, and yielded more (Fig. 5.3). Conversely, lines expressing a sense version of the *StGA20ox1* coding sequence had elongated stems, set tubers later and had lower yields. Analogous mutations underpin agronomically-viable dwarfism alleles in rice (Spielmeyer et al., 2002) and maize (Paciorek et al., 2022). While the results of Carrera et al. (2000) were encouraging, their work was done in a photoperiod-sensitive potato landrace, which only sets tubers and flowers in short day conditions (8 hours light, 16 hours of dark). GAs are known to be major inhibitors of this transition, which suggests a direct causal link between their lowered levels in antisense lines and the associated early tuberization and elevated yields (Jackson & Prat, 1996; X. Xu et al., 1998). Modern commercial varieties have been bred to tuberize independent of photoperiod (Ewing & Struik, 1992), so it is unclear if they would benefit from an analogous gene edit. To date, this is the only example of a dwarfing allele being engineered into potato involving the GA-signalling pathway.

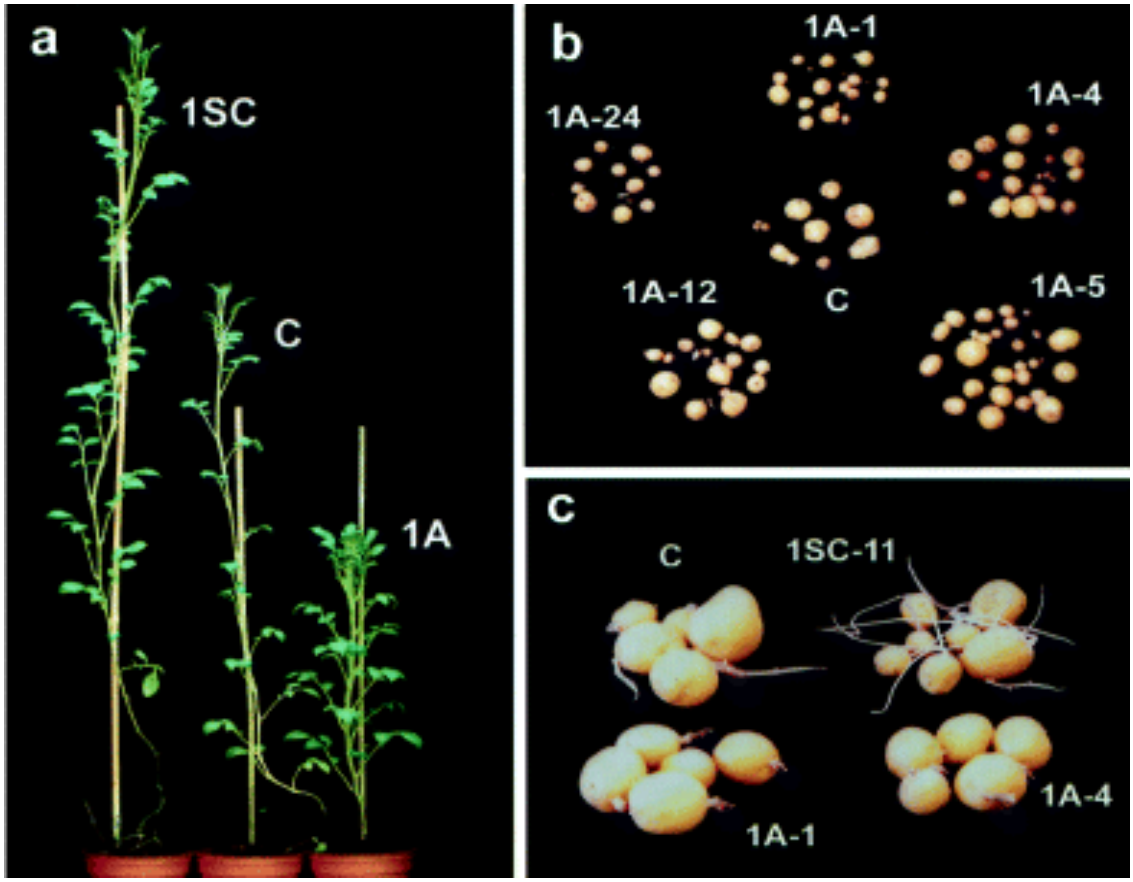


Figure 5.3. Knockdown of *StGA20ox1* results in dwarfed potatoes. **A** Lines expressing an anti-sense *StGA20ox1* transcript (1A) were dwarfed compared to WT lines (C), while lines expressing a sense *StGA20ox1* transcript (1SC) were elongated. **B** Anti-sense lines also had a higher yield compared to WT lines (C). **C** Tuber dormancy was prolonged in anti-sense lines and reduced in sense-expressing lines compared to WT (C). Source: Carrera et al. (2000).

Instead, there are several advantages to generating a dwarfism allele in potato by targeting a DELLA gene. Firstly, mutations in the potato *GAI* orthologue can be induced via GE and any resulting transgene-free plants can be trialled in the field in England (Genetic Technology (Precision Breeding) Act, 2023). Secondly, a gain-of-function mutation like that generated by Tomlinson et al. (2019) should be dominant, which is important for a crop that exists as either an autotetraploid, or a diploid transitioning to a hybrid breeding system. Furthermore, mutations disrupting the degradation of *GAI* have been shown to confer quantitative resistance against necrotrophs in *Arabidopsis* (Navarro et al., 2008; Shahnejat-Bushehri et al., 2016), wheat and barley (Saville et al., 2012), which would be a beneficial trait in potato,

particularly against early blight caused by *Alternaria solani* (Schmey et al., 2024). Lastly, it seems prudent to target the comparatively downstream DELLA proteins, as GA signalling regulates not just plant growth but also tuber development, dormancy, and sprouting (Chen et al., 2022; Di et al., 2024) and affecting GA biosynthesis might have pleiotropic effects. Still, it is not clear how a dwarfism allele in potato might impact tuber yield or shape, especially in the field, and in this project, I set out to address these uncertainties.

Aims

The goal of this project was to test whether introducing a moderate dwarfism allele could enable potato varieties to grow in a more compact form and how that might impact yield or other important harvest characteristics. This project had the following aims:

- Use CRISPR/Cas9 to target the *GAI* orthologue in diploid potato and introduce either an Alanine deletion, as done by Tomlinson et al. (2019), or an analogous mutation with promising agronomic properties
- Phenotype the dwarf mutants in the glasshouse
- Obtain T-DNA-free lines by outcrossing any dwarf mutant alleles in the diploid background and phenotype these in the field, looking at plant architecture, yield, tuber shape and size

5.2 Results

5.2.1 Generating DELLA dwarf potatoes

Targeting the *StGAI* gene

First, I BLAST-searched for the sequence of the *StGAI* gene targeted by Tomlinson et al. (2019) against the Solyntus V1.1 genome. The only high-confidence result was the *StGAI* gene on chromosome 11 (PGSC0003DMG400015692) whose sequence in the region of the DELLA motif was identical. I later PCR-amplified this region from lines B26 and B100 and amplicon sequencing confirmed 100% identity (data not shown). Thus, the protospacer targeted by

Tomlinson et al. (2019) was also present in B26 and the same sgRNA could be used.

I assembled a L2 plasmid (pYDJJ032) containing the *NptII* selection marker, the intron-bearing SpCas9 (Huang et al., 2021), the FAST-Red seed selection marker (Shimada et al., 2010), and the sgRNA targeting the DELLA motif (Fig. 5.4 A). This was transformed into B26 internodes via *Agrobacterium* co-cultivation. As the desired dwarf shoots might require longer to develop on the callus, the tissue culture stage was extended from two to three months, after which every shoot was harvested. This ultimately yielded ca. 780 T_0 plants, of which over 750 were successfully transferred to soil alongside WT B26 controls. Some T_0 plants exhibited a clear dwarfed phenotype (Fig. 5.4 B), but these were so severe as to be unlikely to be agronomically viable. To help identify dwarfs with milder phenotypes, I sprayed the T_0 plants with a solution of GA_3. As a result, many exhibited thin and elongated growth (Fig. 5.4 C) as compared with the unsprayed WT (Fig. 5.4 D), indicating they are either loss-of-function mutants or unedited lines. Other T_0s had a reduced response to the GA_3 (Fig. 5.4 E) and were earmarked as putative mild dwarfs. The severe dwarfs initially identified did not noticeably respond to the GA_3 treatment and retained a highly compact, rosette-like growth habit (Fig. 5.4 F), while a few plants exhibited a mix of growth habits (Fig. 5.4 G) likely due to being chimeras with gain-of-function alleles.

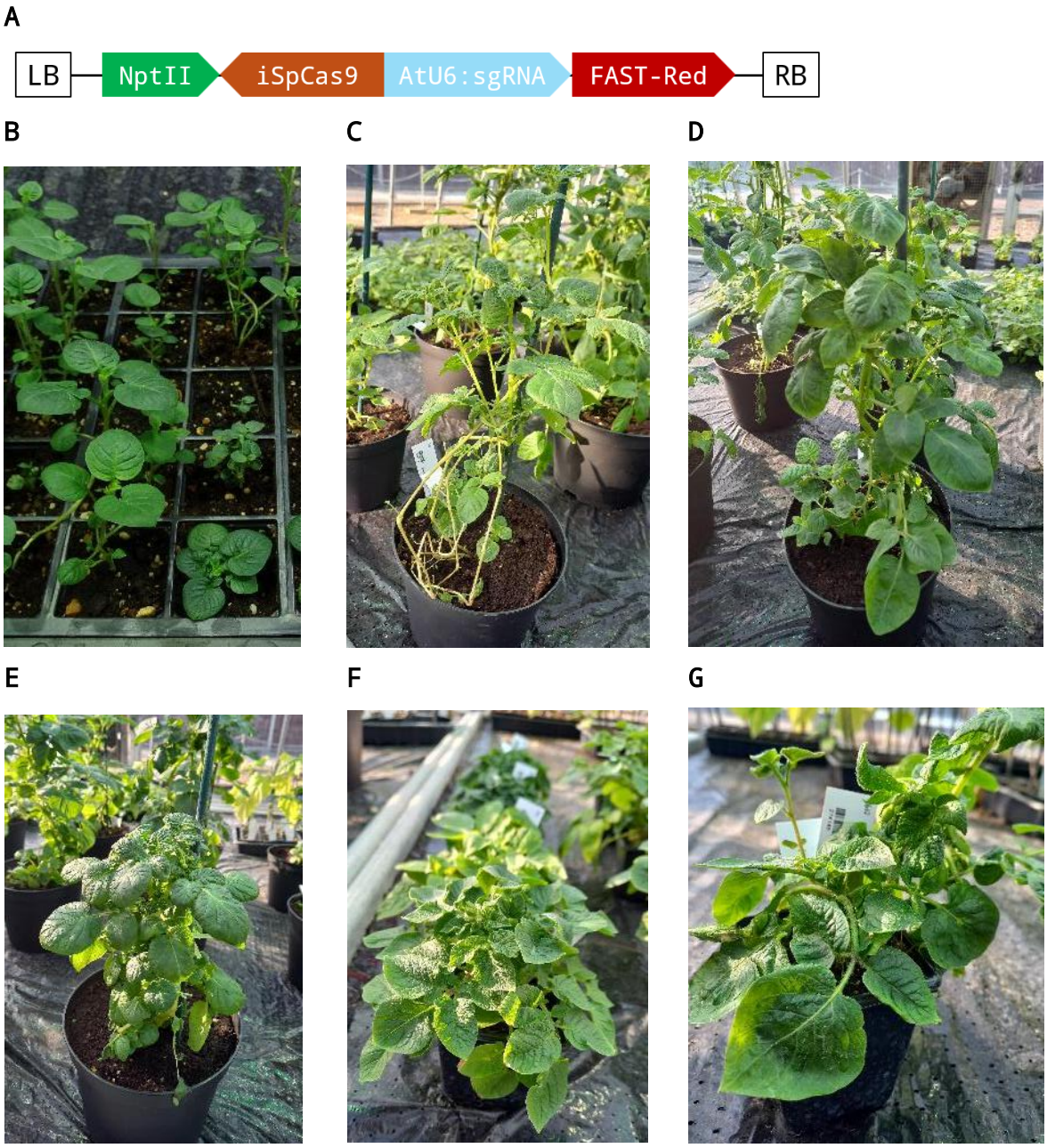


Figure 5.4. Generating DELLA mutants.
A The pYDJJ032 construct used to target the *StGAI* gene. The sgRNA used is the same as used by Tomlinson et al. (2019). **B** Some T_0s exhibited clear signs of dwarfism. **C** A highly elongated T_0, likely a loss-of-function mutant. **D** An unsprayed WT B26 plant. **E** A moderately dwarfed T_0. **F** A severely dwarfed T_0. **G** A T_0 exhibiting a mix of growth habits, likely a chimera with gain-of-function mutations.

Characterising T_0 lines

I selected a panel of T_0 lines with interesting phenotypes to characterise in more detail, including mild-to-severe dwarfs and some highly elongated lines (Fig. 5.5 A). The latter (e.g. line

655) retained an elongated growth habit long after the GA_3 treatment, suggesting that they really carried loss-of-function alleles. Furthermore, these plants failed to set tubers, instead producing numerous elongated stolons. Nor did they set any fruit, as their flowers proved infertile and occasionally reverted back to vegetative growth.

The dwarfed lines differed in the severity of their dwarfism, and sometimes in their leaf morphology (e.g. line 654). The difference in degree of dwarfism was presumably linked to the extent of the indel at the DELLA motif but this was not confirmed by genotyping. Instead, I attempted to outcross all these to WT B100 (used as male parent) but B26's poor fertility made this difficult. I was ultimately successful and recovered F_1 seeds from lines 543, 263, 235, and 457. I recovered tubers from 16 dwarfed T_0 lines, 13 of which were noticeably more elongated than the WT B26 tubers (Fig. 5.5 B). This was concerning, as industry generally prefers round to mildly oblong varieties. Notably, line 235 produced particularly elongated, sickle-shaped tubers, while those of lines 702, 148 and 457 remained globular. The degrees of elongation and dwarfism did not seem to correlate, as lines 235 and 457 were similarly dwarfed yet producing tubers with very different shapes. It seemed likely that at least some of the elongation was due to the GA_3 treatment which the WT B26 control was not subject to. Line 457 was taken forward on account of its round tubers, the availability of F_1 seed, and its moderate degree of dwarfism.

A

B

WT 543 263 551 148 235 457

Figure 5.5. T_0 dwarf phenotypes.
A Phenotypic series of T_0 plants (representative panel). Some lines (655) had highly elongated phenotypes and failed to set fruit or tubers. Others displayed mild-to-severe dwarfing, set tubers, and produced normal-looking flowers. Lines 543, 263, 235, and 457 proved fertile and were successfully outcrossed to B100. **B** Tubers of dwarfed lines tended to be more elongated than the WT control. Among these, the tubers of lines 148 and 457 tended to be relatively globular. This might have been partially caused by the GA_3 treatment.

Line 457 carries a Valine deletion

As some of the T_0 lines appeared to be chimeras (Fig. 5.4 G), I opted to genotype the mutation carried by line 457 in its F_1 progeny (seed collected in seed pack SP10453). As pYDJJ032 contained the FAST-Red marker, I screened against it in the seeds (roughly half of which were FAST-Red-positive). I initially sowed 16 non-fluorescent seeds from SP10453, of which 9 germinated. These seedlings were treated with GA_3 to identify dwarf F1 individuals. As controls, I also treated two lines from a WT B26×B100 cross (SP10016) with GA_3, while another two received a mock treatment. The F1 plants from both populations exhibited considerable variation in morphology and response to treatment, reflecting the high heterozygosity of the parental lines. While GA_3 treatment generally induced elongated growth, two SP10453

lines (PL03070 and PL03071) showed a reduced response (Fig. 5.6A). Amplicon sequencing revealed these were heterozygous for a deletion of a Valine residue (ΔVal) immediately downstream of the DELLA motif (Fig. 5.6 B). Meanwhile, the non-dwarfed lines were found to be heterozygous for a single nucleotide insertion (+T) just downstream of the DELLA motif (Fig. 5.6 C). The former edit appeared to be a dominant gain-of-function allele, which explained the mild dwarfism and reduced response to GA_3 application. The latter edit seemed to result in a recessive frameshift-induced loss-of-function mutation with no obvious phenotype in the heterozygote. Thus, line 457 must have simultaneously carried a gain-of-function and a loss-of-function allele of *GAI*.

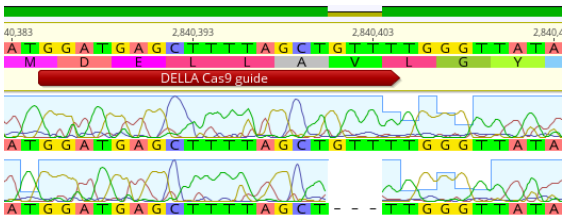
I analysed the differences between dwarfed and non-dwarfed siblings to quantify the effect of the ΔVal allele. This analysis is predicated on the assumption that the +T heterozygotes are phenotypically indistinguishable from the WT. This assumption is consistent with previous reports of analogous loss-of-function alleles in tomato (Livne et al., 2015; Lor et al., 2014). This point is further discussed in Section 5.3.

Lastly, I also compared the shape of tubers from GA_3 and mock-treated control plants and found the former were elongated, which confirmed that the sprays were at least partially the cause of the change in tuber morphology (Fig. 5.6 D).

A

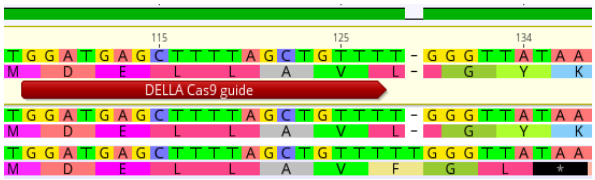
B

PL03071



C

PL03069



D

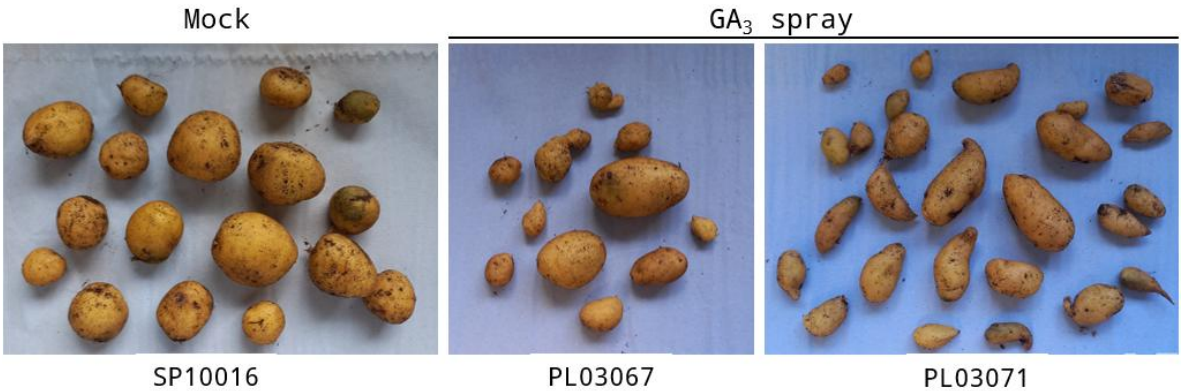


Figure 5.6. Phenotype and genotype of F_1 dwarfs.

A FAST-Red-negative F_1 lines from SP10453 and WT controls (SP10016) were treated with a GA_3 foliar spray or a mock solution.

Most plant produced elongated growth following GA_3-treatment (PL03067-9) but two stood out as low responders (PL03070-1). **B** The low responders were genotyped at the DELLA motif via amplicon sequencing, which revealed both were heterozygous for a Valine deletion immediately downstream of the DELLA motif. **C** Both dwarfed (PL03071) and WT (PL03067) plants set elongated tubers after GA_3 treatment, indicating that this change in morphology might not be a result of the dwarfism allele.

5.2.2 Dwarf performance in the field

More thorough phenotyping of the F_1s was planned in a field environment, where the results might be more relevant to potential applications.

Selecting F_1 lines for a field trial

To obtain additional ΔVal lines, I sowed out all remaining non-fluorescent SP10453 seeds. This yielded 72 F_1 plants, some of which grew noticeably more compact than their siblings (Fig. 5.7). The putative DELLA dwarfs remained ca. 30% shorter but were vigorous and set flowers, berries, and tubers as usual. As the population was large, and required considerable care, I reduced it by 22 non-dwarfed individuals.

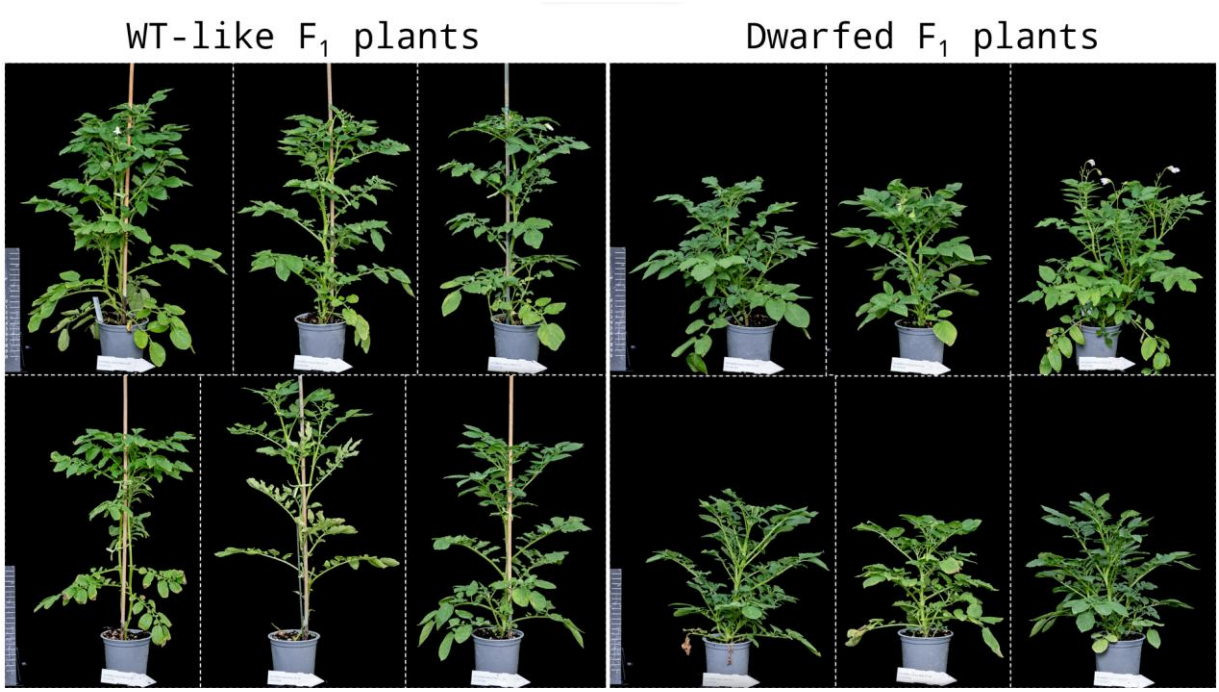
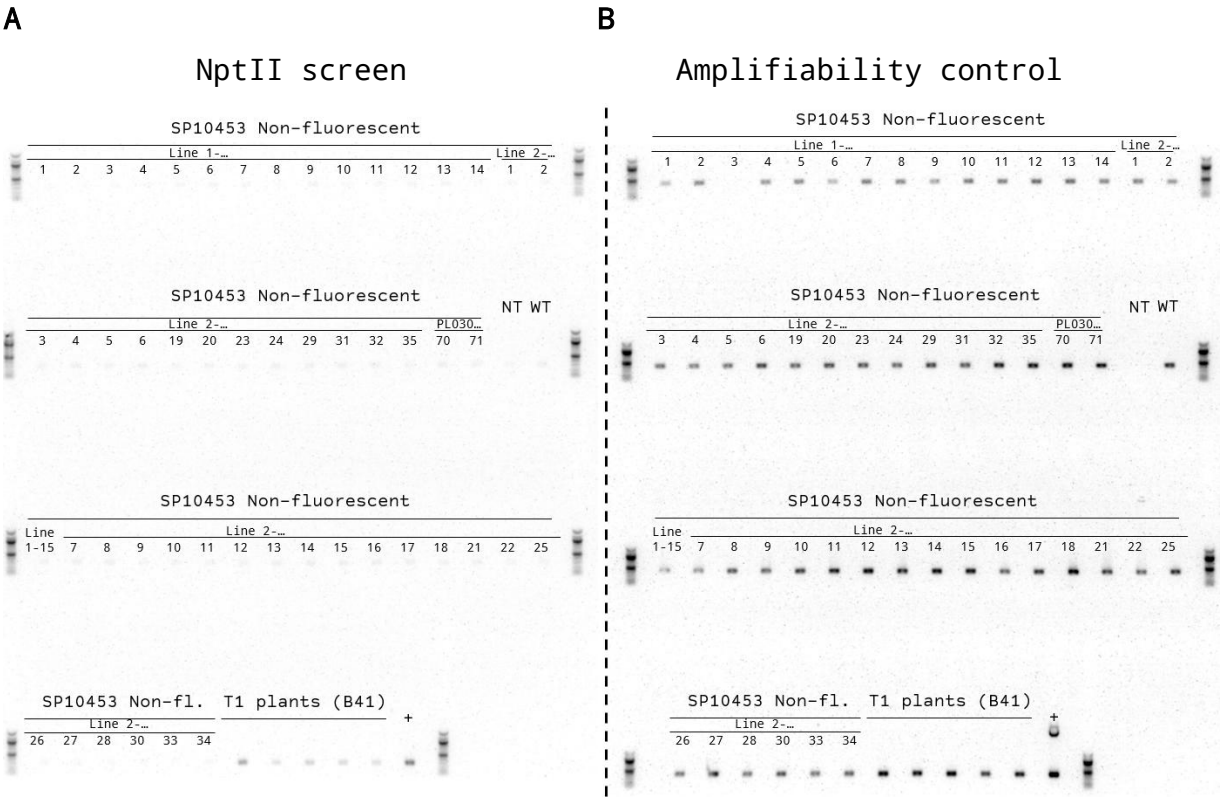


Figure 5.7. F_1 phenotypes in the glasshouse.

Current regulations in England required that for plants to be designated as Qualifying Higher Plants, and thus be eligible for field trials under the new flexible regulatory regime, the lines need to be T-DNA-free (Genetic Technology (Precision Breeding) Act, 2023). To further ensure that these FAST-Red-negative segregants did not carry any T-DNA insertions, I PCR-screened them for two regions of the T-DNA – the *NptII* gene and the Fast-Red terminator – as well as a slightly longer region of the endogenous *Cai1* locus to confirm the DNA was amplifiable. As positive controls, I also screened five T_0 transformants of the pYDJJ032 plasmid in a different genetic background, which should carry a T-DNA. Neither T-DNA screen produced a convincing amplicon from the SP10453 samples (Fig. 5.8 A&C), while the *Cai1* PCR confirmed that all samples except one were amplifiable (Fig. 5.8 B; line 1-3 which failed to amplify was discarded). Thus, the SP10453 lines were confirmed to be T-DNA-free.



C

Fast-Red terminator screen

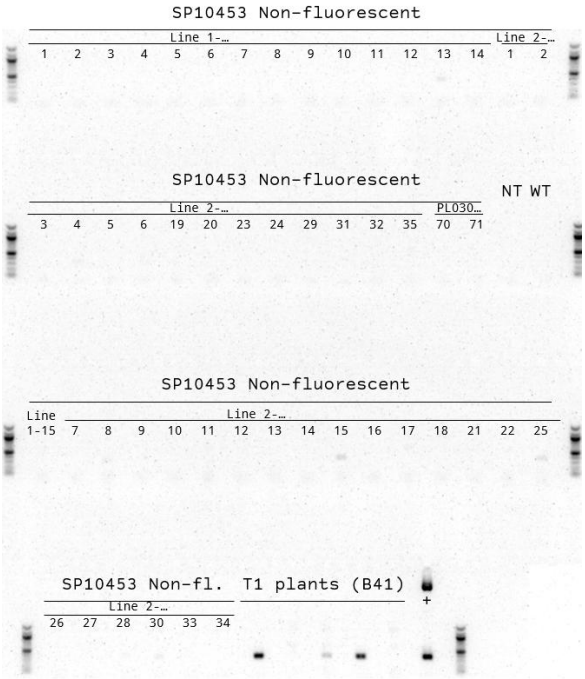
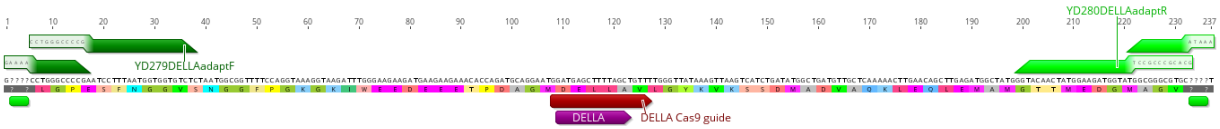


Figure 5.8. The non-fluorescent F_1 lines were T-DNA-free.
A & C FAST-Red-negative lines were screened for two T-DNA amplicons and neither produced products consistently. The T_1 plants on the bottom row were expected to carry a T-DNA and generally yielded the expected PCR product. **B** All DNA preps proved amplifiable except one (line 1-3), which was not taken forward. Positive control was the pYDJJ032 plasmid (+).

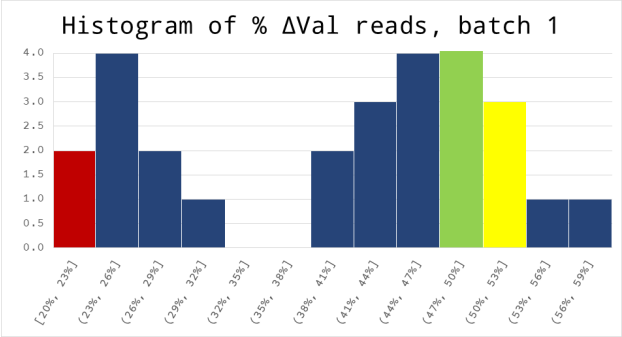
Next, I genotyped the F_1 lines using a custom genotyping pipeline to confirm that the putative dwarfs carried the ΔVal mutation and *vice versa*. To do this, I extracted crude DNA from each line and

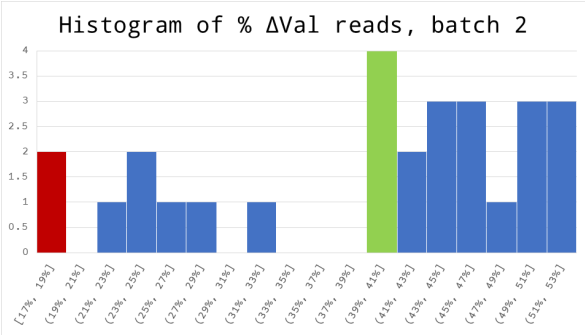
PCR-amplified the DELLA motif region with adapter primers (Fig. 5.9 A). The PCR products were then re-amplified with custom barcoding primers before being purified, pooled together, and submitted for next-generation amplicon sequencing in two batches. The demultiplexed results from all samples contained some ΔVal reads, including the WT and NT controls, which was likely caused either by contamination, sequencing errors in the read barcodes, or some combination of these. Still, when the proportion of ΔVal reads was quantified, the samples followed a bimodal distribution, with the WT control clustering with the samples containing fewer ΔVal reads, and the PL03071 and NT controls clustering with the samples containing more ΔVal reads (Fig. 5.9 B). I concluded that the two clusters correspond to +T and ΔVal heterozygotes, respectively, and the assigned genotypes matched the phenotypes observed earlier. I took forward 12 dwarf lines with >49.5% ΔVal reads, and 12 non-dwarf lines with <28% ΔVal reads. As a final confirmation, I sampled the selected 24 lines again, PCR-amplified across the target site and ran the products on a 4% agarose gel, which should result in a wider PCR band in case the plants carried a deletion as small as 3-bp (Bhattacharya & Van Meir, 2019) (Fig. 5.9 C). As expected, all dwarf lines produced a wide band, while the non-dwarf lines did not.

A



B





C

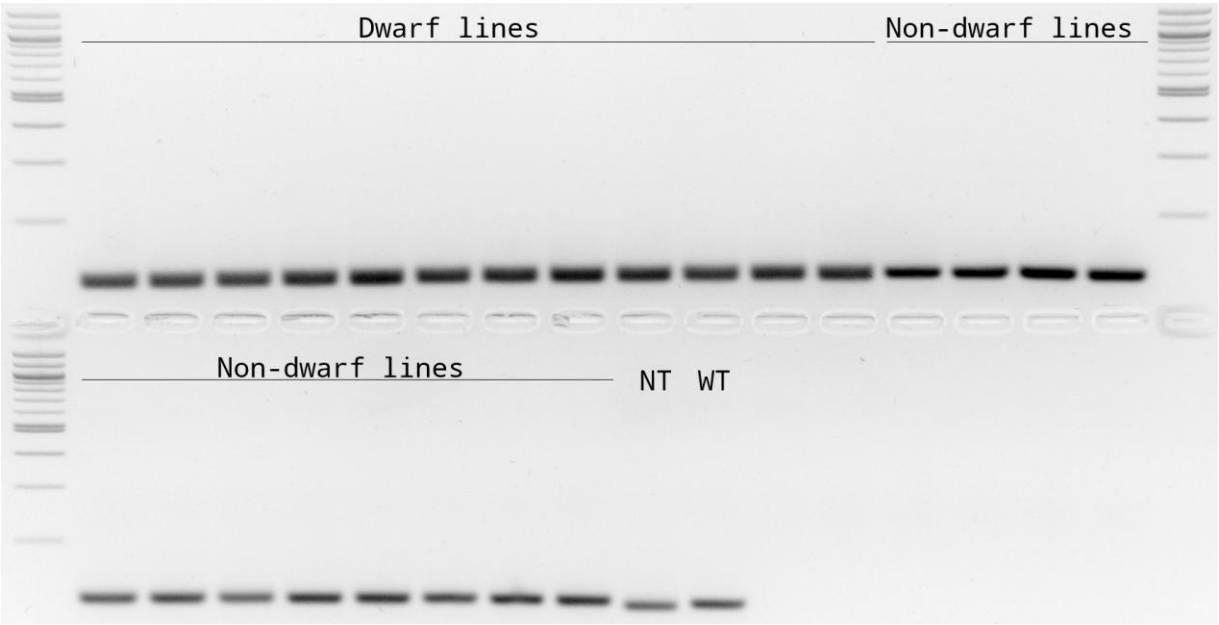


Figure 5.9. F_1 dwarf and non-dwarf line genotyping.
A Diagram of the region amplified with the adapter primers (YD279/YD280) and then re-amplified with barcoding primers.
B A histogram of the proportion of ΔVal reads, in percent, in the different samples after demultiplexing. Red bars contain the WT

control (PL03067), green bars contain the ΔVal control (PL03071), while the yellow bar contained the NT control.
C Primers YD188/YD205 should produce a 135 bp product from the WT sequence and 132 bp product from ΔVal sequence. While those products are too similar to be distinguishable, a heterozygous plant should result in a thicker, more diffuse band when run on a 4% agarose gel (Bhattacharya & Van Meir, 2019). All predicted dwarf lines produced a thick diffuse band, while all the predicted non-dwarf lines produced a sharp band like the WT control. A smaller band was produced in the NT control, indicating that there was some contamination, but this did not invalidate the result.

Field trial design

To test if a more compact growth habit can be introduced in potato without negatively affecting the yield and tuber characteristic, I planned a field trial with three plots of dwarf lines and another three with non-dwarf lines (Fig. 5.10). Each plot contained four distinct lines, with four plants per line, totalling 16 test plants per plot. Each plot was then enclosed with guards (*Solanum tuberosum* cv. Swift). Additionally, three plots were planted with the potato variety Swift, which is naturally short, and another three with a variety called Cara, which is naturally tall. This was done with the intention of comparing the growth habits of dwarfs and non-dwarfs to contemporary varieties. However, this approach was later abandoned this due to the difference in seed tuber size between the glasshouse-grown diploid varieties and the commercially sourced varieties.

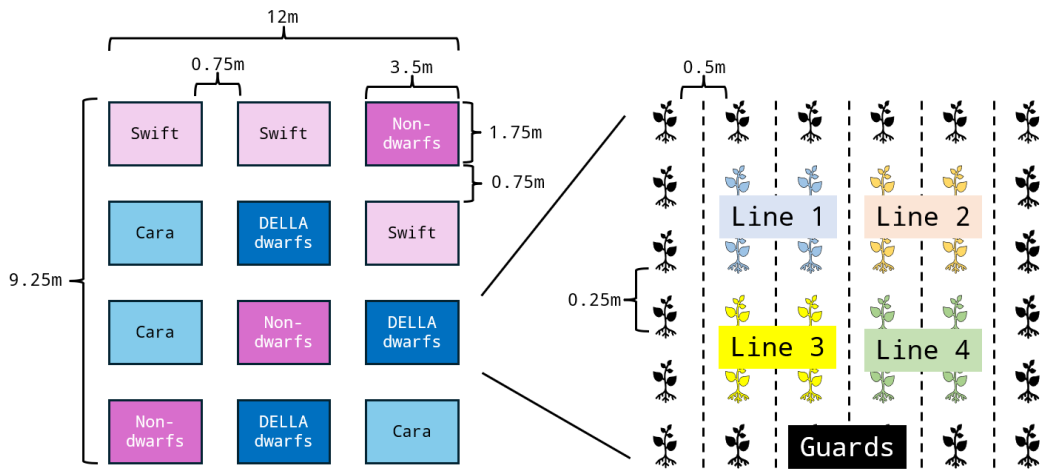


Figure 5.10. Field trial plan.

The four largest tubers were picked from each selected dwarf and non-dwarf line (Fig. 5.11 A). 12 tubers from each group were picked at random and weighted to confirm the average weight of seed did not significantly differ between dwarfs and non-dwarfs (One-way ANOVA, $F_{1,22}$=0.18, P-value = 0.7) (Fig. 5.11 B).

A

B

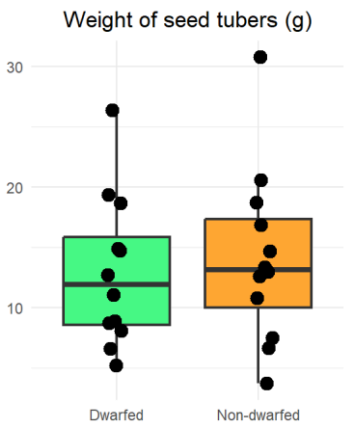


Figure 5.11. Field trial seed tubers.
A Photos of the dwarf (left) and non-dwarf (right) seed tubers. Scale bars are 5 cm. **B** The weight of 12 randomly selected dwarf and non-dwarf seed tubers did not differ significantly. In the boxes, the band indicates the median, the top and the bottom correspond to the first and third quartiles. Whiskers extend to values within 1.5 × of the interquartile range. Dots represent individual data points.

Comparing plant canopies

To quantify the plots' canopy growth, a drone equipped with a multispectral camera and a precise GPS sensor was flown overhead every week during the field trial. Every flight resulted in hundreds of images which were then algorithmically combined into an orthomosaic image. Importantly, the location of every pixel in an orthomosaic is triangulated in 3D space. This yielded a time course of red-green-blue (RGB) images (Fig. 5.12 A), as well as

digital surface models (DSMs) which captured the height of every pixel (Fig. 5.12 B). Crucially, by subtracting the plots' DSM values on any day from the DSM values of the bare soil, I could calculate the growth by that day (Fig. 5.12 C). Plotted against time, the mean plant height of dwarfs and non-dwarf plots followed a Gaussian ("bell-shaped") curve (Fig. 5.12 D). In this, the plant height of dwarf plots seemed modestly but consistently reduced.

To quantify this difference, I analysed the data with a modified Gaussian Non-linear Mixed Effects model (for details, see Section 2.3.2). According to this model, around 95 days after planting the mean plant height of non-dwarf plots peaked at 57.9 ± 2.0 cm (SE), while the dwarf plots' mean height peaked at 51.7 ± 2.1 cm (SE). This difference of ca. 11% (6 ± 2.9 cm [SE]) was statistically significant (P-value = 0.03). The model fit the data very well, with a Residual Standard Deviation of ca. 4 cm (the average vertical distance between data points and their predicted value).

Additionally, I manually measured the plants' height at 94 days after planting, which coincided with the growth curves' optimum (Fig. 5.12 D). A vertical ruler was used to measure the distance from the base of each plant to the tip of its tallest stem. The non-dwarf plants' stems reached an average length of 70.7 ± 1.8 cm (SE), while dwarfs' stems reached 62.1 ± 1.8 cm (SE) on average, which was significantly shorter by ca. 12% (8.5 ± 2.6 cm (SE); P-value = 0.004) (Fig. 5. 13).

While the drone-derived and manually measured datasets quantified growth in different ways (mean canopy height versus lengths of the longest stems), both were consistent with the dwarfs being 11-12% shorter that their non-dwarf siblings. This field phenotype was much milder than that observed in the glasshouse, highlighting the importance of field trials when working with agronomically-relevant traits.

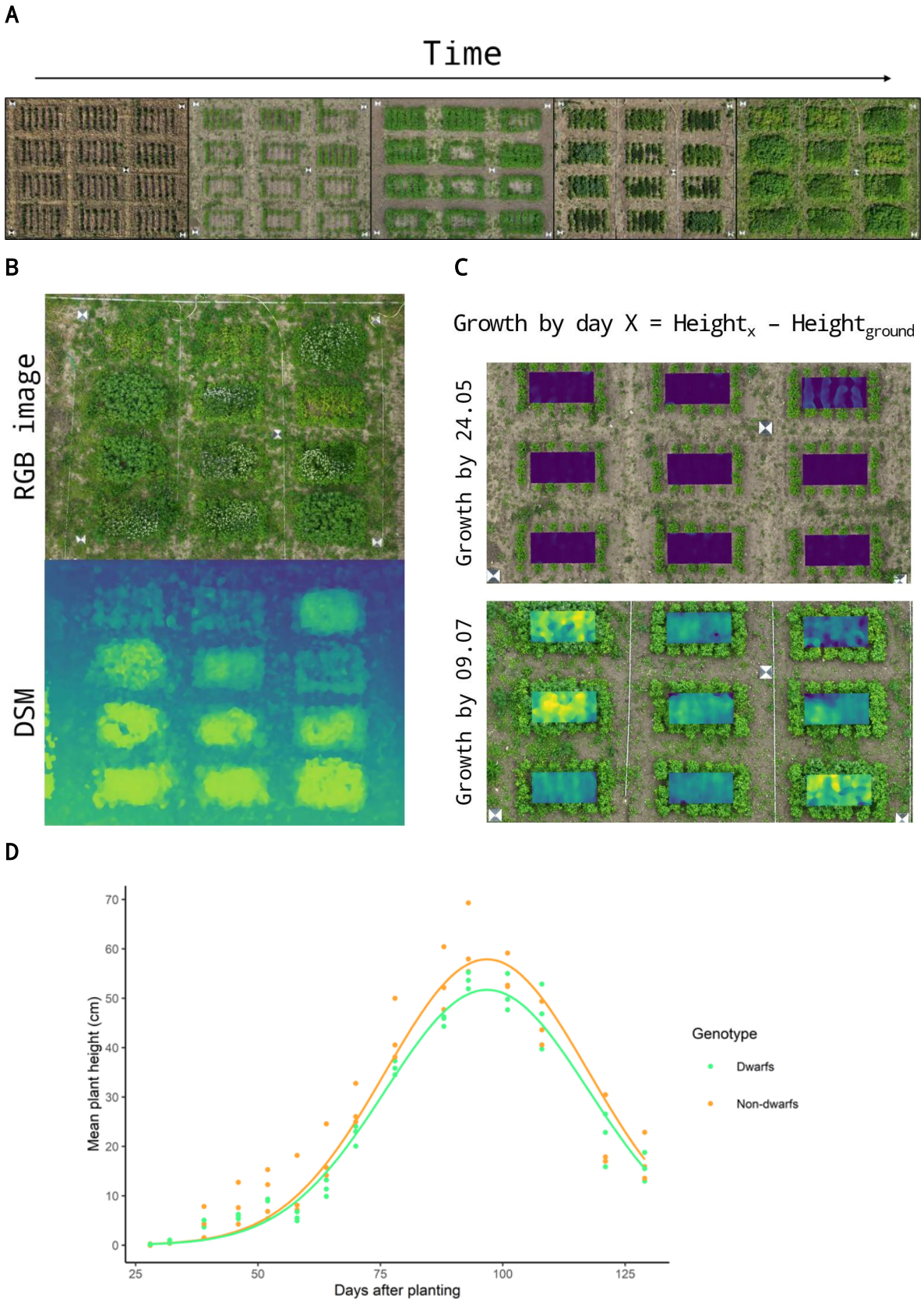


Figure 5.12. Canopy growth in the field.
A Growth of the plots, captured via drone. **B** RGB orthomosaic and its corresponding DSM. DSM height values are displayed in pseudo-colour. **C** Subtracting the soil height from the canopy height at any date yields the growth by that day. **D** Mean canopy height over

the growing period. The canopy development followed a bell-shaped curve. Plotted model curves do not account for the random effect of plot.

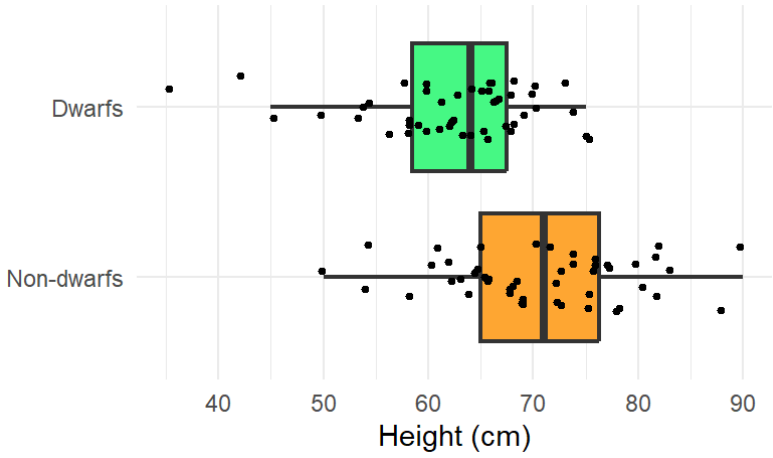


Figure 5.13. Longest stem length.
The dwarfs' mean stem length of 62.1 ± 1.8 cm (SE) was significantly shorter by 12.1% (8.5 ± 2.6 cm [SE]; P-value = 0.004) compared to that of the non-dwarf siblings at 70.7 ± 1.8 cm (SE). The band inside each box is the median; the top and the bottom of the box are the first and third quartiles. The whiskers extend to values within 1.5 × of the interquartile range. Dots represent individual data points.

The field-grown dwarfs and non-dwarfs were also monitored for differences in traits such as morphology, flowering time, flower abundance, and time of senescence (the plants were grown for a total of 132 days to allow them to naturally senesce) but no obvious discrepancies were observed.

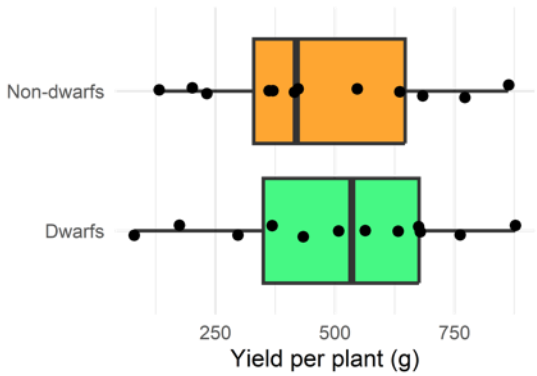
Tuber trait phenotyping

At 132 days after planting, the above-ground foliage was cut off and about two weeks later the tubers were harvested, taking care to harvest each line separately (Fig. 5.14 A). The fresh tuber yield of each line was weighed and divided by the number of plants (three for line 1-1, four for all others) (Fig. 5.14 B). The average yield per dwarf plant (504 ± 69.6 g [SE]) was higher than the average yield per non-dwarf plant (470 ± 67.0 g [SE]) but not significantly (two-sample Student's t-test, $t_{(22)}$ = 0.35, P-value = 0.73). Thus, the more compact growth did not come at the

expense of overall yield. After weighing, some tubers which were affected with what appeared to be *Fusarium* dry rot, a hemibiotrophic disease, were discarded (Fig.5.14 C).

A

B



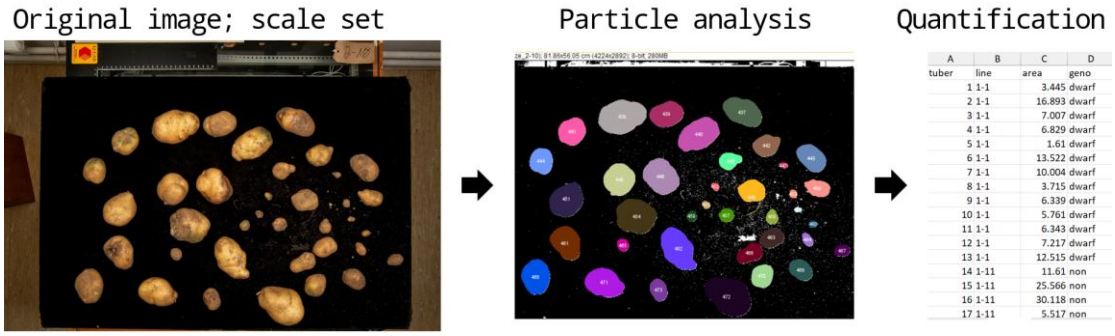
c

Figure 5.14. Dwarf and non-dwarf field trial yields.
A A representative panel of tubers harvested from dwarf and non-dwarf lines. **B** The yield per plant was compared, as for line 1-1 only three of four tubers sprouted. The mean yield per plant of dwarfs (504 ± 69.6 g [SE]) and non-dwarfs (470 ± 67.0 g [SE]) did not differ significantly (two-sample Student's t-test, $t_{(22)} = 0.35$, P-value = 0.73). The band inside each box is the median; the top and the bottom of the box are the first and third quartiles. The whiskers extend to values within 1.5 × of the interquartile range.

Dots represent individual data points. **C** Some tubers were infected with what appeared to be *Fusarium* dry rot.

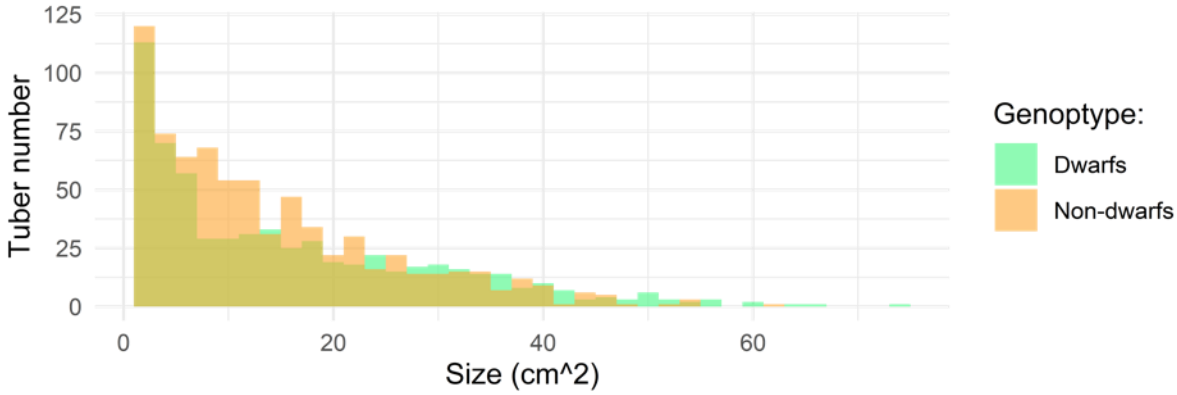
Next, I compared the size distributions of tubers from dwarf and non-dwarf plants. To do this, I photographed the harvested material and used ImageJ to quantify the size (area) of each tuber, with a lower limit of 1 cm^2 (Fig. 5.15 A). A histogram revealed what seemed to be different distributions (Fig. 5.15 B). Both had peaks at the start of the X axis, indicating dwarfs and non-dwarfs produced numerous tiny tubers up to several squared centimetres in size. Beyond this, the non-dwarf's histogram declined gradually to the right. In contract, the dwarf's histogram appeared multimodal – declining rapidly after the initial peak before rising and gradually decreasing again. Thus, the dwarf plants produced fewer tubers with a size of ca. 10 cm^2 and more with sizes of ca. 40 cm^2 and above. To precisely compare the shapes of the histograms, I calculated their empirical cumulative distribution functions (eCDFs; Fig. 5.15 C). The difference between these was significant (Fig. 5.15 D; two-sample Cramer-von Mises test, T = 2.72, P-value = 0.014), confirming the ΔVal dwarfism allele influenced tuber size and shifted the balance towards larger tubers. Individual lines within both genotypic groups varied noticeably in their tuber size distribution (Fig. 5.15 E), indicating that this population was simultaneously segregating for other relevant genetic factors.

A

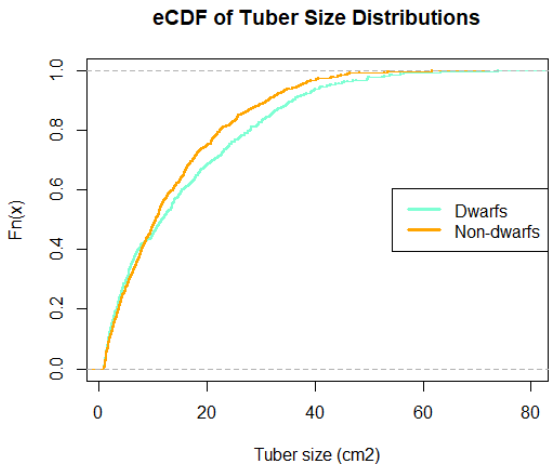


A	B	C	D
tuber	line	area	geno
1	1-1	3.445	dwarf
2	1-1	16.893	dwarf
3	1-1	7.007	dwarf
4	1-1	6.829	dwarf
5	1-1	1.61	dwarf
6	1-1	13.522	dwarf
7	1-1	10.004	dwarf
8	1-1	3.715	dwarf
9	1-1	6.339	dwarf
10	1-1	5.761	dwarf
11	1-1	6.343	dwarf
12	1-1	7.217	dwarf
13	1-1	12.515	dwarf
14	1-11	11.61	non
15	1-11	25.566	non
16	1-11	30.118	non
17	1-11	5.517	non

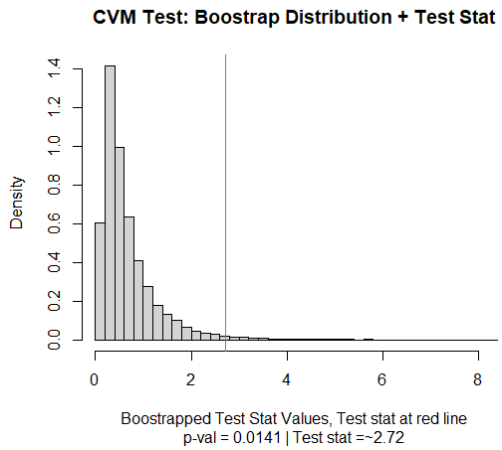
B



C



D



E

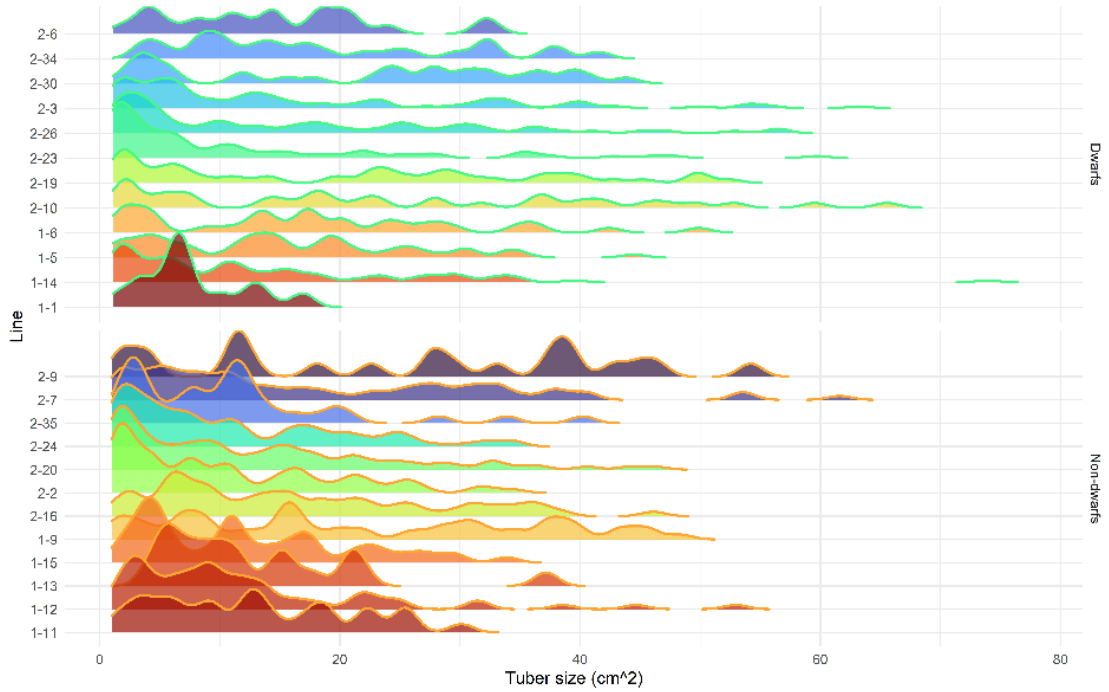


Figure 5.15. Dwarf tuber size distribution.
A Tuber size measuring process. The harvested tubers were photographed, and the resulting images analysed with ImageJ. The size (area) of tubers above 1 cm^2 was recorded. **B** A histogram of the sizes of harvested dwarf and non-dwarf tubers. Note that the two groups have distributions with different shapes. **C** Empirical cumulative distribution functions (eCDFs) of the size of dwarf and non-dwarf tubers. The Y axis value describes the probability [0;1] that a value from a distribution is less than or equal to a given point (X). This allows for the quantitative comparison of the shape of two distributions. **D** The Cramer-von Mises test probes whether the differences between the eCDFs are consistent with the null hypothesis: that the two distributions originate from a single underlying distribution. The differences of a size beyond what can be expected given the sample sizes (red line) are grounds for refuting it. **E** A ridgeline density plot illustrating the size distribution of individual lines.

Lastly, I investigated whether the ΔVal dwarfism allele influenced tuber shape. To test this, I selected eight large tubers from each line, photographed them on top of a lightbox, and analysed their shape using ImageJ. This was done by finding the best-fitting ellipse and calculating its Aspect Ratio (AR, the ratio between an

ellipse's longest and shortest axis) (Fig. 5.16 A). The average dwarf tuber proved to be rounder (AR = 1.38 ±0.02) than the average non-dwarf tuber (AR = 1.42 ±0.02) (Fig. 5.16 B), a marginally significant difference (two-way ANOVA, $F_{1,168}$ = 3.37, P-value = 0.07). Here again there was some additional variance within the genotypic groups and between the individual lines (Fig. 5.16 C). Line 2-23 stood out as having particularly elongated tubers (Fig. 5.16 D). Thus, it seems the dwarfing allele might have a mild effect on tuber shape, promoting rounder tubers.

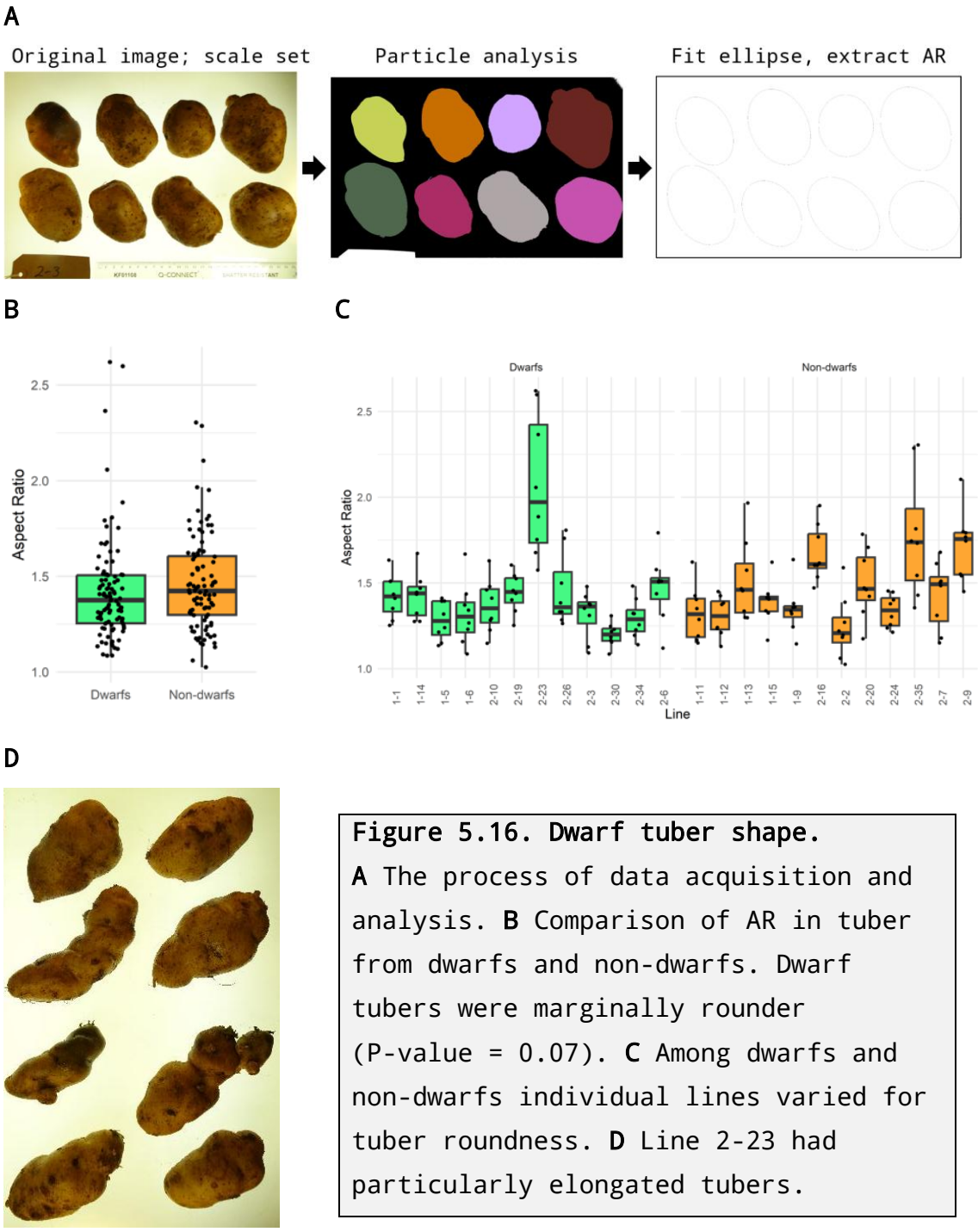


Figure 5.16. Dwarf tuber shape.
A The process of data acquisition and analysis. **B** Comparison of AR in tuber from dwarfs and non-dwarfs. Dwarf tubers were marginally rounder (P-value = 0.07). **C** Among dwarfs and non-dwarfs individual lines varied for tuber roundness. **D** Line 2-23 had particularly elongated tubers.

5.3 Discussion

In this chapter, my goal was to generate a moderate dwarfism allele in diploid potato and test how that impacted growth, yield and other tuber characteristics. The project was successful, with one promising allele (ΔVal) being characterised in the glasshouse and in the field. Below, I discuss the results to date, the assumptions made in the analysis, and what work remains to be done.

Generating dwarfism alleles via *StGAI* mutagenesis

To generate a dwarfism allele, I targeted the DELLA motif of the *StGAI* gene using CRISPR/Cas9 as done by Tomlinson et al. (2019). The transformation yielded hundreds of plants, most of which exhibited a hyper-elongated *PROCERA*-like phenotype. Among these, a few appeared to have normal-to-dwarfed stature, indicating that they might carry gain-of-function *GAI* alleles. To help identify these, I used a GA_3 foliar spray after all the plants were transferred to soil but in hindsight this was unnecessary. In the future, it should be possible to include a higher amount of GA_3 in the LSR2 (shoot-inducing) media, followed by the selective harvesting of shoots that exhibit a muted response.

Of the 16 dwarfed T_0 lines whose tubers were harvested, 13 had elongated tubers. Curiously, the degrees of elongation and dwarfism did not seem to correlate. For example, lines 235 and 457 were similarly dwarfed yet producing tubers with very different shapes. Gibberellins have long been known to influence tuber development (Vreugdenhil & Struik, 1989) but the molecular mechanisms involved remain unclear. Since the current approach to generating dwarfs relies on disrupting the interaction between DELLA protein and GA receptor, and the mutations were found to affect the canopy and tubers differently, it can be hypothesised that in these tissues the DELLA motif is being recognised by different proteins with different tolerances to mutations. The classic GA receptor in potato *GA-insensitive dwarf 1* (*StGID1*; Soltu.DM.01G037830.1) has two uncharacterised paralogues, which I will refer to as GID1-like GA receptor 1 (Soltu.DM.09G022610.1) and GID1-like GA receptor 2 (Soltu.DM.06G000730.1). While all

three genes can be detected throughout the plant, GID1-like GA receptor 1 is expressed at a higher level in the tubers (Hamilton et al., 2024; Zhao et al., 2023). If a mutation can be introduced at the DELLA motif such that it disrupts the interaction with one GA receptor but not another, it might be possible to target the effects of a dwarfism allele. Recent advances in structural biology like AlphaFold Multimer (Evans et al., 2022) might also allow for the prediction of the effects of a mutation, while the dominant nature of gain-of-function *GAI* alleles should enable easy testing of these predictions. Ultimately this could pave the way for a more informed approach to mutating the DELLA motif.

Lessons for future edits could also be gleaned from the three dwarfed lines with round tubers. Of these 457's dwarfism was intermediate, while lines 148 and 702 had a less or more severe phenotype, respectively (Fig. 5.5 A; SFig. 13). This might reveal how to modulate the severity of the dwarfism phenotype, while minimizing the effects on the tubers. To support both lines of enquiry, I plan on genotyping additional T_0 lines with both elongated and round tubers.

The ΔVal allele

To characterise line 457, I genotyped its FAST-Red-negative F_1 progeny. This revealed 457 simultaneously carried a gain-of-function and a loss-of-function allele of *GAI*, which were then inherited by the dwarfed and non-dwarfed progeny. A larger FAST-Red-negative F_1 population was then screened for the presence of T-DNA sequences and confirmed as T-DNA-free. This contained both seemingly dwarfed and non-dwarfed individuals, and it was genotyped as scale using a custom pipeline to confirm the lines' phenotype matched their genotype. The genotyping was successful but some ΔVal reads were present in all samples, including the WT control. This could have simply been caused by sample cross-contamination during PCR setup, but it is likely at least partially due to poor barcode design, as some barcodes differed by a single nucleotide. Also, the barcodes were too close to the ends of the amplicons, where sequencing quality is often lower. In the future, better design of barcoding primers might help mitigate this issue.

After phenotyping the F_1 progeny, I quantified the effects of the ΔVal allele by comparing the dwarf and non-dwarf groups. This analysis assumed that the non-dwarf lines heterozygous for a +T insertion, resulting in a frameshift and *GAI* KO, are phenotypically indistinguishable from WT lines. This assumption is based on the recessive nature of loss-of-function mutants in orthologues in barley (C. A. Foster, 1977), tomato (Livne et al., 2015; Lor et al., 2014), Arabidopsis (Peng et al., 1997; Silverstone et al., 1997), and rice (Ikeda et al., 2001). To test this assumption, I collected F_2 selfed seed from non-dwarf lines. This seed will be germinated and the phenotypes of the heterozygous and null segregants compared.

In the glasshouse, I found the dwarfs exhibited a noticeably more compact growth habit than their non-dwarf siblings, with their main stem height reduced by 30-40%. While this is nominally greater than the 11-12% difference in height seen in the field, this is likely due to plants in the glasshouse being staked (Fig. 5.7) while those in the field were allowed to sprawl, thus reducing the vertical difference between dwarfs and non-dwarfs. All the dwarfs used in this report were heterozygous and I have yet to characterise the phenotype of a homozygous ΔVal line, although prior reports of similar mutants in tomato found no difference (Tomlinson et al., 2019). To test this, I collected F_2 selfed seed from dwarf lines which I will germinate and characterise.

Field performance

The purpose of the field trial was to test whether a more compact growth habit can be introduced in potato without negatively affecting the yield and tuber characteristics. To answer this, I compared the canopy development of dwarf and non-dwarf plots via drone-captured imagery and later compared their harvest using image analysis.

The drone-derived data confirmed the ΔVal allele did result in a significantly more compact canopy. Beyond this, the images also amount to a detailed record of each line’s development, rate of flowering and senescence, which did not seem to differ between the two groups. Notably, these results were in close agreement with

the independently collected manual height measurements. Still, this data has its limitations as plant height measurements cannot comprehensively capture the structure of the canopy. It remains unclear if the dwarfs and non-dwarfs produced a similar amount of leaf area, and the height difference is a result of shorter internodes, or if a reduction in node number and leaf area occurred. It is also unclear if there was a reduction in photosynthetic biomass, or if this was constant but distributed in a different way. To answer such questions, more extensive field trials would be required, with multiple techniques employed simultaneously to analyse the plants' structure. Such experiments are outside the scope of this project.

The plants' canopy development was found to follow a Gaussian ("bell-shaped") curve, necessitating a modified Gaussian Mixed Effect model. While complex and technically challenging to implement, the model fit the data well, and similar approaches have previously been used to model canopy growth in the field (L. Su et al., 2015). In this model, the effect of the individual plots was accounted for, and this was shown to add or subtract up to four centimetres from the expected peak mean plot height. This might be due to the variability between the individual lines, as with only four lines per plot it is possible that one plot got relatively tall lines by chance. Conversely, during harvest it was noticed that some plots were harder to dig up than others, hinting that some might have been tilled better than others which might have contributed additionally to the variability.

Following harvest, the two groups were compared based on yield, tuber shape (elongation), and size. The dwarf plants yielded more on average, but the difference was not statistically significant and there was much more variability within groups than there was between groups. Over the growing season, the plants were fertilised with N_2 at 174 kg/ha, which is within the recommended range of 120-200 kg/ha (Sinclair et al., 2022). This is notable as dwarf varieties in crops like wheat were found to have reduced nitrogen use efficiency, requiring elevated fertiliser use to achieve superior yields (Q. Liu et al., 2021). The dwarf tubers also appeared to be marginally rounder, which conforms to industry

preferences. Lastly, the tuber size distributions between the groups differed significantly, with the harvest of dwarf lines biased towards larger tubers. This is important, since grower decide how densely to plant based on the predicted yield in their target size class, rather than the predicted overall yield (van Dijk et al., 2022).

In summary, in this chapter I demonstrated dwarfism alleles can be generated in potato by targeting the DELLA motif of the *GAI* gene, and I confirmed this results in more compact growth in the field with no yield penalty. Following this proof-of-concept, the next step would be to grow the dwarfs and a suitable control at different planting densities and compare how the yield changes. Based on the current results, it can be predicted that the dwarf's yield should plateau at higher planting densities, thus boosting the maximum yield per unit area.

CHAPTER 6. GENERAL DISCUSSION

In this PhD project, I employed modern biotechnological methods to help address two major challenges in potato breeding, namely their susceptibility to *P. infestans* and stagnating rates of crop improvement. Firstly, I established an efficient method for targeted gene KI, a long-standing goal of the plant genome editing field, and used it to deliver R-genes from a wild potato into an elite diploid line. I later built upon this to achieve efficient targeted gene replacements and introduced TIIGER: a method for iterative gene stacking at the same locus. Notably, TIIGER's ultimate goal is the generation of a transgene-free potato line with multiple R-genes stacked at the same locus. While this has not yet been completed, the first phase of the process was conducted efficiently, successfully delivering large TIIGER cassettes. Lastly, I generated compact growth alleles in potato using CRISPR/Cas9 editing of the *StGAI* gene. One of these seemed agronomically valuable and was characterised in the field, with promising results. These projects' progress to date, outstanding questions, and directions of future work are discussed below.

6.1 Targeted Gene Insertions

The aim of this chapter was to establish an efficient method for gene KI in potato. To this end, I adapted an anthocyanin KI reporter system from tomato to potato by targeting the orthologous *StCai1* locus and later used it to rapidly optimise several aspects of the KI process. Factors which affected KI rates included the presence of introns in the Cas12a gene, presumably affecting Cas12a expression levels, and the choice of plasmid backbone. Conversely, changing the crRNA expression cassette or employing small molecule effectors proved unsuccessful or counterproductive. In the process of improving KIs, I also generated the valuable diploid potato line YD1 which combines high fertility with high transformability and greatly facilitated subsequent research. Lastly, I leveraged the optimised KI approach to insert *Rpi-vnt1* at *Cai1*, which should confer resistance against *P. infestans*.

Future directions

While the KI method arrived at is efficient, some aspects of it remain unclear. One major unknown is exactly how much of an improvement in KI rates was achieved. To answer this, I am in the process of cloning a pair of comparable constructs, only one of which will include the various improvements identified in Chapter 3. These will then be benchmarked against each other in B26 and YD1, allowing for the separate quantification of progress made through plasmid design and through breeding a superior line to work in.

The interpretations of why some factors increased KI rates rest on assumptions which remain untested. Notably, while I found adding introns to Cas12a boosts KI frequency, and the further addition of the UL12 exonuclease did not, but I still need to measure the nuclease protein levels in stable transformants to confirm the difference in performance is due to nuclease protein accumulation. Similarly, while changing the crRNA-expressing promoters, terminators, and flanking cassettes made no difference to the KI rate, suggesting all configurations produce more than enough crRNAs, I still need to confirm they produced different crRNA transcript amounts.

Further variations in plasmid design could be made to try to boost KI rates. For example, additional copies of the crRNA target site could be included on either side of the KI cassette to partially account for protospacer spoilage. However, the KI efficiency already achieved seems broadly sufficient, at least when proceeding via stable transformation. A major limitation of this approach is that most recovered events carry a T-DNA insertion. Thus, I will focus on increasing the efficiency of KIs achieved via transient expression and the putative one-step process, as this would be of crucial importance to applying KIs in species where breeding T-DNAs away isn’t an option, such as tetraploid potato.

Finally, the reasons why some lines (YD1, YD3) readily yield KI sectors, while others (YD2) don’t still need to be established. One working hypothesis posits that “template competition” is the key factor, and that substantial heterozygosity at the target

locus can preclude KIs. Alternatively, YD2 could simply be deficient in some factor necessary for anthocyanin overaccumulation, rendering the KI sectors invisible. Both ideas will be tested as a thorough understanding of the prerequisites for efficient KIs is crucial for their future implementation in other species.

6.2 TIIGER

Building on my work on KIs, I pursued a method for the targeted iterative insertion of genes via recombination (TIIGER). The first phase of the process involves the delivery of one or more genes of interest plus a conditional counter-selection cassette. This was achieved, delivering even large, multigene cassettes at frequencies similar to those of simple KIs. In parallel, I attempted to combine the TIIGER approach with the classic recombinase-mediated cassette exchange (RMCE) method. To do this, I framed a cassette containing the 35S promoter with heterospecific LoxP sites and targeted this upstream of the *Cai1* gene, resulting in an RMCE destination locus. Lastly, to help facilitate TIIGER, and GT in general, I attempted to establish GT-specific selection (GTSS) – a KI reporter that results in chemical resistance, rather than anthocyanin accumulation, with promising early results.

Future directions

While the first phase of TIIGER was demonstrated, the second phase has still not been attempted, involving the replacement of the 35S promoter and *codA* gene with further genes of interest and the original proximal *Cai1* promoter. The success of the first phase shows that targeted replacements are achievable, so it is likely that the success of phase two hinges on effective selection for the desired event. One particular concern is that a simple deletion of the 35S-*codA* cassette would be phenotypically indistinguishable from a genuine secondary TIIGER event. To quantify the frequency of such confounding events, and attempt to minimise them, I will attempt phase two TIIGER events while varying the cassette size and mode of targeting. I will also attempt to transiently implement both TIIGER phases one and two,

as the ability to iteratively deliver genes without having to segregate away T-DNAs at every step would greatly accelerate the speed with which genes can be stacked. Furthermore, outcrosses or selfs could disrupt valuable traits, such as the transformability and fertility which distinguish YD1.

The RMCE phase of the TIIGER-RMCE approach also remains to be completed. In this, the destination locus of the LoxP lines and a cassette bearing R-genes from a Cre-expressing construct will be exchanged; both stable and transient expression approaches will be attempted.

Lastly, the GTSS method requires further development. This will include selection cassettes based on the *NptII* gene and promoter-less cassettes targeted downstream of *StALS1* gene, both of which should result in a more stringent selection.

6.3 Engineering Compact Growth into Potatoes

The aim of this chapter was to target the DELLA motif of the potato *GAI* gene with CRISPR/Cas9 and generate a gain-of-function allele with an agronomically valuable compact growth phenotype and evaluate it in the field as a Precision Bred Organism. This was successfully done, producing a phenotypic series of gain-of-function mutants with varying degrees of dwarfism. A dwarfed line carrying a Valine deletion immediately downstream of the DELLA motif was found to have a promising phenotype and its progeny were assessed in the glasshouse and the field. The dwarfed plants displayed compact growth while retaining their yield and tuber characteristics compared to the control group. This allele seems likely to permit productive growth at higher planting densities, thus potentially increasing potato yield per unit area.

Future directions

The promising field performance of the ΔVal dwarfs is relative to siblings heterozygous for a +T recessive loss-of-function allele at the DELLA locus. The latter were assumed to be phenotypically identical to WT plants. This assumption will be tested by quantitatively comparing the phenotypes of heterozygous +T and null F_2 segregants.

Another unknown is whether the dwarfism phenotype is more severe when homozygous. To answer this, I will quantitatively compare the phenotypes of heterozygous and homozygous ΔVal F_2 segregants.

Beyond stature, the phenotypic series of dwarfs independently varied in tuber and leaf shape too. The effects of the DELLA mutants are thought to stem from their weakened interactions with *StGID1*. However, potato has two other highly similar paralogues which differ in their pattern of expression. If parallels are found between the observed phenotypes, and the predicted interactions with different *StGID1* paralogues, it might be possible to predict the phenotype of *StGAI* alleles and thus gain more control over the crop's morphology.

6.4 Concluding remarks

This thesis has tackled two key challenges in potato cultivation: susceptibility to Late Blight and stagnating productivity. Targeted gene delivery provides a powerful tool to harness the rich genetic diversity of wild relatives, offering durable resistance through precise gene stacking. TIIGER enhances this approach by giving breeders unprecedented control over gene selection and placement, streamlining the integration of resistance genes into elite cultivars at defined locations. Meanwhile, compact growth alleles which have proven valuable in other crops hold the potential to boost potato yields. These projects exemplify how Genome Editing can drive agricultural innovation. As global populations rise and climate pressures intensify, gene editing will be an essential part of the solution – an idea increasingly reflected in evolving regulations and public perceptions.

BIBLIOGRAPHY

Adamczyk, M., & Jagura-Burdzy, G. (2003). Spread and survival of promiscuous IncP-1 plasmids. *Acta Biochimica Polonica*, *50*(2), 425–453. https://doi.org/10.18388/abp.2003_3696

Adams, J. R., de Vries, M. E., Zheng, C., & van Eeuwijk, F. A. (2022). Little heterosis found in diploid hybrid potato: The genetic underpinnings of a new hybrid crop. *G3 Genes|Genomes|Genetics*, *12*(6). https://doi.org/10.1093/G3JOURNAL/JKAC076

Agha, H. I., Shannon, L. M., & Morrell, P. L. (2023). Unloading potatoes: Potato breeding moves forward with only half the genome. *Cell Genomics*, *3*(6), 100343. https://doi.org/10.1016/J.XGEN.2023.100343

Albert, H., Dale, E. C., Lee, E., & Ow, D. W. (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *The Plant Journal*, *7*(4), 649–659. https://doi.org/10.1046/j.1365-313X.1995.7040649.x

Albert, N. W., Butelli, E., Moss, S. M. A., Piazza, P., Waite, C. N., Schwinn, K. E., Davies, K. M., & Martin, C. (2020). Discrete bHLH transcription factors play functionally overlapping roles in pigmentation patterning in flowers of *Antirrhinum majus*. *New Phytologist*, nph.17142. https://doi.org/10.1111/nph.17142

Allen, E. J. (1978). Plant density. In *The Potato Crop* (pp. 278–326). Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7210-1_7

Anderson, T. W. (1962). On the Distribution of the Two-Sample Cramer-von Mises Criterion. *The Annals of Mathematical Statistics*, *33*(3), 1148–1159. https://doi.org/10.1214/aoms/1177704477

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

Armario Najera, V., Twyman, R. M., Christou, P., & Zhu, C. (2019). Applications of multiplex genome editing in higher plants. In *Current Opinion in Biotechnology* (Vol. 59, pp. 93–102). Elsevier Ltd. https://doi.org/10.1016/j.copbio.2019.02.015

Armstrong, M. R., Whisson, S. C., Pritchard, L., Bos, J. I. B., Venter, E., Avrova, A. O., Rehmany, A. P., Böhme, U., Brooks, K., Cherevach, I., Hamlin, N., White, B., Fraser, A., Lord, A., Quail, M. A., Churcher, C., Hall, N., Berriman, M., Huang, S., … Birch, P. R. J. (2005). An ancestral oomycete locus contains late blight avirulence gene *Avr3a*, encoding a protein that is recognized in the host cytoplasm. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(21), 7766–7771. https://doi.org/10.1073/pnas.0500113102

Azizi-Dargahlou, S., & Pouresmaeil, M. (2024). *Agrobacterium tumefaciens*-Mediated Plant Transformation: A Review. *Molecular Biotechnology*, *66*(7), 1563–1580. https://doi.org/10.1007/s12033-023-00788-x

Bally, J., Nakasugi, K., Jia, F., Jung, H., Ho, S. Y. W., Wong, M., Paul, C. M., Naim, F., Wood, C. C., Crowhurst, R. N., Hellens, R. P., Dale, J. L., & Waterhouse, P. M. (2015). The extremophile *Nicotiana benthamiana* has traded viral defence for early vigour. *Nature Plants 2015 1:11*, *1*(11), 1–6. https://doi.org/10.1038/nplants.2015.165

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*(1), 151–163. https://doi.org/10.1105/TPC.113.119792

Barandalla, L., Ritter, E., & Ruiz De Galarreta, J. I. (2006). Oryzalin treatment of potato diploids yields tetraploid and chimeric plants from which euploids could be derived by callus induction. *Potato Research*, *49*(2), 143–154. https://doi.org/10.1007/s11540-006-9014-1

Beck, E., Ludwig, G., Auerswald, E. A., Reiss, B., & Schaller, H. (1982). Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. *Gene*, *19*(3), 327–336. https://doi.org/10.1016/0378-1119(82)90023-3

Begemann, M. B., Gray, B. N., January, E., Gordon, G. C., He, Y., Liu, H., Wu, X., Brutnell, T. P., Mockler, T. C., & Oufattole, M. (2017). Precise insertion and guided editing of higher plant genomes using *Cpf1* CRISPR nucleases. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-11760-6

Bethke, P. C., Halterman, D. A., & Jansky, S. (2017). Are We Getting Better at Using Wild Potato Species in Light of New Tools? *Crop Science*, *57*(3), 1241–1258. https://doi.org/10.2135/CROPSCI2016.10.0889

Bevan, M. (1984). Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Research*, *12*(22), 8711–8721. https://doi.org/10.1093/NAR/12.22.8711

Bhattacharya, D., & Van Meir, E. G. (2019). A simple genotyping method to detect small CRISPR-Cas9 induced indels by agarose gel electrophoresis. *Scientific Reports*, *9*(1), 1–7. https://doi.org/10.1038/s41598-019-39950-4

Bi, G., Su, M., Li, N., Liang, Y., Dang, S., Xu, J., Hu, M., Wang, J., Zou, M., Deng, Y., Li, Q., Huang, S., Li, J., Chai, J., He, K., Chen, Y. hang, & Zhou, J. M. (2021). The *ZAR1* resistosome is a calcium-permeable channel triggering plant immune signaling. *Cell*, *184*(13), 3528-3541.e12. https://doi.org/10.1016/J.CELL.2021.05.003

Bickmann, L., Rodriguez, M., Jiang, X., & Makalowski, W. (2023). TEclass2: Classification of transposable elements using Transformers. *BioRxiv*, 2023.10.13.562246. https://doi.org/10.1101/2023.10.13.562246

Bimboim, H. C., & Doly, J. (1979). A rapid alkaline extraction

procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, *7*(6), 1513–1523. https://doi.org/10.1093/NAR/7.6.1513

Birhman, R. K., & Hosaka, K. (2000). Production of inbred progenies of diploid potatoes using an S -locus inhibitor (*Sli*) gene, and their characterization . *Genome*, *43*(3), 495–502. https://doi.org/10.1139/g00-012

Bollinedi, H., S., G. K., Prabhu, K. V., Singh, N. K., Mishra, S., Khurana, J. P., & Singh, A. K. (2017). Molecular and Functional Characterization of GR2-R1 Event Based Backcross Derived Lines of Golden Rice in the Genetic Background of a Mega Rice Variety Swarna. *PLOS ONE*, *12*(1), e0169600. https://doi.org/10.1371/journal.pone.0169600

Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim-Van Dillen, P. M. E., & Van Der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology*, *28*(3), 495–503. https://doi.org/10.1128/JCM.28.3.495-503.1990

Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., & Zaretskaya, I. (2013). BLAST: a more efficient report with usability improvements. *Nucleic Acids Research*, *41*(W1), W29–W33. https://doi.org/10.1093/nar/gkt282

Bower, K. (2003). When to Use Fisher's Exact Test. *American Society for Quality*, *2*(4), 35–37.

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *26*(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Boyd, N. S., Gordon, R., & Martin, R. C. (2002). Relationship between leaf area index and ground cover in potato under different management conditions. *Potato Research*, *45*(2–4), 117–129. https://doi.org/10.1007/BF02736107

Burstall, L., & Harris, P. M. (1983). The estimation of percentage light interception from leaf area index and percentage ground cover in potatoes. *The Journal of Agricultural Science*, *100*(1), 241–244. https://doi.org/10.1017/S0021859600032676

Butler, N. M., Atkins, P. A., Voytas, D. F., & Douches, D. S. (2015). Generation and Inheritance of Targeted Mutations in Potato (*Solanum tuberosum* L.) Using the CRISPR/Cas System. *PLOS ONE*, *10*(12), e0144591. https://doi.org/10.1371/journal.pone.0144591

Butler, N. M., Baltes, N. J., Voytas, D. F., & Douches, D. S. (2016). Geminivirus-Mediated Genome Editing in Potato *(Solanum tuberosum* L.) Using Sequence-Specific Nucleases. *Frontiers in Plant Science*, *7*, 1045. https://doi.org/10.3389/fpls.2016.01045

Cai, D., Rodríguez, F., Teng, Y., Ané, C., Bonierbale, M., Mueller, L. A., & Spooner, D. M. (2012). Single copy nuclear

gene analysis of polyploidy in wild potatoes (*Solanum* section Petota). *BMC Evolutionary Biology*, *12*(1), 70. https://doi.org/10.1186/1471-2148-12-70

Carputo, D., Monti, L., Werner, J. E., & Frusciante, L. (1999). Uses and usefulness of endosperm balance number. *Theoretical and Applied Genetics*, *98*(3–4), 478–484. https://doi.org/10.1007/s001220051095

Carrera, E., Bou, J., Garcia-Martínez, J. L., & Prat, S. (2000). Changes in GA 20-oxidase gene expression strongly affect stem length, tuber induction and tuber yield of potato plants. *Plant Journal*, *22*(3), 247–256. https://doi.org/10.1046/j.1365-313X.2000.00736.x

Castel, B., Tomlinson, L., Locci, F., Yang, Y., & Jones, J. D. G. (2019). Optimization of T-DNA architecture for Cas9-mediated mutagenesis in *Arabidopsis*. *PLOS ONE*, *14*(1), e0204778. https://doi.org/10.1371/journal.pone.0204778

Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., & Voytas, D. F. (2015). High-frequency, precise modification of the tomato genome. *Genome Biology*, *16*(1), 232. https://doi.org/10.1186/s13059-015-0796-9

Ceulemans, E., Ibrahim, H. M. M., De Coninck, B., & Goossens, A. (2021). Pathogen Effectors: Exploiting the Promiscuity of Plant Signaling Hubs. *Trends in Plant Science*, *26*(8), 780–795. https://doi.org/10.1016/j.tplants.2021.01.005

Chandler, P. M., Marion-Poll, A., Ellis, M., & Gubler, F. (2002). Mutants at the *Slender1* Locus of Barley cv Himalaya. Molecular and Physiological Characterization. *Plant Physiology*, *129*(1), 181–190. https://doi.org/10.1104/PP.010917

Chen, P., Yang, R., Bartels, D., Dong, T., & Duan, H. (2022). Roles of Abscisic Acid and Gibberellins in Stem/Root Tuber Development. *International Journal of Molecular Sciences 2022, Vol. 23, Page 4955*, *23*(9), 4955. https://doi.org/10.3390/IJMS23094955

Chetty, V. J., García, D. J., Narváez-Vásquez, J., & Orozco-Cárdenas, M. L. (2020). Poly (ADP-ribose) polymerase inhibitor 3-methoxybenzamide enhances in vitro plant growth, microtuberization, and transformation efficiency of blue potato (*Solanum tuberosum* L. subsp. *andigenum*). *In Vitro Cellular & Developmental Biology - Plant*, *56*(6), 833–841. https://doi.org/10.1007/s11627-020-10086-9

Choi, J., Chen, W., Suiter, C. C., Lee, C., Chardon, F. M., Yang, W., Leith, A., Daza, R. M., Martin, B., & Shendure, J. (2022). Precise genomic deletions using paired prime editing. *Nature Biotechnology*, *40*(2), 218–226. https://doi.org/10.1038/s41587-021-01025-z

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A. J., & Voytas, D. F. (2010). Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics*, *186*(2), 757–761. https://doi.org/10.1534/genetics.110.120717

Cooper, M., Messina, C. D., Podlich, D., Totir, L. R., Baumgarten, A., Hausmann, N. J., Wright, D., & Graham, G. (2014). Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. *Crop and Pasture Science*, *65*(4), 311–336. https://doi.org/10.1071/CP14007

Cornu, T. I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichtinger, M., Joung, J. K., & Cathomen, T. (2008). DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Molecular Therapy*, *16*(2), 352–358. https://doi.org/10.1038/sj.mt.6300357

Cronn, R., Knaus, B. J., Liston, A., Maughan, P. J., Parks, M., Syring, J. V., & Udall, J. (2012). Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany*, *99*(2), 291–311. https://doi.org/10.3732/ajb.1100356

Dean, C., Jones, J., Favreau, M., Dunsmuir, P., & Bedbrook, J. (1988). Influence of flanking sequences on variability in expression levels of an introduced gene in transgenic tobacco plants. *Nucleic Acids Research*, *16*(19), 9267–9283. https://doi.org/10.1093/NAR/16.19.9267

Dendy, A. (1909). ON THE PRINCIPLES OF HEREDITY AS APPLIED TO THE PRODUCTION OF NEW FORMS OF PLANTS AND ANIMALS. *Journal of the Royal Society of Arts*, *57*(2947), 525–534. http://www.jstor.org/stable/41338622

Di Matteo, J. A., Ferreyra, J. M., Cerrudo, A. A., Echarte, L., & Andrade, F. H. (2016). Yield potential and yield stability of Argentine maize hybrids over 45 years of breeding. *Field Crops Research*, *197*, 107–116. https://doi.org/10.1016/j.fcr.2016.07.023

Di, X., Wang, Q., Zhang, F., Feng, H., Wang, X., & Cai, C. (2024). Advances in the Modulation of Potato Tuber Dormancy and Sprouting. *International Journal of Molecular Sciences 2024, Vol. 25, Page 5078*, *25*(10), 5078. https://doi.org/10.3390/IJMS25105078

Diamos, A. G., & Mason, H. S. (2018). Chimeric 3’ flanking regions strongly enhance gene expression in plants. *Plant Biotechnology Journal*, *16*, 1971–1982. https://doi.org/10.1111/pbi.12931

Dima, O., & Inzé, D. (2021). The role of scientists in policy making for more sustainable agriculture. *Current Biology*, *31*(5), R218–R220. https://doi.org/10.1016/j.cub.2021.01.090

Ditta, G., Stanfield, S., Corbin, D., & Helinski, D. R. (1980). Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proceedings of the National Academy of Sciences*, *77*(12), 7347–7351. https://doi.org/10.1073/pnas.77.12.7347

Djaman, K., Allen, S., Djaman, D. S., Koudahe, K., Irmak, S., Puppala, N., Darapuneni, M. K., & Angadi, S. V. (2022). Planting date and plant density effects on maize growth, yield and water use efficiency. *Environmental Challenges*, *6*, 100417. https://doi.org/10.1016/J.ENVC.2021.100417

Does, M. P., Dekker, B. M. M., de Groot, M. J. A., & Offringa, R. (1991). A quick method to estimate the T-DNA copy number in transgenic plants at an early stage after transformation, using inverse PCR. *Plant Molecular Biology*, *17*(1), 151–153. https://doi.org/10.1007/BF00036819

Douches, D. S., Maas, D., Jastrzebski, K., & Chase, R. W. (1996). Assessment of potato breeding progress in the USA over the last century. *Crop Science*, *36*(6), 1544–1552. https://doi.org/10.2135/cropsci1996.0011183X003600060024x

Du, Z., Yang, L., Zhang, D., Cui, T., He, X., Xiao, T., Li, H., Xing, S., & Xie, C. (2024). Optimizing maize planting density based on soil organic matter to achieve synergistic improvements of yield, economic benefits, and resource use efficiency. *Science of The Total Environment*, *906*, 167597. https://doi.org/10.1016/J.SCITOTENV.2023.167597

Duvick, D. N. (1999). Heterosis: Feeding People and Protecting Natural Resources. In *The Genetics and Exploitation of Heterosis in Crops* (pp. 19–29). John Wiley & Sons, Ltd. https://doi.org/10.2134/1999.geneticsandexploitation.c3

Duvick, D. N. (2001). Biotechnology in the 1930s: the development of hybrid maize. *Nature Reviews Genetics 2001 2:1*, *2*(1), 69–74. https://doi.org/10.1038/35047587

Eijlander, R., & Stiekema, W. (1994). Biological containment of potato (*Solanum tuberosum*): outcrossing to the related wild species black nightshade (*Solanum nigrum*) and bittersweet (*Solanum dulcamara*). *Sexual Plant Reproduction*, *7*(1), 29–40. https://doi.org/10.1007/BF00241885

El-Kharbotly, A., Jacobsen, E., Stiekema, W. J., & Pereira, A. (1995). Genetic localisation of transformation competence in diploid potato. *Theoretical and Applied Genetics*, *91*(4), 557–562. https://doi.org/10.1007/BF00223279

Ellis, J. G., Rafiqi, M., Gan, P., Chakrabarti, A., & Dodds, P. N. (2009). Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. *Current Opinion in Plant Biology*, *12*(4), 399–405. https://doi.org/10.1016/j.pbi.2009.05.004

Engler, C., & Marillonnet, S. (2014). Golden Gate Cloning. In *Methods in Molecular Biology* (Vol. 1116, pp. 119–131). Humana Press Inc. https://doi.org/10.1007/978-1-62703-764-8_9

European Parliament. (2024). *Plants obtained by certain new genomic techniques and their food and feed* (2023/0226(COD)). Legislative Observatory. https://oeil.secure.europarl.europa.eu/oeil/en/document-summary?id=1784332

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., … Hassabis, D. (2022). Protein complex prediction with AlphaFold-Multimer. *BioRxiv*, 2021.10.04.463034. https://doi.org/10.1101/2021.10.04.463034

Evdokimov, A. G., Pokross, M. E., Egorov, N. S., Zaraisky, A. G., Yampolsky, I. V., Merzlyak, E. M., Shkoporov, A. N., Sander, I., Lukyanov, K. A., & Chudakov, D. M. (2006). Structural basis for the fast maturation of Arthropoda green fluorescent protein. *EMBO Reports*, *7*(10), 1006–1012. https://doi.org/10.1038/sj.embor.7400787

Ewing, E. E., & Struik, P. C. (1992). Tuber formation in potato: induction, initiation, and growth. *Horticultural Reviews*, *14*, 89–198.

Fauser, F., Roth, N., Pacher, M., Ilg, G., Sánchez-Fernández, R., Biesgen, C., & Puchta, H. (2012). In planta gene targeting. *Proceedings of the National Academy of Sciences*, *109*(19), 7535–7540. https://doi.org/10.1073/pnas.1202191109

Firman, D. M., & Allen, E. J. (1989). Relationship between light interception, ground cover and leaf area index in potatoes. *The Journal of Agricultural Science*, *113*(3), 355–359. https://doi.org/10.1017/S0021859600070040

Food and Agriculture Organisation of the United Nations. (2019). *World Food and Agriculture Statistical Pocketbook*.

Forsyth, A., Weeks, T., Richael, C., & Duan, H. (2016). Transcription Activator-Like Effector Nucleases (TALEN)-Mediated Targeted DNA Insertion in Potato Plants. *Frontiers in Plant Science*, *7*, 1572. https://doi.org/10.3389/fpls.2016.01572

Foster, C. A. (1977). Slender: an accelerated extension growth mutant of barley. *Barley Genetics Newsletter*, *7*, 24–27.

Foster, S. J., Park, T. H., Pel, M., Brigneti, G., Sliwka, J., Jagger, L., Van Der Vossen, E., & Jones, J. D. G. (2009). *Rpi-vnt1.1*, a Tm-2(2) homolog from *Solanum venturii*, confers resistance to potato late blight. *Molecular Plant-Microbe Interactions : MPMI*, *22*(5), 589–600. https://doi.org/10.1094/MPMI-22-5-0589

Fry, W. (2008). *Phytophthora infestans*: the plant (and R gene) destroyer. *Molecular Plant Pathology*, *9*(3), 385–402. https://doi.org/10.1111/j.1364-3703.2007.00465.x

Fu, X., Richards, D. E., Ait-ali, T., Hynes, L. W., Ougham, H., Peng, J., & Harberd, N. P. (2002). Gibberellin-Mediated Proteasome-Dependent Degradation of the Barley DELLA Protein SLN1 Repressor. *The Plant Cell*, *14*(12), 3191–3200. https://doi.org/10.1105/TPC.006197

Game, J. C., & Mortimer, R. K. (1974). A genetic study of X-ray sensitive mutants in yeast. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *24*(3), 281–292. https://doi.org/10.1016/0027-5107(74)90176-6

Gao, Y., & Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *Journal of Integrative Plant Biology*, *56*(4), 343–349. https://doi.org/10.1111/jipb.12152

Gelvin, S. B. (2010). Plant proteins involved in *Agrobacterium*-mediated genetic transformation. *Annual Review of*

Phytopathology, *48*, 45–68. https://doi.org/10.1146/ANNUREV-PHYTO-080508-081852

Genetic Technology (Precision Breeding) Act (2023). https://bills.parliament.uk/bills/3167

Ghislain, M., Byarugaba, A. A., Magembe, E., Njoroge, A., Rivera, C., Román, M. L., Tovar, J. C., Gamboa, S., Forbes, G. A., Kreuze, J. F., Barekye, A., & Kiggundu, A. (2019). Stacking three late blight resistance genes from wild species directly into African highland potato varieties confers complete field resistance to local blight races. *Plant Biotechnology Journal*, *17*(6), 1119–1129. https://doi.org/10.1111/pbi.13042

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*(6), 1251–1262. https://doi.org/10.1111/TPJ.13446

Gilbertson, L. (2003). Cre–lox recombination: Cre-ative tools for plant biotechnology. *Trends in Biotechnology*, *21*(12), 550–555. https://doi.org/10.1016/j.tibtech.2003.09.011

Gordon, R., Brown, D. M., & Dixon, M. A. (1997). Estimating potato leaf area index for specific cultivars. *Potato Research*, *40*(3), 251–266. https://doi.org/10.1007/BF02358007

Grandgeorge, L. (2021). *Targeted DNA insertion in tomato using RNA-guided nucleases* [University of East Anglia]. https://ueaeprints.uea.ac.uk/id/eprint/84992/

Grant, G. N., Jessee, J., Bloom, F. R., & Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proceedings of the National Academy of Sciences*, *87*(12), 4645–4649. https://doi.org/10.1073/PNAS.87.12.4645

Grützner, R., Martin, P., Horn, C., Mortensen, S., Cram, E. J., Lee-Parsons, C. W. T., Stuttmann, J., & Marillonnet, S. (2021). High-efficiency genome editing in plants mediated by a *Cas9* gene containing multiple introns. *Plant Communications*, *2*(2), 100135. https://doi.org/10.1016/J.XPLC.2020.100135

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M. V., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., … Nusbaum, C. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature*, *461*(7262), 393–398. https://doi.org/10.1038/nature08358

Hameed, A., Zaidi, S. S. e. A., Shakir, S., & Mansoor, S. (2018). Applications of new breeding technologies for potato improvement. In *Frontiers in Plant Science* (Vol. 9, p. 925). Frontiers Media S.A. https://doi.org/10.3389/fpls.2018.00925

Hamilton, J. P., Brose, J., & Buell, C. R. (2024). SpudDB: a database for accessing potato genomic data. *Genetics*, *2024*, 205. https://doi.org/10.1093/GENETICS/IYAE205

Hanin, M., Mengiste, T., Bogucki, A., & Paszkowski, J. (2001). Elevated levels of intrachromosomal homologous recombination in *Arabidopsis* overexpressing the *MIM* gene. *The Plant Journal*, *24*(2), 183–189. https://doi.org/10.1046/J.1365-313X.2000.00867.X

Haverkort, A. J., Boonekamp, P. M., Hutten, R., Jacobsen, E., Lotz, L. A. P., Kessel, G. J. T., Visser, R. G. F., & Van Der Vossen, E. A. G. (2008). Societal costs of late blight in potato and prospects of durable resistance through cisgenic modification. In *Potato Research* (Vol. 51, Issue 1, pp. 47–57). Springer. https://doi.org/10.1007/s11540-008-9089-y

He, Y., Zhang, T., Sun, H., Zhan, H., & Zhao, Y. (2020). A reporter for noninvasively monitoring gene expression and plant transformation. *Horticulture Research*, *7*(1), 152. https://doi.org/10.1038/s41438-020-00390-1

Hedden, P., & Sponsel, V. (2015). A Century of Gibberellin Research. *Journal of Plant Growth Regulation 2015 34:4*, *34*(4), 740–760. https://doi.org/10.1007/S00344-015-9546-1

Hermsen, J. G. T. (1994). Introgression of genes from wild species, including molecular and cellular approaches. In J. E. Bradshaw & D. R. Mackay (Eds.), *Potato genetics.* (pp. 515–538). CAB INTERNATIONAL.

Hochholdinger, F., & Yu, P. (2025). Molecular concepts to explain heterosis in crops. *Trends in Plant Science*, *30*(1), 95–104. https://doi.org/10.1016/j.tplants.2024.07.018

Hoess, R. H., Wierzbicki, A., & Abremski, K. (1986). The role of the loxP spacer region in PI site-specific recombination. *Nucleic Acids Research*, *14*(5), 2287–2300. https://doi.org/10.1093/NAR/14.5.2287

Hori, K., & Watanabe, Y. (2007). Context Analysis of Termination Codons in mRNA that are Recognized by Plant NMD. *Plant and Cell Physiology*, *48*(7), 1072–1078. https://doi.org/10.1093/pcp/pcm075

Horsman, K., Bergervoet, J. E. M., & Jacobsen, E. (1997). Somatic hybridization between *Solanum tuberosum* and species of the *S. nigrum* complex: Selection of vigorously growing and flowering plants. *Euphytica*, *96*(3), 345–352. https://doi.org/10.1023/A:1003038419126

Horsman, K., Fratini, R., Huigen, D. J., & Jacobsen, E. (1999). Successful first and second backcrosses of *S. nigrum* (+) *S. tuberosum* somatic hybrids with both *Solanum* parents. *Sexual Plant Reproduction*, *12*(3), 144–151. https://doi.org/10.1007/s004970050185

Hosaka, K., & Hanneman, R. E. (1998). Genetics of self-compatibility in a self-incompatible wild diploid potato species *Solanum chacoense*. 2. Localization of an S locus inhibitor (*Sli*) gene on the potato genome using DNA markers. *Euphytica*, *103*(2), 265–271. https://doi.org/10.1023/A:1018380725160

Hosmani, P. S., Flores-Gonzalez, M., Geest, H. van de, Maumus, F.,

Bakker, L. V., Schijlen, E., Haarst, J. van, Cordewener, J., Sanchez-Perez, G., Peters, S., Fei, Z., Giovannoni, J. J., Mueller, L. A., & Saha, S. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *BioRxiv*, 767764. https://doi.org/10.1101/767764

Hougas, R. W., & Peloquin, S. J. (1958). The potential of potato haploids in breeding and genetic research. *American Potato Journal*, *35*(10), 701–707. https://doi.org/10.1007/BF02855564

Hougas, R. W., & Peloquin, S. J. (1960). Crossability of *Solanum tuberosum* haploids with diploid *Solanum* species. *European Potato Journal 1960 3:4*, *3*(4), 325–330. https://doi.org/10.1007/BF02365602

Huang, T. K., Armstrong, B., Schindele, P., & Puchta, H. (2021). Efficient gene targeting in *Nicotiana tabacum* using CRISPR/SaCas9 and temperature tolerant LbCas12a. *Plant Biotechnology Journal*, *19*(7), 1314–1324. https://doi.org/10.1111/PBI.13546

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

Hutten, R. C. B. (1994). *Basic aspects of potato breeding via the diploid level* [Agricultural University]. https://library.wur.nl/WebQuery/wurpubs/27413

Ifenkwe, O. P., & Allen, E. J. (1978). Effects of row width and planting density on growth and yield of two maincrop potato varieties. 1. Plant morphology and dry-matter accumulation. *The Journal of Agricultural Science*, *91*(2), 265–278. https://doi.org/10.1017/S0021859600046359

Ikeda, A., Ueguchi-Tanaka, M., Sonoda, Y., Kitano, H., Koshioka, M., Futsuhara, Y., Matsuoka, M., & Yamaguchi, J. (2001). slender Rice, a Constitutive Gibberellin Response Mutant, Is Caused by a Null Mutation of the *SLR1* Gene, an Ortholog of the Height-Regulating Gene *GAI/RGA/RHT/D8*. *The Plant Cell*, *13*(5), 999–1010. https://doi.org/10.1105/TPC.13.5.999

Inglis, P. W., Pappas, M. de C. R., Resende, L. V., & Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLOS ONE*, *13*(10), e0206085. https://doi.org/10.1371/journal.pone.0206085

Ingram, L. C., Richmond, M. H., & Sykes, R. B. (1973). Molecular Characterization of the R Factors Implicated in the Carbenicillin Resistance of a Sequence of *Pseudomonas aeruginosa* Strains Isolated from Burns. *Antimicrobial Agents and Chemotherapy*, *3*(2), 279–288. https://doi.org/10.1128/AAC.3.2.279

Isalan, M., Choo, Y., & Klug, A. (1997). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proceedings of the National Academy of Sciences of the United States of*

America, *94*(11), 5617–5621. https://doi.org/10.1073/pnas.94.11.5617

Ivins, J. D., & Bremner, P. M. (1965). Growth, Development and Yield in the Potato. *Outlook on Agriculture*, *4*(5), 211–217. https://doi.org/10.1177/003072706500400503

Jackson, S. D., & Prat, S. (1996). Control of tuberisation in potato by gibberellins and phytochrome B. *Physiologia Plantarum*, *98*(2), 407–412. https://doi.org/10.1034/J.1399-3054.1996.980224.X

Jacobsen, E., & Schouten, H. J. (2007). Cisgenesis strongly improves introgression breeding and induced translocation breeding of plants. *Trends in Biotechnology*, *25*(5), 219–223. https://doi.org/10.1016/j.tibtech.2007.03.008

Jansky, S., & Hamernik, A. (2009). The introgression of 2× 1EBN *Solanum* species into the cultivated potato using *Solanum verrucosum* as a bridge. *Genetic Resources and Crop Evolution*, *56*(8), 1107–1115. https://doi.org/10.1007/s10722-009-9433-3

Janssen, B. J., & Gardner, R. C. (1990). Localized transient expression of GUS in leaf discs following cocultivation with *Agrobacterium*. *Plant Molecular Biology*, *14*(1), 61–72. https://doi.org/10.1007/BF00015655

Jefferies, R. A., & MacKerron, D. K. L. (1993). Responses of potato genotypes to drought. II. Leaf area index, growth and yield. *Annals of Applied Biology*, *122*(1), 105–112. https://doi.org/10.1111/j.1744-7348.1993.tb04018.x

Jensen, M. A., & Straus, N. (1993). Effect of PCR conditions on the formation of heteroduplex and single-stranded DNA products in the amplification of bacterial ribosomal DNA spacer regions. *PCR Methods and Applications*, *3*(3), 186–194. https://doi.org/10.1101/GR.3.3.186

Jin, F., Li, S., Dang, L., Chai, W., Li, P., & Wang, N. N. (2012). PL1 fusion gene: A novel visual selectable marker gene that confers tolerance to multiple abiotic stresses in transgenic tomato. *Transgenic Research*, *21*(5), 1057–1070. https://doi.org/10.1007/s11248-012-9592-4

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*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jo, K. R., Kim, C. J., Kim, S. J., Kim, T. Y., Bergervoet, M., Jongsma, M. A., Visser, R. G. F., Jacobsen, E., & Vossen, J. H. (2014). Development of late blight resistant potatoes by cisgene stacking. *BMC Biotechnology*, *14*(1), 1–10. https://doi.org/10.1186/1472-6750-14-50

Johnston, S. A., & Hanneman, R. E. (1982). Manipulations of Endosperm Balance Number Overcome Crossing Barriers Between Diploid *Solanum* Species. *Science*, *217*(4558), 446–448. https://doi.org/10.1126/SCIENCE.217.4558.446

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323–329.

https://doi.org/10.1038/nature05286

Jones, J. D. G., Gilbert, D. E., Grady, K. L., & Jorgensen, R. A. (1987). T-DNA structure and gene expression in petunia plants transformed by *Agrobacterium tumefaciens* C58 derivatives. *Molecular and General Genetics MGG*, *207*(2–3), 478–485. https://doi.org/10.1007/BF00331618

Joung, J. K., & Sander, J. D. (2013). TALENs: A widely applicable technology for targeted genome editing. In *Nature Reviews Molecular Cell Biology* (Vol. 14, Issue 1, pp. 49–55). NIH Public Access. https://doi.org/10.1038/nrm3486

Jupe, F., Witek, K., Verweij, W., Śliwka, J., Pritchard, L., Etherington, G. J., Maclean, D., Cock, P. J., Leggett, R. M., Bryan, G. J., Cardle, L., Hein, I., & Jones, J. D. G. (2013). Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant Journal*, *76*(3), 530–544. https://doi.org/10.1111/tpj.12307

Khan, M., Subramaniam, R., & Desveaux, D. (2016). Of guards, decoys, baits and traps: Pathogen perception in plants by type III effector sensors. *Current Opinion in Microbiology, 29,* 49-55. https://doi.org/10.1016/j.mib.2015.10.006

Khurana, S. C., & McLaren, J. S. (1982). The influence of leaf area, light interception and season on potato growth and yield. *Potato Research*, *25*(4), 329–342. https://doi.org/10.1007/BF02357290

Kim, H. K., Min, S., Song, M., Jung, S., Choi, J. W., Kim, Y., Lee, S., Yoon, S., & Kim, H. (2018). Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nature Biotechnology*, *36*(3), 239–241. https://doi.org/10.1038/nbt.4061

Kim, J. H., Lee, S. R., Li, L. H., Park, H. J., Park, J. H., Lee, K. Y., Kim, M. K., Shin, B. A., & Choi, S. Y. (2011). High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice. *PLOS ONE*, *6*(4), e18556. https://doi.org/10.1371/JOURNAL.PONE.0018556

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: Zinc finger fusions to *Fok* I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(3), 1156–1160. https://doi.org/10.1073/pnas.93.3.1156

Kondrák, M., van der Meer, I. M., & Bánfalvi, Z. (2006). Generation of Marker- and Backbone-Free Transgenic Potatoes by Site-Specific Recombination and a Bi-Functional Marker Gene in a Non-Regular One-Border *Agrobacterium* Transformation Vector. *Transgenic Research*, *15*(6), 729–737. https://doi.org/10.1007/s11248-006-9021-7

Kornacki, J. A., West, A. H., & Firshein, W. (1984). Proteins encoded by the trans-acting replication and maintenance regions of broad host range plasmid RK2. *Plasmid*, *11*(1), 48–

57. https://doi.org/10.1016/0147-619X(84)90006-4

Kralemann, L. E. M., de Pater, S., Shen, H., Kloet, S. L., van Schendel, R., Hooykaas, P. J. J., & Tijsterman, M. (2022). Distinct mechanisms for genomic attachment of the 5' and 3' ends of *Agrobacterium* T-DNA in plants. *Nature Plants*. https://doi.org/10.1038/s41477-022-01147-5

Kralemann, L. E. M., van Tol, N., Hooykaas, P. J. J., & Tijsterman, M. (2024). Molecular analysis of the role of polymerase theta in gene targeting in *Arabidopsis thaliana*. *The Plant Journal*, *118*(1), 255–262. https://doi.org/10.1111/tpj.16689

Kwon, T., Huq, E., & Herrin, D. L. (2010). Microhomology-mediated and nonhomologous repair of a double-strand break in the chloroplast genome of *Arabidopsis*. *Proceedings of the National Academy of Sciences*, *107*(31), 13954–13959. https://doi.org/10.1073/pnas.1004326107

Labroo, M. R., Studer, A. J., & Rutkoski, J. E. (2021). Heterosis and Hybrid Crop Breeding: A Multidisciplinary Review. *Frontiers in Genetics*, *12*, 643761. https://doi.org/10.3389/fgene.2021.643761

Lawrenson, T., Clarke, M., Kirby, R., Forner, M., Steuernagel, B., Brown, J. K. M., & Harwood, W. (2024). An optimised CRISPR Cas9 and Cas12a mutagenesis toolkit for Barley and Wheat. *Plant Methods*, *20*(1). https://doi.org/10.1186/S13007-024-01234-Y

Lawrenson, T., Hinchliffe, A., Clarke, M., Morgan, Y., & Harwood, W. (2021). In-planta Gene Targeting in Barley Using Cas9 With and Without Geminiviral Replicons. *Frontiers in Genome Editing*, *3*, 663380. https://doi.org/10.3389/fgeed.2021.663380

Lazo, G. R., Stein, P. A., & Ludwig, R. A. (1991). A DNA Transformation–Competent *Arabidopsis* Genomic Library in *Agrobacterium*. *Bio/Technology*, *9*(10), 963–967. https://doi.org/10.1038/nbt1091-963

Lee, G., & Saito, I. (1998). Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene*, *216*(1), 55–65. https://doi.org/10.1016/S0378-1119(98)00325-4

Li, B., Shang, Y., Wang, L., Lv, J., Wang, F., Chao, J., Mao, J., Ding, A., Wu, X., Cui, M., Sun, Y., & Dai, C. (2022). Highly efficient multiplex genome editing in dicot plants using improved CRISPR/Cas systems. *BioRxiv*, 2022.01.10.475635. https://doi.org/10.1101/2022.01.10.475635

Li, J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., Li, J., & Gao, C. (2016). Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. *Nature Plants*, *2*(10), 1–6. https://doi.org/10.1038/nplants.2016.139

Li, J., Zhang, S., Zhang, R., Gao, J., Qi, Y., Song, G., Li, W., Li, Y., & Li, G. (2020). Efficient multiplex genome editing by CRISPR/Cas9 in common wheat. *Plant Biotechnology Journal*, *19*(3), 427. https://doi.org/10.1111/PBI.13508

Li, J., Zhang, X., Sun, Y., Zhang, J., Du, W., Guo, X., Li, S.,

Zhao, Y., & Xia, L. (2018). Efficient allelic replacement in rice by gene editing: A case study of the *NRT1.1B* gene. *Journal of Integrative Plant Biology*, *60*(7), 536–540. https://doi.org/10.1111/jipb.12650

Li, S., Li, J., Zhang, J., Du, W., Fu, J., Sutar, S., Zhao, Y., & Xia, L. (2018). Synthesis-dependent repair of Cpf1-induced double strand DNA breaks enables targeted gene replacement in rice. *Journal of Experimental Botany*, *69*(20), 4715–4721. https://doi.org/10.1093/jxb/ery245

Liang, Z., & Tzfira, T. (2013). *In vivo* formation of double-stranded T-DNA molecules by T-strand priming. *Nature Communications*, *4*(1), 2253. https://doi.org/10.1038/ncomms3253

Lin, Q., Jin, S., Zong, Y., Yu, H., Zhu, Z., Liu, G., Kou, L., Wang, Y., Qiu, J.-L., Li, J., & Gao, C. (2021). High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nature Biotechnology*, 1–5. https://doi.org/10.1038/s41587-021-00868-w

Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., Wang, Y., Anzalone, A. V., Raguram, A., Doman, J. L., Liu, D. R., & Gao, C. (2020). Prime genome editing in rice and wheat. *Nature Biotechnology*, *38*(5), 582–585. https://doi.org/10.1038/s41587-020-0455-x

Lin, X., Armstrong, M., Baker, K., Wouters, D., Visser, R. G. F., Wolters, P. J., Hein, I., & Vleeshouwers, V. G. A. A. (2020). RLP/K enrichment sequencing; a novel method to identify receptor-like protein (*RLP*) and receptor-like kinase (*RLK*) genes. *New Phytologist*, *227*(4), 1264–1276. https://doi.org/10.1111/nph.16608

Lin, X., Olave-Achury, A., Heal, R., Pais, M., Witek, K., Ahn, H.-K., Zhao, H., Bhanvadia, S., Karki, H. S., Song, T., Wu, C., Adachi, H., Kamoun, S., Vleeshouwers, V. G. A. A., & Jones, J. D. G. (2022). A potato late blight resistance gene protects against multiple *Phytophthora* species by recognizing a broadly conserved RXLR-WY effector. *Molecular Plant*, *15*(9), 1457–1469. https://doi.org/10.1016/j.molp.2022.07.012

Lin, X., Song, T., Fairhead, S., Witek, K., Jouet, A., Jupe, F., Witek, A. I., Karki, H. S., Vleeshouwers, V. G. A. A., Hein, I., & Jones, J. D. G. (2020). Identification of *Avramr1* from *Phytophthora infestans* using long read and cDNA pathogen-enrichment sequencing (PenSeq). *Molecular Plant Pathology*, *00*, 1–11. https://doi.org/10.1111/MPP.12987

Lindhout, P., De Vries, M., Ter Maat, M., Ying, S., Viquez-Zamora, M., & Van Heusden, S. (2018). Hybrid potato breeding for improved varieties. In G. Wang-Pruski (Ed.), *Achieving sustainable cultivation of potatoes* (First edit, Vol. 1). Burleigh Dodds Science Publishing Limited.

Lindhout, P., Meijer, D., Schotte, T., Hutten, R. C. B., Visser, R. G. F., & van Eck, H. J. (2011). Towards F1 Hybrid Seed Potato Breeding. *Potato Research*, *54*(4), 301–312. https://doi.org/10.1007/s11540-011-9196-z

Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., & Ma, W.

(2019). Methodologies for Improving HDR Efficiency. *Frontiers in Genetics*, *9*(JAN), 422460. https://doi.org/10.3389/fgene.2018.00691

Liu, P., Panda, K., Edwards, S. A., Swanson, R., Yi, H., Pandesha, P., Hung, Y.-H., Klaas, G., Ye, X., Collins, M. V, Renken, K. N., Gilbertson, L. A., Veena, V., Hancock, C. N., & Slotkin, R. K. (2024). Transposase-assisted target-site integration for efficient plant genome engineering. *Nature*, *631*(8021), 593–600. https://doi.org/10.1038/s41586-024-07613-8

Liu, Q., Wu, K., Harberd, N. P., & Fu, X. (2021). Green Revolution DELLAs: From translational reinitiation to future sustainable agriculture. *Molecular Plant*, *14*(4), 547–549. https://doi.org/10.1016/j.molp.2021.03.015

Liu, X., Xiang, L. li, Yin, X. ren, Grierson, D., Li, F., & Chen, K. song. (2015). The identification of a MYB transcription factor controlling anthocyanin biosynthesis regulation in *Chrysanthemum* flowers. *Scientia Horticulturae*, *194*, 278–285. https://doi.org/10.1016/j.scienta.2015.08.018

Liu, Y., Lin-Wang, K., Espley, R. V., Wang, L., Yang, H., Yu, B., Dare, A., Varkonyi-Gasic, E., Wang, J., Zhang, J., Wang, D., & Allan, A. C. (2016). Functional diversification of the potato R2R3 MYB anthocyanin activators AN1, MYBA1, and MYB113 and their interaction with basic helix-loop-helix cofactors. *Journal of Experimental Botany*, *67*(8), 2159–2176. https://doi.org/10.1093/jxb/erw014

Livne, S., Lor, V. S., Nir, I., Eliaz, N., Aharoni, A., Olszewski, N. E., Eshed, Y., & Weiss, D. (2015). Uncovering DELLA-Independent Gibberellin Responses by Characterizing New Tomato *procera* Mutants. *The Plant Cell*, *27*(6), 1579. https://doi.org/10.1105/TPC.114.132795

Lor, V. S., Starker, C. G., Voytas, D. F., Weiss, D., & Olszewski, N. E. (2014). Targeted mutagenesis of the tomato *PROCERA* gene using transcription activator-like effector nucleases. *Plant Physiology*, *166*(3), 1288–1291. https://doi.org/10.1104/PP.114.247593

Lyznik, L. A., Gordon-Kamm, W. J., & Tao, Y. (2003). Site-specific recombination for genetic engineering in plants. *Plant Cell Reports*, *21*(10), 925–932. https://doi.org/10.1007/s00299-003-0616-7

Ma, S., Lapin, D., Liu, L., Sun, Y., Song, W., Zhang, X., Logemann, E., Yu, D., Wang, J., Jirschitzka, J., Han, Z., Schulze-Lefert, P., Parker, J. E., & Chai, J. (2020). Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science*, *370*(6521). https://doi.org/10.1126/science.abe3069

Mait, I. B., Dar, S. D., Me, A., Ow Da, G., Kiernan, J. N., Ghosh, S. K., & Shepherd, R. J. (1997). Promoter/leader deletion analysis and plant expression vectors with the figwort mosaic virus (FMV) full length transcript (FLt) promoter containing single or double enhancer domains. *Transgenic Research*, *6*, 143–156.

https://doi.org/https://doi.org/10.1023/A:1018477705019

Maréchal, A., Parent, J. S., Véronneau-Lafortune, F., Joyeux, A., Lang, B. F., & Brisson, N. (2009). Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(34), 14693. https://doi.org/10.1073/PNAS.0901710106

Marinus, J., & Bodlaender, K. B. A. (1978). Growth and yield of seed potatoes after application of gibberellic acid to the tubers before planting. *Netherlands Journal of Agricultural Science*, *26*(4), 354–365. https://doi.org/10.18174/NJAS.V26I4.17078

Martin, R., Qi, T., Zhang, H., Liu, F., King, M., Toth, C., Nogales, E., & Staskawicz, B. J. (2020). Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science*, *370*(6521). https://doi.org/10.1126/science.abd9993

Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., & Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature Biotechnology*, *33*(5), 538–542. https://doi.org/10.1038/nbt.3190

Masuelli, R. W., & Camadro, E. L. (1997). Crossability relationships among wild potato species with different ploidies and Endosperm Balance Numbers (EBN). *Euphytica*, *94*(2), 227–235. https://doi.org/10.1023/A:1002996131315

Mathews, H., Clendennen, S. K., Caldwell, C. G., Liu, X. L., Connors, K., Matheis, N., Schuster, D. K., Menasco, D. J., Wagoner, W., Lightner, J., & Wagner, D. R. (2003). Activation Tagging in Tomato Identifies a Transcriptional Regulator of Anthocyanin Biosynthesis, Modification, and Transport. *The Plant Cell*, *15*(8), 1689–1703. https://doi.org/10.1105/TPC.012963

McGuffie, M. J., & Barrick, J. E. (2021). pLannotate: engineered plasmid annotation. *Nucleic Acids Research*, *49*(W1), W516–W522. https://doi.org/10.1093/NAR/GKAB374

McLellan, H., Harvey, S. E., Steinbrenner, J., Armstrong, M. R., He, Q., Clewes, R., Pritchard, L., Wang, W., Wang, S., Nussbaumer, T., Dohai, B., Luo, Q., Kumari, P., Duan, H., Roberts, A., Boevink, P. C., Neumann, C., Champouret, N., Hein, I., … J Birch, P. R. (2022). Exploiting breakdown in nonhost effector–target interactions to boost host disease resistance. *Proceedings of the National Academy of Sciences*, *119*(35), e2114064119. https://doi.org/10.1073/PNAS.2114064119

Meyer, R. J., & Helinski, D. R. (1977). Unidirectional replication of the P-group plasmid RK2. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, *478*(1), 109–113. https://doi.org/10.1016/0005-2787(77)90249-0

Mikhailov, A. (2019). *Turbo, An Improved Rainbow Colormap for Visualization*. Google Research Blog. https://research.google/blog/turbo-an-improved-rainbow-colormap-for-visualization/

Milner, M. J., Sharma, M., Bates, R. E., Whiting, M., Craze, M. S., Miller, P., Brooks, J., Kouidri, A., & Wallington, E. J. (2024). Differential editing efficiencies in cereal crops: a comparative analysis of tRNA and ribozyme multiplexed guide delivery. *Frontiers in Plant Science*, *15*, 1426184. https://doi.org/10.3389/fpls.2024.1426184

Miroshnichenko, D., Klementyeva, A., Sidorova, T., Pushin, A. S., & Dolgov, S. (2024). Genetic Transformation of Potato without Antibiotic-Assisted Selection. *Horticulturae*, *10*(3), 222. https://doi.org/10.3390/horticulturae10030222

Mlynárová, L., & Nap, J.-P. (2003). A self-excising Cre recombinase allows efficient recombination of multiple ectopic heterospecific lox sites in transgenic tobacco. *Transgenic Research*, *12*(1), 45–57. https://doi.org/10.1023/A:1022112221427

Murase, K., Hirano, Y., Sun, T., & Hakoshima, T. (2008). Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature*, *456*(7221), 459–463. https://doi.org/10.1038/nature07519

Mussolino, C., Morbitzer, R., Lütge, F., Dannemann, N., Lahaye, T., & Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Research*, *39*(21), 9283–9293. https://doi.org/10.1093/nar/gkr597

Nakayasu, M., Akiyama, R., Lee, H. J., Osakabe, K., Osakabe, Y., Watanabe, B., Sugimoto, Y., Umemoto, N., Saito, K., Muranaka, T., & Mizutani, M. (2018). Generation of α-solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the *St16DOX* gene. *Plant Physiology and Biochemistry*, *131*, 70–77. https://doi.org/10.1016/j.plaphy.2018.04.026

Navarro, L., Bari, R., Achard, P., Lisón, P., Nemri, A., Harberd, N. P., & Jones, J. D. G. (2008). DELLAs Control Plant Immune Responses by Modulating the Balance of Jasmonic Acid and Salicylic Acid Signaling. *Current Biology*, *18*(9), 650–655. https://doi.org/10.1016/j.cub.2008.03.060

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G., & Kamoun, S. (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature Biotechnology 2013 31:8*, *31*(8), 691–693. https://doi.org/10.1038/nbt.2655

Nelson, J. W., Randolph, P. B., Shen, S. P., Everette, K. A., Chen, P. J., Anzalone, A. V., An, M., Newby, G. A., Chen, J. C., Hsu, A., & Liu, D. R. (2021). Engineered pegRNAs improve prime editing efficiency. *Nature Biotechnology 2021 40:3*, *40*(3), 402–410. https://doi.org/10.1038/s41587-021-01039-7

New England Biolabs. (2020). Electroporation Protocol (C2986) v2. *Protocols.Io*. https://doi.org/10.17504/protocols.io.bccfistn

Nicolia, A., Fält, A.-S., Hofvander, P., & Andersson, M. (2021). Protoplast-Based Method for Genome Editing in Tetraploid Potato. In Tripodi, P. (eds) *Crop Breeding. Methods in Molecular Biology* (pp. 177–186). Humana, New York, NY.

Nicolia, A., Proux-Wéra, E., Åhman, I., Onkokesung, N., Andersson, M., Andreasson, E., & Zhu, L. H. (2015). Targeted gene mutation in tetraploid potato through transient TALEN expression in protoplasts. *Journal of Biotechnology*, *204*, 17–24. https://doi.org/10.1016/j.jbiotec.2015.03.021

Nishiguchi, R., Takanami, M., & Oka, A. (1987). Characterization and sequence determination of the replicator region in the hairy-root-inducing plasmid pRiA 4b. *Molecular and General Genetics MGG*, *206*(1), 1–8. https://doi.org/10.1007/BF00326529

Nishizawa-Yokoi, A., Nonaka, S., Saika, H., Kwon, Y. I., Osakabe, K., & Toki, S. (2012). Suppression of Ku70/80 or Lig4 leads to decreased stable transformation and enhanced homologous recombination in rice. *New Phytologist*, *196*(4), 1048–1059. https://doi.org/10.1111/J.1469-8137.2012.04350.X

Odell, J. T., Nagy, F., & Chua, N. H. (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature*, *313*(6005), 810–812. https://doi.org/10.1038/313810a0

Oh, S., Kim, S., oPark, H. J., Kim, M. S., Seo, M. K., Wu, C. H., Lee, H. A., Kim, H. S., Kamoun, S., & Choi, D. (2023). Nucleotide-binding leucine-rich repeat network underlies nonhost resistance of pepper against the Irish potato famine pathogen *Phytophthora infestans*. *Plant Biotechnology Journal*. https://doi.org/10.1111/PBI.14039

Ohgane, K., & Yoshioka, H. (2019). Quantification of Gel Bands by an Image J Macro, Band/Peak Quantification Tool. *Protocols.Io*. https://doi.org/dx.doi.org/10.17504/protocols.io.7vghn3w

Oltmanns, H., Frame, B., Lee, L. Y., Johnson, S., Li, B., Wang, K., & Gelvin, S. B. (2009). Generation of Backbone-Free, Low Transgene Copy Plants by Launching T-DNA from the *Agrobacterium* Chromosome. *Plant Physiology*, *152*(3), 1158. https://doi.org/10.1104/PP.109.148585

Ordoñez, B. (2014). *Brochure: Pollen viability assessment.* https://cgspace.cgiar.org/bitstream/handle/10568/65227/78083.pdf

Ow, D. W., Wood, K. V., Deluca, M., De Wet, J. R., Helinski, D. R., & Howell, S. H. (1986). Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science*, *234*(4778), 856–859. https://doi.org/10.1126/SCIENCE.234.4778.856

Paciorek, T., Chiapelli, B. J., Wang, J. Y., Paciorek, M., Yang, H., Sant, A., Val, D. L., Boddu, J., Liu, K., Gu, C., Brzostowski, L. F., Wang, H., Allen, E. M., Dietrich, C. R., Gillespie, K. M., Edwards, J., Goldshmidt, A., Neelam, A., & Slewinski, T. L. (2022). Targeted suppression of gibberellin biosynthetic genes *ZmGA20ox3* and *ZmGA20ox5* produces a short stature maize ideotype. *Plant Biotechnology Journal*, *20*(6), 1140–1153. https://doi.org/10.1111/PBI.13797

Parent, J. S., Lepage, E., & Brisson, N. (2011). Divergent Roles for the Two PolI-Like Organelle DNA Polymerases of *Arabidopsis*. *Plant Physiology*, *156*(1), 254–262.

https://doi.org/10.1104/PP.111.173849

Paszkowski, J., Baur, M., Bogucki, A., Potrykus, I., & Rochaix, J.-D. (1988). Gene targeting in plants. *The EMBO Journal*, *7*(13), 4021–4026. https://doi.org/10.1002/J.1460-2075.1988.TB03295.X

Peach, C., & Velten, J. (1991). Transgene expression variability (position effect) of *CAT* and *GUS* reporter genes driven by linked divergent T-DNA promoters. *Plant Molecular Biology*, *17*(1), 49–60.

Pel, M. A., Foster, S. J., Park, T.-H., Rietman, H., van Arkel, G., Jones, J. D. G., Van Eck, H. J., Jacobsen, E., Visser, R. G. F., & Van der Vossen, E. A. G. (2009). Mapping and Cloning of Late Blight Resistance Genes from *Solanum venturii* Using an Interspecific Candidate Gene Approach. *Molecular Plant-Microbe Interactions*, *22*(5), 601–615. https://doi.org/10.1094/MPMI-22-5-0601

Peloquin, S. J., Gabert, A. C., & Ortiz, R. (1996). Nature of "pollinator" effect in potato (*Solanum tuberosum* L.) haploid production. *Annals of Botany*, *77*(5), 539–542. https://doi.org/10.1006/anbo.1996.0064

Peng, J., Carol, P., Richards, D. E., King, K. E., Cowling, R. J., Murphy, G. P., & Harberd, N. P. (1997). The *Arabidopsis GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes & Development*, *11*(23), 3194–3205. https://doi.org/10.1101/GAD.11.23.3194

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M. D., & Harberd, N. P. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature*, *400*(6741), 256–261. https://doi.org/10.1038/22307

Pereman, I., Melamed-Bessudo, C., Dahan-Meir, T., Herz, E., Elbaum, M., & Levy, A. A. (2019). Single Molecule Imaging of T-DNA Intermediates Following *Agrobacterium tumefaciens* Infection in *Nicotiana benthamiana*. *International Journal of Molecular Sciences*, *20*(24), 6209. https://doi.org/10.3390/ijms20246209

Pérez-González, A., & Caro, E. (2018). Effect of transcription terminator usage on the establishment of transgene transcriptional gene silencing. *BMC Research Notes 2018 11:1*, *11*(1), 1–8. https://doi.org/10.1186/S13104-018-3649-2

Perroud, P. F., Guyon-Debast, A., Veillet, F., Kermarrec, M. P., Chauvin, L., Chauvin, J. E., Gallois, J. L., & Nogué, F. (2022). Prime Editing in the model plant *Physcomitrium patens* and its potential in the tetraploid potato. *Plant Science*, *316*, 111162. https://doi.org/10.1016/J.PLANTSCI.2021.111162

Pham, G. M., Hamilton, J. P., Wood, J. C., Burke, J. T., Zhao, H., Vaillancourt, B., Ou, S., Jiang, J., & Robin Buell, C. (2020). Construction of a chromosome-scale long-read reference genome assembly for potato. *GigaScience*, *9*(9). https://doi.org/10.1093/GIGASCIENCE/GIAA100

Phumichai, C., Mori, M., Kobayashi, A., Kamijima, O., & Hosaka, K. (2005). Toward the development of highly homozygous diploid potato lines using the self-compatibility controlling *Sli* gene. *Genome*, *48*(6), 977–984. https://doi.org/10.1139/g05-066

Pok Man Ngou, B., Ahn, H.-K., Ding, P., & G Jones, J. D. (2021). Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature*, 1–6. https://doi.org/10.1038/s41586-021-03315-7

Port, F., & Bullock, S. L. (2016). Augmenting CRISPR applications in *Drosophila* with tRNA-flanked sgRNAs. *Nature Methods*, *13*(10), 852–854. https://doi.org/10.1038/NMETH.3972

Powell, A. F., Feder, A., Li, J., Schmidt, M. H. W., Courtney, L., Alseekh, S., Jobson, E. M., Vogel, A., Xu, Y., Lyon, D., Dumschott, K., McHale, M., Sulpice, R., Bao, K., Lal, R., Duhan, A., Hallab, A., Denton, A. K., Bolger, M. E., … Usadel, B. (2022). A *Solanum lycopersicoides* reference genome facilitates insights into tomato specialized metabolism and immunity. *The Plant Journal*, *110*(6), 1791–1810. https://doi.org/10.1111/TPJ.15770

Puchta, H. (2017). Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Current Opinion in Plant Biology*, *36*, 1–8. https://doi.org/10.1016/J.PBI.2016.11.011

Puchta, H., Dujon, B., & Hohn, B. (1996). Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proceedings of the National Academy of Sciences*, *93*(10), 5055–5060. https://doi.org/10.1073/PNAS.93.10.5055

Puchta, H., Swoboda, P., & Hohn, B. (1995). Induction of intrachromosomal homologous recombination in whole plants. *The Plant Journal*, *7*(2), 203–210. https://doi.org/10.1046/j.1365-313X.1995.7020203.x

Qi, Y., Zhang, Y., Zhang, F., Baller, J. A., Cleland, S. C., Ryu, Y., Starker, C. G., & Voytas, D. F. (2013). Increasing frequencies of site-specific mutagenesis and gene targeting in *Arabidopsis* by manipulating DNA repair pathways. *Genome Research*, *23*(3), 547–554. https://doi.org/10.1101/gr.145557.112

Rakosy-Tican, E., Thieme, R., König, J., Nachtigall, M., Hammann, T., Denes, T.-E., Kruppa, K., & Molnár-Láng, M. (2020). Introgression of Two Broad-Spectrum Late Blight Resistance Genes, *Rpi-Blb1* and *Rpi-Blb3*, From *Solanum bulbocastanum* Dun Plus Race-Specific R Genes Into Potato Pre-breeding Lines. *Frontiers in Plant Science*, *11*, 501103. https://doi.org/10.3389/fpls.2020.00699

Romano, A., Raemakers, K., Visser, R., & Mooibroek, H. (2001). Transformation of potato (*Solanum tuberosum*) using particle bombardment. *Plant Cell Reports*, *20*(3), 198–204. https://doi.org/10.1007/s002990000314

Rommens, C. M., Richael, C. M., Yan, H., Navarre, D. A., Ye, J., Krucker, M., & Swords, K. (2008). Engineered native pathways for high kaempferol and caffeoylquinate production in potato.

Plant Biotechnology Journal, *6*(9), 870–886. https://doi.org/10.1111/J.1467-7652.2008.00362.X

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Transformation of *E. coli* by High-voltage Electroporation (Electrotransformation). In *Molecular cloning: A laboratory manual. Volume 1* (Second Edi, p. 1.74). Cold Spring Harbor Laboratory Press.

Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., Kaneko, T., Nakamura, Y., Shibata, D., Aoki, K., Egholm, M., Knight, J., Bogden, R., Li, C. C., Shuang, Y., Xu, X., Pan, S., Cheng, S., Liu, X., … Gianese, G. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, *485*(7400), 635–641. https://doi.org/10.1038/nature11119

Saville, R. J., Gosman, N., Burt, C. J., Makepeace, J., Steed, A., Corbitt, M., Chandler, E., Brown, J. K. M., Boulton, M. I., & Nicholson, P. (2012). The 'Green Revolution' dwarfing genes play a role in disease resistance in *Triticum aestivum* and *Hordeum vulgare*. *Journal of Experimental Botany*, *63*(3), 1271–1283. https://doi.org/10.1093/JXB/ERR350

Schell, J., & Van Montagu, M. (1977). The Ti-Plasmid of *Agrobacterium Tumefaciens*, A Natural Vector for the Introduction of NIF Genes in Plants? *Basic Life Sciences*, *9*, 159–179. https://doi.org/10.1007/978-1-4684-0880-5_12

Schenk, J. J., Becklund, L. E., Carey, S. J., & Fabre, P. P. (2023). What is the "modified" CTAB protocol? Characterizing modifications to the CTAB DNA extraction protocol. *Applications in Plant Sciences*, *11*(3), e11517. https://doi.org/10.1002/APS3.11517

Schindele, P., Merker, L., Schreiber, T., Prange, A., Tissier, A., & Puchta, H. (2023). Enhancing gene editing and gene targeting efficiencies in *Arabidopsis thaliana* by using an intron-containing version of tt*Lb*Cas12a. *Plant Biotechnology Journal*, *21*(3), 457–459. https://doi.org/10.1111/pbi.13964

Schindele, P., & Puchta, H. (2020). Engineering CRISPR/*Lb*Cas12a for highly efficient, temperature-tolerant plant gene editing. *Plant Biotechnology Journal*, *18*(5), 1118–1120. https://doi.org/10.1111/pbi.13275

Schmey, T., Tominello-Ramirez, C. S., Brune, C., & Stam, R. (2024). *Alternaria* diseases on potato and tomato. *Molecular Plant Pathology*, *25*(3), e13435. https://doi.org/10.1111/MPP.13435

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

Schober, P., & Vetter, T. R. (2018). Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare. *Anesthesia and Analgesia*, *127*(3), 792. https://doi.org/10.1213/ANE.0000000000003653

Schreiber, T., Prange, A., Schäfer, P., Iwen, T., Grützner, R., Marillonnet, S., Lepage, A., Javelle, M., Paul, W., & Tissier, A. (2024). Efficient scar-free knock-ins of several kilobases in plants by engineered CRISPR-Cas endonucleases. *Molecular Plant*, *17*(5), 824–837. https://doi.org/10.1016/j.molp.2024.03.013

Schulte-Geldermann, E., Gildemacher, P. R., & Struik, P. C. (2012). Improving Seed Health and Seed Performance by Positive Selection in Three Kenyan Potato Varieties. *American Journal of Potato Research*, *89*(6), 429–437. https://doi.org/10.1007/s12230-012-9264-1

Shahnejat-Bushehri, S., Nobmann, B., Devi Allu, A., & Balazadeh, S. (2016). JUB1 suppresses *Pseudomonas syringae*-induced defense responses through accumulation of DELLA proteins. *Plant Signaling & Behavior*, *11*(6). https://doi.org/10.1080/15592324.2016.1181245

Shaked, H., Melamed-Bessudo, C., & Levy, A. A. (2005). High-frequency gene targeting in *Arabidopsis* plants expressing the yeast *RAD54* gene. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(34), 12265–12269. https://doi.org/10.1073/pnas.0502601102

Shalev, G., Sitrit, Y., Avivi-Ragolski, N., Lichtenstein, C., & Levy, A. A. (1999). Stimulation of homologous recombination in plants by expression of the bacterial resolvase *RuvC*. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(13), 7398–7402. https://doi.org/10.1073/pnas.96.13.7398

Shang, Q., Zhang, D., Li, R., Wang, K., Cheng, Z., Zhou, Z., Hao, Z., Pan, J., Li, X., & Shi, L. (2020). Mapping quantitative trait loci associated with stem-related traits in maize (*Zea mays* L.). *Plant Molecular Biology*, *104*(6), 583–595. https://doi.org/10.1007/s11103-020-01062-3

Sharma, N., Kaur, N., & Gupta, A. K. (1998). Effects of gibberellic acid and chlorocholine chloride on tuberisation and growth of potato (*Solanum tuberosum* L). *Journal of the Science of Food and Agriculture*, *78*(4), 466–470. https://doi.org/10.1002/(SICI)1097-0010(199812)78:4<466::AID-JSFA140>3.0.CO;2-1

Shimada, T. L., Shimada, T., & Hara-Nishimura, I. (2010). A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of *Arabidopsis thaliana*. *The Plant Journal*, *61*(3), 519–528. https://doi.org/10.1111/j.1365-313X.2009.04060.x

Shukla, V. K., Doyon, Y., Miller, J. C., Dekelver, R. C., Moehle, E. A., Worden, S. E., Mitchell, J. C., Arnold, N. L., Gopalan, S., Meng, X., Choi, V. M., Rock, J. M., Wu, Y. Y., Katibah, G. E., Zhifang, G., McCaskill, D., Simpson, M. A., Blakeslee, B., Greenwalt, S. A., … Urnov, F. D. (2009). Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature*, *459*(7245), 437–441. https://doi.org/10.1038/nature07992

Shull, G. H. (1908). The Composition of a Field of Maize. *Journal of Heredity*, *4*(1), 296–301. https://doi.org/10.1093/JHERED/OS-4.1.296

Silverstone, A. L., Mak, P. Y. A., Martínez, E. C., & Sun, T. P. (1997). The New RGA Locus Encodes a Negative Regulator of Gibberellin Response in *Arabidopsis thaliana*. *Genetics*, *146*(3), 1087–1099. https://doi.org/10.1093/GENETICS/146.3.1087

Sinclair, A., Wale, S., & Bingham, I. (2022). *SRUC Technical Note: Nitrogen recommendations for cereals, oilseed rape and potatoes*. https://www.sruc.ac.uk/media/eeelqian/tn731-nitrogen-recommendations-for-cereals-oilseed-rape-and-potatoes-update-31-10-2022.pdf

Singer, K., Shiboleth, Y. M., Li, J., & Tzfira, T. (2012). Formation of Complex Extrachromosomal T-DNA Structures in *Agrobacterium tumefaciens*-Infected Plants. *Plant Physiology*, *160*(1), 511. https://doi.org/10.1104/PP.112.200212

Smigocki, A. C., & Owens, L. D. (1989). Cytokinin-to-Auxin Ratios and Morphology of Shoots and Tissues Transformed by a Chimeric Isopentenyl Transferase Gene. *Plant Physiology*, *91*(3), 808–811. https://doi.org/10.1104/PP.91.3.808

Spielmeyer, W., Ellis, M. H., & Chandler, P. M. (2002). Semidwarf (sd-1), “green revolution” rice, contains a defective *gibberellin 20-oxidase* gene. *Proceedings of the National Academy of Sciences*, *99*(13), 9043–9048. https://doi.org/10.1073/pnas.132266399

Spooner, D. M., Ghislain, M., Simon, R., Jansky, S. H., & Gavrilenko, T. (2014). Systematics, Diversity, Genetics, and Evolution of Wild and Cultivated Potatoes. *The Botanical Review*, *80*(4), 283–383. https://doi.org/10.1007/s12229-014-9146-y

Sprague, G. F., & Tatum, L. A. (1942). General vs. Specific Combining Ability in Single Crosses of Corn. *Agronomy Journal*, *34*(10), 923–932. https://doi.org/10.2134/AGRONJ1942.00021962003400100008X

Stockem, J., de Vries, M., van Nieuwenhuizen, E., Lindhout, P., & Struik, P. C. (2020). Contribution and Stability of Yield Components of Diploid Hybrid Potato. *Potato Research*, 1–22. https://doi.org/10.1007/s11540-019-09444-x

Stougaard, J. (1993). Substrate-dependent negative selection in plants using a bacterial *cytosine deaminase* gene. *The Plant Journal*, *3*(5), 755–761. https://doi.org/10.1111/J.1365-313X.1993.00755.X

Stuttmann, J., Barthel, K., Martin, P., Ordon, J., Erickson, J. L., Herr, R., Ferik, F., Kretschmer, C., Berner, T., Keilwagen, J., Marillonnet, S., & Bonas, U. (2021). Highly efficient multiplex editing: one-shot generation of 8× *Nicotiana benthamiana* and 12× *Arabidopsis mutants*. *The Plant Journal*, *106*(1), 8–22. https://doi.org/10.1111/tpj.15197

Su, L., Wang, Q., Wang, C., & Shan, Y. (2015). Simulation Models of Leaf Area Index and Yield for Cotton Grown with Different

Soil Conditioners. *PLOS ONE*, *10*(11), e0141835. https://doi.org/10.1371/JOURNAL.PONE.0141835

Su, Y., Viquez-Zamora, M., den Uil, D., Sinnige, J., Kruyt, H., Vossen, J., Lindhout, P., & van Heusden, S. (2020). Introgression of Genes for Resistance against *Phytophthora infestans* in Diploid Potato. *American Journal of Potato Research*, *97*(1), 33–42. https://doi.org/10.1007/s12230-019-09741-8

Sun, C., Lei, Y., Li, B., Gao, Q., Li, Y., Cao, W., Yang, C., Li, H., Wang, Z., Li, Y., Wang, Y., Liu, J., Zhao, K. T., & Gao, C. (2023). Precise integration of large DNA sequences in plant genomes using PrimeRoot editors. *Nature Biotechnology*. https://doi.org/10.1038/s41587-023-01769-w

Sun, T. P. (2011). The Molecular Mechanism and Evolution of the GA–GID1–DELLA Signaling Module in Plants. *Current Biology*, *21*(9), R338–R345. https://doi.org/10.1016/J.CUB.2011.02.036

Szarzanowicz, M. J., Waldburger, L. M., Busche, M., Geiselman, G. M., Kirkpatrick, L. D., Kehl, A. J., Tahmin, C., Kuo, R. C., McCauley, J., Pannu, H., Cui, R., Liu, S., Hillson, N. J., Brunkard, J. O., Keasling, J. D., Gladden, J. M., Thompson, M. G., & Shih, P. M. (2024). Binary vector copy number engineering improves *Agrobacterium*-mediated transformation. *Nature Biotechnology 2024*, 1–9. https://doi.org/10.1038/s41587-024-02462-2

Tang, D., Jia, Y., Zhang, J., Li, H., Cheng, L., Wang, P., Bao, Z., Liu, Z., Feng, S., Zhu, X., Li, D., Zhu, G., Wang, H., Zhou, Y., Zhou, Y., Bryan, G. J., Buell, C. R., Zhang, C., & Huang, S. (2022). Genome evolution and diversity of wild and cultivated potatoes. *Nature 2022 606:7914*, *606*(7914), 535–541. https://doi.org/10.1038/s41586-022-04822-x

Tang, X., Lowder, L. G., Zhang, T., Malzahn, A. A., Zheng, X., Voytas, D. F., Zhong, Z., Chen, Y., Ren, Q., Li, Q., Kirkland, E. R., Zhang, Y., & Qi, Y. (2017). A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nature Plants*, *3*(3), 17018. https://doi.org/10.1038/nplants.2017.18

Taylor, T. E., & Veilleux, R. E. (1992). Inheritance of competencies for leaf disc regeneration, anther culture, and protoplast culture in *Solanum phureja* and corrleations among them. *Plant Cell, Tissue and Organ Culture*, *31*(2), 95–103. https://doi.org/10.1007/BF00037692

Tomlinson, L., Yang, Y., Emenecker, R., Smoker, M., Taylor, J., Perkins, S., Smith, J., MacLean, D., Olszewski, N. E., & Jones, J. D. G. (2019). Using CRISPR/Cas9 genome editing in tomato to create a gibberellin-responsive dominant dwarf *DELLA* allele. *Plant Biotechnology Journal*, *17*(1), 132–140. https://doi.org/10.1111/pbi.12952

Tóth, E., Weinhardt, N., Bencsura, P., Huszár, K., Kulcsár, P. I., Tálas, A., Fodor, E., & Welker, E. (2016). Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian

cells. *Biology Direct*, *11*(1), 46. https://doi.org/10.1186/s13062-016-0147-0

Townsend, J. A., Wright, D. A., Winfrey, R. J., Fu, F., Maeder, M. L., Joung, J. K., & Voytas, D. F. (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature*, *459*(7245), 442–445. https://doi.org/10.1038/nature07845

Troyer, A. F. (2006). Adaptedness and Heterosis in Corn and Mule Hybrids. *Crop Science*, *46*(2), 528–543. https://doi.org/10.2135/cropsci2005.0065

Uitdewilligen, J. G. A. M. L., Wolters, A. M., Van der Vlies, P., Visser, R. G. F., & Van Eck, H. J. (2011). SNP and haplotype identification from targeted next-generation re-sequencing in a set of 83 potato cultivars. *The 18th Triennial Conference of the European Association for Potato Research*, 97.

van Dijk, L. C. M. M., de Vries, M. E., Lommen, W. J. M. M., & Struik, P. C. (2022). Transplanting Hybrid Potato Seedlings at Increased Densities Enhances Tuber Yield and Shifts Tuber-Size Distributions. *Potato Research*, *65*(2), 307–331. https://doi.org/10.1007/s11540-021-09522-z

van Kregten, M., de Pater, S., Romeijn, R., van Schendel, R., Hooykaas, P. J. J., & Tijsterman, M. (2016). T-DNA integration in plants results from polymerase-θ-mediated DNA repair. *Nature Plants*, *2*(11), 16164. https://doi.org/10.1038/nplants.2016.164

van Lieshout, N., van der Burgt, A., de Vries, M. E., ter Maat, M., Eickholt, D., Esselink, D., van Kaauwen, M. P. W., Kodde, L. P., Visser, R. G. F., Lindhout, P., & Finkers, R. (2020). Solyntus, the new highly contiguous reference genome for potato (*Solanum tuberosum*). *G3: Genes, Genomes, Genetics*, *10*(10), 3489–3495. https://doi.org/10.1534/g3.120.401550

Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E. F., & Hellens, R. P. (2007). Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods*, *3*(1), 12. https://doi.org/10.1186/1746-4811-3-12

Vartak, S. V., & Raghavan, S. C. (2015). Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing. *The FEBS Journal*, *282*(22), 4289–4294. https://doi.org/10.1111/FEBS.13416

Visser, R. G. F., Bachem, C. W. B., Borm, T., de Boer, J., van Eck, H. J., Finkers, R., van der Linden, G., Maliepaard, C. A., Uitdewilligen, J. G. A. M. L., Voorrips, R., Vos, P., & Wolters, A. M. A. (2014). Possibilities and Challenges of the Potato Genome Sequence. *Potato Research*, *57*(3–4), 327–330. https://doi.org/10.1007/s11540-015-9282-8

Vleeshouwers, V. G. A. A., Raffaele, S., Vossen, J. H., Champouret, N., Oliva, R., Segretin, M. E., Rietman, H., Cano, L. M., Lokossou, A., Kessel, G., Pel, M. A., & Kamoun, S. (2011). Understanding and Exploiting Late Blight Resistance in the Age of Effectors. *Annual Review of Phytopathology*, *49*(1), 507–531. https://doi.org/10.1146/annurev-phyto-072910-095326

Vreugdenhil, D., & Struik, P. C. (1989). An integrated view of the hormonal regulation of tuber formation in potato (*Solanum tuberosum*). *Physiologia Plantarum*, *75*(4), 525–531. https://doi.org/10.1111/J.1399-3054.1989.TB05619.X

Vu, T. Van, Nguyen, N. T., Kim, J., Vu, M. H., Song, Y. J., Tran, M. T., Sung, Y. W., & Kim, J. Y. (2024). Enhancing CRISPR-Cas-based gene targeting in tomato using a dominant-negative *ku80*. *Horticulture Research*, *12*(2). https://doi.org/10.1093/HR/UHAE294

Wang, E. S., Kieu, N. P., Lenman, M., & Andreasson, E. (2020). Tissue culture and refreshment techniques for improvement of transformation in local tetraploid and diploid potato with late blight resistance as an example. *Plants*, *9*(6), 695. https://doi.org/10.3390/plants9060695

Wang, J., Hu, M., Wang, J., Qi, J., Han, Z., Wang, G., Qi, Y., Wang, H.-W., Zhou, J.-M., & Chai, J. (2019). Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science*, *364*(6435). https://doi.org/10.1126/science.aav5870

Wang, M., Lu, Y., Botella, J. R., Mao, Y., Hua, K., & Zhu, J. kang. (2017). Gene Targeting by Homology-Directed Repair in Rice Using a Geminivirus-Based CRISPR/Cas9 System. *Molecular Plant*, *10*(7), 1007–1010. https://doi.org/10.1016/j.molp.2017.03.002

Wang, S., Zhang, S., Wang, W., Xiong, X., Meng, F., & Cui, X. (2015). Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. *Plant Cell Reports*, *34*(9), 1473–1476. https://doi.org/10.1007/s00299-015-1816-7

Wang, X., Wang, X., Sun, S., Tu, X., Lin, K., Qin, L., Wang, X., Li, G., Zhong, S., & Li, P. (2022). Characterization of regulatory modules controlling leaf angle in maize. *Plant Physiology*, *190*(1), 500–515. https://doi.org/10.1093/PLPHYS/KIAC308

Watson, D. J. (1952). The physiological basis of variation in yield. *Advances in Agronomy*, *4*, 101–145.

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE*, *6*(2), 16765. https://doi.org/10.1371/journal.pone.0016765

Wei, H., Zhao, Y., Xie, Y., & Wang, H. (2018). Exploiting SPL genes to improve maize plant architecture tailored for high-density planting. *Journal of Experimental Botany*, *69*(20), 4675–4688. https://doi.org/10.1093/JXB/ERY258

Weinthal, D. M., Taylor, R. A., & Tzfira, T. (2013). Nonhomologous End Joining-Mediated Gene Replacement in Plant Cells. *Plant Physiology*, *162*(1), 390–400. https://doi.org/10.1104/PP.112.212910

Werner, S., Engler, C., Weber, E., Gruetzner, R., & Marillonnet, S. (2012). Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioengineered*, *3*(1), 38–43. https://doi.org/10.4161/BBUG.3.1.18223

Wickham, H. (2009). Build a plot layer by layer. In *ggplot2* (pp. 41–64). Springer New York. https://doi.org/10.1007/978-0-387-98141-3_4

Witek, K., Jupe, F., Witek, A. I., Baker, D., Clark, M. D., & Jones, J. D. G. (2016). Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. *Nature Biotechnology*, *34*(6), 656–660. https://doi.org/10.1038/nbt.3540

Witek, K., Lin, X., Karki, H. S., Jupe, F., Witek, A. I., Steuernagel, B., Stam, R., van Oosterhout, C., Fairhead, S., Heal, R., Cocker, J. M., Bhanvadia, S., Barrett, W., Wu, C. H., Adachi, H., Song, T., Kamoun, S., Vleeshouwers, V. G. A. A., Tomlinson, L., … Jones, J. D. G. (2020). A complex resistance locus in S*olanum americanum* recognizes a conserved *Phytophthora* effector. *Nature Plants*, *7*(2), 198–208. https://doi.org/10.1101/2020.05.15.095497

Xie, K., Minkenberg, B., & Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), 3570–3575. https://doi.org/10.1073/pnas.1420294112

Xu, R., Qin, R., Li, H., Li, D., Li, L., Wei, P., & Yang, J. (2017). Generation of targeted mutant rice using a CRISPR-Cpf1 system. *Plant Biotechnology Journal*, *15*(6), 713–717. https://doi.org/10.1111/pbi.12669

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., Zhang, G., Yang, S., Li, R., Wang, J., Orjeda, G., Guzman, F., Torres, M., Lozano, R., Ponce, O., Martinez, D., De La Cruz, G., Chakrabarti, S. K., Patil, V. U., … Visser, R. G. F. (2011). Genome sequence and analysis of the tuber crop potato. *Nature*, *475*(7355), 189–195. https://doi.org/10.1038/nature10158

Xu, X., Van Lammeren, A. A. M., Vermeer, E., & Vreugdenhil, D. (1998). The Role of Gibberellin, Abscisic Acid, and Sucrose in the Regulation of Potato Tuber Formation in Vitro. *Plant Physiology*, *117*(2), 575–584. https://doi.org/10.1104/PP.117.2.575

Yasumoto, S., Umemoto, N., Lee, H. J., Nakayasu, M., Sawai, S., Sakuma, T., Yamamoto, T., Mizutani, M., Saito, K., & Muranaka, T. (2019). Efficient genome engineering using Platinum TALEN in potato. *Plant Biotechnology*, *36*(3), 167–173. https://doi.org/10.5511/plantbiotechnology.19.0805a

Youles, M., Marillonnet, S., Jones, J., Egan, L., Witek, A. I., & Alam, M. (2024). A pRK290-derived binary vector suitable for assembling large T-DNAs for plant transformation using Golden Gate cloning. *Zenodo*. https://doi.org/10.5281/ZENODO.14514683

Yuan, G., Martin, S., Hassan, M. M., Tuskan, G. A., & Yang, X. (2022). PARA: A New Platform for the Rapid Assembly of gRNA Arrays for Multiplexed CRISPR Technologies. *Cells*, *11*(16), 2467. https://doi.org/10.3390/CELLS11162467/S1

Yuan, L. (1966). A preliminary report on male sterility in rice. *Science Bulletin*, *4*, 32–34.

Yuan, L. (2017). Progress in super-hybrid rice breeding. *The Crop Journal*, *5*(2), 100–102. https://doi.org/10.1016/J.CJ.2017.02.001

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., & Zhang, F. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, *163*(3), 759–771. https://doi.org/10.1016/J.CELL.2015.09.038

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*(1), 1–18. https://doi.org/10.1186/s13059-017-1164-8

Zhang, Y., Butelli, E., & Martin, C. (2014). Engineering anthocyanin biosynthesis in plants. *Current Opinion in Plant Biology*, *19*, 81–90. https://doi.org/10.1016/j.pbi.2014.05.011

Zhang, Y., Zhang, F., Li, X., Baller, J. A., Qi, Y., Starker, C. G., Bogdanove, A. J., & Voytas, D. F. (2012). Transcription Activator-Like Effector Nucleases Enable Efficient Plant Genome Engineering. *Plant Physiology*, *161*(1), 20. https://doi.org/10.1104/PP.112.205179

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology*, *7*(1–2), 203–214. https://doi.org/10.1089/10665270050081478

Zhao, L., Zou, M., Jiang, S., Dong, X., Deng, K., Na, T., Wang, J., Xia, Z., & Wang, F. (2023). Insights into the Genetic Determination of the Autotetraploid Potato Plant Height. *Genes*, *14*(2), 507. https://doi.org/10.3390/GENES14020507/S1

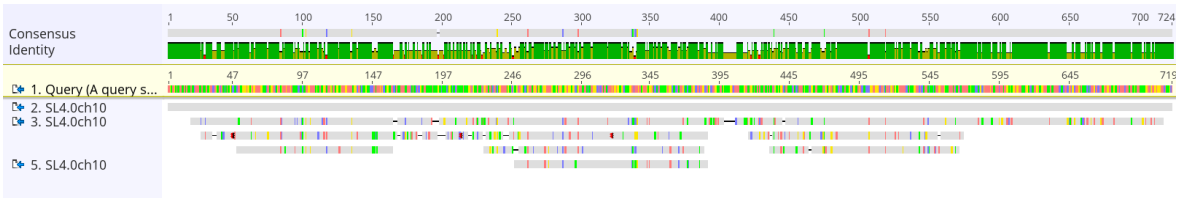
Zhi, L., TeRonde, S., Meyer, S., Arling, M. L., Register III, J. C., Zhao, Z.-Y., Jones, T. J., & Anand, A. (2015). Effect of *Agrobacterium* strain and plasmid copy number on transformation frequency, event quality and usable event quality in an elite maize cultivar. *Plant Cell Reports*, *34*(5), 745–754. https://doi.org/10.1007/s00299-014-1734-0

APPENDIX

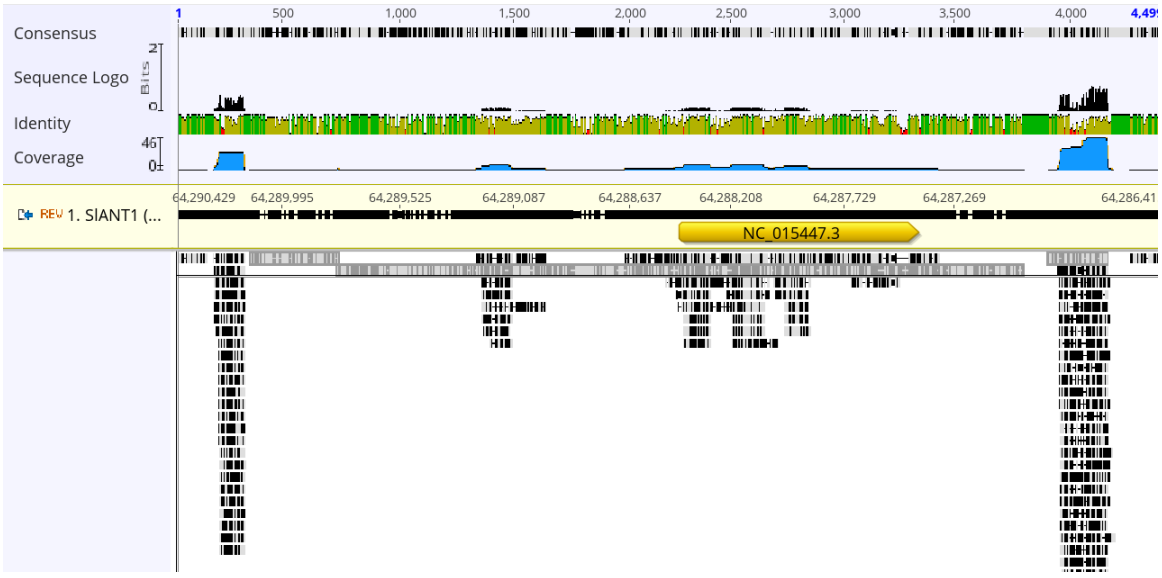
Appendix 1. Supplementary Figures

Supplementary Figure 1. *ANT1* homologue in Solyntus.
A Query-centric view of the regions Čermák et al. (2015) used as homology arms BLAST-searched against the SL4.0 tomato genome. Top hit (64287737-64288455) is a perfect match. **B** Query-centric view of *SlANT1* ORF (NC_015447.3, 65260733-65261744), plus 2 kb upstream and 1kb downstream, BLAST-searched against the Solyntus V1.1 genome. Top hit and two adjacent-mapping hits highlighted in grey. The two peaks in coverage mapping to both ends of the query are reads from other chromosomes, likely mapping to transposable elements.

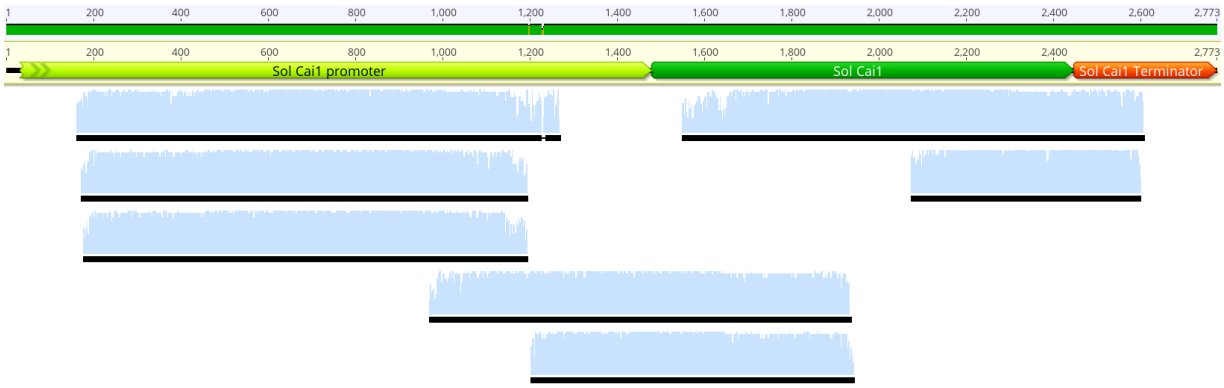
A



B

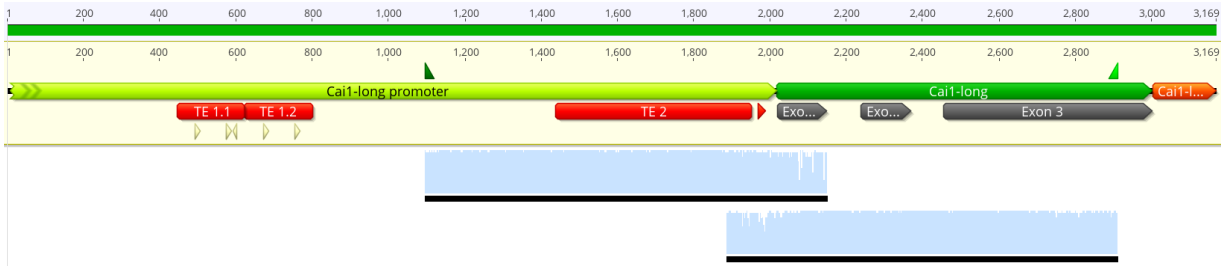


Supplementary Figure 2. *Cai1* allele in B99, B100, Solyntus. Sanger sequencing reads from the cloned B99, B100, and Solyntus *Cai1* amplicons mapping to the *Cai1* allele in the Solyntus V1.1 genome.

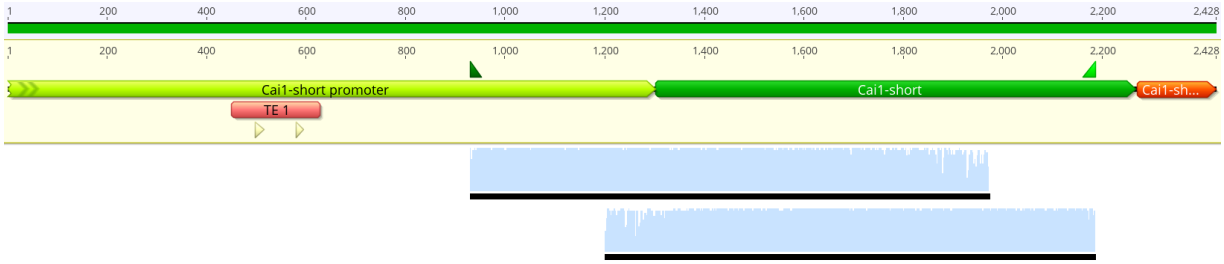


Supplementary Figure 3. Sequencing *Cai1* allele amplicons Sanger sequencing reads of cloned 1.81 kb (**A**) and 1.26 kb (**B**) bands mapped to the *Cai1-long* and *Cai1-short* alleles, respectively. In both cases, the reads map perfectly.

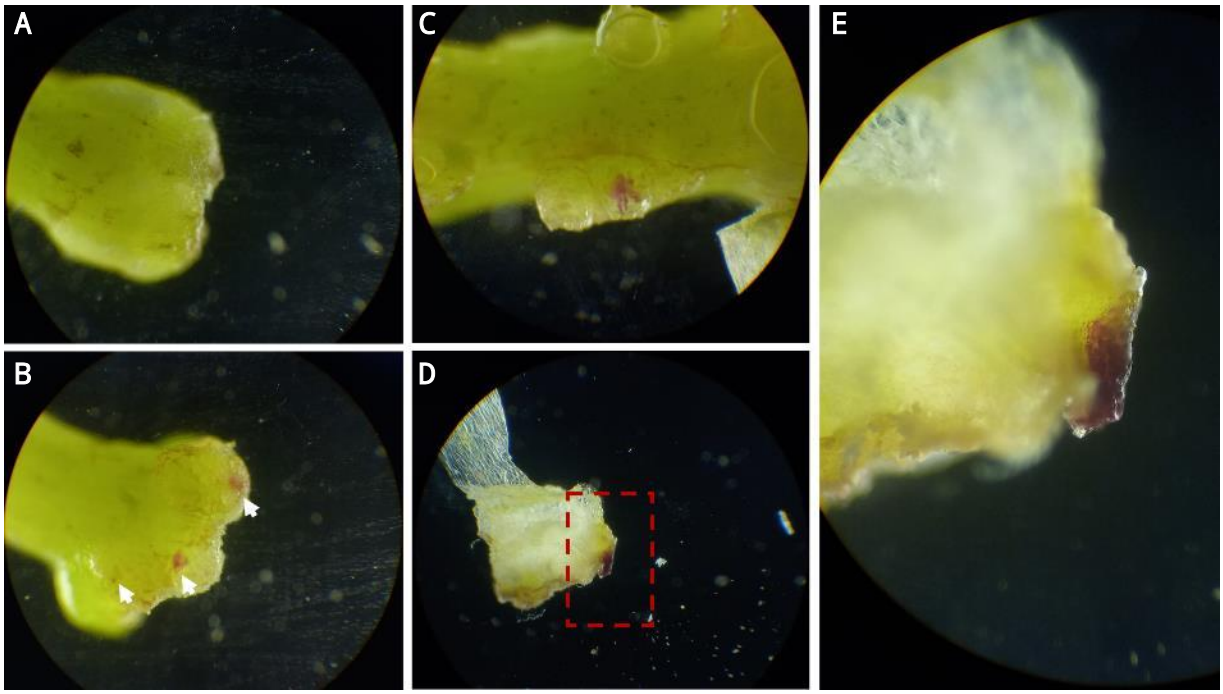
A



B



Supplementary Figure 4. KI sectors viewed under a microscope.
A Callus terminus without KI sector. **B** Callus terminus with three, seemingly independent KI sectors (white arrows). **C-D** Single KI sector. **E** Higher magnification of KI sector in red rectangle. Magnification is 100X in A-C, 40X in D, and 200X in E.

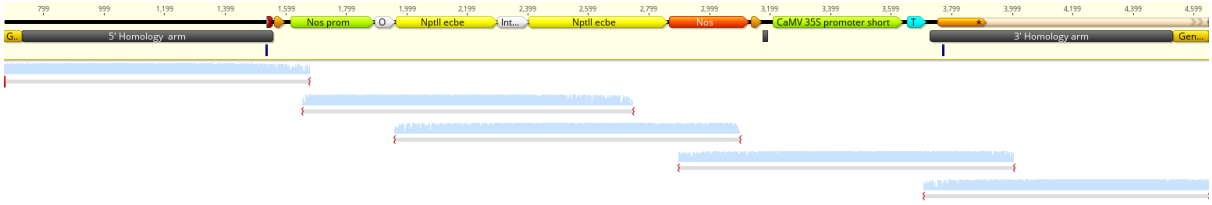


Supplementary Figure 5. Mosaicism in some T_0 lines.

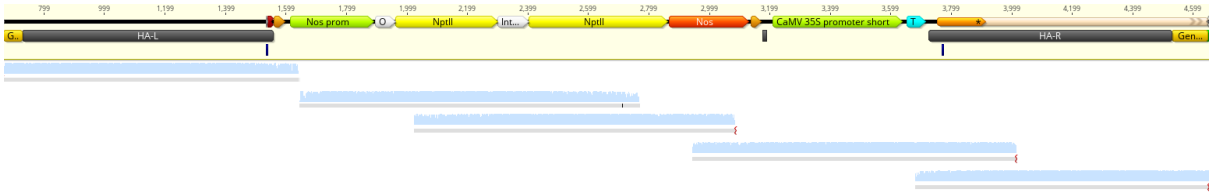
Supplementary Figure 6. Sequencing the 4kb bands of pKI1-1, pKI1-3, and pRAD54-3.

Sanger sequencing reads mapped to the expected KI locus. The yellow annotations on the outside of the two homology arms are genomic DNA sequence.

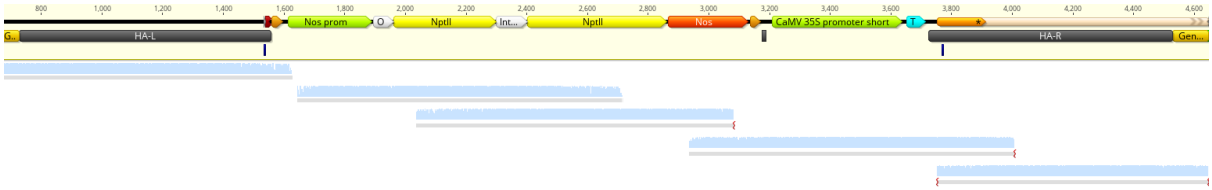
pKI1-1



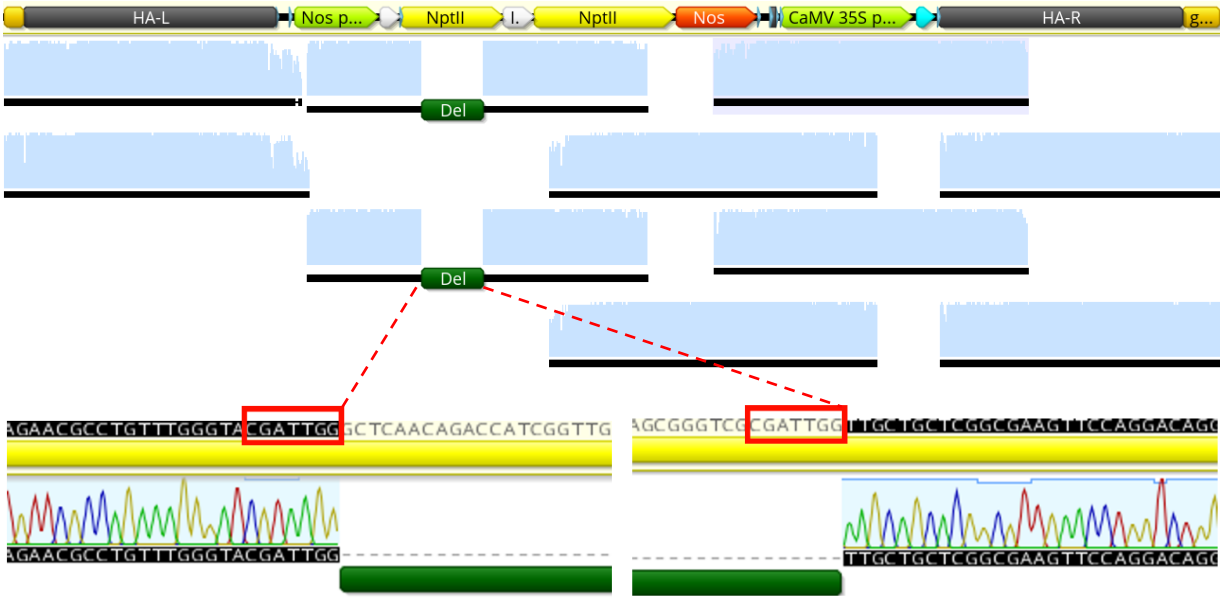
pKI1-3



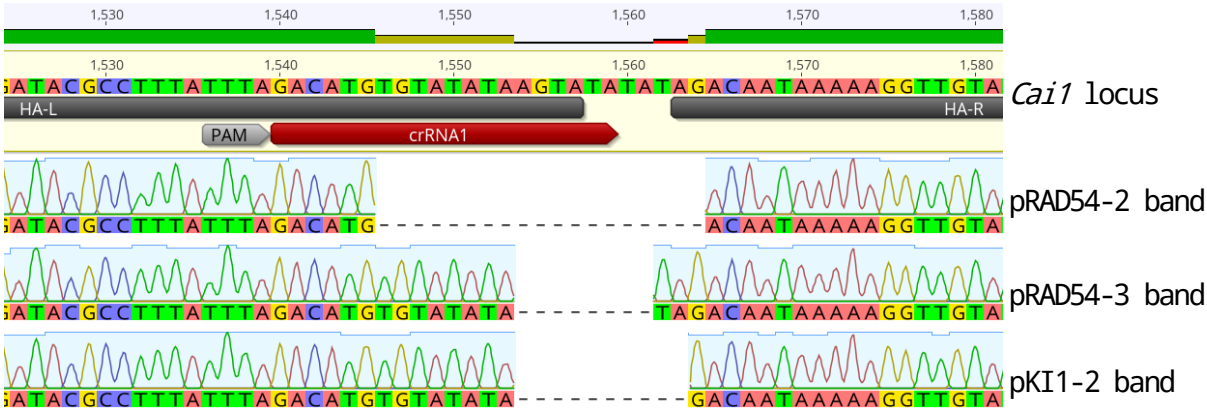
pRAD54-3



Supplementary Figure 7. pRAD54-2 KI locus sequencing.
Sanger sequencing reads mapped to the expected KI locus for pRAD54. The yellow annotations on the outside of the two homology arms are genomic DNA sequence. The reads map to the predicted locus except for a 201 bp deletion (Del) in the first exon of the *NptII* gene. The borders of the deletion contain 7 bp microhomologies (red boxes). Suggesting that the deletion might be the result of a spontaneous rearrangement.

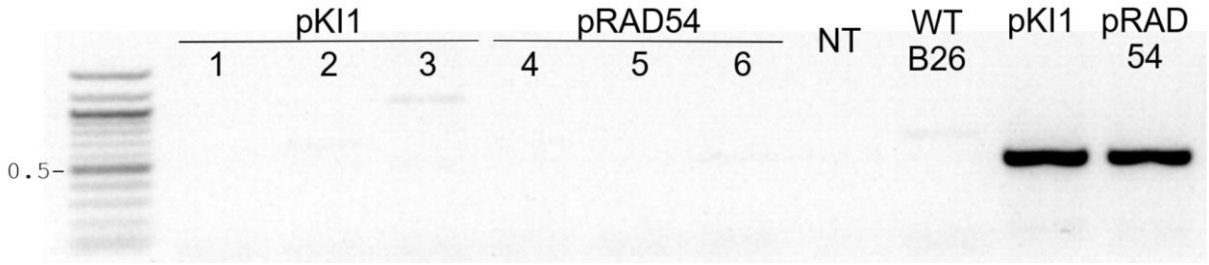


Supplementary Figure 8. KI bottom bands sequenced.
Bottom bands of three T_0 KI lines cloned and Sanger-sequenced.



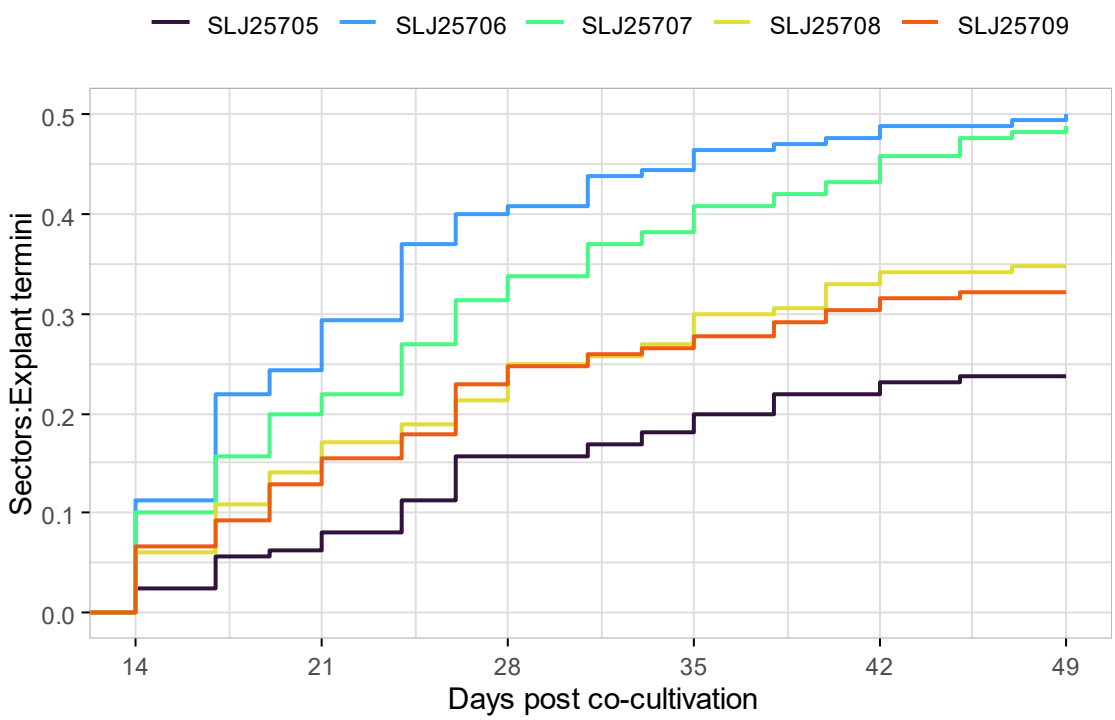
Supplementary Figure 9. KI T_0 lines screened for the plasmid backbone.

The lines were PCR-screened for the presence of the plasmid backbone using primers YD190/YD191 which amplify a 540 bp region ca. 100 bp to the left of the left T-DNA border. The expected amplicon was not observed in any T_0 line, the NT, or WT controls, while the plasmid controls produced the expected band. Thus, the T_0 lines do not have any detectable plasmid backbone integrations. Ladder is 100 bp (NEB).



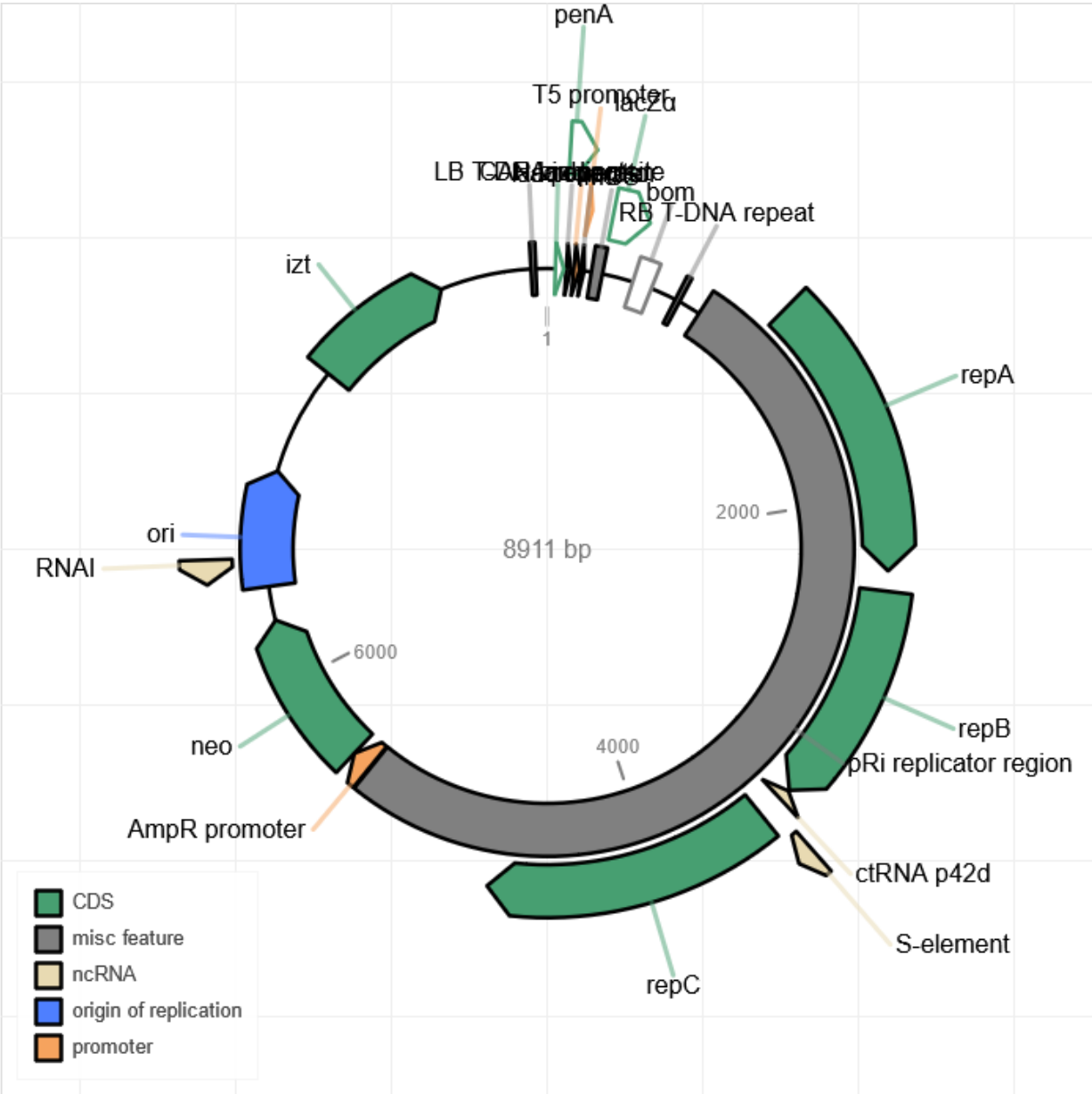
Supplementary Figure 10. KI sector emergence over the 7 weeks after co-cultivation.

By 28 dpcc, 74% of the final KI sector count had already occurred. By 35 dpcc, that figure was 87%. SLJ25705 - pCas12a; SLJ25706 - piCas12a; SLJ25707 – pUL12-iCas12a (also called pU6:Ribo); SLJ25708 – pU6:tRNA; SLJ25709 – pFMV:tRNA.

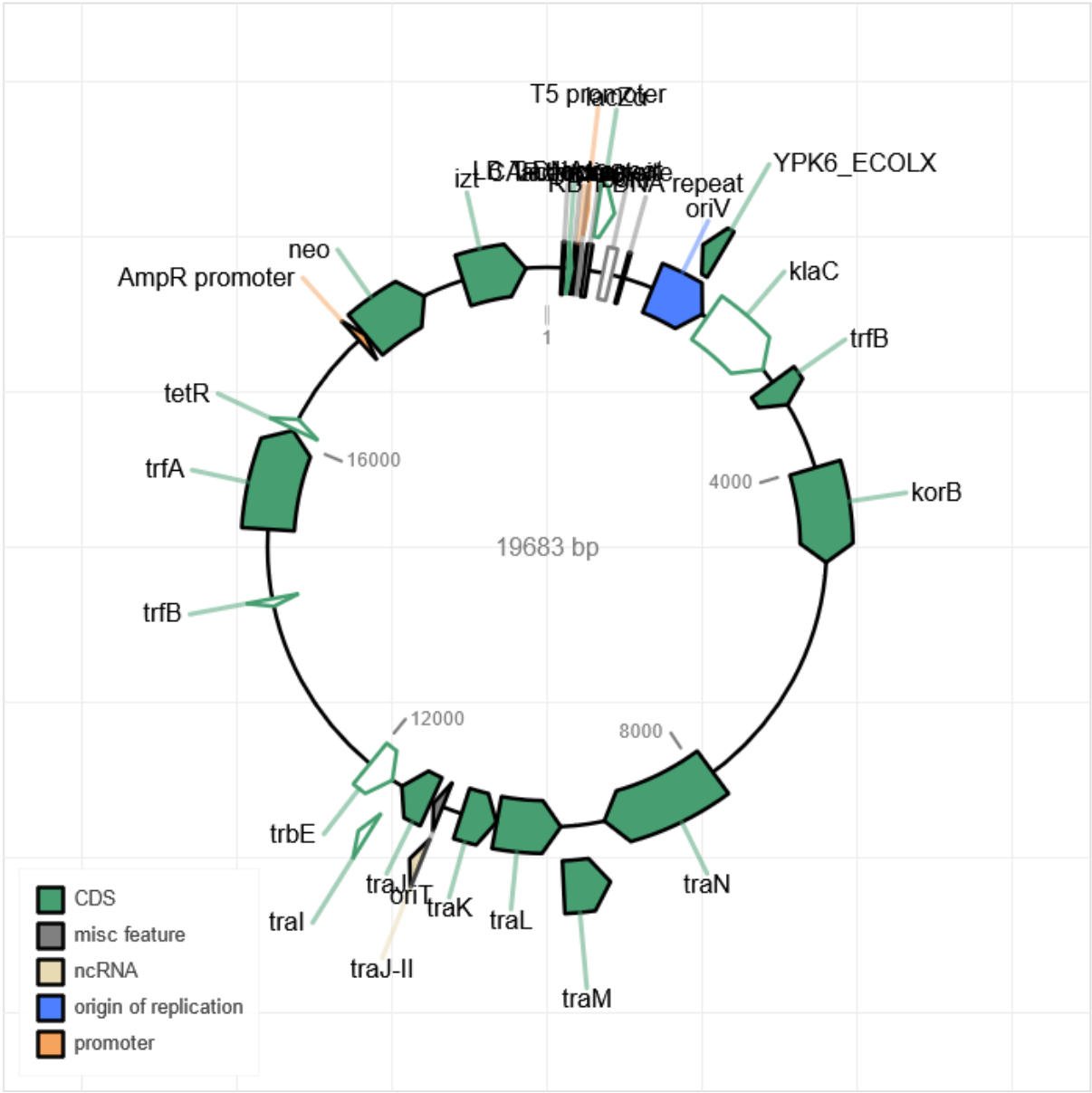


Supplementary Figure 11. Plasmid maps of binary vectors used. Annotations automatically generated with pLannotate (McGuffie & Barrick, 2021).

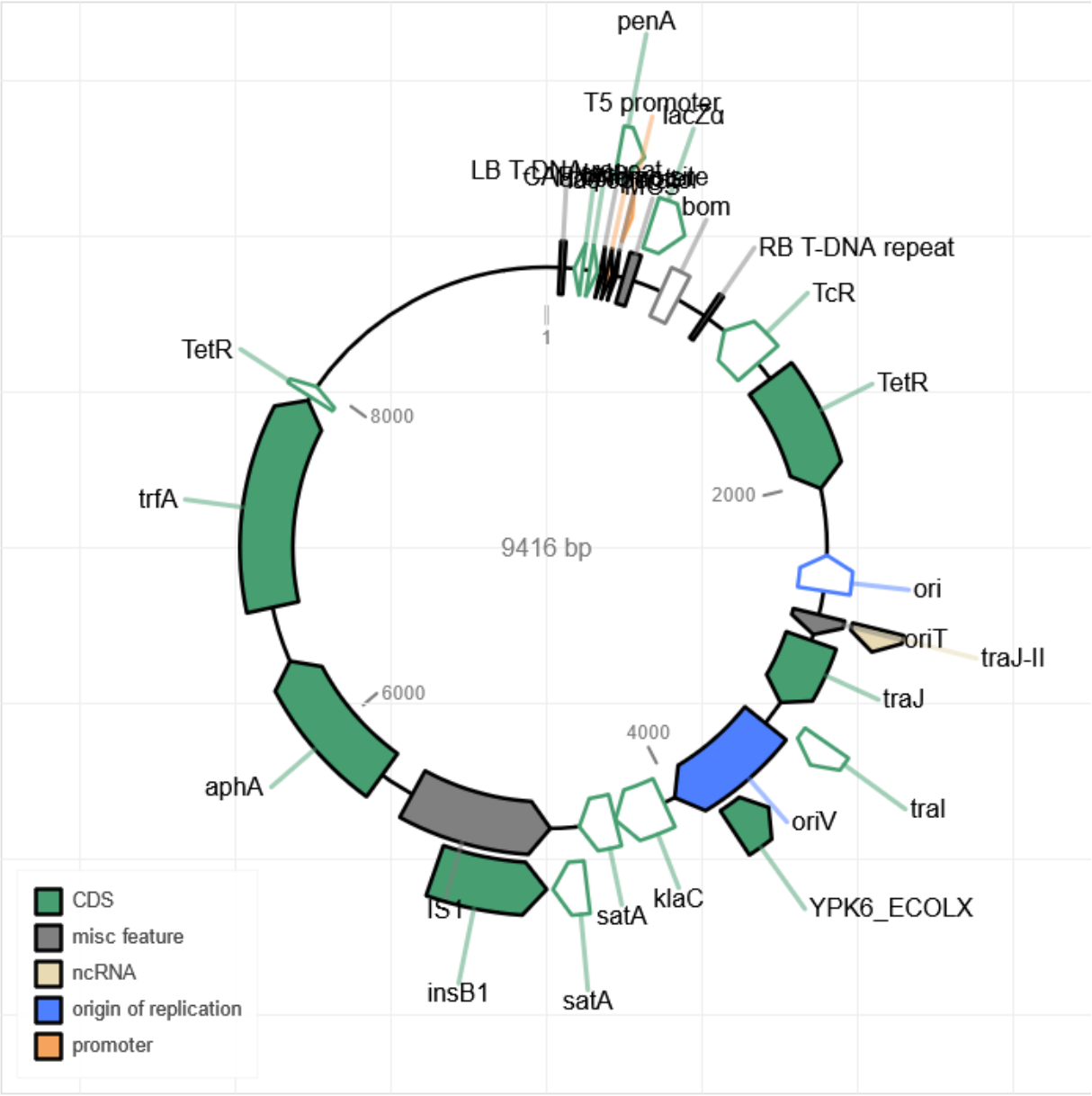
pAGM31195



LP_pRKan290

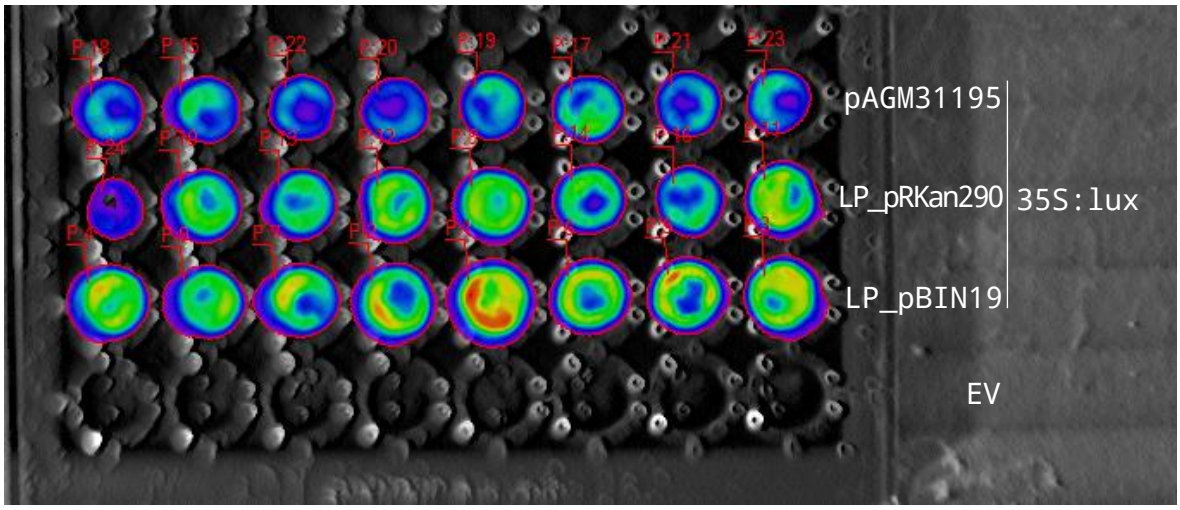

T5 promoter
YPK6_ECOLX
repeat
oriV
klaC
trfB
korB
traN
traM
traL
traK
traJ-II
oriT
traJ
traI
trbE
trfB
trfA
tetR
AmpR promoter
neo
izt
1
4000
8000
12000
16000
19683 bp
CDS
misc feature
ncRNA
origin of replication
promoter

LP_pBIN19


penA
T5 promoter
lacZα
LB T-DNA repeat
bom
RB T-DNA repeat
TcR
TetR
TetR
trfA
1
8000
2000
9416 bp
ori
oriT
traJ-II
traJ
6000
4000
aphA
tral
oriV
YPK6_ECOLX
klaC
satA
IS1
insB1
satA
CDS
misc feature
ncRNA
origin of replication
promoter

Supplementary Figure 12. The luciferase signal strength in Fig. 3.27 quantified as photon counts.

Label	Index	Counts [cts]
P:1	1	1.381E+07
P:2	2	9.581E+06
P:3	3	1.010E+07
P:4	4	9.225E+06
P:5	5	9.019E+06
P:6	6	8.977E+06
P:7	7	8.316E+06
P:8	8	8.512E+06
P:9	9	7.912E+06
P:10	10	7.431E+06
P:11	11	8.898E+06
P:12	12	8.324E+06
P:13	13	6.313E+06
P:14	14	6.302E+06
P:15	15	5.304E+06
P:16	16	5.432E+06
P:17	17	6.001E+06
P:18	18	4.631E+06
P:19	19	5.416E+06
P:20	20	3.926E+06
P:21	21	4.537E+06
P:22	22	4.131E+06
P:23	23	4.070E+06
P:24	24	2.501E+06



Supplementary Figure 13. Tubers of T_0 line 702.

Appendix 2. R code analysing KI sector counts

```
# Time-to-Event analysis of the time course sector counts of constructs with 1 or 2 crRNAs and introns
in Cas12a or not (YDJJ143-46)

# Install and load relevant packages:
install.packages("survival", repos = "https://cloud.r-project.org")
install.packages("survminer", repos = "https://cloud.r-project.org")
install.packages("dplyr", repos = "https://cloud.r-project.org")
install.packages("readxl", repos = "https://cloud.r-project.org")
library(survival)
library(survminer)
library(dplyr)
library(readxl)

##############################################################################
#Importing data

rawdata <- read_excel("counts.xlsx")
glimpse(rawdata)

# The two ends of an explant are treated independently. Thus, the KI sectors
# should be counted relative to the total number of transformation points
# (TF Points=2*Explants)
rawdata$tfpoints=2*rawdata$Explants

# Restructuring data into long format
long=data.frame()
for (i in 1:(length(rawdata$Construct))){
  const=c(rep(rawdata$Construct[i],rawdata$tfpoints[i]))
  #The following bits needs to be edited if changing the time points!!!!!
  stat=c(rep(1,rawdata$'28'[i]), rep(0,(rawdata$tfpoints[i]-rawdata$'28'[i])))
  dpcc=c(rep(14,rawdata$'14'[i]),
         rep(16,(rawdata$'16'[i]-rawdata$'14'[i])),
         rep(18,(rawdata$'18'[i]-rawdata$'16'[i])),
         rep(21,(rawdata$'21'[i]-rawdata$'18'[i])),
         rep(23,(rawdata$'23'[i]-rawdata$'21'[i])),
         rep(25,(rawdata$'25'[i]-rawdata$'23'[i])),
         rep(28,(rawdata$tfpoints[i]-rawdata$'25'[i])))
  temp=data.frame(dpcc,stat,const)
  long=rbind(long, temp)
}
rm("temp", "const","dpcc","i","stat")

##############################################################################
# Fit survival data using the Kaplan-Meier method

surv_object <- Surv(time = long$dpcc, event = long$stat)
surv_object

# Fit the Kaplan-Meier curve
```

```
fit1 <- survfit(surv_object ~ const, data = long)
summary(fit1)

################################################################################
# Log-rank tests
# The Bonferroni critical value for 2 comparisons is 0.05/3=0.017

# Comparing 1 crRNA in the middle (SLJ25734/YDJJ145) with 2 crRNAs, one on either side
(SLJ25732/YDJJ143)
survdiff(Surv(dpcc, stat) ~ const, data = subset(long, long$const %in% c("YDJJ143","YDJJ145")))
# Chisq= 18.1  on 1 degrees of freedom, p= 2e-05    <- data DOES disprove null hypothesis! ***
# Bonferroni-corrected P-value = 6e-05
# 2 crRNAs on either side of the substitution facilitate significantly more KIs than
# using just 1 crRNA in the middle.

# Comparing Cas12a with 2 crRNAs and either with (SLJ25733/YDJJ144) or w/o introns (SLJ25732/YDJJ143)
survdiff(Surv(dpcc, stat) ~ const, data = subset(long, long$const %in% c("YDJJ143","YDJJ144")))
# Chisq= 21.4  on 1 degrees of freedom, p= 4e-06   <- data DOES disprove null hypothesis! ***
# Bonferroni-corrected P-value = 1.2e-05
# When using 2 crRNAs, more nuclease, and thereby more cleavage, yields more KIs.

# Comparing Cas12a with 1 crRNA and either with (SLJ25735/YDJJ146) or w/o introns (SLJ25734/YDJJ145)
survdiff(Surv(dpcc, stat) ~ const, data = subset(long, long$const %in% c("YDJJ145","YDJJ146")))
# Chisq= 0.2  on 1 degrees of freedom, p= 0.6    <- data does NOT disprove null hypothesis.
# Bonferroni-corrected P-value = 1.0
# When using 1 crRNA, more nuclease, and thereby more cleavage, does not yield more KIs.
# This suggests there is a rate-limiting step downstream of target cleavage.
```

Appendix 3. R code output of Non-linear Mixed Effects model

```
# Nonlinear Mixed Effects model with a Gaussian curve accounting for the effect of genotype and the
# repeated measurements of plots
# Y = (Max+geno_num*Eff)*exp{-0.5[(X - Mid)/b]2}
nlmemod1=nlme(mean~max*exp(-0.5*((days-mid)/b)^2),
              data = deriv_dsm,
              fixed = c(max~geno, mid+b~1),
              random = max~1,
              start=c(max=0.54893, max.genodwarf=0.54893, mid=95.21822, b=20.47498),
              groups=~plot)
summary(nlmemod1)
# Nonlinear mixed-effects model fit by maximum likelihood
# Model: mean ~ max * exp(-0.5 * ((days - mid)/b)^2)
# Data: deriv_dsm
#       AIC       BIC  logLik
# -304.6221 -289.6232 158.311
#
# Random effects:
#   Formula: max ~ 1 | plot
#         max.(Intercept)   Residual
# StdDev:      0.02865833 0.04001715 <- StDev of plot random eff comparable to that of residuals! Worth
# accounting for.
#
# Fixed effects:  c(max ~ geno, mid + b ~ 1)
#                     Value Std.Error DF   t-value p-value
# max.(Intercept)  0.51705 0.0207268 81  24.94575  0.0000
# max.genonon      0.06184 0.0285498 81   2.16611  0.0332 <- Significant difference!
# mid             95.11604 0.4457367 81 213.39067  0.0000    Non-dwarf canopies peaked at ca. 6.2 cm
taller.
# b               20.53971 0.4254164 81  48.28143  0.0000
# Random effects:
#   Formula: max ~ 1 | plot
#                max   Residual
# StdDev: 0.02865724 0.04001723 <- StDev of plot random eff comparable to that of residuals! Worth
accounting for.

# I can add up the max value for the non-dwarfs (0.51704624 + 0.06184196 = 0.5788882)
# but I still need the SE!
# One can get that in 2 steps:
# 1. Get variance-covariance matrix for fixed effects
vc = vcov(nlmemod1)
sqrt(vc["max.(Intercept)", "max.(Intercept)"] + vc["max.genonon", "max.genonon"] +
    2 * vc["max.(Intercept)", "max.genonon"])
# [1] 0.02039337
# The Non-dwarf Max is 0.5788882 (SE=0.02039337).

random.effects(nlmemod1) #<-random coeff that need to be added to max to plot curve for individual plot
#    max.(Intercept)
# 3      0.042290758
# 5     -0.005073314
# 8     -0.013964784
# 9     -0.015366542
# 10    -0.028325974
# 11     0.020439856
```

Appendix 4. R code for analysis of tuber shape. Box-Cox data transformation and linear modelling.

```
install.packages(readxl)
library(readxl)
# Import data
shape <- read_excel("shape.xlsx")
shape$line=as.factor(shape$line)
View(shape)

hist(shape$ar)
shapiro.test(shape$ar)
# Shapiro-Wilk normality test
#
# data:  shape$ar
# W = 0.87553, p-value = 1.718e-11 <- not at all normally distributed
library(MASS)
b <- boxcox(lm(ar~geno+line %in% geno, data = shape))
# Exact lambda
lambda <- b$x[which.max(b$y)]
lambda # -1.111111. For lambda= ca. -1, it is best to normalize with 1/x as that is easier to interpret
hist(1/shape$ar)
shapiro.test(1/(shape$ar))
# Shapiro-Wilk normality test
#
# data:  1/(shape$ar)
# W = 0.98736, p-value = 0.08448 <- getting somewhere! Doing the proper Box-Cox might still be better
tho
shape$new_ar <- (shape$ar ^ lambda - 1) / lambda
shapiro.test(shape$new_ar)
# Shapiro-Wilk normality test
#
# data:  shape$new_ar
# W = 0.98966, p-value = 0.1803 <- very good! I will ultimately proceed with the BoxCox-transformed
data.
hist(shape$new_ar)

################################################################################
# Modelling

# Trying the simplest linear model
mod1 <- lm(new_ar ~ geno, data = shape)
# Trying models which account for the difference in line performance
install.packages("lmerTest")
library(lmerTest)
library(emmeans)
mod2=lm(new_ar~geno+line, data = shape)
mod3=lmer(new_ar~geno+(1|line), data=shape)
AIC(mod1, mod2, mod3)
#         df       AIC
# mod1     3 -316.4046
# mod2    25 -404.6086 <- significantly better than the rest
# mod3     4 -362.8514

summary(mod2)
# Residuals:
#       Min        1Q    Median        3Q       Max
# -0.195711 -0.052578  0.009362  0.045880  0.184470
#
# Coefficients: (1 not defined because of singularities)
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  0.286307   0.027994  10.228  < 2e-16 ***
# genonon      0.113557   0.039589   2.868 0.004655 ** <- not clear which line this is
# line1-11    -0.168063   0.039589  -4.245 3.61e-05 ***
# line1-12    -0.174524   0.039589  -4.408 1.85e-05 ***
# line1-13    -0.079714   0.039589  -2.014 0.045656 *
# line1-14     0.001205   0.039589   0.030 0.975749
# ...
anova(mod2)
# Response: new_ar
# Df  Sum Sq  Mean Sq F value    Pr(>F)
# geno        1 0.02114 0.021140  3.3721   0.06808 .
# line       22 1.04360 0.047436  7.5665 7.256e-16 ***
# Residuals 168 1.05323 0.006269
emmeans(mod2, "geno")
# geno  emmean      SE  df lower.CL upper.CL
# dwarf  0.271 0.00808 168    0.255    0.287
# non    0.292 0.00808 168    0.276    0.308
# not very useful as they are transformed
# Back-transforming the AR:
install.packages("sae")
library(sae)
bxcx(0.271, lambda = lambda, InverseQ = TRUE, type = "BoxCox") #1.380488
bxcx(0.271+0.00808, lambda = lambda, InverseQ = TRUE, type = "BoxCox") #1.396645
bxcx(0.271-0.00808, lambda = lambda, InverseQ = TRUE, type = "BoxCox") #1.36472
# Mean Dwarf AR = 1.38(SE=0.02)
bxcx(0.292, lambda = lambda, InverseQ = TRUE, type = "BoxCox") #1.423328
bxcx(0.292+0.00808, lambda = lambda, InverseQ = TRUE, type = "BoxCox") #1.44057
bxcx(0.292-0.00808, lambda = lambda, InverseQ = TRUE, type = "BoxCox") #1.406517
```

```
# Mean Non-dwarf AR = 1.42(SE=0.02)
pairs(emmeans(mod2, "geno"))
# contrast    estimate     SE  df t.ratio p.value
# dwarf - non   -0.021 0.0114 168  -1.836  0.0681 <-marginally significant!
# Results are averaged over the levels of: line

par(mfrow=c(2,2))
plot(mod2)
# Diagnostic plots look great!
```

Appendix 5. Primers and oligonucleotides used.

ID	OLIGO NAME	OLIGO SEQUENCE
-	B26 Cai1 ter L1 R	TGGTCTCAAGCGTGATAACAATCGGCTTAGGATAAC
-	AN73	TTATCGTCCAGGAACTCAGGCA
-	AN79	GTGCGATCGAATTTGTCGATTAATCT
-	LG14	ATGGTCTCAGGAGGTCAACATGGTGGAGCACGACAC
-	LG15	ATGGTCTCATACTGTATCGATAATTGTAAATGTAATTGT
-	LG266	ATGGTCTCATACTACTCGAGCGTGTCCTCTCCAA
YD1	GG-target-5'Hom arm	TGAAGACATTGTGTTTAGACATGTGTATATAAGTATATATCAATACATGCTCAGTTAAATATTGTCAT A
YD2	GG-5'Hom arm	TGAAGACATGCTCTACTTATATACACATGTCTAAATAAAGGCG
YD3	GG-3'Hom arm	TGGTCTCTAGTAAGACAATAAAAAGGTTGTATAATATATCATG
YD4	GG-target-3'Hom arm	AGGTCTCAAGCGTTTAGACATGTGTATATAAGTATATATCATTGCAATAATTTTCCAGTAAATTTG
YD24	YD24_DELLA_guide_F	TGGTCTCTATTGGATGAGCTTTTAGCTGTTTGTTTAGAGACGA
YD25	YD25_DELLA_guide_R	TGGTCTCTAAACAAACAGCTAAAAGCTCATCCAATAGAGACGA
YD27	YD27 B26 Cai1 distal Hom arm F	TGAAGACATTGTGTTTAGACATGTGTATATAAGTATATATTTTAGTTTTCCAAATTCAGTTATGATAT TATG
YD28	YD28 B26 Cai1 distal Hom arm R	TGAAGACATGCTCCAGGCCAAGTATTTTCAAAGAAC
YD29	YD29_35S_-106_F	ATGGTCTCAGGAGGTGGATTGATGTGACATC
YD35	YD35_Cai1_KI_junction _F2	CATGATTATGGCCATTTAGCAGC
YD36	YD36_Cai1_KI_junction _R1	CCAACTTTCATGACTTATTTGTCC
YD38	YD38_custom_dummy_F	TGAAGACTTGAGCCTCTACCAAAGGTCATTACAAGTCTTCA
YD39	YD39_custom_dummy_R	TGAAGACTTGTAATGACCTTTGGTAGAGGCTCAAGTCTTCA
YD40	YD40_MH1+target_F	TGAAGACTTTGTGTTTAGACATGTGTATATAAGTATATATTTATATAGACATGTGTATATAAGTAGAG CAAGTCTTCA
YD41	YD41_MH1+target_R	TGAAGACTTGCTCTACTTATATACACATGTCTATATAAATATATACTTATATACACATGTCTAAACAC AAAGTCTTCA
YD44	YD44_MH2_F	TGGTCTCTAGTAAGACAATAAAAAGGTTGTATAATATATATATACTTATATACACATGTCTAAACGCT AGAGACCA
YD45	YD45_MH2_R	TGGTCTCTAGCGTTTAGACATGTGTATATAAGTATATATATATTATACAACCTTTTTATTGTCTTACT AGAGACCA
YD46	YD46_NHEJ_L_cut_F	TGAAGACTTTGTGTTTAGACATGTGTATATAAGTATATAGAGCAAGTCTTCA
YD47	YD47_NHEJ_L_cut_R	TGAAGACTTGCTCTATATACTTATATACACATGTCTAAACACAAAGTCTTCA
YD48	YD48_NHEJ_R_cut_F	TGGTCTCTAGTAATATATACTTATATACACATGTCTAAACGCTAGAGACCA
YD49	YD49_NHEJ_R_cut_R	TGGTCTCTAGCGTTTAGACATGTGTATATAAGTATATATTACTAGAGACCA
YD61	YD61_tRNA_F	TGGTCTCTATTGAACAAAGCACCAGTGGTCTAG
YD62	YD62_guide1_R	TGGTCTCTAGTATCTACACTTAGTAGAAATTATGCACCAGCCGGG
YD63	YD63_guide1_F	TGGTCTCTTACTGAGCATGTATTGACAGGCCAAACAAAGCACCAGTGGTCTAG
YD64	YD64_guide2_R	TGGTCTCTTGTCATCTACACTTAGTAGAAATTATGCACCAGCCGGG
YD65	YD65_guide2_F	TGGTCTCTGACATGTGTATATAAGTATATATAACAAAGCACCAGTGGTCTAG
YD66	YD66_tRNA_R	TGGTCTCTAAGCTGCACCAGCCGGG
YD127	YD127_YD35_GG	TGGTCTCTGGAGCATGATTATGGCCATTTAGCAGC
YD128	YD128_YD36_GG	TGGTCTCTAGCGCCAACTTTCATGACTTATTTGTCC
YD131	YD131_RAD54_seq1	GTGCCAATCAAGGGATATGTAC
YD158	YD158	TGGTCTCTGGAGCGGAGAATTAAGGG

YD159	YD159	TGGTCTCTGCTCCCAATCTCAGTCACAGGAAGAAC
YD160	YD160	TGGTCTCTACCTATGAATAGTACACATATGACTTCATTG
YD161	YD161	TGGTCTCTAGCGTTTAGACATGTGTATATAAG
YD162	YD162	TGGTCTCTAGCGTGATAACAATCGGC
YD168	YD168_TIKI_B_F1	TGAAGACTTTGTGTTTAGACACAGAGGCTTATCTAGAGGGTTTTAGTTTTCCAAATTCAGTTATGATA TTATG
YD170	YD170_TIKI_B_R	TGAAGACTTTCTGTTTAGACACAGAGGCTTATCTAGAGGGCATTGCAATAATTTTCCAGTAAATTTG
YD181	YD181_551_f	TGGTCTCTGGAGATAACTTCGTATAATGTATACTATACGAAGTTATCGCTAGAGACCA
YD182	YD182_551_r	TGGTCTCTAGCGATAACTTCGTATAGTATACATTATACGAAGTTATCTCCAGAGACCA
YD185	YD185_35S-TACT	TGGTCTCAAGTAACTCGAGCGTGTCCTCTCC
YD186	YD186_66_f	TACTATAACTTCGTATAGCATACATTATACGAACGGT
YD187	YD187_66_r	TACTACCGTTCGTATAATGTATGCTATACGAAGTTAT
YD192	YD192_GG_134	TGGTCTCTGGAGACGTTAGTGTTGAGTAGTGGC
YD193	YD193_Npt/ecbe_F	GATGAAGCAGCTCGTTTGTCC
YD194	YD194_Npt/ecbe_R	TCTCTCAATTCGGTGCTTTGC
YD204	YD204DELLAadaptF	CCTGGGCCCCGAGTCTCTAATGGCGGTTTTCC
YD205	YD205DELLAadaptR	GCACGCCCGCCACCCATAGCCATCTCAAGCTG
YD212	YD212_barcd_F1	GAAAACCTGGGCCCCGA
YD213	YD213_barcd_F2	GAAATCCTGGGCCCCGA
YD214	YD214_barcd_F3	GAATTCCTGGGCCCCGA
YD215	YD215_barcd_F4	GATTTCCTGGGCCCCGA
YD216	YD216_barcd_F5	GTTTTCCTGGGCCCCGA
YD217	YD217_barcd_F6	GTTTACCTGGGCCCCGA
YD218	YD218_barcd_F7	GTTAACCTGGGCCCCGA
YD219	YD219_barcd_F8	GTAAACCTGGGCCCCGA
YD220	YD220_barcd_R1	AAATAGCACGCCCGCCA
YD221	YD221_barcd_R2	AATAAGCACGCCCGCCA
YD222	YD222_barcd_R3	ATTATGCACGCCCGCCA
YD223	YD223_barcd_R4	ATATTGCACGCCCGCCA
YD232	YD232_GFP_CAGA_F	TGGTCTCTCAGAGGATCTGGATCTGAGTC
YD233	YD233_GFP_GAGC_R	TGGTCTCTGCTCTTATCATTCCTCACCAGCATC
YD246	YD246	GGGTCTTTCTTCCTGGCTATTC
YD279	YD279DELLAadaptF	CCTGGGCCCCGAATCCTTTAATGGTGGTGTCTC
YD280	YD280DELLAadaptR	GCACGCCCGCCATACCATCTTCCATAGTTGTACC
YD233	YD233_GFP_GAGC_R	TGGTCTCTGCTCTTATCATTCCTCACCAGCATC
YD246	YD246	GGGTCTTTCTTCCTGGCTATTC
YD279	YD279DELLAadaptF	CCTGGGCCCCGAATCCTTTAATGGTGGTGTCTC
YD280	YD280DELLAadaptR	GCACGCCCGCCATACCATCTTCCATAGTTGTACC