Targeted DNA insertion in tomato using RNA-guided nucleases

By Lila Grandgeorge

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

The Sainsbury Laboratory

July 2021

Abstract

Insertion of novel DNA sequences at defined locations in plant genomes, known as a knock-in (KI), is highly desirable due to its potential for crop trait improvement. However, KIs are difficult to achieve. Previous KI attempts showed variable efficiencies (0.1% to 25%) in different plant species. I aimed to establish a high efficiency KI protocol in tomato and tested a range of variables for their ability to improve rates of KIs by homologous recombination. A 35S promoter or 35S enhancer was targeted upstream of the tomato *ANT1* gene, leading to purple pigmentation of tissues upon successful insertion and this was scored to measure KI efficiency. Variables tested included induction of double-stranded breaks (DSB) at the genomic target using CRISPR/Cas9 or a nickase allele to deliver single stranded breaks. Two viral replicons based on different strains (acute or mild) of a Bean Yellow Dwarf virus (BeYDV) Geminivirus were tested to provide a high copy number of the donor template, and for assessing replicon cargo size impact on KI efficiency. In these experiments, KI efficiencies were low and did not reach above 3%. Regeneration of edited plants from excised purple sectors was challenging due to overgrowth of surrounding wild-type cells. To limit the growth of "escape" background tissues and improve the regeneration of plant material containing a KI, I tested the use of a temperature-dependent selectable marker (Degron-NptII) to eliminate cells enabled to survive on selective medium through transient *NptII* expression rather than stable insertion of the T-DNA. Experiments demonstrated the degron-NptII strategy improved the selection precision of transgenic tissues at the callus, shoot and root formation stages and reduced the occurrence of escapes. Incorporating the degron-NptII strategy and removing the viral replicon (suspected to interfere with the regeneration process) from transformation vectors, further variables were tested for high efficiency KI. Compared to Cas9, a temperature tolerant *Lb*Cas12a (ttCas12a) allele achieved higher KI rates (26.2% compared to 20.7%). Inducing three DSBs (one DSB at target site and two at extremities of the donor fragment on T-DNA) instead of one DSB (at genomic target) increased rates of KIs when using Cas9 (12% compared to 22%). The three DSBs approach improved rates of KIs to a lesser extent compared to one DSB when utilising tt*Lb*Cas12a (24.2% compared to 28.2%). Additionally, the blight resistance gene *Rpi-vnt1* was knocked in alongside a 35S promoter, making a 7.3 kb DNA insert and resulted in a mean of 27.8% KI efficiency. True KI events with full *R* gene insertion were confirmed by PCR and Sanger sequencing in several samples.

Table of contents

List of tables

List of figures

List of appendices

Acknowledgements

I would like to begin by thanking my supervisor Prof Jonathan Jones for giving me the opportunity to undertake this PhD project in his lab at an exciting time for the development of genome editing technologies. I am thankful for his scientific ingenuity which led to many "cunning plans" and fruitful experiments. It has been a rich experience and I am grateful for the many scientific and interpersonal skills I have been able to perfect over the years. Likewise, I would like to sincerely thank my supervisor Prof Wendy Harwood for her insightful scientific guidance and precious support and encouragements throughout this project. Thank you to the other members of my PhD supervisory team, Dr Tom Lawrenson, and finally to Dr Laurence Tomlinson for her continued support, friendship, mentorship over the years and close supervision during the first year of my PhD.

The Sainsbury laboratory has provided me with a stimulating work environment, and I feel very fortunate for the help I received from our dedicated support teams, with special thanks to Matthew Smoker and Jodie Taylor for their tissue culture work and Mark Youles for providing Golden Gate cloning modules and troubleshooting advice. Thanks to Phil Robinson from JIC for photographing my tomato explants. The completion of this project would not have been possible without the help of other talented individuals; thanks to Dr Nicola Atkinson for sharing her tissue culture knowledge and help with fine-tuning the regeneration of purple sectors into shoots. Thanks to Hsuan Pai for her help assembling the final constructs used in this research project. Thanks to Dr Hee-Kyung Ahn for her help with devising the Image J analysis protocol.

I would like to extend my thanks to all the members - past and present - of JJ's lab for their cheerful companionship, with special thanks to Joanna, Aga and Hee-Kyung. Their friendship has been an immense source of joy and support over the last four years. Equally, being part of the TSL student community has been a delightful source of laughter, kindness, and support.

Last, but not least, I would like to express my deepest gratitude to my partner and to my family, especially to my dad, for their loving support. Thanks for bringing back assurance in myself when confidence lacked and for reviving my motivation and determination when times were challenging. Your pride in my PhD research endeavour certainly gave me a sense of accomplishment which has been a glorious driving force.

Abbreviations

ANT1	*Anthocyanin 1* gene
BeYDV	Bean yellow dwarf virus
Bp	Base pair
Cas	CRISPR-associated
CDS	Coding sequence
CRISPR	Clustered regularly interspaced palindromic repeats
crRNA	CRISPR RNA
Cycb1	Cyclin B1;1
Degron	Degradation signal
Diglig	Digestion ligation
DNA	Deoxyribonucleic acid
DSB	Double-stranded break
gRNA	guide RNA
HR	Homologous recombination
ICE	Inference of CRISPR Edits
Kan	Kanamycin
Kb	Kilobase
KI	Knock-in
LIR	Long intergenic region
MMEJ	Micro-homology end joining
MRN	MRE11, RAD50 and NBS1 complex
MRE11	Meiotic Recombination 11
NBS1	Nijmegen Breakage Syndrome 1
NptII	*Neomycin phosphotransferase* type II
NHEJ	Non-homologous end joining
NPBT	New plant breeding techniques
nt	Nucleotide
PAM	Protospacer Adjacent Motif
PCR	Polymerase chain reaction
Pol	Polymerase
RAD50	Radiation sensitive 50
Rep	Replicase gene
RPA	Replication Protein A

Rpi-vnt1	Resistance to *Phytophthora infestans* from *Solanum venturii*
SDN	Site-directed nuclease
SDSA	Synthesis-dependent strand annealing
SIR	Short intergenic region
SSB	Single stranded break
T-DNA	Transfer DNA
TALEN	Transcription Activator-like Effector Nuclease
TIDE	Tracking of Indels by DEcomposition
Tt	temperature-tolerant
ZFN	Zinc-finger nuclease

Chapter 1. General introduction

Environmental stress pressure faced by crops in the fields is ever increasing due to climate change, with more frequent extreme weather events, such as drought, and changing weather patterns which promote the incidence and spread of pests and diseases. Yet, we need to produce more food for a rapidly growing population. This poses a serious challenge to the plant breeding industry who need to produce new crop varieties more resilient to abiotic and biotic stresses to curb and prevent yield losses. Classical breeding programmes involve lengthy and complex paths to produce new crop varieties. The advent of site-directed nucleases (SDN), such as CRISPR Cas nucleases, is offering unprecedented opportunities for plant breeding *via* genome editing. Targeted insertion of novel genes at predefined locations in a crop's genome to confer a beneficial trait, referred to as knock-in (KI), or gene targeting, is a promising application of SDNs. But, as any technology still in its infancy, it requires concerted efforts to elaborate and refine protocols to advance crop trait improvement, as KIs are challenging to achieve in most plants because they occur at low rates. My thesis research aims to test a series of technical and biological variables for their potential to boost the frequency of KIs in the solanaceous crop, tomato. This introduction chapter will present the current challenges of crop trait improvement and the advantages of targeted genome manipulations. I will also review the gene targeting field and what underpins a successful KI and, reciprocally, what factors contribute to low success of KI endeavours.

1.1 The long road to crop trait improvement

Crop production is facing unprecedented challenges. Not only does crop production need to double to meet the demand of a predicted world population of 9.6 billion by 2050 (Tilman et al., 2011), it needs to achieve this utilising fewer resources and under increased, unpredictable environmental pressures. In the face of climate change, irregular weather patterns and extreme

weather events i.e. drought, floods, frosts, will become more frequent and all have a destructive potential for crop production leading to crop loss. Likewise, the occurrence and spread of pest and disease outbreaks will also be altered by climate change and will provoke additional crop losses (Mbow et al., 2019). Moreover, changing weather patterns are threatening up to a third of the global crop production, as currently arable lands may become unsuitable for agriculture if global warming is not limited to 1.5 °C / 2 °C (Kummu et al., 2021). Importantly, the food system (from cropping to retail) is accountable for 21-37% of the global greenhouse gases (GHG) emissions (carbon dioxide, methane and nitrous oxide) which are contributing to climate change, 9-14% of which are attributed to crop and livestock production (Mbow et al., 2019). Furthermore, applications of fertilisers lead to environmental nitrogen and phosphorous pollution, causing eutrophication of lakes and rivers and causing a detrimental cost to wildlife and ecosystems (Adegbeye et al., 2020). Similarly, pesticide transfer from the field into the environment causes damages to other living organisms (e.g. birds, beneficial insects, aquatic life) due to their toxicity and results in air, soil and waterway pollution (Tudi et al., 2021). Additionally, the manufacturing and transport of fertilisers and pesticides bears an important carbon footprint (Audsley et al., 2009). In a concerted effort to promote sustainable agricultural practices, the European Commission (EC) recently released the Farm to Fork Strategy (EC, 2020), central lever to the European Green Deal, a political commitment to respond to the climate crisis by reaching net zero GHG emissions by 2050 (EC, 2019). The Strategy sets the target of a 30% reduction in fertiliser use and 50% reduction of harmful pesticide usage by 2030. Meeting this complex challenge will require a profound transformation of crop production systems and will likely rely on interdisciplinary innovations, amongst which plant genetic improvement has a key role. Indeed, breeding of improved cultivars will help build crop resilience against abiotic and biotic stresses to prevent yield losses whilst reducing reliance on fertilisers and pesticides. Thus, plant breeders require an efficient and robust toolkit to meet that challenge.

Plant breeding seeks to produce improved plant cultivars to meet the needs of farmers and consumers by re-combining beneficial alleles to achieve 'superior' phenotypes. Important traits that receive much attention include yield, nutritional quality and shelf-life. Although the domestication of crops started around 10,000 years ago with the selection of plants with phenotypes that facilitated harvest or had high productivity, a more conscious effort for trait improvement began in the 20th century. Using the concept of genetic trait inheritance laid out

by Gregor Mendel, plant breeders have been able to produce new cultivars by introgression of useful alleles of genes from the same species or closely related species harbouring desirable traits *via* crosses, followed by the screening and selection of progeny for desired phenotypes - a breeding method referred to as traditional or classical breeding. Classical plant breeding therefore relies on existing mutations within the genetic pool of elite cultivars or sexually compatible species for the generation of improved traits as a result of genomic recombination after crosses (Taagen et al., 2020), which is less amenable in vegetatively propagated crops (e.g. potato). Moreover, the occurrence of a genetic bottleneck during crop domestication has caused the loss a lot of genetic diversity along the way, reducing the potential for improving traits. Therefore, an important aspect of plant breeding is to increase genetic diversity in elite cultivars through the deliberate generation of mutations or through introgression of natural variation.

Various methods have been developed to do so. For instance, mutation breeding, which relies on plant tissue exposure to mutagenic agents, has been extensively used since the 1920s, when the application of mutagens such as radiation e.g. x-rays (Stadler, 1928) and chemicals e.g. ethyl methane-sulfonate, EMS were shown to produce heritable mutations. Within 30 years, mutation breeding became a well-established breeding technique and has since produced over 3200 crop varieties which are listed in the Mutant Variety Database (https://mvd.iaea.org/). Nevertheless, mutation breeding is hindered by the stochastic nature of mutations created, which incurs creation of large mutant populations followed by extensive screening, making this breeding method laborious and time-consuming (Mba, 2013). Indeed, breeding programs using traditional breeding and mutation breeding can take between seven to twelve years to complete i.e. until cultivar release (Acquaah, 2012). Although technical advancement such as marker-assisted selection (MAS) – where endogenous DNA markers can be tracked with co-segregating genes conferring that desired phenotype – have improved the speed of these breeding programmes, classical and mutation breeding may still not be efficient enough to meet the predicted demand of improved crops (Scheben et al., 2017). However, the recent development of speed breeding, which significantly accelerates completion of a plant life cycle, shortening duration between generations may be a game-changer and is recognised as accelerating breeding programs (Hickey et al., 2019).

The ability to deliver defined DNA sequences into plant cells since the 1980s (Fraley et al., 1983) was pivotal for more rapid manipulation of plant genomes and the development of transgenic breeding, where desirable traits are conferred *via* the integration of exogeneous genes. Three methods are available for DNA delivery to transform plants:

(1) *Agrobacterium tumefaciens*, the causative agent of the crown gall disease in plants (Smith and Townsend, 1907), can be genetically engineered to contain on its transfer-DNA (T-DNA) sequences of interest that will be transferred inside the plant's nucleus as a single stranded DNA (ssDNA) molecule *via* the bacterial type IV secretion system, with the potential for the T-DNA to become stably integrated in the genome (Lacroix and Citovsky, 2019), thereby inserting material contained on the latter in the plant genome. So far, *Agro*-mediated plant transformation has been the prevalent technology used over the last 30 years for the production of genetically modified (GM) crops, such as the widely grown insect-resistant *Bt*-cotton (Raman, 2017).

(2) Biolistic bombardment uses DNA-coated metal particles to deliver DNA directly inside plant cells, thereby resulting in transient and stable transformation of the recipient cells (Altpeter et al., 2005). The cellular damage caused by the pressure during particle bombardment is destructive for the exposed tissues, and the DNA that is integrated is often concatenated and fragmented (Altpeter et al., 2016).

(3) Protoplast transformation by polyethylene glycol (PEG-transformation) or electroporation enables uptake of DNA through the cell membrane, made temporarily permeable (Mathur and Koncz, 1998). Although protoplast isolation and transformation is applicable to most plants, not all plants are amenable to full plant regeneration from protoplast and this requires laborious and time consuming protocols (Baltes et al., 2017).

Despite providing a faster and more versatile approach compared to classical and mutation breeding, transgenic breeding has its own shortcomings. Limitations arise from the randomness of the integration of the T-DNA in the genome. This can incur transgene expression variability based on insertion site (Butaye et al., 2005) and also has the potential to disrupt endogenous gene expression upon insertion (Bouché and Bouchez, 2001). Targeted transgene insertion may prevent such issues by directing the transgene to a genomic location that can be tested for transgene expression level. Moreover, due to random T-DNA insertion, performing sequential transformations to confer multiple novel traits would result in scattered

transgenes insertions across the genome. Because these will not co-segregate as one locus in the progeny, it renders more complex the tracking of the transgene segregation in the progeny to obtain a genetically stable line. Besides, the development of transgenic breeding has been limited by the necessities required for complying with the stringent regulation of GM crops which requires costly and lengthy tests and safety assessment (Prado et al., 2014) and transgenic varieties have faced public mistrust (Sikora and Rzymski, 2021).

Current methods available to breeders to produce improved crop varieties have clear limitations which include lengthy breeding program (for classical and mutation breeding) or safety assessment (transgenic breeding). This hurdle is compounded to a lack of precision over the genetic changes promoted by classical and mutation breeding, and a lack of control and precision over exogenous transgene insertion site with transgenic breeding.

1.1.1.1 Precise and targeted DNA integration by homologous recombination

Being able to insert exogenous DNA material at precise locations in genomes has been the Holy Grail of plant genetic engineering for about 30 years (Puchta, 2016). Relying on the cell's endogenous homologous recombination (HR) pathway, exogenous DNA sequences can be inserted at chosen locations provided the sequence of interest is flanked by homology regions to the intended genomic target to enable the genetic information contained on the latter to be copied in *via* DNA polymerase-mediated synthesis. This process is referred to as gene knock-in (KI), or gene targeting (GT) (Atkins and Voytas, 2020; Huang and Puchta, 2019). This method has the powerful potential to enable precise genome engineering by the insertion of single genes, but also gene stacking at one locus and gene/allele replacements. Moreover, having control over the insertion location of exogenous genes of interest means that plant genomes can be scanned for 'landing pads' or 'safe harbours', where the HR-mediated insertion of the transgene will not trigger detrimental effects on the plant (Dong et al., 2020; Gao et al., 2020).

Additionally, having control over the insert site with KIs makes possible sequential transgene insertions by HR at one locus to create a gene stack where all knocked-in genes co-segregate together in the progeny which largely improves the breeding of varieties with different traits conferred by several transgenes. Besides, relying on targeted KIs to insert genes of interest enables the production of edited lines which only contain the desired DNA sequence and no T-DNA-contained sequences e.g. left/right border, selectable marker, as T-DNA inserts may be crossed out the genome from a KI line provided these T-DNA insertions are not genetically

linked to the KI target site. This is an advantage over *Agrobacterium*-mediated classical transgenic methods which generally produce lines that contain T-DNA specific sequences alongside the transgene of interest. Production of T-DNA/marker-free plants for commercialisation is preferred from a regulatory stand-point as unwanted DNA sequences can trigger regulatory concerns (Wolt and Wolf, 2018).

The first reports of GT in mammalian cells occurred in the mid-1980s (Smithies et al., 1985) and earned the 2007 Nobel Prize in Medicine to the pioneers of this field (Doetschman et al., 1987; Thomas and Capecchi, 1987). Shortly after, the first GT experiments in plants were reported: Paszkowski et al. (1988) demonstrated GT in *Nicotiana tabacum* protoplasts (Paszkowski et al., 1988). The frequency of recovery of clones carrying a KI varied between 10^{-4} and 10^{-5}. Offringa et al. (1990) also obtained GT frequency around 10^{-4} in *N. tabacum* (Offringa et al., 1990). Although the feasibility of GT in plants was confirmed 30 years ago, the deployment of this method as a tool for genome engineering has been largely hampered by the low rates of KI event recovery (Puchta, 2002).

1.1.1.2 Explaining the low rates of gene targeting in plants

Homologous recombination is not the predominant DNA repair pathway

Formation of DNA double stranded breaks (DSBs) may induce cell death if not repaired, hence intricate and robust DNA repair mechanisms have evolved in living organisms to ensure genome stability in the face of DNA damage. DSBs are sensed by ATM (ataxia telangiectasia mutated) and ATR (ataxia telangiectasia and Rad3 related) kinases and trigger cell cycle arrest and signal transduction cascade *via* phosphorylation of the histone variant H2AX around the break site to recruit downstream DNA repair factors (Manova and Gruszka, 2015). Although the main DNA DSB repair pathway is HR in bacteria and yeast, the choice of repair pathway is more complex in multicellular eukaryotic organisms such as plants. DBS repair is composed of two major pathways, the HR-mediated repair and the non-homologous end-joining repair (NHEJ), each branching into several subtypes (Figure 1.1). NHEJ, the predominant repair pathway in somatic cells (Knoll et al., 2014), usually creates small changes to the DNA sequence during the repair process i.e. deletions or insertion of few base pairs (bp) (Puchta and

Fauser, 2013). NHEJ is constitutively active throughout the cell cycle and is very efficient at repairing DSB due to the high affinity of its repair proteins, Ku70 and Ku80, for exposed DNA ends (Shibata et al., 2018). The heterodimer Ku70/Ku80 binds to the broken ends and acts a tethering platform between the two ends, protecting them from end-resection and concomitantly recruiting downstream complexes such as DNA ligase 4 (LIG4)-XRCC4 for the repair of the phosphodiester bond between the broken ends (Manova and Gruszka, 2015).

NHEJ is classed as a non-conservative, mutagenic pathway and has therefore received limited attention to mediate the precise integration of exogenous DNA sequences, despite its prevalence as a repair mechanism in cells. The main reason stems from the likely generation of mutations on either ends of the inserted DNA and/or at the target locus, making for a less precise insertion compared to sequences inserted by HR. Additional mutations to the genome aside from the intended KI may elevate regulatory scrutiny over the final product. For some applications, precise insertion of DNA may be required for a successful KI, i.e. in frame insertion of a coding sequence. Formation of NHEJ-derived mutations on the template or at the target site would impede KI success rate.

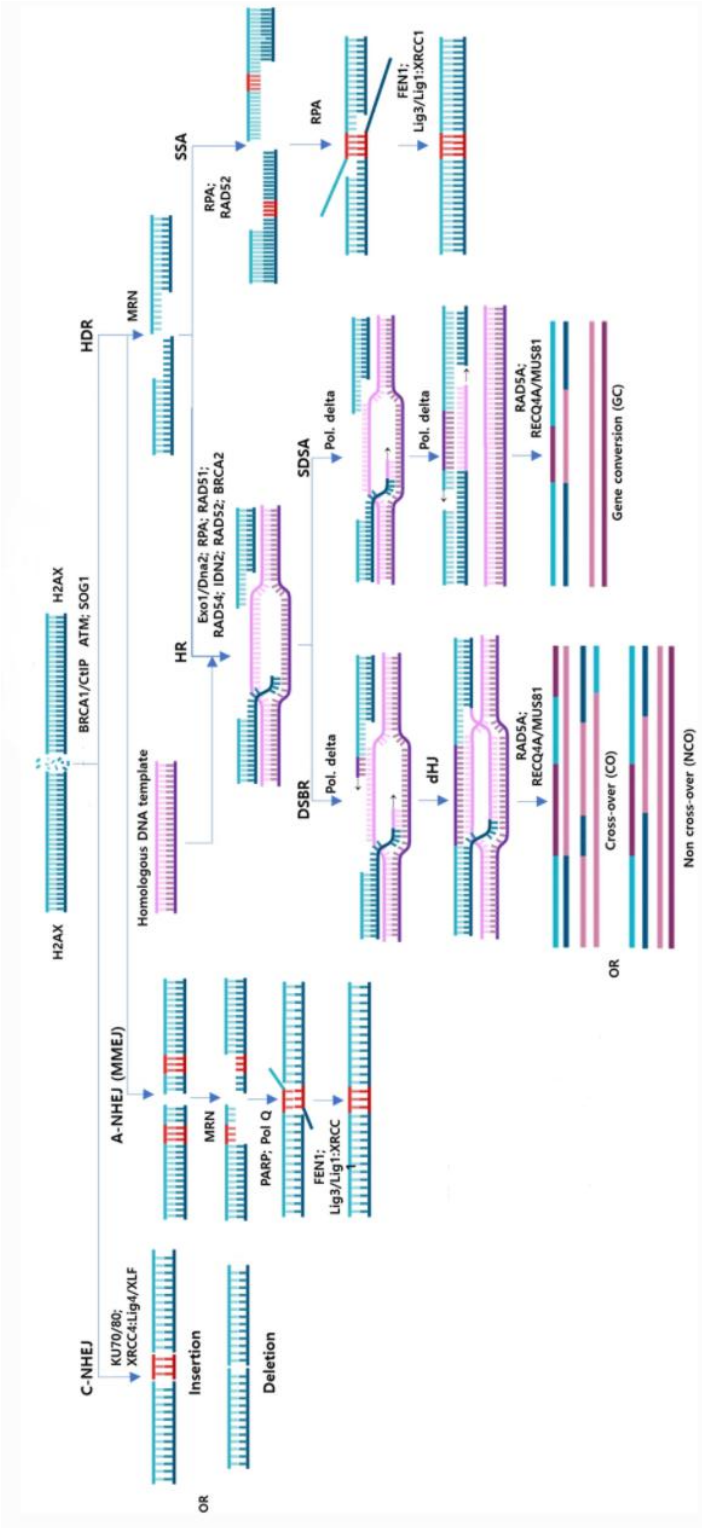


Figure 1.1 Diagram representing the subtypes of DNA DSB repair pathways branching from the two main categories, HR and NHEJ. Classical NHEJ (c-NHEJ) ligates DNA ends often creating indels. The alternative NHEJ (alt-NHEJ), also called microhomology mediated end joining (MMEJ), proceeds by strand annealing between microhomologies (represented in green) exposed after short end resection by the MRN complex, followed by break repair with PARP and Pol Q. Single strand annealing (SSA) operates similarly to MMEJ but with

extended DNA resection with the intervention of additional HR-specific components like RPA. Break repair via HR necessitates formation of elongated 3' ssDNA nucleoprotein filaments involving RPA, BRCA2 followed by RAD51 and RAD54-mediated homology search. Upon annealing to a homologous sequence, strand displacement (D-loop) and invasion of the filament into the homologous dsDNA take place and are followed by DNA synthesis by Pol delta. After branch migration and D-loop resolution, DNA content from the homologous template has been precisely inserted at original break site, resulting in a non-cross over (NCO) outcome, or gene conversion (GC), during Synthesis dependent strand annealing (SDSA). During DBS repair (DSBR), two Holliday junctions are formed and can be resolved to result either in a cross-over (CO) or a non-cross over (NCO) repair outcome. Figure adapted from Van vu et al. (2019).

HR repair pathways begin with CtIP-mediated (C-terminal binding protein-interacting proteins), or SOM1 in plants (Uanschou et al., 2007), followed by activation of the MRN complex (composed of Meiotic Recombination 11 (MRE11), Radiation sensitivity 50 (RAD50) and Nijmegen Breakage Syndrome (NBS1)) to perform 5' end resection of the DNA ends to expose 3' ssDNA overhangs. Four different pathways can subsequently be recruited to complete the repair. Microhomology-mediated end joining (MMEJ), although considered as an alternative NHEJ pathway, relies on exposure of microhomologies (2-20 bp) within the resected 3' ssDNA overhangs, strand annealing between the homologies, gap filling by Polymerase Q (Pol Q), and ligation to seal the break. Likewise, single-strand annealing (SSA), anneals regions of microhomology between the 3' ssDNA overhangs produced after resection without relying on a repair template (Orel et al., 2003). As genetic material is lost during this process, it is generally seen as non-conservative and not considered as a true HR mechanism. Other HR-based pathways which are considered as “true HR” include the synthesis-dependent strand-annealing (SDSA). The resected 3' ends are coated by Replication Protein A (RPA) proteins to protect the DNA from degradation. Subsequently, Radiation sensitive 51 (RAD51) replaces RPA, helped by Breast Cancer 2 (BRCA2), to form a nucleoprotein filament that will proceed to the homology search step in the nucleus (Van Vu et al., 2019). When homology is found, one of the 3' ends invades the homologous dsDNA molecule (the repair template) by displacing one strand (D-loop formation) and annealing to the complementary strand. Next, DNA is synthesised by Pol delta, copying in the genetic

information contained on the repair template at the break site. The second true HR pathway, often referred to as classical HR, will proceed from the involvement of both resected 3' ends in strand invasion and D-loop formation, creating two Holliday junctions which can lead to a non-cross over outcome (genetic information is transferred from the template DNA to the repaired DNA molecule) or a cross-over outcome (genetic exchange between the repaired DNA and the template DNA) depending on the resolution of the DNA junctions by Radiation sensitive 54 (RAD54) and the helicase RECQ4. SDSA was shown to be the main repair pathway recruited in somatic cells to generate KIs (Puchta, 1998). HR is restricted to cell cycle stages from mid-S phase until the end of the G2 phase, during which HR and NHEJ repair enzymes compete against each other to repair DSBs, as Ku heterodimers have been shown to have an inhibitory effect on HR processes (Fukushima et al., 2001).

Taken together, these findings demonstrate that the complex and hierarchical nature of DSB repair in plants predominantly leads to non-true HR events, which explains the low rates of KI observed previously. Strategies for boosting GT efficiency are therefore highly desirable and the last 30 years have seen multiple attempts with mixed success (reviewed in Puchta, 2002; Puchta and Fauser, 2013).

Chromatin structure influences DNA repair outcome and editing efficiency

Plants, like other eukaryotes, have a highly structured genome organised into nucleosomes, chromatin further divided in two types, and chromosome territories inside their nucleus. The occurrence of DSBs therefore takes place in the context of chromatin. The encounter between the broken DNA and its intact homologue is a prerequisite of HR (Lieberman-Lazarovich and Levy, 2011). Therefore, physical distance in the nucleus between the targeted locus and the provided homologous template could influence GT success rates. It was indeed shown in yeast that the physical distance between a broken molecule and its homologous sequence is a limiting factor of HR since recombination rates are negatively correlated with increased distance (Lee et al., 2016). In maize, frequent interchrosomal recombination between endogenous homologous sequences occurring in close proximity was observed (Liu et al., 2020c). Moreover, in a recent study in barley, it was shown that in the F2 progeny from KI lines, individuals containing a KI also had a T-DNA insertion which happened to be genetically linked to the targeted locus, suggesting that KI success and donor template proximity are related (Lawrenson et al., 2021). Therefore, finding the homologous partner in

the genome is challenging requiring extensive chromatin remodelling and nucleosome shifting to undertake homology search, further limiting gene targeting success.

An additional layer of regulation of HR is achieved through chromatin conformation. Several studies have shown that chromatin folding at the break site determines which repair pathway will be recruited. In human cell lines, it was established that transcriptionally active regions are preferentially repaired by HR, which is determined by specific histone marks. The active chromatin mark histone 3 lysine 36 tri-methylation (H3K36me3), which is associated with transcription, attracts the chromatin-interacting protein LEDGF which subsequently recruits the key HR initiator factor, CtIP, that will recruit MRN to perform the DNA resection step (Daugaard et al., 2012). Moreover, a genome-wide analysis of DSBs indicated that actively transcribed regions have a higher propensity to be repaired by HR, shown *via* RAD51-bound DSBs in regions enriched in H3K36me3 and with high Polymerase II levels (indicative of transcription elongation) (Aymard et al., 2014). Moreover, a DSB, even if less than 1 kb away from a gene, if the latter is not transcribed, the DSB is less likely to be bound by RAD51, and demonstrates the strong influence of transcription on favouring HR. Another study revealed a role for histone acetylation in repair choice (Tang et al., 2013). Histone 4 Lysine 16 acetylation is enriched at the 5' end of transcribed genes and correlates with elevated levels of BRCA1 (an HR-specific factor) at DSBs in such regions. This acetylation mark prevents the recruitment of 53BPI, an anti-resection factor, at that break site. Taken together, these findings demonstrate the crucial role of chromatin structure in specifically recruiting HR factors to active euchromatin-located DSBs. Therefore, it may be possible to use this knowledge in gene targeting experiments to select landing pads based on their genomic location and chromatin structure in order to increase the rates of knock-ins. Although such experiments are lacking in a plant system, these data from human cell lines may also apply to plants based on the broadly conserved pathways of DSB repair between mammals and plants (Spampinato, 2017). Nevertheless, some evidence exists that chromatin structure also controls HR in plants. Arabidopsis mutants in the chromatin assembly factor 1 (*CAF-1*) involved in nucleosome formation exhibit a 40-fold increase in somatic homologous recombination (Endo et al., 2006). This increase in somatic HR in *caf-1* mutants is explained by a general open chromatin state and not by increased interactions between homologous chromosomes (Kirik et al., 2006).

To summarise, several cellular factors can affect rates of KI. Relying exclusively on the spontaneous occurrence of somatic HR to promote the insertion of an exogenous gene at a

chosen locus is bound to be a low efficiency approach because of two prerequisites: (1) incidence of a DSB at the desired genomic locus and (2) recruitment of HR to repair the break. Yet, DSB repair regulation is complex, HR restricted to specific phases of the cell cycle, and several other pathways can be more rapidly recruited to repair the break i.e. NHEJ, MMEJ, SSA. Furthermore, distance between the homologous donor template and the genomic target in the nucleus can impact the success rate of a potential KI, and inherent properties of the genomic target, such as chromatin structure, further influence the likelihood of a KI. Nevertheless, advances in molecular biology tools have improved the feasibility of KIs in plants.

1.2 Three kinds of site directed nucleases for genome editing

A significant improvement in GT efficiency came with the controlled and targeted induction of DSBs in the genome using SDNs. Induced DSB at the genomic target site primes the DNA repair response at the desired locus, elevating the possibility for HR to be recruited. In 1993, homing endonucleases, also called meganucleases, e.g. *I-SceI,* were used in tobacco protoplasts to deliver DSBs *in vivo* and increased rates of KI events were observed at the targeted genomic loci (Puchta et al., 1993). However, the use of meganucleases is limited to the number and location of their recognition sites occurring in the genome of the studied organism, which restricts the application of this technique. The real breakthrough came with the advent of programmable nucleases in the 2000s. This section will look at the three types of programmable nucleases available for genome editing nowadays.

1.2.1 Zinc Finger Nucleases

Zinc Finger Nucleases (ZFN) are the first discovered form of programmable nucleases and were developed during the 1990's. ZFNs work in pairs, each monomer being composed of 3 to 4 zinc finger domains, represented as the coloured boxes F1, F2 and F3 in Figure 1.2. Each recognise and bind 3 consecutive base pairs (Pabo et al., 2001) and provide the sequence-specificity of the nuclease. Each ZNF monomer is fused to a bacterial type IIS restriction enzyme, *FokI,* which cuts the DNA in a non-sequence-specific manner (Kim et al., 1996; Smith, 2000). Designing ZFN has proven complicated, time consuming and expensive and not always reliable, which limited the universal use of this tool.

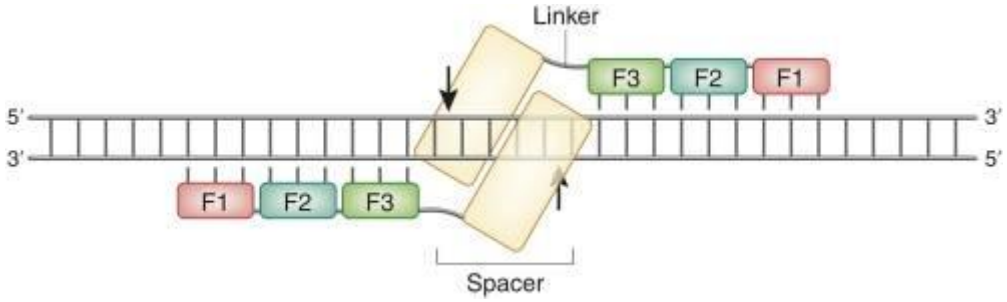


Figure 1.2 Representation of a ZFN dimer at its recognition site. The coloured boxes (F1, F2 and F3) represent the fused zinc finger proteins (usually 3 to 4 are used), each specifically recognising 3 successive base pairs in the DNA. Each ZFN monomer has a 9 to 12 bp recognition site. This DNA recognising protein complex is fused to the non-specific type IIS restriction enzyme *FokI* (yellow square) with the aid of a linker. Working as a dimer, a DSB is created by the cooperative activity of the two *FokI* nuclease domain. Reprinted with permissions from Oxford University Press, Copyright 2011. (Carroll, 2011).

1.2.2 Transcription activator-like effector nucleases

The second class of programmable nucleases are transcription activator-like effector nucleases (TALENs). These also are DNA-binding proteins, isolated from a family of effectors (referred to as AvrBs3 or TALE) secreted by the type III secretion pathway of the plant pathogen *Xanthomonas* (Boch and Bonas, 2010; Moscou and Bogdanove, 2009). Composed of direct repeats of 34 to 35 amino acids, TALE's DNA binding domain contains two amino acids, called the repeat variable residues (RVD), at position 12 and 13 within each of the repeats that recognise and bind a specific DNA base, as shown in Figure 1.3. TALEN monomers are usually designed with 15 to 20 RVDs chosen for their capacity to recognise and bind 30 bp at the target locus (Baltes and Voytas, 2015). The main drawback of this tool is the complexity of the design (requiring 2 TALEs, programmed to recognize several nucleotides apart on opposite strand) and its large size.

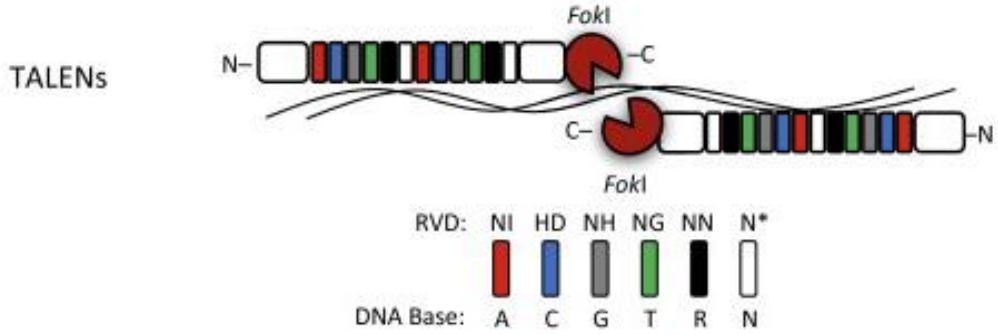


Figure 1.3 Configuration of a pair of TALENs. Each monomer consists of 10 to 15 repeat variable residues (RVD) that recognise different DNA bases in a one-to-one correspondence e.g. RVD NI recognises adenine (A) bases. TALENs DNA recognition domains are customised accordingly to the DNA target sequence. Each monomer is fused to the non-specific restriction enzyme *FokI*, which will cut the DNA at the target site. Adapted from *Enabling plant synthetic biology through genome engineering*, Baltes and Voytas, Copyright (2015), with permission from Elsevier.

1.2.3 Clustered regularly interspaced short palindromic repeat-associated nucleases

1.2.3.1 The discovery of the CRISPR system

The most recent addition to the array of programmable genome editing tools is the Clustered Regularly Interspaced Short Palindromic Repeat system (CRISPR) and its associated nucleases. It is the most widely adopted programmable nuclease due to its ease of use and has been referred to as a revolutionising discovery with a far-reaching potential, from medical to agricultural applications, and earned a Nobel Chemistry prize to the scientists who pioneered the discovery (Ledford and Callaway, 2020; Jinek et al., 2012). The earliest clue towards unravelling the CRISPR system dates back to 1987 when CRISPR sequences were first observed in *E. coli* (Ishino et al., 1987), but it took another two decades before deciphering their role in prokaryotes. These CRISPR loci, represented in Figure 1.4, are present in all prokaryotes and harbour conserved features. Within CRISPR arrays, short direct repeats of 14-20 bp (CRISPR repeats) are flanked by short stretches of 25-40 bp of variable sequence (CRISPR spacers) and with neighbouring CRISPR-associated genes (*Cas*) although the number and combination of *Cas* genes differ between bacterial species (Jansen et al., 2002). The spacer sequences show homology to sequences in bacteriophages (Pourcel et al., 2005; Mojica et al., 2005) and their function is associated with immune defence against these

infecting agents, as bacterial strains carrying a given spacer are immune to the phage from which that spacer sequence derives (Mojica et al., 2005). In accordance with this initial finding, the involvement of CRISPR spacers and *Cas* genes in bacterial immunity against invading nucleic acids was experimentally demonstrated (Barrangou et al., 2007). The presence of a proto-spacer in a CRISPR array provides immunity against infectious agents with corresponding DNA sequences in their genome, the neighbouring *Cas* genes encode protein components for immunity and new spacers are added in the CRISPR array post-phage infection, providing an immune memory to potentiate immune defence upon secondary infection by that same infectious agent (a step called adaptation, or immunisation). The defence response, called the interference step, is mediated by Cas proteins, such as Cas9, guided by a spacer (referred to as a CRISPR RNA, crRNA). Upon base pairing of the crRNA to its complementary proto-spacer, the nuclease undergoes conformational changes which exposes its two catalytic sites, RuvC, which cleaves the non-complementary strand of the target DNA, and the HNH site, cleaving the complementary strand (Jinek et al., 2012), producing a double-stranded cut 3 bp upstream of a proto-spacer adjacent motif (PAM), i.e. NGG in the case of Cas9, a feature in the target DNA that is essential for Cas nuclease activity (Garneau et al., 2010). The cleavage of the DNA of the invading agent (bacteriophage or plasmid) interrupts its proliferation inside the bacterium, thus promoting immunity.

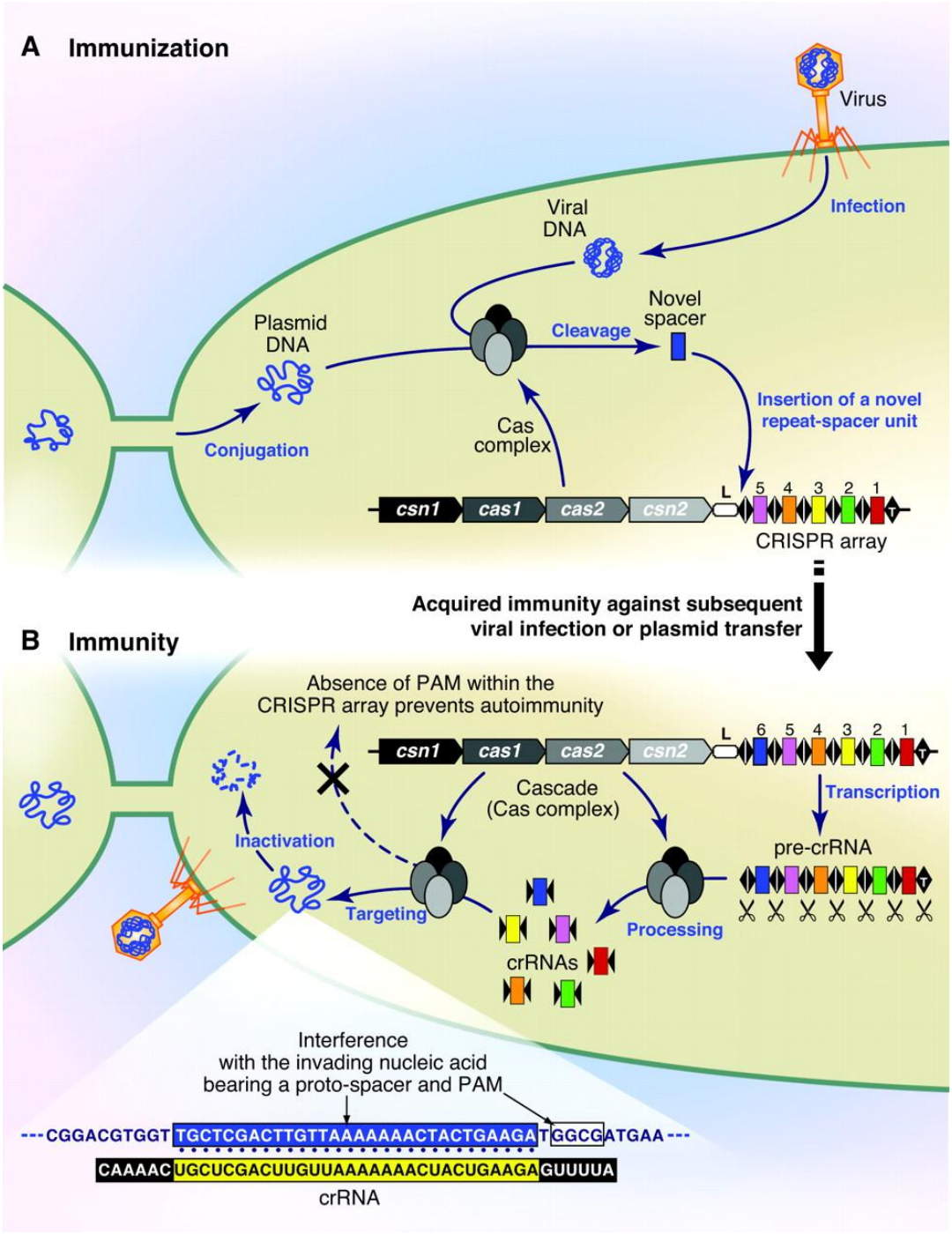


Figure 1.4 CRISPR/Cas-mediated immunity in prokaryotes. **(A) Immunisation, or adaptation step**. The proteins Cas1 and Cas2 are involved in the adaptation step by cleaving novel invading DNA and integrating it (acquisition step) as a new spacer unit within the CRISPR array. **(B) Immunity, or interference step**. CRISPR arrays containing the spacers are transcribed into long pre-crRNA before being processed into individual mature crRNAs by trans-activating CRISPR RNAs (tracrRNA) and endoribonuclease RNase III. (Deltcheva et al., 2011). The mature crRNAs then associate with Cas proteins to form an active CRISPR-associated complex for antiviral defence (Cascade) (Brouns, 2008). The Cascade will scan the

cell in search of a target site, first by searching for proto-adjacent motifs (PAM) which are 3 to 4 bp motifs, and then by complementary base-paring of the crRNA to the proto-spacer which cause conformational changes in the nuclease to expose its two catalytic sites (RuvC and HNH) to the DNA and induce a DSB to inactivate the proliferation of the invading DNA. From Horvath et al., 2010, reprinted with AAAS permission.

The great potential for RNA-programmable genome editing of the CRISPR/Cas9 system was realised when it was demonstrated that a crRNA can be fused to a partial tracrRNA to create an artificial, chimeric guide RNA (gRNA) molecule that retains its ability to bind to the Cas9 nuclease and to cleave its target DNA sequence (Figure 1.5) (Jinek et al., 2012). Thus, it is possible to engineer a synthetic gRNA with a crRNA sequence complementary to a locus of interest in a given organism. The year 2013 marks a milestone in CRISPR/Cas experiments as two proof of concept studies confirmed the applicability of the CRISPR/Cas system as a tool for genome editing in higher eukaryotes (Jinek et al., 2013; Le Cong, et al 2013). Although the first CRISPR-Cas targeted cleavage experiments were carried out using the *Streptococcus pyogenes* Cas9, *Sp*Cas9, several other orthologues have been isolated from the following bacteria, *Staphylococcus aureus*, *Neisseria meningitides*, *Streptococcus thermophilus*, which has increased the number of Cas9 variants available in the CRISPR toolkit (Cebrian-Serrano and Davies, 2017).

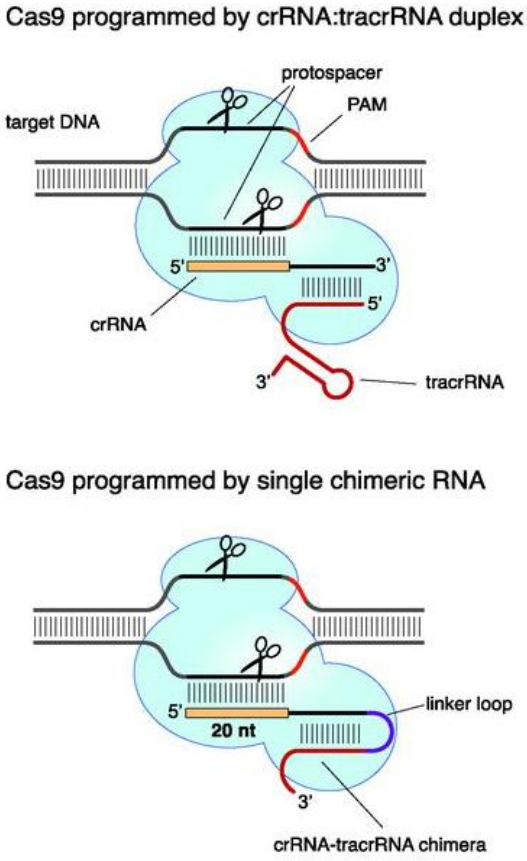


Figure 1.5 Programmability of CRISPR/Cas9. The Cas9 endonuclease is guided to its target site by a duplex RNA structure composed of the crRNA – with base pair complementarity to the protospacer – and the tracrRNA hybridised to the 3' end of the crRNA, which is essential for target recognition (top image). The crRNA:tracrRNA duplex is engineered into one chimeric RNA molecule containing around 20 bp complementary to the target DNA fused to the tracrRNA (bottom image) Adapted from Jinek et al., 2012, with permission from AAAS.

Furthermore, the repertoire of CRISPR nucleases available for RNA-directed genome editing is growing. Notable is the discovery of Cas12a (also known as Cpf1), isolated from the bacterium *Francisella* and the archea *Prevotella* (Zetsche, et al. 2015). Belonging to a similar class of CRISPR systems as Cas9, but belonging to type III sub-class (review on CRISPR system classification Makarova et al., 2020), Cas12a only requires a short crRNA of 43 nt (half the size of that of Cas9) and no tracrRNA is required as Cas12a contains an endoribonucleolytic domain to process pre-crRNA into mature, individual crRNA for RNA-guided DNA cleavage (Fonfara et al., 2016). Cas12a recognises T rich PAMs (TTTN) positioned at the 5' of the protospacer (as opposed to 3' end for Cas9) and delivers a staggered cut (unlike Cas9 which creates blunt DSBs) producing 5 bp overhangs and cleaves the DNA 23 bp downstream of the PAM sequence on the target strand and 18 bp downstream of the PAM sequence on the non-target strand (Zetsche et al., 2015). Several orthologues have been isolated and used for genome editing, *Lachnospiraceae bacterium* Cas12a, *Francisella novicida* Cas12a and *Acidaminococcus* sp. Cas12a (Zetsche, et al. 2015). Furthermore, belonging to the class V type V system, a new enzyme coined Cms1 (CRISPR from *Microgenomates* and *Smithella*) has been discovered and its editing potential demonstrated in rice (Begemann et al., 2017b). The CRISPR nucleases described above have the specificity to target DNA molecules guided by their gRNA. A novel enzyme, Cas13 (also known as C2c2), from the class 2 type VI system was isolated from *Listeria seeligeri* and shown to perform RNA-guided RNA cleavage using a single crRNA (Shmakov et al., 2015). This is an important discovery for studying RNA biology as this new tool provides a new method for RNA interference (e.g. immunity to RNA viruses) and post-transcriptional gene repression. Moreover, a catalytically inactive Cas13, dCas13, fused to an adenosine deaminase base could successfully produce mRNA base pair editing in mammalian cells, which carries important implications for therapeutics (Cox et al., 2017).

1.2.3.2 Optimisation and engineering of the CRISPR nucleases for versatile applications

Recent innovations have broadened the spectrum of application of CRISPR nucleases. Variants of Cas9 have been engineered to overcome some limitations of the technique (Table 1.1). For instance, the necessity for NGG PAM motif at the target site greatly restricts the number of potential genomic target sites. Recently, a Cas9 allele, xCas9, was engineered to exhibit broad PAM sequence recognition (Hu et al., 2018). A similar outcome was achieved for Cas12a, whose targeting range was estimated to increase by 3-fold (Gao et al., 2017). Remarkably, a SpRY Cas9 variant was recently engineered to be PAM-less, i.e. NNN, which opens up access to any genomic sequence (Ren et al., 2021). Besides, functional conversions of Cas9 have diversified its cellular applications. A catalytically dead Cas9 (dCas9) caused by two mutations, Asp10Ala and His840Ala, has been used to develop single-base editing systems, called base editors. Because the nuclease retains its RNA-guided DNA binding activity, various types of molecules have been fused to the dCas9 for different purposes. For example, a cytidine deaminase has been fused to the N-terminal of a dCas9 to promote an amino-acid change substitution (C to T) at predefined locations (Komor et al., 2016). A different reaction can be catalysed (A to G) by fusing an artificial adenine deaminase (Gaudelli et al, 2017). Alongside several proof of concept studies which confirmed the feasibility of base pair editing in plants (mainly in rice, Li et al., 2017; Lu and Zhu, 2017; Ren et al., 2018; potato, Veillet et al., 2019; Brassicaceae, Kang et al., 2018), there are now reports of base-editing being used for crop trait improvement (Bastet et al., 2019). Furthermore, a technique called 'prime-editing' can now be added to the CRISPR-Cas toolbox. Prime editing has been demonstrated in human cell lines to correct alleles associated with genetic diseases. A Cas9 nickase was fused to a reverse transcriptase and combined with a prime editing gRNA (pegRNA) which serves two functions: (1) directs the nickase at the target locus (2) provides the RNA template encoding the desired edits that will be transcribed at the nicked site by the reverse transcriptase. Prime editing enables unprecedented editing versatility (DNA insertion, deletion, base pair changes) while bypassing the need for a double-stranded break which often causes collateral damage in the genome (Anzalone et al., 2019). Prime editing has since been applied in rice and wheat (Lin et al., 2020; Butt et al., 2020). Additionally, CRISPR-Cas technology can also modify gene expression. For instance, a dCas9 can be fused to transcriptional activator (e.g. VP64) or repressor domains (e.g. SRDX) to modify the expression of the targeted gene (Lowder et al., 2015). Gene transcription has also been altered through epigenetic modifications of the target

site using a dCas9 fused to chromatin modifying enzyme domains (e.g. demethylase TET1) (Gallego-Bartolomé et al., 2018).

In recent years, scientists have endeavoured to improve the efficiency of CRISPR/Cas for genome editing and many toolkits with optimised protocols are already available for diverse organisms (for plants, see Cermak et al., 2017). CRISPR/Cas9 genome editing has been performed in a wide range of plant species and proof of concept studies have now progressed to applied research. Excellent recent reviews have detailed recent progress in crop improvement using programmable nucleases (Bhat et al., 2020; Zhu et al., 2020). Crop traits that have received attention include yield, disease resistance (reviews by Veillet et al., 2020; Zaidi et al., 2020; Vu et al., 2020a focus on tomato), nutritional value and product quality (review by Ku and Ha, 2020), breeding capabilities of elite lines (Liu et al., 2020a; Yao et al., 2018; Enciso-Rodriguez et al., 2019). In-depth description of CRISPR-Cas genome editing studies for crop trait improvement mediated by gene knock-outs, base-editing (review by Mishra et al., 2020; Bharat et al., 2020) and prime-editing will not be provided in this introduction chapter as it falls outside the scope of this research project which focuses on genome editing *via* KIs. Next, we will look at how programmable nucleases have been used for genome editing by gene KI.

Table 1.1. List of Cas variants that have been used for genome editing purposes.

Nuclease	Description	Pros	Cons	Reference
Cas9	Nuclease making a blunt DSB at the targeted locus	Robust and well researched cleavage activity in many living species	Limited to occurrence of PAM near bp to be modified Some concerns with off-target activity	(Jinek et al., 2013)
Cas12a	Nuclease making a staggered DSB at the target locus	Staggered cut may be preferred for some application DSB occurs away from the PAM Lower off-target activity compared to Cas9	Limited to occurrence of PAM near bp to be modified T rich PAM restricts this enzyme from GC reach regions	(Zetsche et al., 2015)
Cas13	Nuclease cutting RNA molecules	Widens RNA-based fundamental and	Clear off-targeted activity	(Shmakov et al.,

		applied research		2015)
xCas9	Engineered to have broader PAM recognition i.e. NG, NNG, GAA, GAT and CAA	Additional genomic target sites available and lower off target activity	Not apparent from primary studies	(Hu et al., 2018)
Cas12a S542R/K607R and S542R/K548V/ N552R	Engineered to have broader PAM recognition i.e. TYCV and TATV, where Y= C or T and V=A, C, or G.	3 times more genomic target sites available and retains high on-target efficiency	Not apparent from primary studies	(Gao et al., 2017)
SpRY Cas9	Engineered to have no PAM restriction e.g. PAM = NNN	Opens editing to previously inaccessible loci	Secondary off-target effect	(Ren et al., 2021)
DNA Base editors	Dead Cas9 fused to cytidine deaminase / adenine deaminase	Generates precise one bp modification in the DNA instead of DSB	Limited to occurrence of PAM near bp to be modified	(Komor et al., 2016)
RNA Base editors	Dead Cas13 fused to adenine deaminase	Generates precise one bp modification in the RNA instead of DSB	Limited to occurrence of PAM near bp to be modified	(Cox et al., 2017)
Epigenome editors	Dead Cas9 fused to chromatin modifying enzyme domains	Widens epigenetic-based fundamental and applied research	Limited to occurrence of PAM near chromatin mark to be modified	(Gallego-Bartolomé et al., 2018)
Prime editors	Cas9 nickase fused to a reverse transcriptase guided by a pegRNA containing desired edit	Promotes short edits i.e. <20 bp (insertion/allele switch) without formation of a DSB	May not be suited to longer modifications i.e. >100 bp	(Anzalone et al., 2019)

1.3 Genome engineering by targeted insertion of novel genetic material

Induction of DSBs significantly increases GT efficiency by triggering the activation and recruitment of DNA the repair pathway at the genomic target. The advent of programmable nucleases accelerated GT investigations. First and second-generation SDNs i.e. ZFN and TALEN have both successfully been employed to generate KIs (Wright et al., 2005; Zhang et al., 2013b; Budhagatapalli et al., 2015; Shukla et al., 2009) although these experiments will not be discussed here as most of the recent reports involve Cas9. Of note, high rates of KIs can be obtained using ZFNs i.e. 28% KI efficiency of a 5 kb insert in tobacco BY-2 cells (Schiermeyer et al., 2019). In this section, we will review KI attempts using CRISPR-Cas9 nuclease.

1.3.1 CRISPR-Cas9 mediated gene targeting

GT frequencies vary greatly between experiments using CRISPR/Cas enzymes as the DSB-generating tool (reviewed by (Zhang et al., 2017b; Li et al., 2017; Dong and Ronald, 2021). Herbicide resistance by allele replacement (i.e. allele switch), whereby an endogenous gene allele is replaced by the allele provided on the donor template, has been used in KI experiments in various plant species such as *Arabidopsis thaliana* (Schiml et al., 2014), *Zea mays* (Svitashev et al., 2015), *Glycine max* (Li et al., 2015), *Linum usitatissimum* (Sauer et al., 2016), *Solanum tuberosum* (Butler et al., 2016a) and *Oryza sativa* (Endo et al., 2016; Sun et al., 2016b). Other traits have also been successfully modified such as purple anthocyanin pigmentation of tissues as a phenotypic screen for successful GT in tomato explants. Danilo et al. (2018) obtained 1.29% KI efficiency when restoring the wild-type allele from a 1013 bp deletion in the *DFR* gene. *DFR* encodes an enzyme from the anthocyanin biosynthesis pathway. Anthocyanin pigmentation is abolished in *dfr* plants, without any detectable impact on plant growth or fertility, making this an attractive locus to assay KI events. *dfr* explants did not display any purple tissue during tissue culture, whereas explants with a KI had restored production of anthocyanins which were visible in stem and leaf tissues (Danilo et al., 2018). Tissue-specific control of Cas9 (*Sa*Cas9) expression produced rates of heritable KIs averaging 1% efficiency in Arabidopsis in the T1 when under the control of an egg cell specific promoter EC1 (Wolter et al., 2018). Restricting Cas9 activity to the germline is thought to increase the likelihood of producing heritable KI events, increasing the rates of KIs. Boosting KI efficiency in

Arabidopsis was also achieved using a sequential transformation process. The parental line contained a stable T-DNA with the Cas9 expression cassette (Miki et al., 2018). This parental line was subsequently transformed with a binary vector that contains the gRNA expression cassette and the donor template. In the T2, between 6 to 9% KI efficiency at different loci were detected by PCR screens. This study also compared several promoters for Cas9 expression and found that the egg cell and early embryo-specific promoter, *DD45*, gave the highest rates of KI. The same authors recently published another KI attempt using a "all-in-one" plasmid transformation approach (as opposed to sequential transformation of CRISPR and KI reagents in their previous publication). KI efficiency was ten times lower, 0.68% (Peng et al., 2020), than in their previous experiment (Miki et al., 2018). Nevertheless, KI efficiency could be boosted to 2.4% by using a viral translational enhancer to improve Cas9 expression, highlighting the scope of improvement that can be achieved by optimizing key parameters underlying KIs.

High efficiency allele replacement was achieved in tomato at the *ALS1* locus; 31 explants out of 244 (12.7%) acquired resistance to the herbicide chlorsulfuron after allele switch (Danilo et al., 2019). Interestingly, no T-DNA insertion could be detected in 38% of the edited T0 following pre-selection for cells with transient *NptII* expression on kanamycin, giving an overall rate of 4.9% transgene-free KI efficiency.

Gene targeting experiments are progressing from the proof-of-concept studies to deployment. One example shows potential for the food production sector: the shelf-life of the tomato cultivar M82 was improved by inducing alcobaca mutant genotype (*alc*) *via* an amino acid change from Val to Asp in the Non-ripening gene (NOR). The frequency of the *alc* mutation relying on CRISPR/Cas9 and the HR pathway reached 7.6% in T0 (Yu et al., 2017). By inserting the maize native promoter *Zm*GOS2 by homologous recombination in the 5' UTR of the endogenous *ZmARGOS8* (a negative regulator of ethylene response) to increase its expression levels (KI frequency, 1%), edited maize plants showed increased grain yield under drought stress conditions compared to wild type plants (Shi et al., 2017). In another example, an allele replacement was performed at the *Sl High affinity* K^+ *transporter* gene, as a Asn217Asp amino acid change variant confers salt tolerance in tomato. Designed without a visual nor biological selectable marker, this strategy produced one successful allele replacement in the transformants analysed (Vu et al., 2020b), averaging to a 0.66% success rate. Exact detail of the approach, i.e. homology arms and donor template length were not communicated. Remarkably,

a 5.2kb targeted insert has been achieved for nutritional quality improvement in rice (Dong et al., 2020). A carotenoid cassette containing two carotenoid biosynthetic genes was inserted at a predefined locus in 1 out of 16 regenerated rice plants (6% efficiency). This insertion took place via NHEJ and resulted in an inverted insertion of the carotenoid cassette alongside the flanking homology arms, which interestingly occurred on two independent occasions (same KI strategy at a different locus, no KI efficiency estimations given for this experiment). This is the first study to demonstrate the feasibility of targeted insertion (albeit through unintended NHEJ) of large DNA molecules i.e. two individual genes in plants which resulted in another Golden Rice variety with elevated carotenoid content in the grains.

Overall, the rates of KIs achieved through the employment of Cas9 without a complementary approach are relatively low. Several strategies have been elaborated to further elevate the rates of KIs and will be discussed in the following section.

1.4 Strategies for enhancing gene targeting frequency in plants and animals

1.4.1 Modulating the endogenous DNA repair machinery

1.4.1.1 Increasing HR efficiency by modulating availability of HR enzymes

Since HR occurs at low frequency in plants, a strategy to increase KI sought to use HR-specific enzymes from organisms with an efficient HR pathway such as *Saccharomyces cerevisiae.* By stably inserting the gene *ScRAD54* – which promotes strand invasion - for heterologous expression in Arabidopsis, KI frequency one to two orders of magnitudes higher than wild type plants were observed (Shaked et al., 2005; Even-Faitelson et al., 2011). In a recent GT experiment in tomato, a 35S promoter-driven expression cassette for Sl*RAD51* or Sl*RAD54* was incorporated on the T-DNA alongside editing reagents and homologous template. Overexpression of these two key players of HR did not significantly increase rates of KI. Curiously, overexpression of *SlRAD54* decreased the rates of KI compared to the control (Vu et al., 2020c). Although constitutive heterologous expression of HR-specific genes was shown to significantly increase the rates of HR in tobacco, such an approach can have damaging pleiotropic effects on plant development and morphology (Barakate et al., 2020). Simultaneous overexpression of CtIP and MRE11 combined to the downregulation of the NHEJ component XRCC4 led to a 3.4-fold improvement in GT frequency in poplar (Movahedi et al., 2021).

Alternatively, Cas9 can be fused with HR precursors, i.e. CtIP, to increase their availability at the DBS site to bias the repair pathway choice. This led to a two-fold increase or more of GT event at independent loci, in different cell lines (Charpentier et al., 2018). In a similar approach, a short Cas9 C-terminus fusion to a binding motif, Brex27, acts a recruiter for RAD51 and promotes its accumulation at the target site to form nucleoprotein filaments. A two to three-fold enhancement in KI efficiency of insert size reaching 3 kb was observed in human cells (Ma et al., 2020). Interestingly, it was reported in rice that supplementing the tissue culture regeneration medium with RAD51-stimulatory compound 1 (RS-1) consistently improved the rates of KIs, although the overall KI rates obtained in this study were low, <1% (Nishizawa-Yokoi et al., 2020).

1.4.1.2 Dampening the activity of HR competitor or suppressor pathways

Alternatively, key enzymes of the NHEJ pathway have been targeted to enable higher HR activity. For example, *AtKu70* mutant Arabidopsis attained a 16-fold GT increase, and *Atlig4* mutants showed a three to four-fold increase (Qi et al., 2013). Instead of relying on knock-out mutants, another method uses inhibiting molecules such as Scr7, which has the ability to limit the binding affinity between Ligase 4 and DNA by binding to the ligase's DNA interacting domain. This strategy increased KI efficiency 19-fold in mammalian cells (Maruyama et al., 2015). In line with this approach, some have investigated the rates of KIs in mutant Arabidopsis lines for HR repressors (RTEL1, RMI2, FANCM1) but no elevated rates of KI were observed (Wolter and Puchta, 2019b). However, there is evidence in the moss *Physcomitrella patens* that the DNA helicase RecQ4 represses homologous recombination and that *Recq4* mutants showed a two-fold increase in gene targeting efficiency (Wiedemann et al., 2018). Interestingly, this study also identified an HR enhancer, RecQ6, whose mutants had severely decreased KI efficiency (14.6% against 65.9% KI efficiency in the wild type).

1.4.2 Increasing donor template availability at the break site for repair

1.4.2.1 Liberation of the stably integrated donor template with "in planta" gene targeting

Assuring the availability of the donor template at the DSB is one of the many challenges for successful KI. The idea of "*in planta*" GT, coined by Puchta's lab in 2012, consists of an initial T-DNA-mediated stable insertion of the editing reagents (nuclease and gRNA) and donor template. The donor template is designed such that is contains two target sites at both ends (represented by the vertical black arrows in Figure 1.6) to enable its excision from the genome post nuclease recognition and cleavage. In theory, the excision of the donor template

synchronises DNA repair activation at the target site and at the template and could facilitate interaction between these two repair foci. Moreover, another benefit from this strategy is that it allows for genome editing in every daughter cell of the dividing cell population with a T-DNA insertion, increasing the chances of recovery of the desired event. However, GT efficiencies obtained with this method are low: 0.14% (Schiml et al., 2014), 0.12% (Hahn et al., 2018a), 1.47% (Wolter and Puchta, 2019b) in Arabidopsis, 2% in rice (Li et al., 2016). Recently, the *in planta* approach led to 4% KI efficiency in maize when combined to a heat inducible Cas9 expression (Barone et al., 2020). Organisation of the elements on the T-DNA enabled reconstruction of a split selectable marker gene upon Cas9-mediated excision of the donor template. This way, only cells which have an increased chance to undergo a KI due to successful donor excision event are selected for. Also performed in maize, these sets of experiments demonstrated elevated rates of KI efficiency when using the *in planta* approach (9%) compared to the standard approach (2.4%) (Peterson et al., 2021). The genomic locus targeted there had previously been assessed to carry medium-to-high HR repair potential.

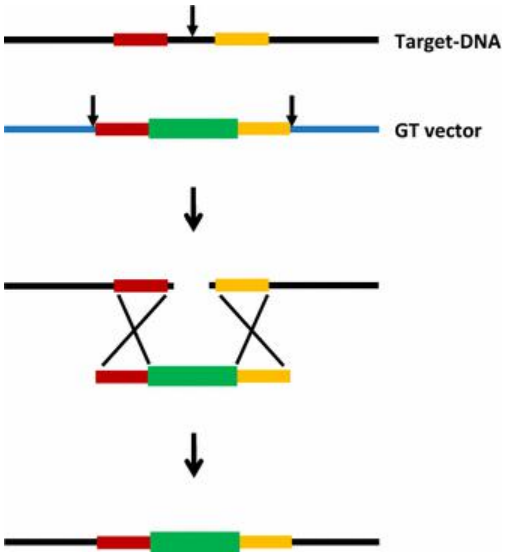


Figure 1.6 Schematics of *in planta* gene targeting. The nuclease is directed to the genomic target (Target-DNA) and to the extremities of the donor template contained on the GT vector (cleavage sites are represented by the vertical black arrows). The free dsDNA repair template can become available at the break site to be used as a template for sequence insertion. (Diagram from Steinert et al., 2016). Reprinted with permission from Springer Nature, Copyright 2016.)

1.4.2.2 Elevated abundance of donor template in the nucleus with geminivirus vectors

Increased copy number of the donor template correlates with higher rates of KI frequency (Sun et al., 2016a). Plant DNA viruses, such as geminiviruses, have been engineered to produce high copy number of the homologous DNA repair template through viral replication inside the nucleus for KI purposes. In one instance, copy number of the viral replicon peaked at 6000 copies, five days post infection (Baltes et al., 2014). In an experiment conducted in tomato, a constitutive Cauliflower Mosaic Virus 35S promoter was inserted upstream of a MYB transcription factor encoding gene, *anthocyanin 1* (*ANT1*), resulting in purple, high-anthocyanin tomato plants (Čermák et al., 2015). The KI frequency of a 2 kb donor template reached 11.6% when expressed on a viral replicon using CRISPR/Cas9, compared to 1.2% when using a standard T-DNA approach. At the tomato *carotenoid isomerase (CRTISO)* locus, KI efficiency reached 25% using the replicon strategy, which is currently the highest published KI rate in plants (Dahan-Meir et al., 2018). In this study, the *tangerine* allele of *CRTISO* (causing orange fruit colour) was replaced by the wild type allele by knocking-in the deleted 281 bp fragment which restores red fruit colour. Rates of GT in potato are reported to be increased 10 to 100-fold with this strategy (Baltes et al., 2014; Butler et al., 2016b). Confirming the potential of geminivirus as a universal genome engineering tool, KI of a 1.2 kb insert was successfully achieved in the complex genomic context of hexaploid wheat, at two different endogenous loci and at similar frequencies, 5.7% and 6.4% (Gil-Humanes et al., 2017). Impressively, in some cases the GT occurred in all three homoeoalleles (A, B and D) at the targeted ubiquitin locus (frequency of this event is not mentioned by the authors), which demonstrates that the 94-95% identity between the homology arms of the donor and homoeoalleles A and B is sufficient to enable KI by HR. Combining both the *in planta* and replicon strategies, researchers were not able to produce significant and heritable KIs in barley, however, 2 out of 14 T0 plants (i.e. 14% KI efficiency) contained a KI without using the replicon (Lawrenson et al., 2021).

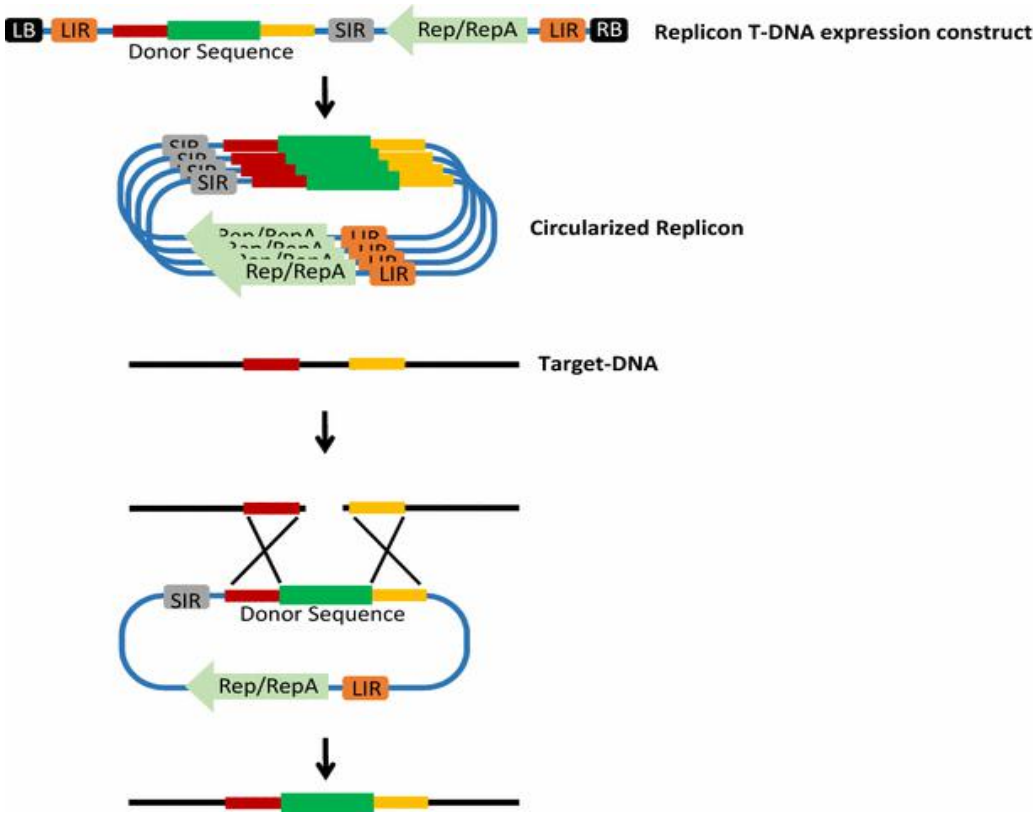


Figure 1.7 Representation of the geminivirus replicon strategy. On a T-DNA, the sequences required for viral replication are included (LIR/SIR/Rep/RepA). The sequences for the donor template or the nuclease gene are cloned between LIR and SIR to be contained on the viral replicon once circularised. (Diagram from Steinert et al., 2016) Reprinted with permission from Springer Nature, Copyright 2016.

1.4.2.3 Covalent linkage of the donor template to the Cas9 nuclease

In somatic cells, even though DSB can be repaired by ectopic homologous sequences at low frequency (Puchta et al., 1999; De Pater et al., 2018), the sister-chromatid, or sometimes the homologous chromosome, is preferentially used as the repair template during HR due to their topologic proximity to the break (Vu et al., 2014). Since the lack of availability of the donor template at the break site can limit KI success, physically bringing the repair template to the DSB could be beneficial. Hence, covalently tethering the donor template to Cas9 was shown to elevate the rates of KIs in human cells ~30-fold (Aird et al., 2018; Savic et al., 2018).

Another approach used a Cas9-monomeric streptavidin fusion delivered in combination with biotinylated ssDNA repair templates showed improved KI rates in mammalian cells (Gu et al., 2018; Ma et al., 2017). This approach has recently been adapted into rice using a fusion

between Cas9 and the *Agrobacterium* VirD2 relaxase (Ali et al., 2020a), a key component of the T-DNA transfer pathway by binding to the right border of the T-DNA (Jasper et al., 1994). This technique promoted 9% KI efficiency at the *ALS* locus, a five-fold improvement compared to the control, non-VirD2-mediated approach (Ali et al., 2020a).

1.4.3 Creating a bias in the cell's repair to favour HR recruitment

1.4.3.1 Synchronising nuclease activity with cell cycle stages.

The DNA repair machinery is under cell cycle control (reviewed by Hustedt and Durocher, 2017). NHEJ is active throughout the cell cycle, whereas HR is only active from mid-S phase until late G2 phase, during which the two pathways are competing (Kass and Jasin, 2010). To limit the formation of DSB during the G1 phase when only NHEJ is active, Cas9 expression can be synchronised with the cell cycle using protein fusion to restrict Cas9 activity to HR-permissive phases and was shown to induce a two-fold increase of KIs in human cells (Gutschner et al., 2016; Vicente et al., 2019). However, the same method did not increase KI frequency in human stem cells (Howden et al., 2016). Alternatively, it is possible to synchronise a mammalian cell population in an HR-favourable phase using the compound nocodazole: by preventing microtubule polymerisation, it arrests the cell cycle at the G2/M border after DNA replication, when HR is functional. Several human cell lines showed increased GT when synchronised at the G2/M border compered to non-synchronised cells (Yang et al., 2016). Cell synchronisation was performed in the green algae *Chlamydomonas reinhardtii* and promoted 35% KI efficiency when cells were transformed when synchronised at the optimal time point to provide high HR/NHEJ events ratio (Angstenberger et al., 2020).

1.4.3.2 Different lesions (SSB, DSB) are repaired by different pathways

As NHEJ is so efficient at repairing DSB, creating a DNA lesion that is not repaired *via* this pathway could help side-tracking NHEJ to enable HR to proceed to act on the break for KI. Studies in mammalian cells have shown that single-stranded breaks, SSB (also called nicks), are not substrates for NHEJ repair but that instead, these are repaired by a pathway related to HR (Davis and Maizels, 2014; Bothmer et al., 2017a). Moreover, a recent study in human cell lines demonstrated that delivering SSBs at the target genomic locus could promote high efficiency KIs (Chen et al., 2017). Besides, DNA lesion structure i.e. blunt, staggered, with 5' or 3' overhangs, differentially engage the repair pathways (Bothmer et al., 2017b; Vriend et al.,

2016; Wolter et al., 2018). Therefore, employing the nuclease Cas12a, which produces 5' overhangs, could improve the rates of KIs. These strategies will be discussed in more detail in Chapter 3 and Chapter 5, respectively.

1.4.3.3 Triggering somatic recombination from tandem repeat structures

It was observed that intrachromosomal somatic recombination can occur at high frequency between tandem repeats (Feng et al., 2014). Taking advantage of this, researchers have designed a short donor template which integrates at the genomic target probably via NHEJ in this instance. Upon insertion of the donor template, a tandem repeat with the flanking genomic sequence is formed. Since the Cas9 target site is recreated following donor insertion, it allows secondary cleavage of the target site and recombination to occur between the tandem repeats to replace the endogenous sequence by the one provided on the donor template. This approach led to 6% KI efficiency of a 130 bp fragment in rice (Lu et al., 2020).

1.4.3.4 Utilising RNA-based homology templates

Studies cited until now relied on dsDNA or ssDNA homologous donor template to repair the break. However, it was shown in human cells that RNA templates can also be utilised for HR-repair to produce a KI (Keskin et al., 2014). The rationale behind RNA-based homology template is that RNA:DNA hybrids are more stable than DNA:DNA duplex (Chein and Davidson, 1978). Therefore, RNA-templated KIs were demonstrated in rice but only reached 0.09% KI efficiency (Li et al., 2019). Higher rates of KI were obtained when a mixture of RNA and DNA templates were available (4%), suggesting the higher potential of DNA templates to promote KIs.

1.4.4 Using chromatin structure to improve editing efficiency

As discussed in the previous section, chromatin structure is an important regulator of DNA repair outcome, and therefore can be modulated to favour gene targeting. Chromatin accessibility influences Cas9 activity and editing efficiency (Horlbeck et al., 2016; Yarrington et al., 2018; Liu et al., 2019), hence could limit GT success. In different human cell lines, 20 gRNAs were designed to target sequences with varying degrees of accessibility based on DNase I hypersensitivity data. Mutagenesis was more efficient for gRNAs targeting open DNA than for those targeting closed DNA (i.e. genomic regions inaccessible to DNAse I and

transcriptionally inactive) (Jensen et al., 2017). In line with these findings, another study established that heterochromatin (in a CpG island context) limits the rates of mutagenesis (Kallimasioti-Pazi et al., 2018). Surprisingly, these authors did not observe a consistent repair bias towards HR or NHEJ when comparing the type of repairs at 3 euchromatic or heterochromatic loci. The discrepancy with previous reports that chromatin state affects the choice of the repair pathway (Tang et al., 2013; Aymard et al., 2014), *c.f.* section 1.2.2, is hypothetically caused by Cas9's ability to locally open chromatin structure over -100/+100bp from its binding site (Barkal et al., 2016). This Cas9-induced remodelling of the chromatin could suppress a pre-existing chromatin-induced repair bias. Others were able to elevate KI efficiency in human cell lines by delivering a chromatin donor template as opposed to naked DNA, because DNA packaged in chromatin DNA, the naturally occurring form of DNA occurring eukaryotes, is supposedly preferred by HR machinery components compared to other DNA substrates (Cruz-becerra and Kadonaga, 2020). This approach was two to seven time more efficiency at producing KIs at different loci compared to when using naked template DNA.

1.5 Project aims and objectives

Plant breeding has an important role to play in delivering biotic and abiotic stress resilient, nutritious, high-yielding crops that can also enable more sustainable agricultural practices. Using currently available breeding techniques, the creation of novel crop varieties is slow e.g. 10-20 years for different species, and lacks the ability to make precise and targeted genetic changes. It is therefore crucial that new molecular tools are developed to enable faster crop genome editing so that the output of breeding efforts is more quickly translated to elevated crop performance in the field to keep up with rapidly evolving environmental conditions. The CRISPR-Cas system is a game-changing tool for bringing novel genetic trait combinations into crops rapidly and precisely i.e. *via* KIs but the low rates of KIs obtained in plants prevents unlocking the full potential of this approach.

The ultimate goal of my project was to establish a high efficiency methodology that facilitates precise and targeted gene insertion in the crop tomato. Based on our understanding of the HR pathway and its limitations, this research project aimed to test a set of variables and assess their potential to elevate KI efficiency. In the first instance, this study attempted to recapitulate the highest rates of KIs published at the time (~10% KI efficiency), utilising a similar methodology used by the authors (Čermák et al., 2015) i.e. same genomic target, same visual marker for KIs,

use of Cas9 and viral replicons. In conjunction, variations of the visual marker, viral replicon and allele were assessed alongside testing environmental features such as temperature.

Based on the difficulty in recovering and regenerating edited callus tissues into viable plantlets, this study also focused on elaborating an improved transgenic tissue selection pipeline to preferentially foster the growth and development of cells containing a KI over non-transformed callus mass.

Finally, employing our improved transgenic selection methodology, further variables were tested for their ability to induce high frequency KI. For instance, CRISPR nucleases with different properties were tested, alongside different strategies for priming the DNA repair response. Importantly, I tested the applicability for trait improvement of the methodology developed during the first half of the project, by attempting to knock in a Resistance gene (*R* gene) at the chosen *ANT1* locus to confer late blight resistance to the tomato cultivar Moneymaker.

Chapter 2. Material and Methods

2.1 Materials

2.1.1 Microbial material

2.1.1.1 Escherichia coli

Escherichia coli strain DH10B genotype F– mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 *rec*A1 *end*A1 araD139 Δ (ara leu) 7697 galU galK rpsL nupG λ– was used for the construction of Golden Gate modules and vectors (level 0, 1, M and 2).

Escherichia coli strain NEB® beta-10, genotype Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (Str^R) rph spoT1 Δ(mrr-hsdRMS-mcrBC), was used for the transformation of cloned PCR products.

Escherichia coli strain Stbl4 (ElectroMAX™) genotype *mcr*A Δ(*mcr*BC-*hsd*RMS-*mrr*) *rec*A1 *end*A1 *gyr*A96 *gal*⁻ *thi*-1 *sup*E44 λ⁻ *rel*A1 Δ(*lac-pro*AB)/F' *pro*AB⁺ *lac*I^qZΔM15 *Tn*10 (Tet^R) was used for the transformation of Golden Gate vectors (level P) containing a geminivirus replicon to avoid plasmid instability and recombination.

2.1.1.2 Agrobacterium tumefaciens

Agrobacterium tumefaciens strain used is AGL1. After transformation, cells were grown on LB medium containing carbenicillin, rifampicin and kanamycin at 28 ˚C for 48h.

2.1.2 Plant material

Solanum lycopersicum cultivars Heinz, Micro-Tom and Moneymaker were used in this study. Seed germination and growth conditions are described in section 2.2.7. below.

Nicotiana benthamiana was used for the leaf infiltration assay to quantify Geminivirus replicon copy number by quantitative PCR. *N. benthamiana* seeds were sown on soil and grown at 24 ˚C under 16h/8h light/dark photoperiod, with 55% humidity.

2.1.3 Genetic material

2.1.3.1 Golden Gate modules available

Some of the modular parts used for cloning plant transformation vectors were already available from the plasmid database of the SynBio platform, based at the The Sainsbury Laboratory,

Norwich. The Golden Gate compatible module containing the degron was kindly provided by authors from Faden et al. (Faden et al., 2016).

Table 2.1 List of Golden Gate compatible modules that were already available or acquired for this study.

Golden Gate Modules	Species of origin	Vector ID	Source	Reference
Ubiquitin 10 promoter	*A. thaliana*	pICSL12015	Mark Youles	N/A
E9 terminator	*Pisum sativum*	pICSL60004	Laurence Tomlinson	N/A
LIR (Long intergenic repeats)	bean yellow dwarf virus (BeYDV)	pICSL11052	Mark Youles	N/A
LIR/SIR/Rep (Short intergenic repeat, Replicase)	bean yellow dwarf virus (BeYDV)	pICSL11053	Mark Youles	N/A
Neomycin phosphotransferase NptII-ecbe	bacterial, codon optimised for *Nicotiana benthamiana*	pAGM31841	IPB Halle	N/A
Dummy modules	N/A	N/A	N/A	N/A
Cas9 CDS	*Streptococcus pyogenes*, *A. thaliana* codon optimised	pICSL90016	Oleg Raitskin	Fauser et al., 2014
35S promoter + Ω leader	promoter from cauliflower mosaic virus (CaMV), leader from tobacco mosaic virus (TMV)	pICH51277	Icon Genetics	N/A
U6-26 promoter	*A. thaliana*	pICSL90002	Laurence Tomlinson	N/A
Degron	N/A, chimeric protein	pICSL11150	Nico Dissmeyer	Faden et al., 2016
ttLbCas12a CDS	*Lachnospiracea bacterium,* temperature tolerant variant allele, Arabidopsis codon-ptimised	pICSL90022	Holger Puchta	Schindele and Puchta, 2019
***Rpi-vnt1* gene**	*Solanum venturii*	LBJJ73	Laurence Tomlinson	N/A
***Cyclin B1* destruction box as a C-terminus tag**	*Solanum lycopersicum*	Synth2_SlCYCB1;1_C-ter	Synthesised by TWIST Bioscience	N/A
Cas12a gRNA_1 (*ANT1* target)	N/A	Cas12a_gRNA-ant1	Synthesised by TWIST Bioscience	N/A

Table 2.2 List of binary vectors used in this study.

Vector ID	Nuclease	gRNA target p*ANT1*	Insert	Insertion pathway	Backbone
LGJJ52	Cas9	-60bp	NptII + 35S promoter-Ω	HR	pICSL4723-P1
LGJJ53	$Cas9^{D10A}$	-60bp	NptII + 35S promoter-Ω	HR	pICSL4723-P1
LGJJ54	Cas9	-60bp	NptII + 35S promoter-Ω	HR	pICSL4723-P1
LGJJ55	$Cas9^{D10A}$	-60bp	NptII + 35S promoter-Ω	HR	pICSL4723-P1
LGJJ58	dCas9	-60bp	NptII + 35S promoter-Ω	HR	pICSL4723-P1
LGJJ96	Cas9	-723bp	NptII + 35S enhancer	HR	pICSL4723-P1
LGJJ97	Cas9	-800bp	NptII + 35S enhancer	HR	pICSL4723-P1
LGJJ112	Cas9	-723bp	NptII + 35S enhancer	HR	pICSL4723-P1
LGJJ113	Cas9	-800bp	NptII + 35S enhancer	HR	pICSL4723-P1
LGJJ139	N/A	N/A	NptII + 35S enhancer::*ANT1*::terminator	T-DNA	pICSL32281
LGJJ151	N/A	N/A	Degron-NptII + 35S promoter::*ANT1*::terminator	T-DNA	pICSL32281
LGJJ156	N/A	N/A	Degron-NptII + 35S enhancer::*ANT1*::terminator	T-DNA	pICSL32281
LGJJ180	Cas9	-60bp	Degron-NptII + 35S promoter-Ω	HR	pICSL4723-P1
LGJJ181	Cas9	-60bp	Degron-NptII + 35S promoter-Ω	HR	pICSL4723-P1
LGJJ182	Cas9-Cycb1	-60bp	Degron-NptII + 35S promoter-Ω	HR	pICSL4723-P1
LGJJ183	Cas9-Cycb1	-60bp	Degron-NptII + 35S promoter-Ω	HR	pICSL4723-P1
LGJJ191	ttCas12a	-60bp	Degron-NptII + 35S promoter-Ω	HR	pICSL4723-P1
LGJJ192	ttCas12a	-60bp	Degron-NptII + 35S promoter-Ω	HR	pICSL4723-P1
LGJJ216	Cas9	-60bp	Degron-NptII + Rpi-Vnt1 + 35S promoter-Ω	HR	pICSL4723-P1
pTC217	Cas9	-60bp	NptII + 35S promoter	HR	pLSLR
pTC147	N/A	N/A	NptII + 35S promoter	T-DNA	pLSLR

2.1.4 Media and buffers

2.1.4.1 Microbiology

Lysogeny broth (LB) medium contains 10 g tryptone, 5 g of yeast extract, 5 g NaCl, 1 g of glucose, pH 7.0. Ten grams of agar was added to make a solid medium.

Lennox (L) medium contains 10 g tryptone, 5 g of yeast extract, 5 g NaCl, pH 7.0. Ten grams of agar was added to make a solid medium.

2.1.4.2 Plant tissue culture

Germination medium (pH 5.8)

For 1 litre

MS medium	1×
Glucose	10 g
Agarose	6 g

Co-cultivation medium (pH 5.7)

For 1 litre

MS medium	1×
Sucrose	30 g
2,4-D	0.5 mg
Agarose	6 g

Regeneration medium (pH 6.0)

For 1 litre

MS salts	1×
Myo-inositol	100 mg
Nitsch's vitamins*	1 ml of 1000× stock
Sucrose	20 g
Agargel	4 g
Zeatin Riboside	2 mg
Timentin	320 mg/l
Kanamycin	100 mg/l

*Nitsch's vitamins, working concentration (mg/l)

Thiamine	0.5
Glycine	2.0
Nicotinic acid	5.0
Pyridoxine HCl	0.5
Folic acid	0.5
Biotin	0.05

Rooting medium (pH 6.0)

For 1 litre

MS medium	0.5×
Sucrose	5 g
Gelrite	2.25 g
Timentin	320 mg/l
*Kanamycin	50 mg/l,

* if using a selective rooting medium.

2.1.4.3 Buffers

Tris Acetic EDTA (TAE) buffer (10×)

For 1 litre

Tris	48.4 g
EDTA	3.72 g

pH 8.0 with acetic acid

2.1.5 Antibiotics

Table 2.3 List of antibiotics used in this study for bacterial cultures.

Antibiotic	Stock concentration	Working concentration
Carbenicillin	100 mg/ml in H_2O	100 µg/ml
Spectinomycin	100 mg/ml in H_2O	100 µg/ml
Kanamycin	150 mg/ml in H_2O	150µg/ml
Rifampicin	10 mg/ml in methanol	10µg/ml

2.2 Methods

2.2.1 2.2.1 Polymerase chain reaction (PCR) and gel imaging

2.2.1.1 Colony PCR

Colony PCR to select transformant *E. coli* colonies was performed using standard *Taq* polymerase (NEB, catalogue #M0273S). PCR reactions were assembled using the manufacturer's protocol. PCR cycling conditions were 95°C for 30 sec, (95°C for 15 sec, 'T_a' for 20 sec, 68°C for 'T_e') × 25, 68°C for 5 min. Annealing temperature (T_a) was adapted for each PCR based on the melting temperature (Tm) of the primers used (determined using the online NEB Tm calculator) and the extension time (T_e) based on the length of the PCR product (1 min/Kb).

2.2.1.2 PCR to generate Golden Gate compatible modules

Amplification of DNA to generate Golden Gate compatible modules was performed using Q5 Hi-Fidelity DNA polymerase (NEB, catalogue #M0491S). PCR reactions were assembled using the manufacturer's protocol. PCR cycling conditions were 98°C for 30 sec, (98°C for 10 sec, 'T_a' for 20 sec, 72°C for 'T_e') × 35, 72°C for 2 min. Annealing temperature (T_a) was adapted for each PCR based on the melting temperature (Tm) of the primers used (determined using the online NEB Tm calculator) and the extension time (T_e) based on the length of the PCR product (30 sec/Kb). Primers used to create these new modules are details in table 2.5 in section 2.2.3.1. After gel imaging, the band containing the sequence of interest was excised from the agarose gel and purified on silica columns according to the manufacturer's protocol (QIAGEN QIAquick Gel Extraction Kit, catalogue #28704).

2.2.1.3 Transgenic material genotyping by PCR

Screening of transgenic tomato shoots was performed using a GoTaq® Hot Start *Taq* polymerase (Promega, catalogue #M5122). PCR reactions were assembled using the manufacturer's protocol. PCR cycling conditions were 95°C for 2 min, (95°C for 30 sec, 'T_a' for 30 sec, 72°C for 'T_e') × 30, 72°C for 5 min. Annealing temperature (T_a) was adapted for each PCR based on the melting temperature (Tm) of the primers used (determined using the online Promega Tm calculator) and the extension time (T_e) based on the length of the PCR product (1 min/Kb).

Knock-in specific PCRs to amplify the left and right junctions of the insert from purple tomato tissue genomic DNA were performed using KAPA HiFi HotStart ReadyMix PCR kit (KAPA Biosystems, catalogue #KK2601). Reactions were set up according to the manufacturer's protocol. PCR cycling conditions were 95°C for 3 min, (98°C for 20 sec, 'T_a' for 15 sec, 72°C for 'T_e') × 30, 72°C for 'T_e'. Annealing temperature (T_a) was adapted for each PCR based on the melting temperature (Tm) of the primers used and the extension time (T_e) based on the length of the PCR product (1 min/Kb).

2.2.1.4 Gel imaging

PCR products were migrated on agarose gel (1.5% agarose gel made with 1× TAE buffer and 0.01 µl/ml of ethidium bromide) by electrophoresis at 150-200 V before being visualised under UV light gel imager (G:box F3-LBF, Syngene).

2.2.2 Quantitative PCR (qPCR)

Following the *Nicotiana benthamiana* leaf infiltration assay (described in section 2.2.6), relative quantification of the Geminivirus replicon copy number *in planta* was measured by qPCR. I used the KAPA SYBR® FAST qPCR (KAPA BIOSYSTEMS, catalogue #KK4600) master mix for the reactions according to the manufacturer's protocol, in triplicates (biological replicates). The qPCR reactions were performed in the BIO-RAD CFX96 Real Time System and C1000 thermal cycler machine. The cycling conditions were 95˚C for 3 min, (95˚C for 10 sec, 58˚C for 10 sec, 72˚C for 30 sec) × 40, 65˚C for 0.5 sec. Copy number weas calculated in relation to the *N. benthamiana Actin* gene copy number.

Table 2.4 List of primers used to perform the qPCR to assay the Geminivirus replicon copy number.

Copy number from vector:	Forward primer	Reverse primer
LGJJ52	LG130 GATAGTGGGTAGTGCCATCTTGA	LG131 TCTCACCGACAAGTGTTGTGG
LGJJ112, LGJJ113		LG133 TACCAATGGACAAGATAGGCTG
ZCM-CTD		LG142 GTCGTGCTCCACCATGTTGA
***Nb Actin* control**	LT899 GCTCCTCGAGCAGTGTTTCC?	LT900 CCTTCTGTCCCATTCCGACCA

2.2.2.1 NptII gene copy number assay by qPCR

Copy number assay of the selectable marker expression cassette was performed by the company IDNA genetics (Norwich, UK). Sample gDNA extraction and preparation was carried out as per their proprietary method. Adapted from previously published protocol, transgene copy number was measured by qPCR in individual plants (Bartlett et al., 2008).

In a multiplex reaction, an amplicon from the Nos terminator of the *NptII* transgene expression cassette (with a FAM reporter-TARMA quencher) and an amplicon from a *Solanum lycopersicum* internal positive control (with a VIC reporter-TARMA quencher) were amplified. The TaqMan primers and probes were used at 100 nM. The cycling conditions were: initial denaturation for 15 minutes at 95˚C, then (95˚C for 15 seconds and 60˚C for 60 seconds) × 40 cycles in an QuantStudio 5 real time PCR machine. They ran two replicate assays per sample. Fluorescence from the FAM and VIC fluorochromes was measured during each 60˚C step and the Ct values obtained. The Delta Ct value was obtained from the

difference between the Ct values for the Nos terminator and the internal positive control. Delta Ct values were used to allocate the samples into groups with the same copy number of the *NptII* cassette.

2.2.3 Golden Gate cloning

Golden Gate cloning enables fast and unidirectional modular assembly of DNA fragments in one reaction to build vectors containing multiple transcription units (Werner et al., 2012; Engler et al., 2014). The Golden Gate method employs the type IIS restriction enzymes *Bsa*I and *Bpi*I which cut outside of their recognition site and generates 4 bp overhangs that are unique for any given position at level 0, 1, 2, M and P vectors.

Digestion-ligation reactions (DigLigs) were performed according to the TSL SynBio platform protocol. A 20 µl total volume reaction contained: 1 µl of the appropriate restriction enzyme (*Bpi*I or *Bsa*I, Thermo Fisher, catalogue #ER0291 and #ER1011) and 2 µl of their corresponding buffer (Thermo Fisher, buffer G), 1 µl of T4 DNA ligase (NEB, catalogue #M0202) and 2 µl T4 DNA ligase buffer (NEB), a volume of acceptor vector corresponding to 20 femtomolar and a volume of each insert module corresponding to 40 femtomolar to obtain a 2:1 ratio of insert to acceptor vector. Each reaction was topped with dH_2O to the final volume. The diglig reaction was carried out in a thermocycler with the following conditions: (3 min at 37°C, 4 min at 16°C) × 26, followed by enzyme denaturation steps, 5 min at 50°C and 5 min at 80°C. The appropriate bacterial strain was transformed with 1 µl of the reaction following the protocol described in section 2.2.4.

2.2.3.1 Creation of new Golden Gate compatible modules

Modules were created by PCR amplification of the DNA sequence of interest prior to cloning in the corresponding Golden Gate destination vector. The primers for the amplification of the sequence of interest were designed to contain a *Bsa*I or *Bpi*I restriction sites and the complementary 4 nucleotides overhang to their destination vector at their 5′ . List of the Golden Gate modules created for this research project is available in table 2.5.

Table 2.5 List of Golden Gate modules that were generated by PCR for this study. The bolded 6 nucleotides correspond to the enzyme restriction site and the underlined 4 nucleotides to the overhangs created after cleavage. MM= cultivar Moneymaker.

Module ID	Module content	Primer ID	Primer sequence (5′ to 3′)	Template	Destination Vector
LGJJ6	Dead Cas9	LT749		LBJJ358-1	pICSL01005
		LT750	ag**GAAGAC**AAAAGCTCAAACCTTCCTCTTCTTCTTAG		
		LT751	ttGAAGACgaTTACGATGTTGATgcTATCGTGC		
		LG8	at**GAAGAC**taCGAAAACCTTCCTCTTCTTCTTAGGATCAGCCCTTGA		
LGJJ13	AtCasD10A	LT749	ga**GAAGAC**aaAATGGATAAGAAGTACTCTATCGGACTCGcTATCG	LBJJ358-1	pICSL01005
		LT450	ag**GAAGAC**AAAAGCTCAAACCTTCCTCTTCTTCTTAG		
LGJJ4	*ANT1* promoter homology arm	LG1	ag**GGTCTC**Aggag**AACCACACAACACTTGTCGGTGAGA**	*S. lycopersicum* var. MM gDNA	pICH47781
		LG2	ag**GGTCTC**Aagcg**CCAACTTGTAGAGAGCCTTCTACA**		
LGJJ5	*ANT1* promoter homology arm with gRNA target site	LG3	ag**GGTCTC**AGGAG**TACAATTTAATACACCTTTTAGGAACCACAAC**ACTTGTCGGTGAGATT	*S. lycopersicum* var. MM gDNA	pICH47781
		LG2	ag**GGTCTC**Aagcg**CCAACTTGTAGAGAGCCTTCTACA**		
LGJJ10	*ANT1* CDS homology arm	LG16	ag**GGTCTC**aAGTA**TAATATATATTATCAAATTATTATGAACAGTACA**TCTATGtccTCATTG	*S. lycopersicum* var. MM gDNA	pICH47732
		LG5	ag**GGTCTC**aAGCG**TGCAATGTTCCTCCTCGTCCAAAATTACATTA**TTG		
	35S promoter-omega leader	LG14	at**GGTCTC**aGGAG**GTCAACATGGTGGAGCACGA**	pICH5127 7	
		LG15	at**GGTCTC**Atact**GTATCGATAATTGTAAATGTAATTGT**		
LGJJ11	*ANT1* CDS homology arm with gRNA target site	LG16	ag**GGTCTC**aAGTA**TAATATATATTATCAAATTATTATGAACAGTACA**TCTATGtccTCATTG	*S. lycopersicum* var. MM gDNA	pICH47732
		LG6	ag**GGTCTC**aAGCG**CCTAAAAGGTGTATTAAATTGTATGCAATGTT**CCTCCTCGTCCAAAATTACAT		
	35S promoter-omega leader	LG14	at**GGTCTC**aGGAG**GTCAACATGGTGGAGCACGA**	pICH5127 7	
		LG15	at**GGTCTC**Atact**GTATCGATAATTGTAAATGTAATTGT**		

Table 2.5 continued

Table 2.5 continued

Module ID	Module content	Primer ID	Primer sequence (5′ to 3′)	Template	Destination Vector
LGJJ102	mild LIR	LG161	aa**GAAGAC**ttGGAGGTTGTTGTGACTCCGAGGGGTTG	pTC217	pICH47761
		LG162	aa**GAAGAC**ttAGCGGAGGGTCGTACGAATAATTCGTATC		
LGJJ105	mild SIR/LIR/Rep	LG163	aa**GAAGAC**ttggagGAGTGTACTTCAAGTCAGTGGG	pTC217	pICH47772
		LG162	aa**GAAGAC**ttAGCGGAGGGTCGTACGAATAATTCGTATC		
LGJJ21	Lox P left	LG33	GAGCATAACTTCGTATAATGTATGCTATACGAAGTTAT	oligo annealing	pICH47791
		LG34	GGCAATAACTTCGTATAGCATACATTATACGAAGTTAT		
LGJJ22	Lox P right	LG35	GCAAATAACTTCGTATAATGTATGCTATACGAAGTTAT	oligo annealing	pICH47742
		LG36	TAGTATAACTTCGTATAGCATACATTATACGAAGTTAT		
LGJJ28	Piggybac left	LG59	GC**GGTCTC**AACATGAGTTAACCCTAGAAAGATAATCATATTG	EC68616	pAGM1311
		LG60	GC**GGTCTC**AACAAGGCACTCAGTAACTTAGCCTATTGAATAA		
LGJJ29	Piggybac right	LG61	GC**GGTCTC**AACATACTATATTCAATAGGCTAAGTTACTGAGA	EC68616	pAGM1311
		LG62	GC**GGTCTC**AACAAGTAATTAACCCTAGAAAGATAGTCTGCGTAA		
LGJJ44	35S double enhancer	LG84	GC**GGTCTC**AACATGCAAtgtgAtctcaggaggtcaac	pICH45089	pAGM1311
		LG58	GC**GGTCTC**AACAATAGTAGGAAGGGTCTTGCGAAGGATAG		
LGJJ62	U6-26p:gRNA_-723bp:ter	LG92	ga**GGTCTC**aATTGTAATACAAAGTGAACCCACAGTTTAAGAGCTATGCTGGAA	pICSL70015	pICH47751
		BC_A3	tgt**ggtctc**aAGCGACCCCAGAAATTGAAC		
LGJJ63	U6-26p:gRNA_-800bp:ter	LG91	ga**GGTCTC**aATTGTGATATAGAGTGAATACATAGTTTAAGAGCTATGCTGGAA	pICSL70015	pICH47751
		BC_A3	tgt**ggtctc**aAGCGACCCCAGAAATTGAAC		

Module ID	Module content	Primer ID	Primer sequence (5′ to 3′)	Template	Destination Vector
LGJJ184	*ANT1* promoter homology arm with Cas12a T.S.	LG246	ag**GGTCTC**AGGAGTTTAATACACCTTTTAGGCACGTGTATAACCACAACACTTGTCGGTGAGATT	LGJJ4	pICH47781
		LG2	ag**GGTCTC**AagcgCCAACTTGTAGAGAGCCTTCTACA		
	35S promoter-Omega leader	LG14	at**GGTCTC**aGGAGGTCAACATGGTGGAGCACGA	LGJJ10	
		LG15	at**GGTCTC**AtactGTATCGATAATTGTAAATGTAATTGT		
LGJJ185	*ANT1* CDS homology arm with Cas12a T.S.	LG16	ag**GGTCTC**AAGTATAATATATTATCAAATTATTATGAACAGTACATCTATGtccTCATTG	LGJJ10	pICH47751
		LG247	Ag**GGTCTC**aAGCGATACACGTGCCTAAAAGGTGTATTAAATGCAATGTTCCTCCTCGTCCAAAATTACAT		
	35S promoter-Omega leader	LG14	at**GGTCTC**aGGAGGTCAACATGGTGGAGCACGA		
		LG15	at**GGTCTC**AtactGTATCGATAATTGTAAATGTAATTGT		
LGJJ68	5' *ANT1* promoter homology arm (for 35S enhancer KI) for	LG105	gc**GGTCTC**aGGAGTGATATAGAGTGAATACATAAGGCAGCCTATCTTGTCCATTGGTAT	*S. lycopersicum* var. Heinz gDNA	pICH47772
		LG106	gc**GGTCTC**aAGCGCTCTATATCAACTGACTATAAATCAATTCG		
LGJJ70	5' *ANT1* promoter homology arm (for 35S enhancer KI) for	LG109	gc**GGTCTC**aGGAGTAATACAAAGTGAACCCACAAGGCAGCCTATCTTGTCCATTGGTAT	*S. lycopersicum* var. Heinz gDNA	pICH47772
		LGJJ110	gcG**GGTCTC**aAGCGGTTCACTTTGTATTAAGTGTATGAACG		
LGJJ83	3' *ANT1* promoter homology (for gRNA_-800bp)	LG107	gc**GGTCTC**aGGAGAACATATACAGTTGATACAATTGTATAATTCG	*S. lycopersicum* var. Heinz gDNA	pICH47761
		LG116	gc**GGTCTC**aAGCGATAATTTGATAATATATTATACTAC		
LGJJ85	3' *ANT1* promoter homology (for gRNA_723bp)	LG111	gc**GGTCTC**aGGAGAACATATACACTTAATATAATTGTATTCCTTGA	*S. lycopersicum* var. Heinz gDNA	pICH47761
		LG122	gcG**GGTCTC**aAGCGCCTTGTGGGTTCACTTTGTATTAATAATTTGATAATATATTATACTACTTTTTTATTGTCTATATATACT		

2.2.4 Bacterial transformation

Electro-competent bacterial cells were transformed by electroporation with 1 µl the corresponding plasmid using 1 mm electroporation cuvettes in a MicroPulser (Bio-Rad). Pre-set settings by the manufacturer were used for electroporation (1.8 kV for *E. coli*, and 2.2 kV for *Agrobacterium*). Cells were resuspended in 300 µl of SOC medium and allowed to recover in a 37˚C (*E. coli*) shaker for 1h or in a 28˚C shaker for 2h (*A. tumefaciens, E. coli* Stbl4*)*. 50 µl of the cells were spread on LB plates supplied with the appropriate antibiotic. Additionally, 40 µl of X-Gal (25 mg/ml) and 10 µl of IPTG (1 M) were pipetted onto the plates to allow blue/white selection of the clones. Plates were placed at 37 ˚C overnight for *E. coli*, or 28 ˚C for 48h hours for *A. tumefaciens* or *E. coli* Stbl4.

2.2.5 Plasmid isolation and verification

White colonies confirmed by colony PCR were inoculated to 3 ml of liquid LB with the corresponding antibiotic and grown over night in a 37 ˚C shaker, 220 rpm.

Plasmids were isolated the following day using QIAprep Spin Miniprep Kit (Qiagen, catalogue #27104) according to the provided protocol. Eluted plasmid DNA was verified by restriction enzyme digestion analysis before being sent for Sanger sequencing (GENEWIZ).

2.2.6 Nicotiana benthamiana leaf infiltration assay

Agrobacterium AGL1 glycerol stocks containing the appropriate vectors were streaked onto LB plates containing the correct antibiotic and were incubated at 28 ˚C for 48h. Colonies were scooped with a sterile pipette tip and resuspended in 2 ml of 10 mM MES, 10 mM $MgCl_2$ (pH 5.6) and 150 µM acetosyringone and incubated at room temperature for three hours. O.D. was measured and diluted with the MES solution to an OD_{600} of 0.6. The final infiltration culture contained 1:1 ratio of the tested construct and of p19 (RNAi suppressor), OD_{600} 0.6. The cultures were injected on the abaxial leaf surface of six-week-old *Nicotiana benthamiana* using a blunt syringe. Leaf tissues were harvest two days post infiltration and at five days post infiltration and frozen in liquid nitrogen.

2.2.7 Tomato transformation and regeneration

Tomato transformation and tissue culture was carried out by our specialised plant transformation support team at the Sainsbury Laboratory. Tomato seeds were surfaced sterilised with a 70% EtOH treatment for 2 min, rinsed with sterile water, then soaked in a 10% bleach (sodium hypochlorite) solution for three hours, shaking. Sterilised seeds were sown

on germination medium and vernalised for three weeks at 4 ˚C. Germination was induced by moving the seeds at 22 ˚C.

Cotyledons from seven-day-old seedlings received two transversal cuts with a scalpel to produce explants. Explants were placed on co-cultivation plates with a feeder layer (2 ml of tobacco suspension culture plated onto co-cultivation plates) and immersed in *Agrobacterium* suspension (*Agrobacterium* culture resuspended in MS medium with 3% sucrose to an OD_{600} of 0.6) and were co-cultivated for 48h at room temperature in low light. Explants were then moved to regeneration plates containing kanamycin and placed for three days in a Sanyo cabinet set at 28 ˚C, 16h photoperiod. After the heat treatment, the plates were placed in a growth room set at 22 ˚C, 16h photoperiod, for the rest of the regeneration procedure. Explants were sub-cultured to fresh regeneration plates every two to three weeks. Shoots were cut off from explants and placed in rooting medium eight to ten weeks after transformation.

2.2.7.1 Regeneration of explants transformed with a temperature-sensitive selectable marker (Degron-NptII)

To reduce the amount of non-transformed shoots, the *NptII* gene was replaced with a *Degron-NptII* version. The degron used here is an N terminal tag of the NPTII making the protein unstable at temperatures above 17 ˚C. The germination and co-cultivation procedures were performed in the same way as described above. After the co-cultivation step, explants were placed on non-selective regeneration plates and put in a Sanyo cabinet set at 28 ˚C, 16h/8h day/light for three days. After the heat-mediated NPTII degradation step, explants were moved to kanamycin containing regeneration plates and placed in a Sanyo cabinet set at 17 ˚C, 16h/8h day/light, for two weeks. To increase the growth of the explants, plates were moved in a growth room set at 22 ˚C, 16h/8h day/light for the rest of the regeneration procedure and sub-cultured to fresh kanamycin containing regeneration medium every two weeks. Shoots were cut off from explants and placed in rooting medium between four and half and six months after transformation, when shoots were at least 4 cm tall.

2.2.7.2 Evaluation of shoot growth in the NptII/degron-NptII comparison experiment

Shoot growth was analysed three months after transformation. The first assessment involved scoring every shoot, or shoot cluster, regenerating from each explant and recording the number of green and purple shoots to assess the growth ratio between the two. A diverse phenotypic range of shoots and leaves regenerating from callus was observed, so a selection on which phenotypes to include in the scoring was applied. To best reflect the number of developing

shoots that have a viable potential to become plantlets, leaves and shoots with odd structures were not included i.e. Figure 2.1, lumpy or tubular leaf structures were excluded as assuming that development in these was compromised. Shoots with leaf-resembling shapes and structures were included, regardless of the size, as shown in the right panel of Figure 2.1.

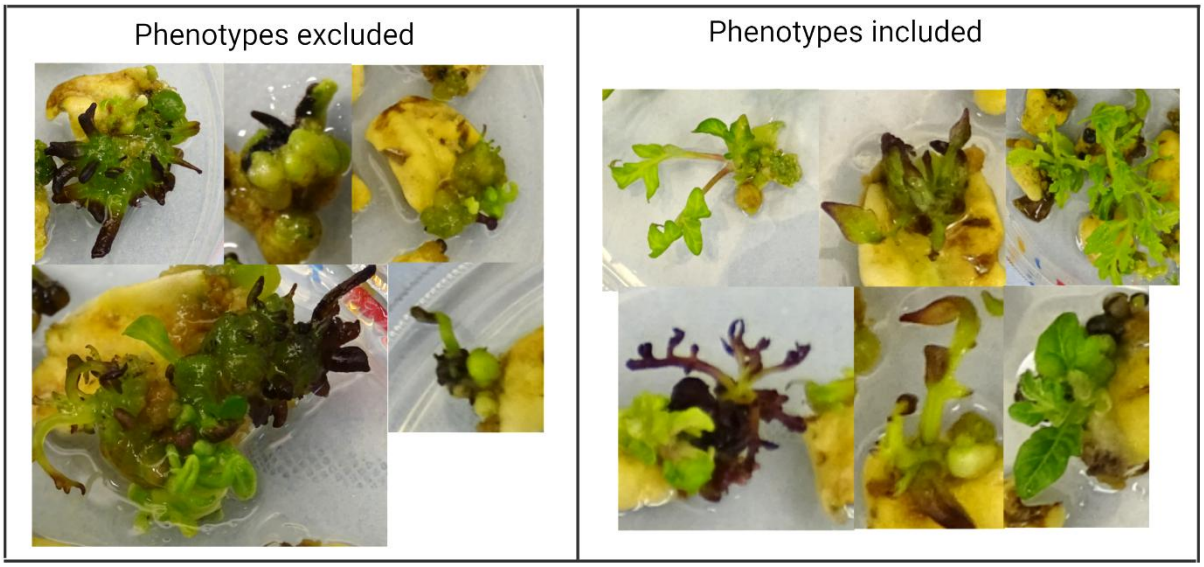


Figure 2.1 Leaf and shoot phenotypes observed when assessing the ratio of green to purple shoots. On the left panel are examples of phenotypes excluded from the assessment due to odd structures, suggesting impaired development. On the right panel are examples of phenotypes that were included in the assessment.

At four months post transformation, the shoot production and regeneration of the explants was assessed. Each explant was allocated to a category based on the size of its tallest regenerating shoot. The categories were: tallest growth between 1) 0-1 cm, 2) 1-3 cm 3) 4 cm and above (Figure 2.2).

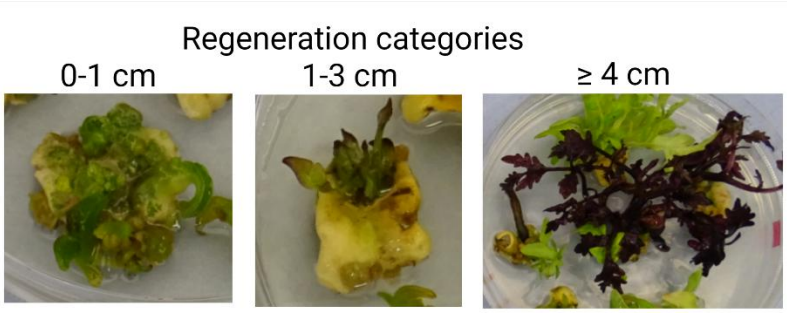


Figure 2.2 Representative example of explants in each regeneration category.

2.2.8 Phenotyping of calli, regeneration of purple sectors and calculation of knock-in rates

Regenerating calli were screened between four weeks (for explants transformed with a plasmid containing a standard *NptII*) to six weeks (for explants transformed with a vector containing a *degron-NptII*) after transformation for the appearance of purple sectors indicating a successful insertion of a 35S promoter upstream of the *ANT1* gene in these cells. Purple sectors were observed under the microscope and photographed. The number of explants showing at least one purple sector was counted and recorded. The rate of KIs was calculated by dividing the number of explants with at least one purple sector by the total number of explants transformed. A KI efficiency was obtained as a percentage by multiplying the frequency of KI events by 100.

The purple sectors were subsequently isolated from the explant alongside a portion of callus to sustain the development of purple tissues. The excised sectors were placed on non-selective regeneration plates and sub-cultured to fresh medium every two weeks until viable purple shoots developed. These were then excised and placed in non-selective rooting medium.

2.2.9 Isolation of plant genomic DNA

Genomic DNA was extracted from leaf tissue by freezing up to 100 mg (fresh weight) contained in a 1.5 ml Eppendorf tube in liquid nitrogen and grinded using a tissue grinding machine (Geno/Grinder®). The following steps of the genomic DNA extraction were performed using the DNeasy Plant Mini Kit (Qiagen, catalogue #69104) according to the manufacturer's protocol.

2.2.10 Molecular characterisation of KI events

PCRs were set up as described in section 2.2.1.3. After verification by gel electrophoresis, the PCR products were used without purification directly for cloning using the Neb® PCR cloning kit (catalogue #E1202S), as per manufacturer's protocol. Ligated PCR products into the NEB cloning vector pTmini 2.0 were subsequently transformed into NEB® beta-10 *E. coli* competent cells. Transformants were screened by colony PCR and between one to four positive colonies were selected per cloned PCR product. Purified plasmid from these positive clones were sent for Sanger sequencing (GENEWIZ sequencing services). Sequencing results were analysed using the software CLC to align the sequencing read to the reference sequence containing the expected KI outcome.

Table 2.6 List of primers and PCR conditions for molecular characterisation of KI events.

Amplicon content	Primer ID and sequence (5′ to 3′)	Amplicon size	Ta	Te
For LGJJ181 samples				
Right Junction	LG141 ATGACGCACAATCCCACTATC	1.3 kb	65°C	1.40 min
	LG185 GTGGATGGTAACCCATTCTAAC			
Left Junction	LG280 ACCATTCGATCGTACGTGCT	1.9 kb	60°C	2.0 min
	LG23 TATGGAACGTCAGTGGAGCA			
For LGJJ216 samples				
Right Junction (PCR B)	LG297 ATGAATTATTGTGTTTACAAGACTTG	5.3 kb	62°C	6.0 min
	LG185 GTGGATGGTAACCCATTCTAAC			
Right junction (PCR C)	LG290 GACACGCTCGAGTATAAGAG	1.2 kb	58°C	2.0 min
	LG140 CAACTTGTAGTTGGGACAC			
Left Junction	LG203 CATTCAATTGCGATGATCTACG	4.2 kb	64°C	4.0 min
	LG283 GGCTGTGAGTTTGGGCTATT			
Across insert	LG280 ACCATTCGATCGTACGTGCT	10.0 kb	60°C	10.0 min
	LG185 GTGGATGGTAACCCATTCTAAC			
For LGJJ191 samples				
Right junction	LG290 GACACGCTCGAGTATAAGAG	1.2 kb	58°C	2.0 min
	LG140 CAACTTGTAGTTGGGACAC			
Left junction	LG203 CATTCAATTGCGATGATCTACG	1.4 kb	60°C	2.0 min
	LG23 TATGGAACGTCAGTGGAGCA			
Across insert	LG203 CATTCAATTGCGATGATCTACG	5.1 kb	62°C	6.0 min
	LG140 CAACTTGTAGTTGGGACAC			
For LGJJ52 sample				
Right junction	LG141 ATGACGCACAATCCCACTATC	1.3 kb	58°C	2.0 min
	LG140 CAACTTGTAGTTGGGACAC			
Left junction	LG203 CATTCAATTGCGATGATCTACG	1.3 kb	60°C	2.0 min
	LG23 TATGGAACGTCAGTGGAGCA			

2.2.11 Assembly of golden gate gRNA expression cassettes

Level 1 Golden gate modules containing Cas9-compatible gRNAs were built by PCR (primer sequences in table 2.5). The forward oligo contained the *BsaI* site and appropriate overhangs, the chosen spacer sequence following by the sequence binding to the template. The template vector contains the EF scaffold sequence of the guide and the 67bp of U6-26 terminator sequence. The PCR product was gel purified and subsequently cloned alongside a U6-26 promoter into its destination level 1 vector.

The Cas12a gRNA expression cassette under the control of a 26-U6 promoter was synthesised by TWIST bioscience and cloned into a level 1 acceptor vector. The gRNA scaffold (cRNA direct repeat) and the guide sequence were flanked by self-cleaving ribozyme sequences (hammerhead ribozyme on the left and HDV ribozyme on the right), followed by a U6-26

terminator. Based on published data (Schindele and Puchta, 2019; Tang et al., 2017), the sequences used to build the expression cassette were as follows:

aaattactgatgagtccgtgaggacgaaacgagtaagctcgtcTAATTTCTACTAAGTGTAGATnnnnn
nnnnnnnnnnnnnnnnnnGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTC
GGCATGGCGAATGGGACTTTTT

Key:

Hammerhead ribozyme

Lb-crRNA direct repeat

23nt variable target sequence

HDV ribozyme

RNA PolIII terminator

The list of gRNAs used with Cas9 and Cas12a are listed in table 2.7.

Table 2.7 List of gRNAs used in this study targeting the ANT1 promoter at proximal or distal regions adapted for Cas9 or Cas12a use.

Cas	Name	Sequence (5′ to 3′)	PAM	Source
Cas9	gRNA_-800bp	TGATATAGAGTGAATACATA	AGG	Lila Grandgeorge
	gRNA_-723bp	TAATACAAAGTGAACCCACA	AGG	Lila Grandgeorge
	gRNA_-60bp	ACAATTTAATACACCTTTT	AGG	Cermak et al. 2015
Cas12a	gRNA_12a_1	ATACACCTTTTAGGCACGTGTAT	TTTA	Van Vu et al. 2018

2.2.12 Testing gRNA mutagenesis efficiency

The gRNAs used to deliver a DSB in the distal region of the *ANT1* promoter (gRNA_-800bp and gRNA_-723bp) were tested for mutagenesis efficiency and compared to that of a published gRNA (gRNA_-60bp, Cermak et al., 2015). Tomato var. Moneymaker explants were transformed with the binary vectors containing gRNA_-800bp (vector ID LGJJ112), gRNA_-723bp (vector ID LGJJ113) or gRNA_-60bp (vector ID LGJJ52). After the three-day heat treatment at 28 °C, material was taken through to rooting stage. Each shoot which fully rooted in kanamycin containing medium was selected for downstream analysis of the Cas9 target site and gDNA was extracted. Each sample was tested for the presence or absence of Cas9 by PCR. A second PCR was performed to amplify across the Cas9 genomic target sites using the primers listed in the Table 2.8. The high fidelity Q5 *Taq* polymerase (NEB,

catalogue #M0491S) was used to limit introduction of mutations during amplification. Bands of the expected size were gel purified and sent for Sanger sequencing (GENEWIZ). The softwares TIDE (Tracking of indels by decomposition, Brinkman et al., 2014) and Synthgo ICE (Inference of CRISPR Edits, Hsiau et al., 2018) were used to analyse the number of mutated samples at the target site. Based on these results, the efficiency of the gRNA was calculated by dividing the number of samples with an edit at the target site by the number of rooted shoots in each group.

Table 2.8 List of oligos used to test gRNA mutagenesis efficiency *in vivo*.

Amplicon content	Primer ID and sequence (5′ to 3′)	Amplicon size
Cas9 screen	LG278 ACTAACTCTGTGGGATGGGC	170 bp
	LG279 AGCGGTTCTCTTGAGTCTGG	
-60bp *ANT1* promoter target site region	LG143 GCTGGCAGGATAGGTACATTG	260 bp
	LG144 GGTTCATGGACTGATGAAGAAG	
-800bp *ANT1* promoter target site region	LG261 GAGTAGTGGCGTAAGTGTAAATAATTAG	532 bp
	LG264 GTTATAACTATCATTAATCGTGAGAGG	

2.2.13 Image J analysis of calli

The abaxial side of the explants was photographed on the same day six weeks after transformation. A Nikon D5 camera was used held on tripod so that the distance between the objective and the plate was identical for all plates. Settings used were ISO-100, F-stop f/11, exposure time 1/30 sec. The raw images (NEF files) were converted to a format compatible with the Image J software. Magnified pictures (x7) of each explant were collated into a document which was saved as a PNG file that can be opened in Image J (Figure 2.3, 2). Once opened in Image J, the Split Channels tool was used to create three 8-bit grayscale images, separately containing the red, green and blue components from the original RGB image. The red wavelengths were only detected from the purple sectors on the red grayscale image hence the red image was used for subsequent analysis step (Figure 2.3, 3). Using the Threshold function, the red grayscale image was adjusted so that regions corresponding to purple sector were highlighted in white and the rest in black, as background (Figure 2.4, right image).

Threshold values were set between 0-70 and 0-82. Next, the Region of Interest manager tool was used on the original RGB image to manually select the regions on the picture containing callus. Callus region was defined as any region of the explant where tissue growth was observed (as callus mass or shoot), usually along the lengthwise cuts on the cotyledon. These regions are delimited in yellow, as seen in Figure 2.4 (left image). The area selected on the original RGB image was then transferred onto its corresponding thresholded red grayscale image (Figure 2.4, right image). The Measure option was then used to calculate the callus area (section contained within the yellow lines, in pixels) and the percentage of this callus area corresponding to purple sectors (i.e., white pixels).

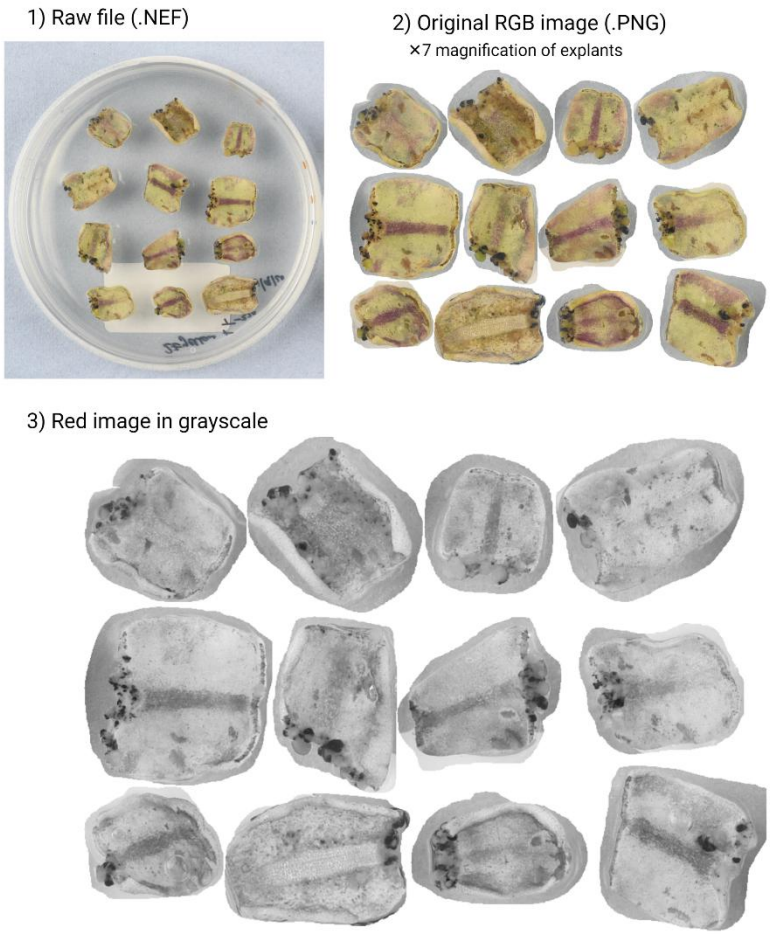


Figure 2.3 Image preparation and conversion. 1) the raw image NEF file was converted to 2) a PNG file after explant magnification (× 7) to create the original RGB image. 3) Grayscale image containing the red elements of the RGB image after the colour split channel step. Dark regions correspond to purple sectors.

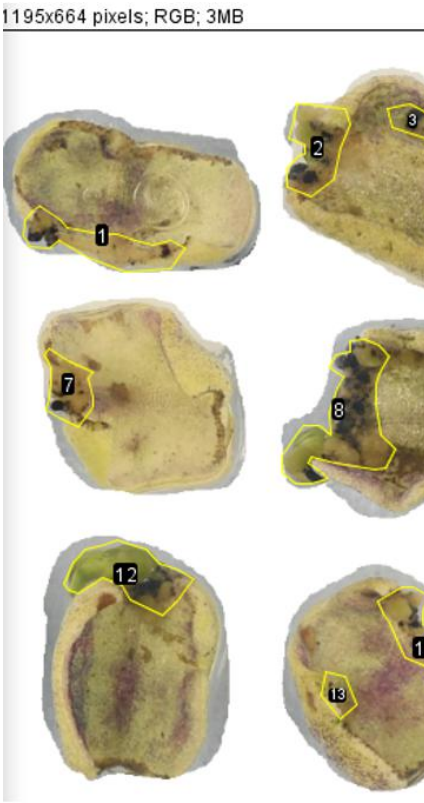


Figure 2.4 Manual selection of the callus area using the Region of Interest manager in Image J. On the left image, callus regions have been manually

selected and are delimited by the yellow line on the original RGB image. On the right, the defined callus regions have been transferred to the corresponding thresholded red grayscale image. The thresholding was adjusted so that highlighted pixels (i.e., white pixels) correspond to the purple sectors seen in the original RGB image.

2.2.14 Data analyses

2.2.14.1Gene targeting frequencies analysis

After scoring of the purple sectors from explants, the data was recorded, analysed and represented graphically as bar charts using Excel.

2.2.14.2 Statistical analyses

Chi-square tests and the Student t test were performed using the statistical analysis software GraphPad Prism 8.4.3. The Chi-square test was performed two-sided p value. The confidence level chosen is the default 0.05%.

2.2.15 Figure mounting and scientific illustrations

Scientific illustrations and figures were created with BioRender (www.biorender.com).

Chapter 3. Development of a phenotypic screen to assess rates of knock-ins and testing of a landing pad at the tomato *ANT1* locus

3.1 Introduction

Targeted and precise insertion of novel genes into plant genome has been a long sought-after goal of genetic engineering. Developing gene targeting, or knock-in-breeding, whereby genes conferring new traits can be inserted at defined locations in a crop's genome for trait improvement requires addressing three points: (1) low rate of KI due to inherently low rates of occurrence of homologous recombination in differentiated somatic cells, (2) measuring rates of occurrence for desired events when testing variables that may increase KI efficiency, (3) isolating and regenerating edited tissues into whole plants. This chapter will describe experiments aimed to address (1) and (2), whereas the experiments detailed in chapter 4 are aimed to address (3).

Boosting the rates of KIs: choosing tools and variables

It was shown that rates of KIs can be increased by an order of magnitude by triggering the activation of the DNA repair machinery via nuclease-induced DSB at the targeted locus (Wright et al., 2005). The CRISPR nuclease Cas9 can be used to deliver a DSB at the genomic target for activation of the DNA repair machinery. As mentioned in chapter 1, one of the biggest challenges to successful KI is the strong competition from NHEJ machinery throughout the cell cycle for repair of the DSB (Knoll et al., 2014; Fukushima et al., 2001).

Many variables have been tested in diverse experimental systems and shown to influence the rate of targeted recombination. Here, I tested (1) a comparison of KI rates after a single stranded break (SSB) compared to a double stranded break (DSB), (2) the influence of using a

viral replicon to increase the donor template copy number, (3) the influence of elevating temperature on KI rates.

Efficacy of SSB initiation of HR

To favour KI by HR over NHEJ-mediated repair of the DSB, a study in human cell lines successfully employed a nickase ($Cas9^{D10A}$) wherein the RuvC catalytic site contained a mutation that allowed delivery of a SSB at its target site (Chen et al., 2017). SSBs are not a substrate for NHEJ (Vriend et al., 2016) and are repaired in animal cells *via* a branch of the HR pathway, although the details of this mechanism have not been fully elucidated yet (Davis and Maizels, 2014; Bothmer et al., 2017b). Importantly, the actors of nick-HR do not compete with NHEJ components for repair (Vriend et al., 2016). Therefore, the $Cas9^{D10A}$ nickase was targeted at three locations: at the genomic target site and at the two extremities of the homology regions flanking the insert on the vector. This is referred to as the three SSBs strategy. Compared to the standard strategy (one Cas9-mediated DSB at the target site), this three SSBs strategy increased KI efficiency by × 3.5 and × 9 in different cell lines. At the time of experimental design, the three SSBs strategy had not been tested in plants, hence we set out to investigate its KI boosting potential in the tomato cultivar Moneymaker. An important underlying condition for successful gene targeting is the on-target cleavage efficiency by Cas9. Variability in the on-target cleavage efficiency can stem from gRNA design (sequence features of the gRNA such as GC content, sequence composition, nucleotide position, secondary structures) and genomic features (chromatin accessibility) (reviewed by Liu et al., 2020). Therefore, many algorithms have been developed to predict gRNA efficiency, however, the accuracy of these predictions is questionable for CRISPR-based studies in plants (Naim et al., 2020). To account for variability in gRNA mutagenic efficiency, it is important to test several gRNAs.

Influence of elevating template copy number using a viral replicon

As discussed in chapter 1, availability of the homologous template carrying the desired sequence at the target site upon DSB induction is another strategy to boost the rates of KI. Deconstructed geminiviruses (plant ssDNA viruses) have been engineered as replicon systems that provide many copies of homologous templates. This system has been used in many plant species such as tomato (Čermák et al., 2015; Vu et al., 2020b; Dahan-Meir et al., 2018) potato (Butler et al., 2016a), rice (Wang et al., 2017), wheat (Gil-Humanes et al., 2017), cassava

(Hummel et al., 2018), and Arabidopsis (Hahn et al., 2018a; De Pater et al., 2018). Bean yellow dwarf virus (BeYDV) sequences have been used to create the replicon in dicot plant species whereas sequences from the wheat dwarf virus is used for monocots. Sequences essential for replication of viral replicons are maintained i.e. the short intergenic region, SIR, the long intergenic region, LIR, and the replicase gene, Rep and introduced on a T-DNA for delivery into plant cells. In contrast, the viral genes coding for the coat protein and movement proteins are removed and replaced by other components on the plant transformation vectors i.e. nuclease and donor template. Once delivered into the plant cell, the LIR sequences promote circularisation of the replicon. Subsequently, the Rep/RepA (a splicing variant of Rep) will induce a replication permissive cellular state through interactions with multiple host factors (Hanley-Bowdoin et al., 2004, 2013) which initiates viral rolling circle replication in the plant's nucleus. Sequences placed between the LIR will therefore be included on the replicon and be replicated to several thousands of copies (Baltes et al., 2014). Two strains of the geminivirus BeYDV exist, a mild strain (GenBank accession number: DQ458791, Halley-Stott et al., 2007) and an acute strain (Liu et al., 1997). Nucleotide sequence comparison between the two genomes presented 81 nt differences in total, located in coding and non-coding regions. These nt polymorphisms gave rise to 20 amino acid substitutions, some of which are of notable significance for the viral biology as amino acids changes are occurring in functional domains (rolling circle replication motifs, Retinoblastoma binding motifs) of the Rep/RepA sequence. Key differences between the two strains revolve around plant cell cycle control and viral replication. Comparison of these two strains *in planta* confirmed that infectious symptoms were milder and appeared later when inoculated with the mild strain, with lower viral DNA accumulation as shown by qPCR (Halley-Stott et al., 2007). Previous gene targeting experiments have based their geminivirus replicon system on the mild strain (Čermák et al., 2015; Hahn et al., 2018a; Vu et al., 2020b; Dahan-Meir et al., 2018).

Influence of temperature on Cas9 activity and knock-ins

Environmental conditions have also been shown to increase rates of KI. Some reports have shown that increased temperature can enhance the enzymatic activity of Cas9 which leads to higher mutagenic rates (Le Blanc et al., 2017). Furthermore, elevated temperature can also increase somatic homologous recombination in Arabidopsis seedlings (Rahavi and Kovalchuk,

2013). Indeed, environmental stresses can induce modifications of the chromatin structure and alter the rates of somatic HR (Yao and Kovalchuk, 2011; Molinier et al., 2005).

The importance of a robust assay to detect KI events

To develop a methodology that enables comparative assessment of KI rates from different experimental set-ups, an effective assay to measure the frequency of desired events occur with a given strategy is essential. Furthermore, as KIs are notoriously low frequency events, a robust screen for such events is necessary to detect all events. For instance, a low sensitivity assay such as GUS for scoring KI events led to an underestimation of KI frequencies in tobacco protoplasts (Wright et al., 2005). As the authors were relying on a double selection system to screen for events i.e. restoration of GUS:NptII reporter upon insertion, the discrepancy between the number of kanamycin resistant transformants with that of GUS+ transformants enabled them to detect the low sensitivity of the GUS assay. For this reason, it is often beneficial to utilise a double selection system for detection of KI events.

Anthocyanins are flavonoids, polyphenolic pigments that can be synthesised in most plant tissues and organs (flowers, fruits, leaves, seeds, roots, tubers) across the plant kingdom (Lee, 2002). Anthocyanins have diverse roles in plants, from attracting pollinators to the reproductive organs with tissue specific pigmentation of petals to protecting against physiological stresses. Due to their antioxidant properties, anthocyanins are able to protect from oxidative stress induced by reactive oxygen species in vegetative tissues caused by abiotic stress such as drought, cold, strong light, etc. (Nakabayashi et al., 2014; Catalá et al., 2011; Steyn et al., 2002; Sharma et al., 2019). The distinct roles of anthocyanins being based on developmental stages or environmental cues means that their production must be specifically regulated (Albert et al., 2014). A protein complex composed of three types of transcription factors, an R2R3-MYB, a basic Helix-Loop-Helix (bHLH) and a WD-repeat (WD40), together making the MBW complex (Xu et al., 2015), is involved in the control and coordination of the transcription of early structural genes (i.e. *chalcone synthase*, *chalcone isomerase*, *flavonoid 3-hydroxylase*) and late structural genes (*dihydroflavonol 4-reductase*, *anthocyanin synthase*) encoding enzymes that catalyse the conversion of precursor compounds into anthocyanin compounds (pelargonidin, cyanidin and delphinidin). The R2R3-MYBs play a key role in the regulation of anthocyanin production as they possess a highly conserved DNA binding domain at the N-terminal (R2R3 motif) and a more variable transcriptional activation

domain at the C-terminal (Pireyre and Burow, 2015). Four homologous R2R3 MYBs involved in anthocyanin regulation have been characterised in tomato, *SlAN2* (*Solyc10g086250*), *SlANT1* (*Solyc10g086260*), *SlANT1-like* (*Solyc10g086270)* and *SlAN2-like/Aft (Solyc10g086290*) all located close to each other on chromosome 10 (Kiferle et al., 2015; Yan et al., 2020).

In cultivated tomato (*Solanum lycopersicum*), although production of anthocyanins mostly occurs in the stems and hypocotyl tissues (Roldan et al., 2014) as a stress response, several studies have shown that ectopic expression of *ANT1* under a 35S promoter leads to strong purple pigmentation in most organs i.e. leaves, stems, anthers, fruits (Schreiber et al., 2012; Kiferle et al., 2015). Thus, anthocyanin pigmentation has been used in several plant species as a visual marker for successful transformation (Kortstee et al., 2011; Jin et al., 2012; Kim et al., 2010; Rommens et al., 2008) and has also been employed as a visual marker for successful KIs in tomato (Čermák et al., 2015; Vu et al., 2020b). As mentioned above, the R2R3-MYB transcription encoded by *ANT1* operates in a protein complex with a bHLH and WD40 protein to promote anthocyanin biosynthesis. Anthocyanin accumulation can be achieved through transgenic overexpression of the R2R3-MYB alone as the latter harbours a master regulator activity allowing it to upregulate its bHLH partner (Kiferle et al., 2015) to form a functional MBW complex, as WD40 protein is actively expressed in all tissues (Ramsay and Glover, 2005). Nevertheless, absence of purple colouration can be observed in *ANT1* overexpressing tissue since the bHLH protein controls the tissue-specific production of anthocyanins (Montefiori et al., 2015) by binding to biosynthetic gene promoters for activation. For example, fruit flesh of tomatoes overexpressing *ANT1* was mostly red whilst pigments accumulated mostly in the fruit skin (Kiferle et al., 2015).

In the gene targeting experiments, a 35S promoter was inserted by homologous recombination upstream of *ANT1* which led to the formation of sectors with strong purple pigmentation on the developing callus in the event of a KI. These purple sectors were then used as a proxy to estimates the rates of KIs. Cermak et al. (2015) measured desired events by counting the number of explants with at least one developing purple sector and divided by the total number of transformed explants to give a frequency. This frequency was multiplied by a 100 to obtain a percentage representing the number of edited explants. Normalisation of the frequency was done by dividing the initial KI frequency by the baseline transformation efficiency (using a vector expressing *35S::ANT1* on a T-DNA) to factor in that some explants are not successfully

transformed and adjust the rates of KI accordingly. In Vu et al (2020), they counted the total number of purple sectors and divided this number by the number of explants transformed to obtain a frequency of KI which was then normalised using the frequency of purple sectors when a transforming a vector expressing *35S::ANT1*. The first method emphasises the occurrence of edited samples in a given experiment, whereas the second focuses on the occurrence of KI events at the cellular level. The different scoring methods employed by Cermak et al (2015) and Vu et al (2020) does not enable straight comparison of KI efficiency between these two experiments. Interestingly, *ANT1* was shown to be upregulated by a 4 × 35S enhancer located 1.8 kb upstream of its CDS in an activation tagging experiment (Mathews, 2003), revealing a potential alternative strategy and landing pad for KIs.

3.2 Aims

The work in this chapter began with the aim of performing an independent repeat of the Cermak et al (2015) experiment to confirm high rates of knock-ins can be achieved using our own vectors. Our read-out for KI success is also purple sector development on regenerating callus after insertion of a 35S promoter upstream of *ANT1* coding sequence (CDS). Additionally, some variables were introduced to observe their influence on knock-in efficiency. I tested (1) Cas9 vs. $Cas9^{D10A}$, delivering one DSB at the genomic target site or three SSBs (one at the genomic target and two simultaneously at the extremities of the homology regions on the vector), (2) the effect of replicon size (Figure 3.1) and (3) the effect of a 28 °C heat treatment early in the tissue culture phase on the recovery of desired events.

35S promoter-Ω::*ANT1* edited cells suffer from poor regeneration. An additional second strategy tested in this chapter entails knocking in a 2 × 35S enhancer and *NptII* at -800 bp from *ANT1*. 35S enhancer-mediated upregulation on *ANT1* is expected to retain a visually detectable phenotype that can be scored whilst triggering a lower *ANT1* expression compared to that of a 35S promoter. Furthermore, two strains of the BeYDV (mild or acute) were compared for their ability to promote KI. To account for gRNA efficiency variability, two gRNAs were used separately to target the *ANT1* promoter.

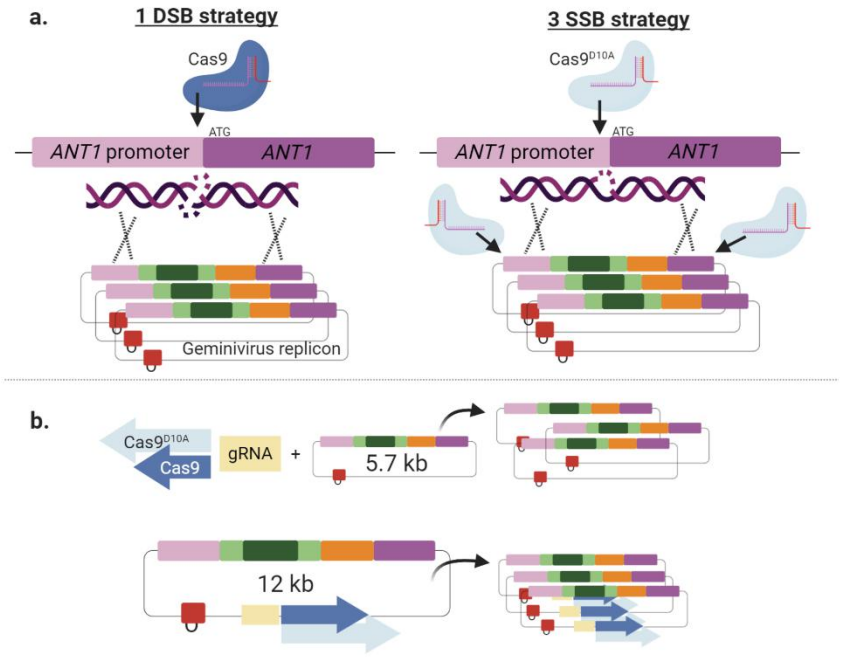


Figure 3.1 Schematics illustrating the two variables tested to increase gene targeting success. A. Comparing the rates of KIs when using the one double stranded break (one DSB) approach or the three SSBs (three SSBs). B. Comparing the rates of KIs with different size and content of replicon. In the first instance, the expression cassettes for Cas9/nickase and the gRNA are placed outside the replicon, whereas they are included and replicated from the replicon in the second case.

3.3 Construction of Golden Gate vectors for 35S promoter knock-in at the ANT1 locus

In this experiment, I assessed different variables for their potential to promote high efficiency KIs. Different sizes and content of the Geminivirus replicon were tested to assess the significance of the trade-off between cargo load, replicon content copy number and KI efficiency. I assessed KI efficiency when the expression cassettes for the nuclease/nickase and the gRNA are placed outside the replicon – therefore generating few copies of the CRISPR reagents – making for a 5.7 kb replicon which has the potential to replicate the donor template (i.e. the sequence to be inserted flanked by homology regions) to a higher copy number. Conversely, efficiencies were assessed when the Cas9 and gRNA expression cassettes are included on the replicon creates a 12 kb replicon, producing more copies of the CRISPR reagents but potentially hindering replication due to the cargo size, and causing an overall copy number reduction of all the components on the replicon.

The Golden Gate cloning method was used to assemble the various parts required to build 5 binary vectors to knock in a 35S promoter alongside the antibiotic resistance marker gene *NptII* at the *ANT1* locus. Details of the golden gate cloning protocol, modules already available and generation of new modules are described in sections 2.1.3.1, 2.2.1.2 and 2.2.3. Based on type II restriction enzymes, *Bpi*I and *Bsa*I, which cleave outside their recognition sequence to produce unique 4 nucleotide overhangs, the Golden Gate cloning method enabled the multipart, unidirectional assembly in all of KI constructs of a total of 11 level 1 modules, which were split into 2 level M acceptor vectors and further assembled into a binary acceptor vector, referred to as level P. The assembly process is described in Figure 3.2. The same principle applies for every construct used in this research project. Shown in Figure 3.3, constructs LGJJ52 and LGJJ53 compared KI frequency between one DSB (*Arabidopsis thaliana*-codon optimised Cas9) delivered at the genomic target (LGJJ52) or when three SSBs (Cas9 nickase, $Cas9^{D10A}$) are delivered at the genomic target and at the extremities of the regions of homology) flanking the donor template (LGJJ53). These two constructs carry the genes for the nuclease or nickase and the gRNA outside the geminivirus replicon - as placed outside of the long intergenic regions (LIRs) - making for a 5.7 kb replicon. Similarly, vectors LGJJ54 and LGJJ55 compared KI efficiency between one DSB and three SSBs approach.

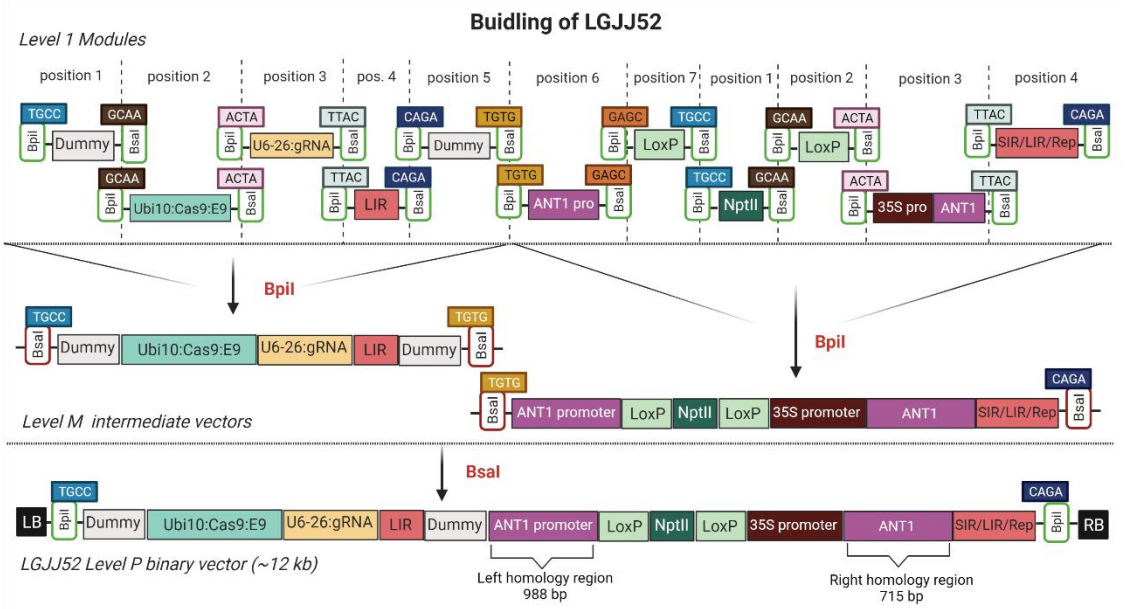


Figure 3.2 Diagram showing the cloning strategy of plant transformation constructs. Level 1 modules for each of the 7 positions have unique 4 nucleotide overhangs released upon digestion of the level 1 vectors with *Bpi*I which allows unidirectional assembly of level M

vectors after ligation by a DNA ligase. Level M vectors are then assembled in level P binary vectors using the restriction enzyme *Bsa*I and a DNA ligase.

However, these two constructs carry the genes for the nuclease or nickase and the gRNA inside the geminivirus replicon LIRs to obtain high copy numbers of Cas9, Cas9^{D10A}, gRNA, making for a 12 kb replicon. LGJJ58 is a negative control with a dead Cas9 (dCas9) which does not cut the genomic target, nor the donor template on the replicon. KI frequency obtained with LGJJ58 represents the endogenous rate of KI when homologous recombination is not primed by nuclease – or nickase - induced DNA lesions. Here, the antibiotic selectable marker gene *NptII* was flanked by Lox P sites to allow for excision of the marker in subsequent generations of edited plants *via* recombination by a Cre recombinase. For consistency, my constructs were made with the same gRNA and homology regions that were used in Cermak et al.'s (2015) experiment. In contrast, the Golden Gate modules encoding for the geminivirus replicon are based on an acute strain of the BeYDV, whereas a mild strain is used in Cermak et al. (2015). Furthermore, the 35S promoter to be knocked-in contains a tobacco mosaic virus omega leader sequence (referred as Ω) at its 3′ end to further enhance translation *in planta* (Gallie, 2002). To compare efficiencies, my constructs were transformed in parallel with the vector pTC217 from Cermak et al's 2015 publication and their pTC147 vector to obtain a baseline rate of purple sectors when phenotype is not relying on a KI.

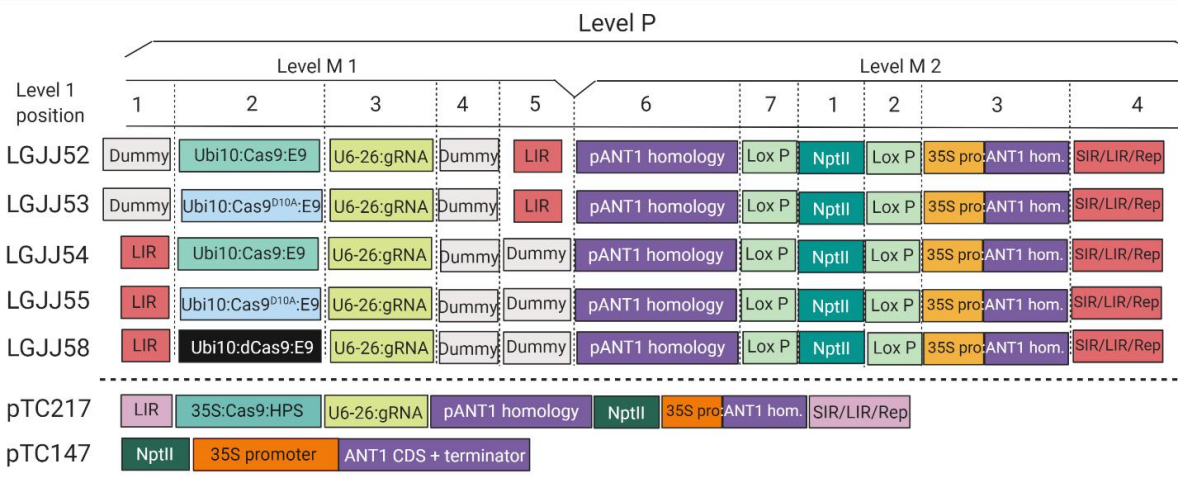


Figure 3.3 Diagram of the first set of constructs to attempt gene targeting at the *ANT1* locus to perform an independent repeat of the Cermak et al. (2015) experiment whilst testing some

variables for their potential to increase rates of KIs. Vector LGJJ52 and LGJJ53 have the gene expression cassettes for the nuclease or nickase and the gRNA outside of the replicon whereas they are included on the replicon for vectors LGJJ54 and LGJJ55. LGJJ53 and LGJJ55 carry a nickase allele, Cas9^{D10A}, whereas LGJJ58 contains a dead Cas9. The homology regions to *ANT1*, represented in purple blocks, are identical to that used by Cermak et al (2015). Left homology region contains 988 bp of *ANT1* promoter and is cloned in position 6 (pANT1 purple block). Right homology region contains 715 bp of *ANT1* CDS starting from -25 bp from the ATG (ANT1 hom. purple block in level M2 position 3). It was cloned in combination to a 35S promoter to make the module in position 3 of the 2nd level M part. The red squares contain the BeYDV geminivirus replicon sequences, SIR short intergenic region, LIR, long intergenic region, Rep, Replicase. Elements of these vectors are not drawn to scale.

Figure 3.4 In silico representation of the DNA sequence at the *ANT1* locus after 35S promoter KI by HR. TATA box elements is highlighted in yellow in the 35S promoter, the start of transcription annotated as +1. The 5' UTR of the transcript is a chimeric sequence between 35S CaMV sequence, TMV omega leader sequence (translation enhancer), and 24 bp of endogenous *ANT1* UTR sequence before the start of *ANT1* coding sequence (exon 1).

3.4 Testing different variables and their effect on gene targeting efficiency

In this KI approach, a 35S promoter and an *NptII* selectable marker are targeted 60 bp upstream of the *ANT1* reading frame which leads to anthocyanin accumulation in plant tissues

and kanamycin-resistant tissues upon integration. The five KI constructs described in section 3.3 and Figure 3.3 were transformed into Moneymaker tomato explants. Vector pTC217 was used as positive control for KIs, as it previously reached 10% KI efficiency (Cermak et al., 2015). This vector contains a 35S:Cas9 expression cassette, a U6-26:gRNA cassette expressing the same gRNA used here, a donor template containing an *NPTII* cassette and a 35S promoter flanked by homology regions to *ANT1* (map of the construct available in Appendix D.1). Between four to six weeks after transformation, explants were observed and the number of explants displaying at least one purple sector was recorded as a putative KI event. KI frequencies were calculated by dividing the number of explants showing at least one purple sector by the total number of transformed explants (Table 3.1, 4th column). KI are represented as a percentage of successfully edited samples (and referred to as KI efficiency) in column 5 of Table 3.1 by multiplying the KI frequency by 100. An additional control vector, pTC147, which contains a T-DNA with a 35S promoter::*ANT1* expression cassette (map of the construct available in Appendix D.2), was transformed into Moneymaker and 85% of the explants harboured at least one purple sector.

Table 3.3.1 Calculations of gene targeting frequencies and efficiencies in Moneymaker. Frequency of KI event is calculated by dividing the number of explants showing at least one purple sector by the total number of explants transformed. The efficiency can be measured as a percentage of the explants with a putative knock-in by multiplying KI frequency by 100. *pTC147 is a control vector carrying on a T-DNA an expression cassette with 35S promoter::ANT1 where purple pigmentation relies on T-DNA insertion.

Cultivar transformed	Vector ID	Explants used for transformation	Explants with at least one purple sector	Frequency of purple calli	KI efficiency (%)
Moneymaker	LGJJ52	252	6	0.024	2.40
	LGJJ53	192	0	0	0.00
	LGJJ54	246	3	0.012	1.20
	LGJJ55	228	1	0.004	0.40
	LGJJ58	206	0	0	0.00
	pTC217	370	37	0.100	10.00
	pTC147	*138*	*117*	*0.848*	*84.80**
Micro Tom	LGJJ52	352	15	0.043	4.30
	pTC217	320	18	0.056	5.60
	LGJJ58	181	1	0.006	0.60

Notably, the delivery of a single DSB at the genomic target by Cas9 led to higher KI efficiencies (LGJJ52 and LGJJ54, 2.38% and 1.2% respectively) compared to the three SSBs strategy (0% and 0.43% for LGJJ53 and LGJJ55 respectively) (Figure 3.5). Furthermore, the high KI efficiency observed for LGJJ52 (Cas9 and gRNA expression cassettes outside replicon) at 2.38% compared to LGJJ54 (Cas9 and gRNA expression cassettes inside the replicon) at 1.2% indicates that a higher copy number of the nuclease and the gRNA does not correlate with an increase in KI efficiency. However, these data are based on one round of transformation and do not provide statistical significance. The published KI vector pTC217 also reached 10% KI efficiency in our hands. Explants transformed with the negative control (LGJJ58) had no purple sectors.

Based on these results, the experiment was repeated in the tomato cultivar Micro-Tom, which has been utilised in other GT experiments (Cermak et al., 2015; Dahan-Meir et al., 2018). Only LGJJ52 was used as performing best in Moneymaker. As shown in Figure 3.5, KI efficiency reached 4.3% in this cultivar. The positive control pTC217 showed 5.6% KI efficiency, which is nearly half the rate obtained in Moneymaker. The negative control LGJJ58 displayed one explant with a purple sector (1/181 transformed explants). These data indicate that the discrepancy observed in the KI efficiencies between my experiment and that of Cermak et al. (2015) and Dahan-Meir et al. (2018) is not dependent on the tomato genotype but rather on experimental design.

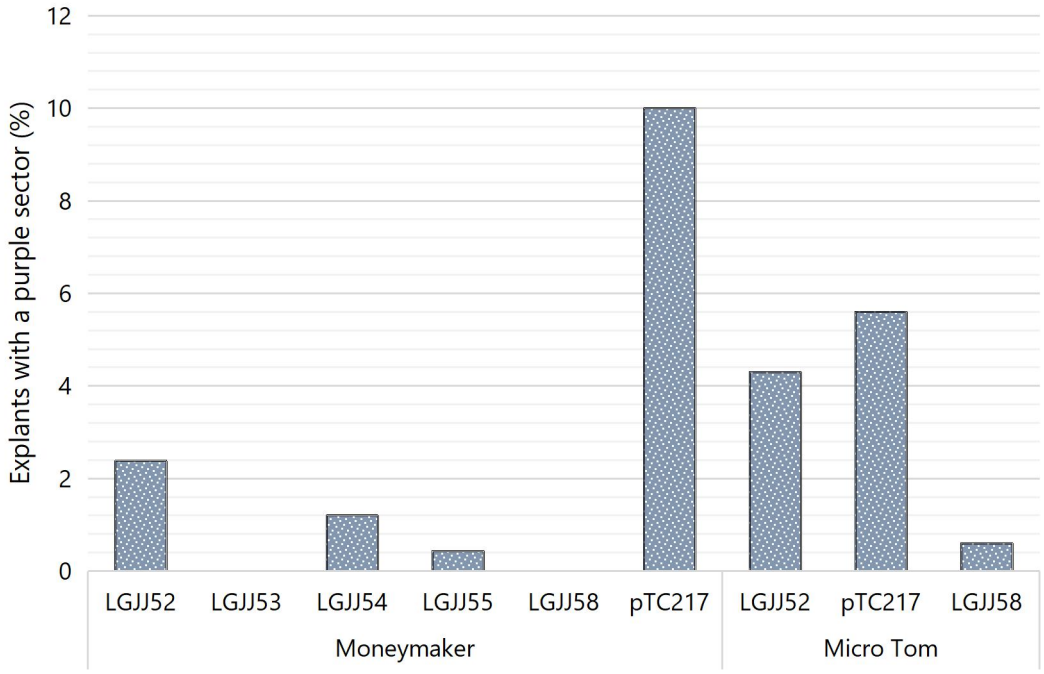


Figure 3.5 Knock-in efficiency of a 35S promoter at the *ANT1* locus in the tomato cultivars Moneymaker and Micro Tom. This bar chat represents the percentage of explants showing at least one purple sector for vectors LGJJ52, LGJJ53, LGJJ54, LGJJ55, and the negative control LGJJ58 and positive control pTC217. These data represent the efficiencies obtained from one round of transformation for each vector.

3.5 The effect of elevated temperature on knock-in rates

Environmental conditions were also tested for their ability to increase rates of KI. Several reports have shown that increased temperature can enhance the enzymatic activity of Cas9 which leads to higher mutagenic rates (Le Blanc et al., 2017). Furthermore, elevated temperature can also increase somatic homologous recombination in Arabidopsis seedlings (Rahavi and Kovalchuk, 2013). Indeed, environmental stresses can induce modifications of the chromatin structure and alter the rates of somatic HR (Yao and Kovalchuk, 2011; Molinier et al., 2005). Therefore, a heat treatment (three days at 28 °C) was included during the initial phase of tomato explant regeneration by tissue culture.

Heat treatment consisted of a three-day incubation at 28 °C after co-cultivation. Based on one round of transformation for the six vectors tested, the number of explants with a KI event was higher in the group that received the heat treatment (Figure 3.6, red bars), compared to

explants staying at 22 ˚C (Figure 3.6, blue bars). These data indicate that heat treatment during early explant regeneration may increase the frequency of KI events.

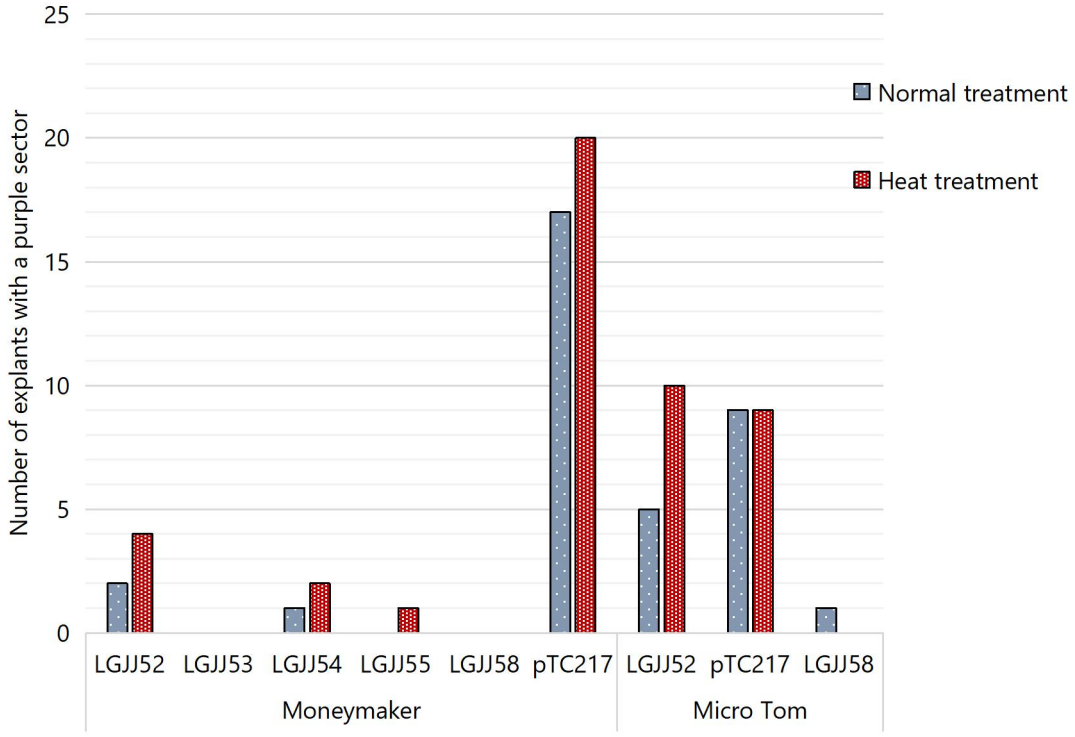


Figure 3.6 Bar chart describing the influence of elevated temperature on recovering gene targeting events. The heat treatment consisted of a three-day period at 28 ˚C after co-cultivation with *Agrobacterium*, before placing explants at 22 ˚C for the rest of the tissue culture phase. These data represent the count of explants with purple sectors obtained from one round of transformation for each vector.

3.6 Verification and characterisation of a knock-in event

To verify the presence of KI events, genomic DNA (gDNA) was extracted from purple leaf tissue derived from an explant transformed with LGJJ52 (Figure 3.7). KI specific PCRs were performed to amplify the DNA that covers these left and right junctions: as shown in Figure 3.7, one primer anneals to the insert-specific sequence (red arrows, LG23, LG141) whereas the second primer anneals to the gDNA (blue arrows, LG203, LG140). This design should only produce an amplicon if there is a KI event at the genomic target site. PCR products of the expected size (1.3 kb) were gel extracted and their purified products sent for Sanger sequencing. Alignment of the sequencing data to the reference sequence of the expected KI product

indicated that no base pair mutations were at the left and right junction of the repaired break (Figure 3.7). Full sequencing coverage of the 1.3 kb PCR amplicons spanning left and right junctions of the KI is available in Appendix E.1 and E.2. These data indicate that repair of the two DNA ends *via* HR after Cas9-induced DSB.

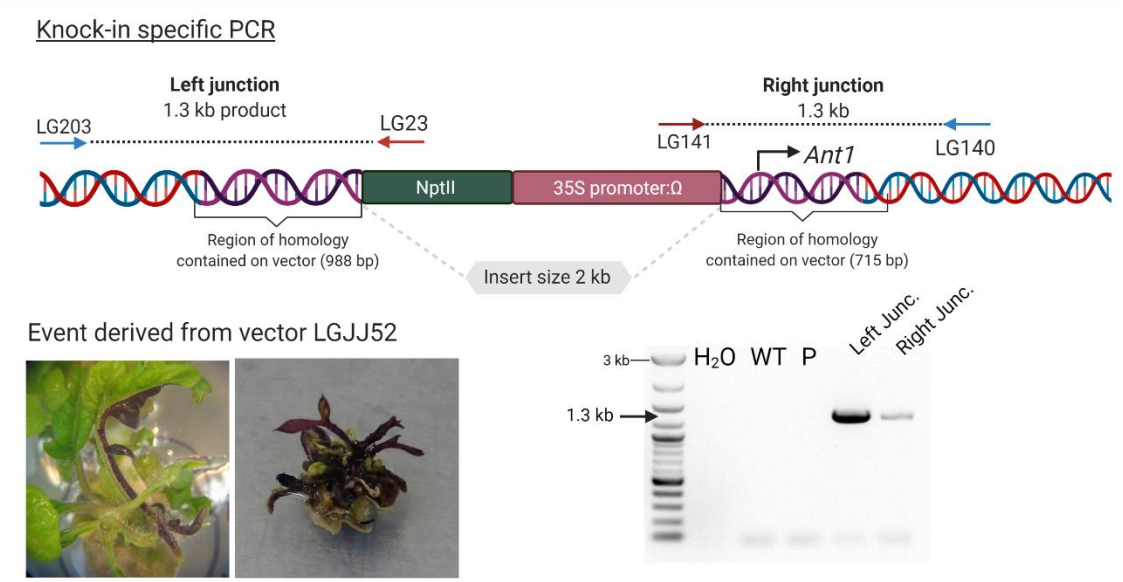


Figure 3.7 Picture of the purple plantlet characterised and workflow for KI characterisation. Description of the knock-in specific PCRs that can only produce an amplicon if a KI has occurred. Left and right junction PCRs were performed in parallel with a blank sample negative control (lane H_2O), wild-type gDNA negative control (WT lane), and LGJJ52 plasmid negative control with primers LG203+LG23 (lane P).

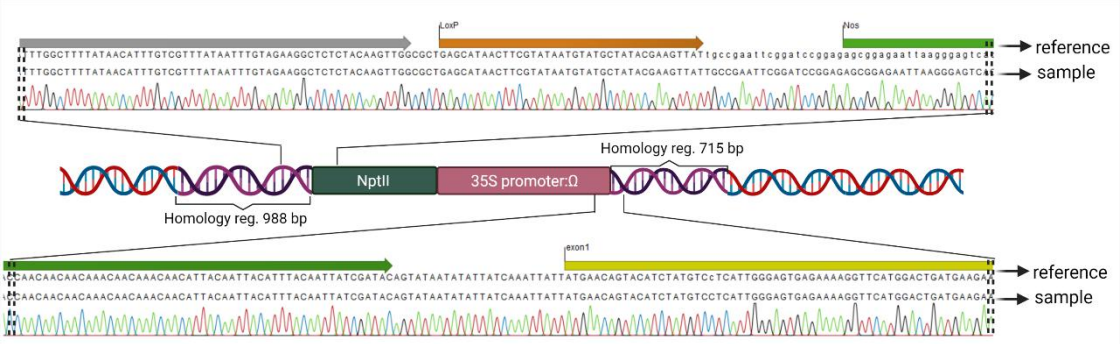


Figure 3.8 Sanger sequencing data of the ANT1 locus extracted from purple leaf tissue (LGJJ52-derived KI) aligned to the reference sequence of the expected KI product. Mutation-free repair at the left and right junction of the break indicates repair and insertion through HR. Dashed black lines at the end of the sequences indicate that only a selection of the sequencing covering the whole 1.3 kb junction is shown for ease of representation.

3.7 Regeneration of purple sectors is challenging

As described by Cermak et al. (2015), purple sectors emerging from regenerating Moneymaker calli were excised from their originating callus to be regenerated on their own. This procedure, however, was only partially successful. During callus development, the purple sectors were often engulfed by neighbouring green tissues before they grew large enough to be excised from callus (Figure 3.9). Out of the 37 purple sectors excised from calli transformed with pTC217, no purple shoot was fully regenerated. As seen in Figure 3.10, even for purple sectors that were excised, only few of them continued to grow and develop (samples P9_1, P12_1) whilst some were overtaken by green tissues (P17_1, P18_1, P11_1). Out of the five vectors transformed into Moneymaker in this experiment, only one purple shoot spontaneously emerged from callus (event from LGJJ52, sequence-verified in Figure 3.7 and 3.8). The limited success in regenerating purple, edited tissues is attributed to compounding factors which are discussed in the discussion section of this chapter (3.13). One likely factor to be noted here is the that the overexpression of *ANT1* following insertion of a 35S promoter-Ω upstream likely incurs a strong fitness cost, perhaps in part through excessive anthocyanin accumulation which hinders the development of the purple tissues (Kortstee et al., 2011; Chawla et al., 1999). Therefore, the second part of this chapter describes an alternative gene targeting strategy tested at the *ANT1* locus to minimise the pleiotropic effects of anthocyanin accumulation in edited cells.

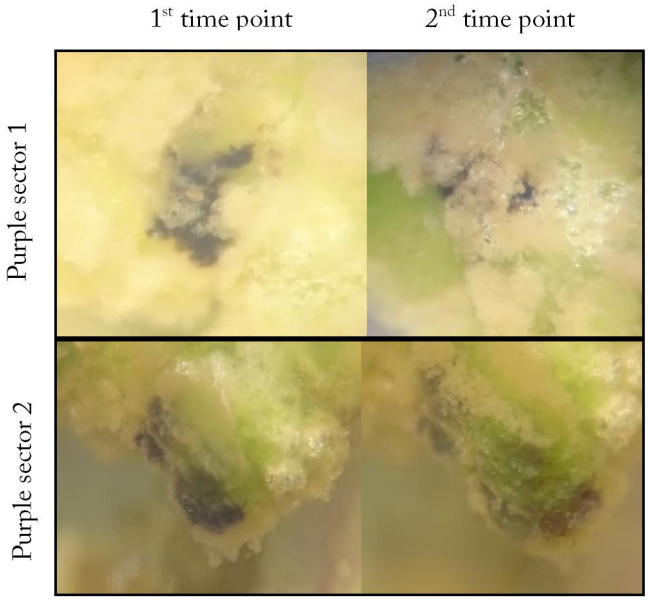


Figure 3.9 Pictures of 2 purple sectors from two individual explants transformed with vector LGJJ52 photographed at 1-week interval. WT cell overgrew the cells with a KI over time.

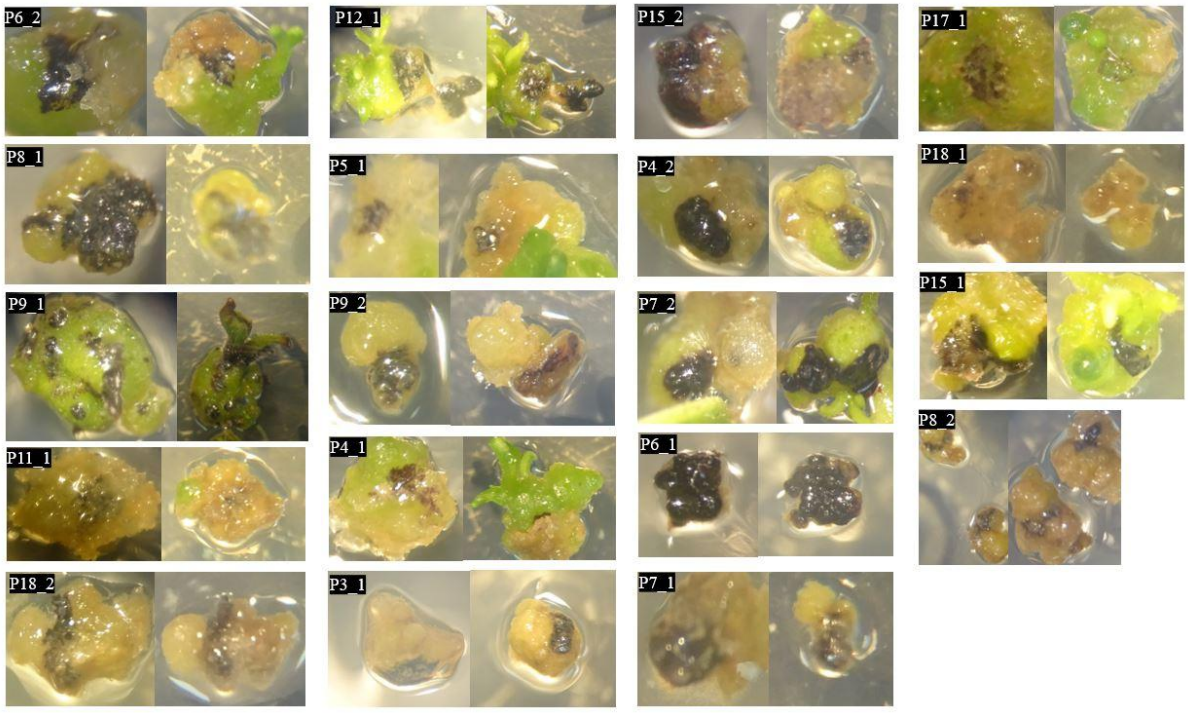


Figure 3.10 Growth of excised purple sectors. For each of these 18 sectors, the first picture was taken the day it was excised and the second picture two weeks later.

3.8 Construction of Golden Gate vectors for 35S enhancer knock-in in the region of -800 bp of *ANT1* CDS

Here, the strategy involves knocking in a 2 × 35S enhancer in the region of -800 bp in the *ANT1* promoter to limit the deleterious effects caused by excessive anthocyanin accumulation upon successful KI. Of note, the 2 × 35S enhancer used here contains the -416 to -31 region of the 35S promoter, containing all elements of the promoter except the core sequence -30 to -0 where the TATA box element is located (Ow et al., 1987) which differs from 35S enhancers used elsewhere (-417 to -86, Mathews, 2003; Weigel et al., 2000). Some sequencing data revealed significant polymorphism in the distal region (beyond -900 bp) of the *ANT1* promoter between the cultivar Heinz and Moneymaker. To simplify the design of the homology region to the distal part of the *ANT1* promoter, this subset of experiment was performed in the tomato cultivar Heinz 1706 as its genome sequence is publicly available (Sato et al., 2012). To account for variability in gRNA efficiency, two individual gRNAs were used to target the region of -800 bp in the *ANT1* promoter.

Using the golden gate cloning system, I built five binary vectors (described in Figure 3.11) to compare KI efficiency of a 2 × 35S enhancer and *NptII* at -800 bp upstream of *ANT1* CDS using Cas9 to deliver a DSB at -723 bp (gRNA_1, LGJJ96 and LGJJ112) or at -800 bp (gRNA_1, LGJJ97 and LGJJ113). Vectors LGJJ96 and LGJJ97 contained a replicon based on an acute strain of BeYDV, whereas the replicon used for LGJJ112 and LGJJ113 was based on a mild strain of the BeYDV (Halley-Stott et al., 2007) used in other GT experiments in tomato (Cermak et al., 2015; Dahan-Meir et al., 2018). The expression cassettes for Cas9 and gRNA were placed outside of the replicon LIR as this proved more efficient in the previous experiment. A positive control (LGJJ139) was built utilising T-DNA transfer of the expression cassettes to establish the phenotype arising from 35S enhancer mediated *ANT1* upregulation at -800 bp. This control vector contained an antibiotic selectable marker *NptII*, 35S enhancer, 800 bp of *ANT1* native promoter, *ANT1* CDS followed by its native terminator. Upon T-DNA insertion of LGJJ139, we expected to observe the associated purple phenotype.

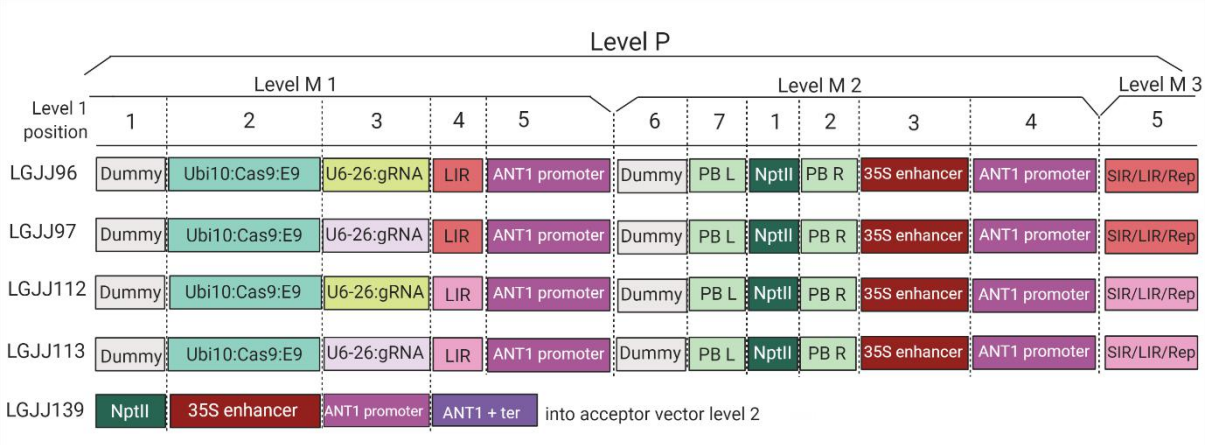


Figure 3.11 Vector design for 35S enhancer knock-in upstream of ANT1. All constructs were built using the Golden Gate technique. Vectors LGJJ96 and LGJJ97 contain the replicon modules based on the acute strain of the BeYDV (LIR, SIR, Rep, red boxes) whereas LGJJ112 and LGJJ113 contain a mild version (pink boxes). Vectors LGJJ96 and LGJJ112 harbour a gRNA targeting at -723 bp and LGJJ97 and LGJJ113 target at -800 bp. LGJJ139 is a positive control for 35S enhancer mediated *ANT1* upregulation.

Table 3.2 Summary of the designed vectors and their experimental purpose.

Vector	Aim	Experimental purpose
LGJJ96	Testing knocking in a 35S enhancer at **-723bp** of *ANT1* ATG using an **acute replicon**	- Knocking in a 35S enhancer instead of a 35S promoter to reduce expression of *ANT1* upon KI to improve purple sector regeneration - Testing whether the mild allele of the replicon enables (1) more KIs to occur (2) better regeneration of transformed tissues due to potentially reduced pleiotropic effects with a mild replicon - Testing 2 different gRNAs to target the -800 bp region of *ANT1* to account for gRNA efficiency variability
LGJJ97	Testing knocking in a 35S enhancer at **-800bp** of *ANT1* ATG using an **acute replicon**	
LGJJ112	Testing knocking in a 35S enhancer at **-723bp** of *ANT1* ATG using a **mild replicon**	
LGJJ113	Testing knocking in a 35S enhancer at **-800bp** of *ANT1* ATG using a **mild replicon**	
LGJJ139	Control vector, T-DNA insertion-mediated expression of *ANT1* under the regulatory control of a 35S enhancer.	Establishes phenotype to look for in case of successful 35S enhancer KI at the endogenous *ANT1* obtained with vectors LGJJ96, 97, 112 and 113.

3.9 The phenotype of *ANT1* upregulation via 35S enhancer is variable

A robust screen to monitor for expected phenotypes mediated by KI helps to recover low frequency events. Therefore, understanding the contribution from *ANT1*-upregulation mediated by a 2 × 35S enhancer upon KI is paramount to accurately detecting and recording these events. Using the phenotype positive control vector LGJJ139, I observed and tracked the evolution of purple sectors arising from *ANT1*-upregulation by a 2 × 35S enhancer in Heinz tomato explants during tissue culture. Sectors with anthocyanin accumulation were visible as early as three weeks after *Agrobacterium*-mediated transformation (Figure 3.12). At the callus stage, the purple sectors had less pigmentation than was observed with 35S promoter and there was variability in intensity between explants. From five weeks after transformation, variability of purple pigmentation continued as leaves developed. In some explants, the abaxial side of the leaf displayed strong purple pigmentation (Figure 3.12). In others, the edges of the leaves accumulated anthocyanins, but the centre of leaves and stems had wild-type pigmentation (i.e. green tissue). Transient production of anthocyanins resulting in pigmentation of plant tissues is known to occur as stress-response mechanism (Chalker-Scott, 1999). Such stress-related anthocyanin pigmentation was indeed observed on regenerating tissue but could clearly be differentiated from the stronger, darker-coloured, discrete sectors induced from transgenic *ANT1* overexpression with LGJJ139 and from KIs with vector LGJJ52 (Figure 3.13).

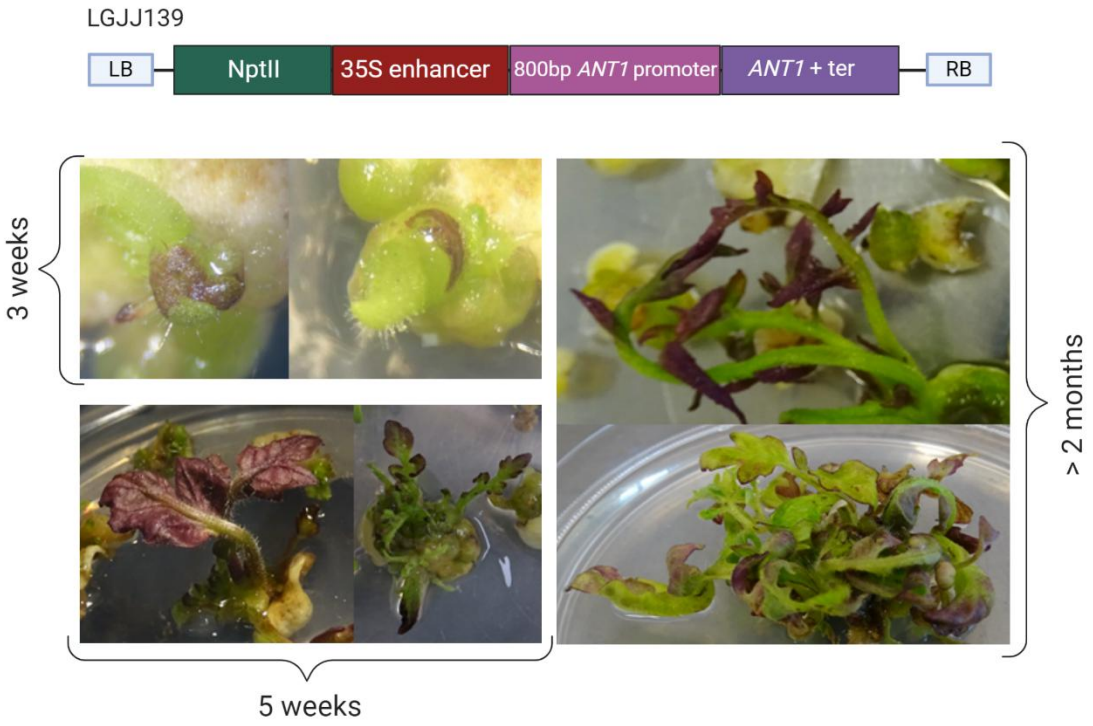


Figure 3.12 Tracking the phenotype generated by 35S enhancer upregulation of ANT1 when ~800 bp upstream of the CDS on a T-DNA. Heinz explants were photographed three weeks, five weeks and over two months after transformation.

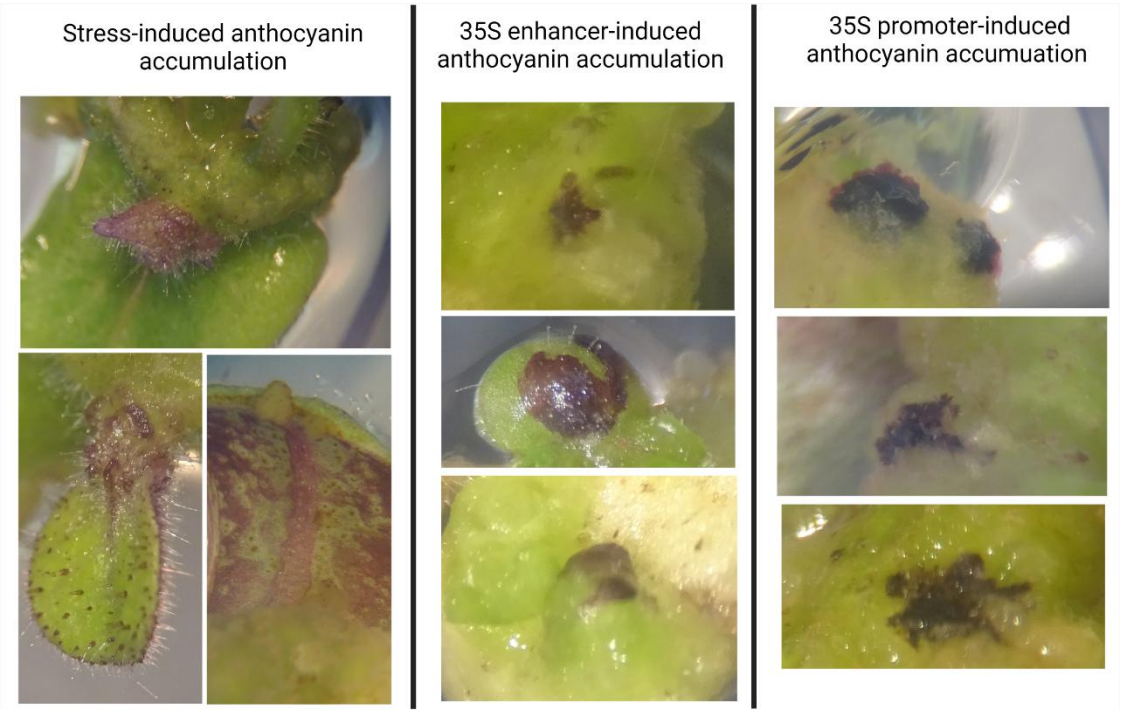


Figure 3.13 Differentiating between stress-induced anthocyanin accumulation and engineered *ANT1* overexpression purple colouration. Tissue culture-associated stress was observed as diffuse, lighter purple colouration in regenerating tissues. 35S enhancer (using vector LGJJ139)

or 35S promoter-induced (using vector LGJJ52) anthocyanin accumulation on the other hand produced discrete, darker sectors that were used to record and measure KI rates.

3.10 The strain of the replicon affects copy number in plant cells

Two strains (mild and acute) of BeYDV exist (Halley-Stott et al., 2007; Liu et al., 1997). Gene targeting experiments using a replicon system were based on the mild allele (Čermák et al., 2015; Dahan-Meir et al., 2018; Vu et al., 2020b; Hahn et al., 2018a). I investigated the difference in replicon copy number when based on a mild or acute strain of the BeYDV by qPCR after a *Nicotiana benthamiana* leaf infiltration assay. Copy number was assessed at two- and five-days post infiltration (dpi) from three biological replicates. Despite carrying an identical cargo load (5.2 kb), the acute strain (LGJJ52) replicated to produce ~ six times more replicons than the mild strain (LGJJ112, LGJJ113) (Figure 3.14). At five dpi (blue bars, Figure 3.14), the mild replicon could reach between 3,000 to 8,000 copies, whereas the acute replicon could rise to 50,000. As expected, a larger cargo load reduced replication. The replicon copy number could be 1.6 times higher at five dpi when the cargo size is decreased from 5.7 kb (LGJJ52) to 2.5 kb (vector ZCM-CTD), although the large error bars indicate important variability between samples.

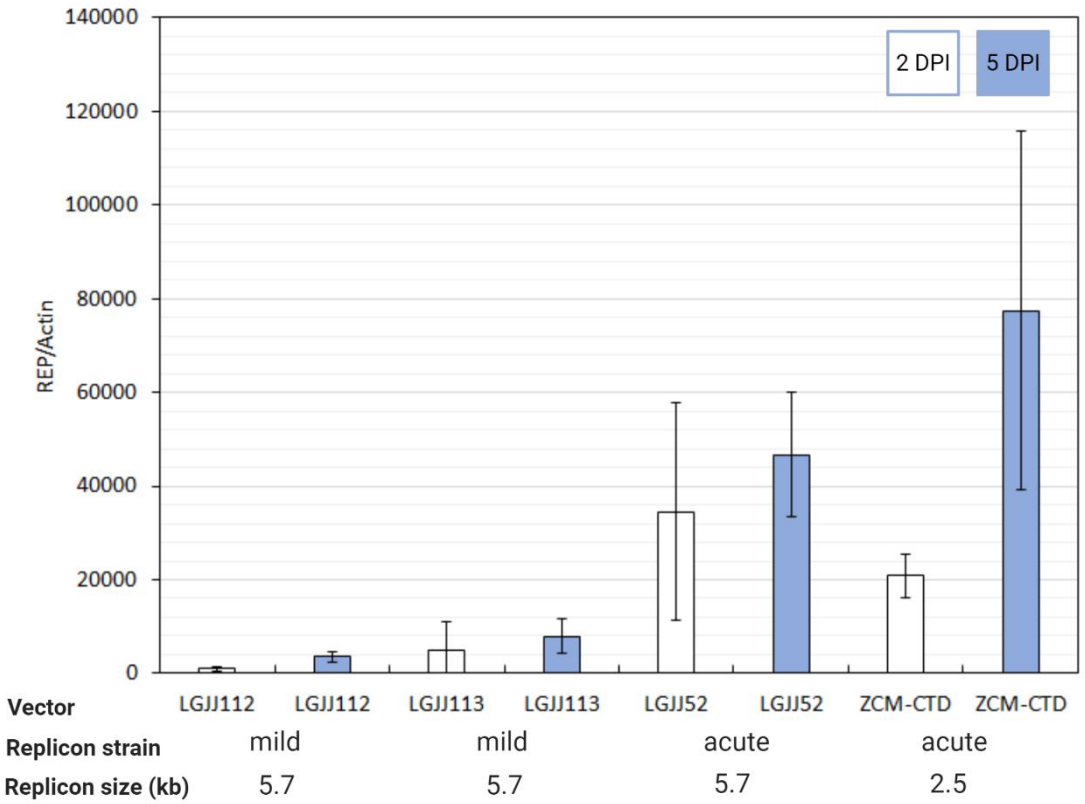


Figure 3.14 Relative quantification of the replicon copy number over time by qPCR from different geminivirus BeYDV strains with varying cargo sizes. Vectors LGJJ112 and LGJJ113 contains a mild strain of the geminivirus whereas LGJJ52 and ZCM-CTD are based on the acute strain. White bars: two days post infiltration (dpi) values; blue bars: five dpi values. Circularised replicon copy number is normalised to *NbActin* copy number. Error bars show the standard deviation of the mean for three biological replicates.

3.11 Testing different geminivirus strains to promote gene targeting of a 35S enhancer in the region of -800 bp of ANT1 CDS.

I next tested the ability of mild and acute BeYDV strains to promote KI by replicating the donor DNA to high copy number. I attempted to knock-in a 35S enhancer in the promoter of *ANT1* either at -723 bp (LGJJ96, LGJJ112: Cas9 + gRNA_1) or at -800 bp (LGJJ97, LGJJ113: Cas9 + gRNA_2) when the donor template is contained on replicon based on the acute strain (LGJJ96 and LGJJ97) or on the mild one (LGJJ112, LGJJ113). All explants were subjected to a three-day period at 28 °C as this seemed to increase KI efficiency in my previous

experiment. No putative KIs were observed using LGJJ96, LGJJ97, LGJJ112 and LGJJ113 out of the 186, 265, 214 and 188 explants transformed, respectively (Table 3.2, Figure 3.15). However, 13 explants from transformation with the positive control vector LGJJ139 showed a visible purple pigmentation (4.5%). Considering this phenotype relied on standard T-DNA insertion, this rate was unexpectedly low. The 35S promoter counterpart of this approach (pTC147, Table 3.1) gave purple sectors in 85% of the transformed explants. Explants transformed with the negative control (LGJJ58) did not show any purple sectors. These data suggest that the 35S enhancer KI strategy is unsuitable for efficient and reliable detection of KIs, which will be discussed in section 3.13.

Table 3.3.3 Gene targeting efficiencies of a 35S enhancer in the *ANT1* promoter of Heinz explants when using two different replicon systems either based on a mild or acute strain of a BeYDV. Two different gRNAs are also used, one targeting at -723 bp (LGJJ96, LGJJ112) and one targeting at -800 bp (LGJJ97, LGJJ113). The control vector LGJJ139 contains on a T-DNA an *ANT1* expression cassette with a 35S enhancer for transcriptional regulation and the negative control LGJJ58 contains a dCas9.

Strain replicon	gRNA target	Vector ID	Explants used for transformation	Explants with a purple sector	Purple calli (%)
Acute	-723 bp	LGJJ96	184	0	0
	-800 bp	LGJJ97	265	0	0
Mild	-723 bp	LGJJ112	214	0	0
	-800 bp	LGJJ113	188	0	0
N/A	n/a	LGJJ139	288	13	4.5
	-60 bp	LGJJ58	189	0	0

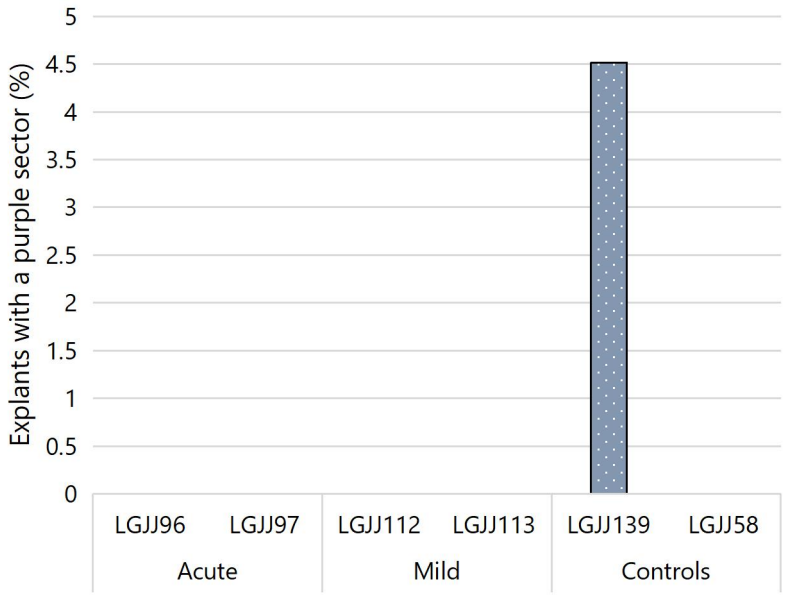


Figure 3.15 *K*nock-in efficiency of a 35S enhancer at -800 bp or -723 bp upstream of ANT1 CDS in the tomato cultivar Heinz using different strains of the Geminivirus. Vectors LGJJ96 and LGJJ97 contain the acute form of the Geminivirus whereas LGJJ112 and LGJJ113 contain a mild strain. Vector LGJJ139 is a positive control, expressing ANT1 under 35S enhancer regulation on a T-DNA and LGJJ58 the negative control, with a dCas9.

3.12 Testing the mutagenesis efficiency of the gRNAs targeting the distal region of the *ANT1* promoter

Despite having used two different gRNAs to account for gRNA efficiency variability and having tested two types of replicon systems (mild or acute), no knock-ins were detected in Heinz explants. To test whether this was caused by poorly performing gRNAs, the efficiency of gRNAs carried on LGJJ112 (gRNA_-723) and LGJJ113 (gRNA_-800) was assessed *in vivo*. LGJJ112 and LGJJ113 were transformed into tomato explants alongside the positive control LGJJ52 (gRNA_-60) (Figure 3.16), as this gRNA generated sufficient activity to achieve 10% KI efficiency in Cermak et al.'s paper (2015). As the previous attempts to regenerate Heinz explants were unsuccessful (data not shown), the cultivar Moneymaker was used for this assay. Regenerating shoots were taken to the rooting stage and all rooted shoots with a well-developed root system (Figure 3.16) were selected for the gRNA mutagenesis assay. Non-rooting shoots in kanamycin rooting medium were considered as non-transformed and

therefore T-DNA free and were subsequently excluded from follow-up analysis. After gDNA extraction from rooted shoots, PCRs were performed to screen for Cas9 presence in the samples and to amplify the DNA sequence spanning 300 bp from the target site of the gRNA. The PCR products of the latter were sequenced by Sanger sequencing method. Analysis of the Sanger sequencing data was performed using two online platforms which perform sequence trace decomposition: Tracking of Indels by DEcomposition (Tide, Brinkman et al., 2014) and Inference of CRISPR Edits (ICE, Hsiau et al., 2019). Algorithms running these two softwares decompose the Sanger traces at the Cas9 genomic target from edited cell populations and compare those traces to that of wild-type cell populations to establish the types of indels and their frequency in the edited cell population. Figure 3.17, panel a, displays the PCR results for Cas9 presence in each sample and whether mutations were observed at the gRNA target site in these samples. A "tick" symbol indicates the mutated samples detected in TIDE while a circled tick symbol indicates mutated samples detected by both TIDE and ICE softwares. As TIDE is able to detect the presence of edits occurring in only 1 to 2% of cells and ICE able to detect edits present in a minimum of 5% of the cells, it was possible to identify an additional couple of mutated samples in each group compared by using TIDE additionally to ICE in the analysis. Panel b of Figure 3.17 shows an example of sequence trace decomposition between sequence reads from a wild-type control and that of an edited sample. Presence of edited, mutated sequences around the cut site creates multiple peaks in the chromatogram of the edited sample. Sequencing trace decompositions for samples detected as mutated by the ICE algorithm have been compiled and are available in the Appendix A.

In summary, out of 24 samples, gRNA_-60 carried on LGJJ52 produced eight samples with a detectable edit (33%), gRNA_-723 carried on LGJJ112 produced five (20%) and gRNA_-800 on LGJJ113 generated ten (41%) (Figure 3.19). A chi-square test was performed on the counts of edited vs non-edited samples so see if the difference in observed proportions between the groups compared was statically significant, but the results indicated otherwise χ^2 (2, N = 72) = 2.428, p = 0.297. These results indicate that all gRNAs are active *in vivo* and generating mutations at approximately the same rate. This leads to the conclusion that the lack of detection of purple sectors in the 35S enhancer KI strategy was not caused by poor gRNA design which would compromise Cas9-mediated genomic target cleavage, and more likely because this assay is prone to false negatives.

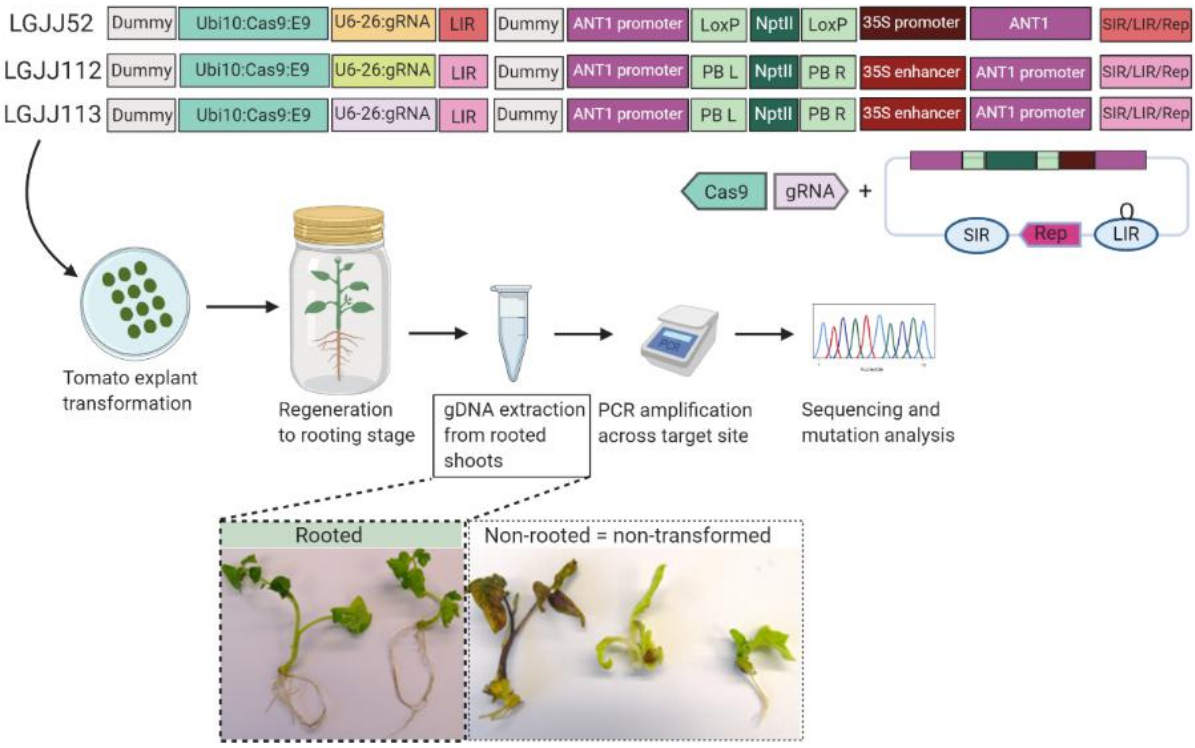


Figure 3.16 Diagram depicting the vectors and workflow used to assess gRNA efficiency *in vivo*. LGJJ52 acts as the control, carrying gRNA targeting -60 bp of ANT1 used by Cermak et al. (2015). Rates of mutagenesis obtained with gRNA_-723 on LGJJ112 and gRNA_-800 on LGJJ113 can be compared to the control.

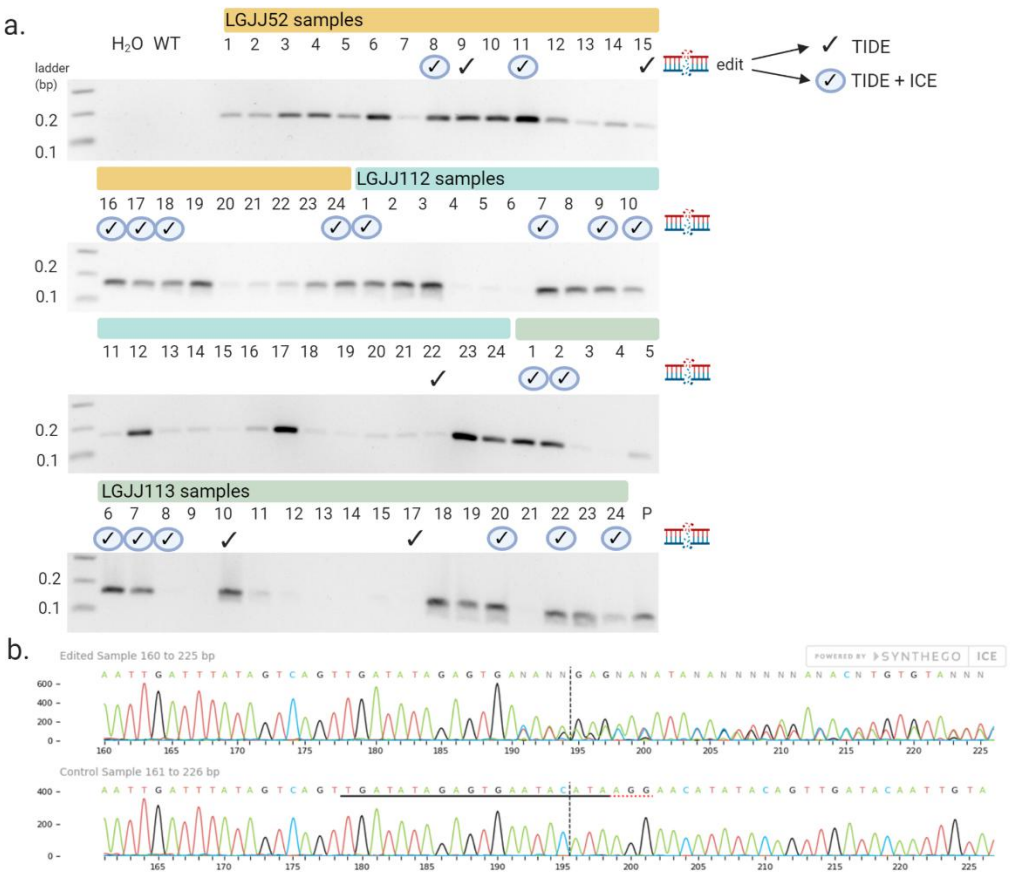


Figure 3.17. Mutation analysis at the *ANT1* gRNA target site. A. PCR results for Cas9 presence in samples transformed either with LGJJ52, LGJJ112 or LGJJ113. Controls used are no template (H2O), WT gDNA (WT) and plasmid LGJJ112 as positive control (P). Samples with a mutation at the target site detected by TIDE or by TIDE and ICE is represented using a tick symbol or a circled tick symbol. B. Representative example of ICE sequencing trace decomposition between WT sample (bottom chromatogram) and edited sample (top chromatogram) to infer occurrence of mutations. Underlined sequence represents the gRNA target sequence (black) and the PAM (red) and the vertical dotted line the expected cut site.

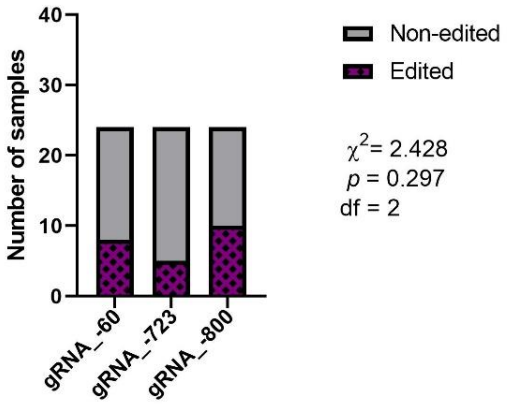


Figure 3.18 Bar chart showing the proportion of samples that harboured a mutation at the gRNA target site (Edited) or displayed WT sequence (Non-edited) in each group of transformed shoots. LGJJ52 contains gRNA_-60, LGJJ112 contains gRNA_-723, LGJJ113 contains gRNA_-800. A Chi-square test was performed to assess if the difference in proportions of edited samples between vectors was statistically significant. Statistical analysis done with the software GraphPad Prism 8.4.3.

3.13 Discussion

Inserting novel DNA material into plants genomes has been a long sought-after goal of genetic engineering due to its powerful potential for crop trait improvements. Nevertheless, the success rates of gene targeting experiments are variable and generally low, with variability within species, between species, and between experimental set ups too. Therefore, efforts are needed to develop methodologies that boost gene targeting efficiency for a given species.

3.13.1 Unexpected low rates of 35S promoter at the *ANT1* locus KIs: key differences between two similar experiments

Despite using a similar set up to Cermak et al (2015), my constructs reached knock-in rates four times lower. Although the similarities between the vectors used in our experiments are important (same Cas9 allele, same gRNA, same homology regions), there are some differences which may have caused this discrepancy. For instance, in the Cermak et al experiment (2015), they used a different strain of the geminivirus to create the replicon, i.e. a mild strain (Halley-Stott et al., 2007), whereas we used an acute strain of the geminivirus. We chose the acute strain as it was previously shown to accumulate to higher copy number inside the nucleus compared to the mild strain (Halley-Stott et al., 2007; present study). There are two purported advantages of using geminivirus-containing vectors in KI experiments. The first is that due to the Rep-mediated rolling circle replication in the nucleus, the replicon produces thousands of

copies of the insert flanked by homology region, which theoretically increases the likelihood of a copy of the insert being in proximity to the Cas9-mediated DSB to be used as a repair template. Because increased physical distance between two homologous DNA molecules negatively correlates with homologous recombination success (Lee et al., 2016), the opportunity to increase the insert copy number, and thus availability, using a replicon certainly has a strong appeal. Secondly, geminiviruses rely on their host cell machinery for DNA replication and transcription, and these host replicative factors are mainly available in proliferating cells. As KIs via HR are thought to occur mostly in late S-phase, the timing of activity of geminiviruses seemed conducive to HR-mediated repair. Indeed, the geminiviral RepA protein was shown to interact with a maize Retinoblastoma-like protein factor, which drives host cells into S-phase, or into an S-phase like state, where replication factors needed by the virus are available (Liu et al., 1999). Moreover, the Rep proteins have been shown to associate with key HR components in plants: Rep, from the mung bean yellow mosaic India virus, with RAD54 in Arabidopsis (Richter et al., 2014) and Rep from the tomato yellow leaf curl virus to a Nucleolin-like protein in tomato (Maio et al., 2020). These data indicate that Rep in these two species of the *Geminiviridae* family potentially increase recombination by hijacking the HR pathway of the host. Enhanced recombination helps generating recombinant geminiviruses in the field, generating genetic diversity for adaptation and sustain population fitness (van der Walt et al., 2009). Although the use of a geminivirus replicon was reported to contribute to significant improvement in KI efficiency in tomato (Čermák et al., 2015; Dahan-Meir et al., 2018; Vu et al., 2020b), *N. benthamiana* (Baltes et al., 2014), potato (Butler et al., 2016a), wheat (Gil-Humanes et al., 2017), rice (Wang et al., 2017), cassava (Hummel et al., 2018). No improvement was seen in Arabidopsis (Hahn et al., 2018b; De Pater et al., 2018). This could be due to the arrangement of expression cassettes on the vector. As the LIR possess bidirectional promoter activity, neighbouring expression cassettes i.e. U6:gRNA:ter might be targeted by the gene silencing machinery from the formation of dsRNA (Atkins and Voytas, 2020; Hahn et al., 2018b). Despite the promising advantages provided by the replicon strategy, it is possible that the presence of the replicon causes different side effects adverse to successful knock-ins, especially at high copy number, as is the case in my study with the use of the acute replicon. Indeed, results from this first set of experiments are in contradiction with Cermak et al. (2015) and Vu et al. (2020), who found that using the replicon increased KI efficiency of a 35S promoter at the *ANT1* locus by an order of magnitude in the tomato cultivar Micro-tom.

Likewise, in another gene targeting experiment in the tomato cultivar Micro-tom (Dahan-Meir, 2018) knock-in efficiencies reached 25% while using a mild BeYDV replicon, which was modified so that the Rep protein is expressed on the T-DNA, outside the replicon, under the expression of a 35S promoter instead of the LIR. Such a set up would lead to even higher numbers of Rep/RepA inside the nucleus, which if causing significant pleiotropic effects on HR success would have hampered gene targeting in this experiment. However, judging by the high rates of KIs (25% edited in T0), it seems unlikely that it is the case in this context. The limitation of our data set is that we have not attempted knocking in a 35S promoter using a mild replicon, therefore we cannot strictly compare our results with that of Cermak's or to other published gene targeting studies as all have based their replicon on the mild strain of BeYDV. Therefore, it is reasonable to assume that the low rates of KI observed here are caused, at least partially, by an over-performing BeYDV replicon. To be noted, the occurrence of geminiviral replicons in the form of supercoiled DNA (Suárez-López and Gutiérrez, 1997) in the nucleus may reduce the availability of the donor template when contained on the replicon to be used for DSB repair. It was shown that linear DNA templates were more prone to be used as an HR template compared to supercoiled plasmids in rats (Remy et al., 2014). It may therefore be beneficial to linearise the donor template by excising it from the replicon in future experiments using a geminiviral replicon strategy.

Another important difference is the 35S promoter that is knocked-in at -60 bp of the *ANT1* CDS. The 35S promoter used in my experiment contains a tobacco mosaic virus omega leader sequence fused at the 3′ end (35S promoter-Ω) which acts a protein translation enhancer (Gallie, 2002), whereas Cermak et al.'s (2015) 35S promoter did not. *ANT1* overexpression accompanied by anthocyanin accumulation in cells can incur a significant fitness cost (Alfenito et al., 1998; Chawla et al., 1999; Kortstee et al., 2011) (Liobikas et al., 2016). Constitutive synthesis of anthocyanin compounds likely imposes metabolic drain on those cells as a lot of energy in the form of ATP is required by several enzyme for catalysis to complete the general phenylpropanoid pathway. Moreover, the ATP-dependent up-taking of cytosol-synthesised anthocyanins into the vacuole by specialised transmembrane transporters (Behrens et al., 2019) adds another energy drain. Reduces cellular energy levels could be slowing down tissue growth. Besides, it may be that strongly pigmented tissues experience reduced photosynthesis level because the large quantity of anthocyanin pigments absorb a significant amount of light

(personal communication with Prof. Cathie Martin, 2021), likely interfering with the normal photosynthetic activity of those tissues.

By inserting a 35S promoter with an omega leader, the anthocyanin production level in edited cells would likely be higher than that of Cermak's. Conceivably, a proportion of successfully edited cells may rapidly die from the cytotoxic effects of high anthocyanin content before being able to progress through several rounds of cell division to create a group of cells with high pigmentation that can be detected and scored. If this is the case, we can assume that the frequencies we observe are in fact under-estimations of the real knock-in rates as a proportion of the events may lead to cell death before they can be recorded, and the occurrence of such 'lethal events' is difficult to measure. The observation of multiple purple sectors with stalled growth and becoming overgrown by neighbouring green tissue certainly supports the claim that anthocyanin accumulation in edited cells has a fitness cost and this will be discussed in a following section.

Cermak et al.'s (2015) and Dahan-Meir et al.'s (2018) experiments were performed in Micro-Tom. In the present study, we favoured using cultivars (Moneymaker and Heinz) with a higher commercial value, yet reasonably easy to transform and cultivate *in vitro*, so that beneficial outcomes from the research would be simple to transfer to breeders. Since the KI rates obtained in Moneymaker were lower (Figure 3.5) than that observed in Micro-Tom by Cermak et al. (2015) and Dahan-Meir et al. (2018), we set out to investigate if a genotype-specific KI performance existed. Because no major differences in KI efficiencies were seen when the same set of vectors was transformed into Moneymaker and Micro-Tom (Figure 3.5), we concluded that the low rates seen in Moneymaker are not genotype-dependent, but rather linked to the experimental set up.

3.13.2 Elevated temperature and DSB at the genomic target are efficient at promoting knock-ins

Here, several variables that may influence the rates of knock-ins were tested in parallel. Rates of knock-ins were compared between delivering a single DSB at the genomic target (one DSB approach) or delivering three SSBs, one SSB at the target site and one at both extremities of the homology regions flanking the insert on the vector (three SSBs approach) were compared. NHEJ is the most active DNA repair pathway in somatic, differentiated cells and competes

against HR machinery to repair DSBs (Knoll et al., 2014; Fukushima et al., 2001). To avoid this competition, the delivery of nickase-mediated SSB at the genomic target site was shown to largely increase the proportion of HR-mediated KI events in human cells lines (Chen et al., 2017) because SSBs are repaired by a pathway related to HR (Davis and Maizels, 2014; Bothmer et al., 2017a) and nicks are not a substrate for NHEJ components (Vriend et al., 2016), at least in mammalian cells. Although a similar set up to Chen et al. (2017) was replicated in my vectors LGJJ53 and LGJJ55 to implement the three SSBs strategy, low rates of KI were detected in Moneymaker explants (0% or 0.4%, respectively). These results are in line with another study in Arabidopsis where the three SSBs approach did not produce any knock-in when the nickase was under the *Pc*Ubi4 constitutive promoter (Wolter et al., 2018). Interestingly, the rates of SSB-induced KI increased to 13% positive T1 lines when the nickase was under an egg cell specific promoter (*At*EC1.1). The difference in efficiency observed between the plant and human system might be due to differences in the DNA repair mechanisms between plant cells and human carcinoma HeLa cells and kidney embryonic stem cells used in Chen et al. (2017) Overall, results from my experiments and those of Wolter et al. (2018) show that the one DSB approach performed better than the three SSBs to induce KI, suggesting that the SSB-induced HR may not be such a common repair pathway in plants. Besides, it is possible that not every cell exposed to the three SSB strategy will have indeed received three nicks. Failure to generate three nicks in any given cell may lower the efficiency of the approach at generating KIs and could explain the low rates observed here but this is difficult to evaluated without data on the number of nicks produced in cells.

Additionally, I assessed whether the size and content of the replicon had an impact on KI success. There might be a trade-off between the copy number of specific elements replicated via replicon, and the total copy number of the replicon itself. For instance, adding the expression cassettes for the nuclease or nickase and the gRNA on replicon would increase the copy number of these reagents which might improve the generation of DSB at the target site and hence improve the chances of knock-ins. On the other hand, increasing the total size of the replicon to 12 kb (when the Cas9 and gRNA are included) might hamper the rolling circle replication and therefore reduce the number of copies of the homology template and of the donor, which might be disadvantageous for KI success. Indeed, we observed higher rates of knock-ins when the Cas9 and gRNA genes were placed outside of the replicon and therefore

not replicated as part of the replicon, suggesting that a lower copy number of these genes is not a limiting factor for KI success in our set up.

Finally, an external, environmental condition was tested for its ability to enhance rates of knock-ins. Elevated temperatures have been shown to increase Cas9 enzymatic activity and to increase the efficiency of CRISPR edits (Le Blanc et al., 2017). Higher temperatures also have the potential to increase somatic homologous recombination activity (Rahavi and Kovalchuk, 2013). By implementing a three-day heat treatment at 28 ˚C to the tomato explants after the *Agrobacterium* co-cultivation step, I observed that the proportion of explants developing at least one purple sector was higher for samples that received a heat treatment compared to the control samples which remained at 22 ˚C. This trend suggest that knock-in experiments performed in the future, or elsewhere, would benefit from implementing a period of elevated temperature early on during tissue culture to boost KI efficiency. A limitation of my dataset is that the efficiencies are assessed based on one round of transformation only. Repeating the experiment would increase the confidence in the trends that are observed. Nevertheless, these findings agree with that of Vu et al. (2020). In their experiment, KI efficiencies increased two-fold when explants received a 31 ˚C heat treatment, compared to 18 ˚C. However, not all plant species can withstand elevated temperatures for several days without suffering significant stress (personal communication with Professor Harwood, W.), hence this strategy might only be applicable to some plant species. The engineering of nucleases with improved mutagenic efficiency at lower temperature is bringing an important tool to the gene targeting field (Schindele and Puchta, 2019).

Despite the low rates of KI, one KI arising from LGJJ52 was molecularly characterised. The sequence alignments covering the right and left junctions of the insert perfectly matched the reference sequence of a predicted KI event, suggesting that the DSB was indeed repaired by homologous recombination using the delivered donor template.

3.13.3 Regeneration of purple sectors is challenging

Two clear observations were noticed when phenotyping the explants six weeks after transformation: (1) the purple sectors, comprising edited cells over-expressing *ANT1*, were frequently overgrown by neighbouring green tissues which were regenerating faster (and will be addressed in chapter 4), and (2) purple sectors that grew large enough to be excised from

their originating callus often exhibited stalled growth at the callus stage, shoot generation stage, or after having developed to a certain size and never produced roots when placed in rooting medium. The difficulty in regenerating edited plant material can likely be attributed both to the presence of the geminivirus replicon and to the substantial accumulation of anthocyanins in the cells. Albeit using a deconstructed geminivirus to build the plant transformation vectors (the movement protein and the coat protein genes have been removed), the parts that are kept (LIR, SIR, Rep/RepA) could be causing pleiotropic effects on callus development and metabolism. Based on a study which analysed the immune response associated with the use of geminivirus (BeYDV) expression systems in *Nicotiana benthamiana*, the Rep/RepA proteins were demonstrated as immune elicitors triggering a hypersensitive response (Diamos and Mason, 2019). Geminiviruses can also be elicitors and targets of the antiviral RNA interference response (Vanitharani et al., 2005). Taken together, these data suggest that the replicon could trigger an immune response in the regenerating tomato cells which may slow down growth. Similar observations of non-regenerating callus or plants have been published: despite an increased rate of knock-ins due to the replicon, edited cells were unable to regenerate into whole plants in some cases (Gil-Humanes et al., 2017) and regenerated whole plants exhibited loss of vigour and a stunted phenotype (Hummel et al., 2018). Possible explanations are that cells containing replicons have a compromised cell division mechanism preventing cells from propagating (Atkins and Voytas, 2020), or that the Rep protein triggers pleiotropic effects, by its presence or by inducing the plant immune response (Hummel et al., 2018). In line with this, the Rep protein of the *Tomato yellow leaf curl virus,* also a member of the *Geminiviridae* family, was shown to interact with at least 54 tomato proteins upon infection (Maio et al., 2020), suggesting that the use of a replicon is likely to cause unknown pleiotropic effects, maybe even more so in my experiments as an acute strain of the geminivirus was used.

Additionally, anthocyanin accumulation to high levels in the cell might be causing the stunted regeneration of edited tissues, as has been observed in previous studies. Anthocyanin accumulation can retard growth and has caused developmental termination in maize tissue overexpressing anthocyanin production transcription factors (Chawla et al., 1999), and the overexpression of an anthocyanin-regulating MYB transcription factor appeared to be lethal for regenerating shoots from apple (Espley et al., 2007) and strawberry explants (Zhang et al., 2020b), but not for potato shoots (Kortstee et al., 2011). Nevertheless, others have found that

overexpression of anthocyanins in potato led to dwarfing and production of mini purple tubers (Rommens et al., 2008). Here, the 35S promoter-Ω inserted upstream of *ANT1* - a key regulator R2R3-MYB transcription factor for anthocyanin production - contains an omega leader sequence which acts at translational enhancer. The combination of the constitutive expression of this anthocyanin activator and its enhanced translation generates quantities of anthocyanins in the cells that are challenging to cope with and cells may incur a strong fitness cost. To alleviate this fitness cost, a different KI strategy has been attempted using a 35S enhancer, which will be discussed in the section below.

3.13.4 The 35S enhancer strategy did enable detection of knock-ins.

Production of anthocyanins as a phenotypic marker for successful knock-ins offers a very robust and rapid screen to measure to rates of desirable events and to isolate the edited cells to be regenerated into plants. However, high accumulation of anthocyanins resulting from the insertion of a 35S promoter-Ω is suspected to hamper cellular regeneration. To mitigate this effect, plant transformation constructs were redesigned to insert a 2 × 35S enhancer in the region of -800 bp in the promoter of *ANT1*. As the presence of the acute form of the replicon was suspected to cause important pleiotropic effects in the 35S promoter KI experiment, the transformation constructs were also redesigned to produce a replicon either based on the mild strain or on the acute strain of the BeYDV to compare KI efficiency between these two replicon systems. The copy number assay by qPCR confirmed that the relative replicon copy number can be up to six times higher when based on the acute BeYDV in *N. benthamiana* leaves. To account for gRNA efficiency variability, two different gRNAs were used individually on each vector.

Although the T-DNA-based positive control experiment to test the phenotype arising from *ANT1* upregulation by a 2 × 35S enhancer (i.e. vector LGJJ139) confirmed that a detectable – albeit weaker than with the 35S promoter - purple pigmentation could occur, no KI were detected with this strategy, regardless of the gRNA or strain of the replicon used, from over 800 Heinz explants transformed. The fact that no KI were observed could be explained by several factors. Firstly, it might be that due to the weaker phenotype, the assay is less reliable. For instance, in the control experiment, only 4.5% of the explants used for transformation exhibited an elevated purple pigmentation. Because this experiment relied on successful T-

DNA transmission to the nucleus to create a phenotype, rates observed are unexpectedly low. Regenerating shoots were not taken to the rooting stage so information on transformation efficiency based on rooting for this experiment is not available. Nevertheless, tomato transformation efficiency is usually between 15% to 30%, therefore, 15%-30% of explants would be expected to show a purple phenotype, instead of 4.5%. This discrepancy suggests a reduced robustness and reliability for this transformation marker. Interestingly, it was shown in Arabidopsis that the sequence 4 × 35S enhancer could be systematically methylated if more than one T-DNA was inserted in the cell (Chalfun-Junior et al., 2003), which largely reduced the recovery of phenotypic mutants in activation tagging experiments due to silencing (Weigel et al., 2000). A similar phenomenon might be occurring to some extent in my experiments, both for the T-DNA-based approach and for the KI approach which relies on a replicon and therefore generates thousands of copies of the 35S enhancer.

Another explanation for the lack of KI detected at -800 bp of *ANT1* might be attributed to the nature of the target site. Testing the gRNAs efficiencies *in vivo* showed that both gRNAs used to target Cas9 to the -800 bp region of *ANT1* were as mutagenic as the gRNA used for 35S promoter KI at -60 bp, indicating that genomic target cleavage by Cas9 is not the limiting factor for KIs to occur. Nevertheless, experiments in human cell lines have shown that chromatin marks and chromatin configuration are key determinants of HR or NHEJ repair (Daugaard et al., 2012; Aymard et al., 2014). DSBs occurring in actively transcribed regions, i.e. with abundant active chromatin marks, H3K36me3, and Polymerase II transcripts, are repaired by HR as this specific chromatin mark is a binding site that recruits HR factors (Aymard et al., 2014). It is possible that the genomic target site for 35S promoter KI at -60 bp of *ANT1* still retains the actively transcribed, HR-permissive chromatin features whereas the chromatin landscape at -800 bp might be different and promote NHEJ-type repairs of DSBs, which might have – at least partially – contributed to no detectable KIs. Transcription analysis and chromatin profiling for H3K36me3 at the *ANT1* locus (covering proximal and distal promoter regions) in cotyledons and callus cell types would enable evaluation of this. To date, no gene targeting experiment has taken into consideration the chromatin configuration of their target sites despite strong evidence that specific histone marks and chromatin profile in general are important determinants of repair outcome (Aymard et al., 2014; Daugaard et al., 2012; Tang et al., 2013; Endo et al., 2006). It would be interesting for future KI experiments to address this unexplored aspect of gene targeting biology which might help narrowing down

suitable ‘landing pads’ in plant genomes with elevated HR activity to further increase the rates of KIs.

Chapter 4. Addressing the challenge of reducing the growth of non-transformed tissue during tomato regeneration

4.1 Introduction

Plant transformation and *in vitro* regeneration of transgenic tissue carrying desired modifications are pillars of plant biotechnology. Genetic transformation protocols usually rely on a selectable marker gene in order to select for transgenic material over non-transformed background tissues. One of the most widely used selectable marker genes is the *Neomycin phosphotransferase* type II (*NptII*), isolated from the *E. coli* transposon Tn5 (Beck et al., 1982). The enzyme NPTII confers resistance to aminoglycoside antibiotics, such as kanamycin, by catalysing the ATP-dependent phosphorylation of the antibiotic molecule (Curtis et al., 1995). Once phosphorylated, the antibiotic can no longer interact with the 30S subunit of the mitochondrial and chloroplast ribosomes, enabling normal protein synthesis by the plant and thereby allowing cell growth despite kanamycin presence (Wilmink and Dons, 1993).

In theory, only cells having received a T-DNA from where *NptII* is expressed should thrive on kanamycin-containing medium, as only these can detoxify the antibiotic. Nevertheless, the occurrence of shoot "escapes" is common. Escapes are defined as background, non-transformed tissues and shoots that can grow despite antibiotic selection (Dan et al., 2009). Shoot escapes are revealed at the rooting stage of tissue culture, as they are unable to grow a fully developed root system in the presence of the antibiotic. Escapes are widespread in the regeneration process and have been reported in several plant species: apple (James et al., 1989), tomato (Dan et al., 2009), flax (Jordan and McHughen, 1988; McHughen and Jordan, 1989), Arabidopsis (Czako and Marton, 1994), *Petunia hybrida* (Renckens et al., 1992), cowpea

(Solleti et al., 2008), citrus (Cervera et al., 1998; Peña et al., 2004), *Medicago sativa* (Rosellini et al., 2007), cassava (Zhang et al., 2000), plum (Padilla et al., 2003) and more. The frequency of escapes in these species ranges between 40% and 90%. This issue is particularly prevalent with the *NptII*/Kanamycin selection system. Despite being a commonly observed phenomenon in plant tissue culture, the reason for escape shoots formation on selective medium is not well studied but several hypotheses for their occurrence have been expressed. Padilla et al. (Padilla et al., 2003) reported that the formation of escapes was linked to the timing of the application of the selecting antibiotic to the medium after co-cultivation. Only 10% of the regenerating plum shoots were escapes when kanamycin was added to the medium right after the co-cultivation step with *Agrobacterium* instead of two weeks after transformation. Additionally, adequate concentration of the selective agents should be applied to the regeneration medium in order to limit the regeneration of escapes (Li et al., 2013). Another study suggested that the presence of escapes could arise from kanamycin depletion in the vicinity of transformed cells which are able to detoxify the antibiotic. Notably, escapes were reduced by half when plants were supplied with the *hemL* mutant allele of the enzyme glutamate 1-semialdehyde aminotransferase (GSA-AT), involved in chlorophyll biosynthesis. *hmL* mutants are insensitive to the phytotoxic compound gabaculine which targets GSA-AT (Rosellini et al., 2007). Authors suggested that this is caused by the different mode of action of these selection systems: kanamycin is detoxified by NPTII enzymes, creating areas depleted with kanamycin in the callus mass where non-transformed cells can survive whereas the mutant GSA-AT insensitive to gabaculine enables transgenic cells to grow without depleting gabaculine, keeping its concentration constant across the plant material. Others have also attributed escapes incidence to a "shielding" effect created by transformed cells in the callus mass which protect non-transformed shoot initials from kanamycin, enabling these to form shoots despite selection (Cervera et al., 1998; McHughen and Jordan, 1989). Furthermore, the importance of early dedifferentiation of cells from the co-cultivation stage *via* auxin (2,4-D) to enter a competent state for stable transformation was highlighted (Peña et al., 2004).

Tomato transformation results in high numbers (80% or above) of escape shoots (Dan et al., 2009; Smoker, M., 2018, personal communication; Zsogon, A., 2018, personal communication, 15th November). Besides causing additional tissue culture work, the growth of background tissue was shown to overtake the growth of edited, anthocyanin-overexpressing cells in my previous work (chapter 3, section 3.7). Given the hypothesis that escapes may be

receiving protection from nearby transformed cells which detoxify the environment, it may also be possible that the resistance to kanamycin in non-transformed cells is provided by transiently expressed NPTII. The number of cells which receives a T-DNA varies between transformation experiments, leading to genetic chimerism of the regenerating material. The number of cells that will undergo stable insertion of the T-DNA after receiving the molecule also varies between experiments. Indeed, in a fraction of cells, the T-DNA will be present for a time during which transient expression takes places (Yoshioka et al., 1996) for a duration varying between one to ten days (Janssen and Gardner, 1990; Werr and Lörz, 1986; Yoshioka et al., 1996). Beyond ten days, T-DNAs have likely been degraded if not stably integrated in the genome (Janssen and Gardner, 1990). Any transient expression of *NptII* at the start of tissue regeneration could provide NPTII enzymes which detoxify kanamycin and enable the formation of shoot primordia from non-stably transformed cells for as long as transient T-DNAs are present in the nucleus.

Eliminating background tissue that does not carry stably integrated T-DNA would not only reduce the tissue culture work, it might also favour the growth of purple sectors by eliminating competition with this population of transiently transformed cells. Living cells control protein metabolic stability to achieve protein homeostasis, a protein quality control pathway which degrades misfolded or damaged proteins in the cell (Varshavsky, 2011). In addition, selective protein degradation is required for proteins whose concentration must vary in time with the cell cycle. This selective protein degradation control is part of the ubiquitin-proteasome pathway and is enabled by degradation signals, called degrons, with the N-degron system discovered in 1986 (Bachmair et al., 1986), which is a conserved pathway across kingdoms (Dissmeyer, 2019). These N-degrons are determined by the presence of degradation signals in the N-terminal residues of a protein which dictate the protein half-life *in vivo*. There are several key determinants to this pathway. First, destabilising amino acids must be present as degradation signal that can be recognised by E3 ubiquitin ligases (N-recognins). Secondly, an internal lysine should be present to be targeted for polyubiquitination (Johnson et al., 1990). Thirdly, configuration of the N-terminal of the protein must allow access to the destabilising amino acid for target recognition (Lévy et al., 1999). The N-terminus structure and conformation can protect these destabilisation amino acids from being recognised, forming a "cryptic" or "dormant" degron until endoprotease-mediated cleavage of some residues exposes the destabilisation residues (Dissmeyer, 2019). When the conditions for selective degradation

are met, a succession of post-translational modifications of the target protein triggers its degradation *via* the N degron pathway.

Artificial heat-activated N-degrons have been engineered to obtain conditional protein expression (Dohmen et al., 1994). This system enables reversible depletion or accumulation of a protein of interest by exposing it to permissive (stable protein) or restrictive temperature (leading to proteolysis). Building on previous experiments, researchers have adapted a degron-mediated temperature dependent selective protein degradation protocol to work in plants (Arabidopsis, tobacco) and animals (*Drosophila melanogaster*) (Faden et al., 2016). As pictured in Figure 4.1, the protein of interest N terminus is fused to the degron cassette which contains 3 elements: (1) a ubiquitin moiety which becomes cleaved co-translationally by deubiquitinating enzymes to expose the destabilising residue of the degron (8.5kDa, 76 amino acids), (2) the amino acid residue, phenylalanine, which is a potent destabilising residue in plants, followed by a short linker, (3) the mouse temperature sensitive (ts) variant of the DIHYDROFOLATE REDUCTASE ($DHFR^{TS}$ "K2" carrying the destabilising substitutions Thr39Ala and Glu173Asp, 22kDa, 222 amino acids) which triggers degradation at temperatures near 30°C and above. As temperature increases, misfolding of the DHFR moiety exposes previously hidden lysine residues. The destabilising residue Phe is likely detected by the N recognin, E3 ligase PROTEOLYSIS1 (PRT1), followed by the Ubiquitin-activating enzyme E1 and Ubiquitin-conjugating enzyme E2 that prime Ubiquitin for polyubiquitination of the DHFR moiety. This targets the degron cassette and its fused protein for degradation *via* the 26S proteasome. Their experiment confirmed the reversible accumulation of the fused protein. Using Degron-NptII in the tomato tissue culture protocol may help reduce the growth of escapes potentially thriving due to transiently expressed NPTII early on during the transformation. Synchronised degradation of the total population of NPTII proteins during the first three days post *Agro* co-cultivation at 28 C is expected from the degron-NPTII approach. Then, when the tissue culture temperature is shifted to 17 C five days after transformation, the number of transient T-DNA molecules is likely to be low. Therefore, when kanamycin selection is applied at +6 days post transformation and stability of the NPTII protein restored, cells able to survive should be those which are transformed. In a context without the degron-NptII, transient T-DNA numbers would gradually decrease over time but would continue supporting NPTII expression transiently for that duration, enabling detoxification of kanamycin in non-transformed cells that may divid and develop into shoot

primordia and give rise to escape shoots. Controlling growth of escape tissue may also be achieved by testing different antibiotic concentrations (Li et al., 2013) but this would not address the issue of the presence of transiently expressed NPTII when selection is applied, which is suspected to promote growth and cell-division of non-transformed cells that eventually grow into non-transformed shoots.

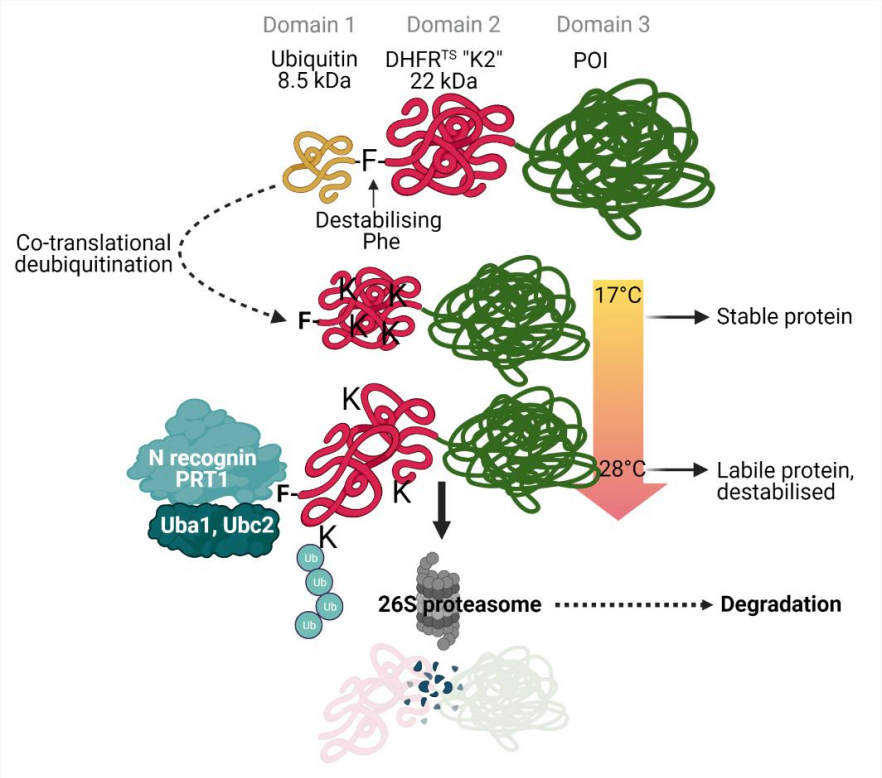


Figure 4.1 Pathway for selective protein degradation *via* the heat-activated N-degron approach. The protein fusion consists of three domains. Domain 1 is a ubiquitin moiety which is co-translationally cleaved by deubiquitinating enzymes. Domain 2 contains the mouse temperature sensitive DHFR^{TS} variant "K2" which contains the destabilising amino acid substitutions Thr39Ala and Glu173Asp, rendering the protein unstable at elevated temperature. Molecular modelling showed that these mutations increase intramolecular flexibility, with several lysine residues gaining in flexibility and accessibility to the ubiquitination machinery (Faden et al., 2016). At elevated temperature, misfolding of the DHFR moiety uncovers lysine residues. Domain 2 is preceded by the destabilising amino acid residue phenylalanine which becomes exposed after domain 1 cleavage. This residue is likely recognised by the N recognin PRT1, and polyubiquitination of the exposed lysine sites is carried out by Ubiquitin activating and conjugating enzymes (Uba1, Ubc2). As a result, the protein fusion including domain 3 (the protein of interest, POI) is labelled for degradation by

the 26S proteasome system. (Diagram adapted from Faden et al., 2016 and Dohmen et al., 1994.)

4.2 Aims

The aim of this chapter is to develop a method that reduces the growth of green, untransformed tissue during tomato *in vitro* regeneration to favour the growth of transformed cells, with a purple phenotype. I hypothesise here that non-transformed tissue can thrive on kanamycin selection medium due to the presence of transiently expressed NPTII. By using a temperature-dependent NPTII, degron-NPTII proteins can be degraded at elevated temperatures, after which only cells with a T-DNA insertion can survive by producing more NPTII.

4.3 Construction of golden gate vectors

To assess the efficacy of a degron-tagged NPTII in reducing the growth of non-transformed tissue, two visual assays were employed. In the first assay, tomato explants were transformed with a T-DNA which either contained a standard *NptII* (pTC147) or a *degron-NptII* (LGJJ151) in addition to an expression cassette containing 35S promoter:*ANT1,* with its full CDS and endogenous terminator (Figure 4.2). Here, strong anthocyanin accumulation in sectors is expected in callus cells that have received a T-DNA. In contrast, green tissue is suspected not to have received a T-DNA and can be identified as "escape" growth. In the second visual assay, the T-DNA still carries a standard *NptII* (LGJJ139) or *Degron-NptII* (LGJJ156) alongside an expression cassette for *ANT1* upregulation from a 35S enhancer at -800 bp upstream of its CDS. This second assay will assess the robustness of the 35S enhancer:*ANT1* phenotypic marker for successful transformation alongside testing the efficacy of the degron-NPTII strategy. The vectors LGJJ139, LGJJ156 and LGJJ151 have been cloned using the Golden Gate cloning method as previously described in the Material and Methods (section 2.2.3). The degron module ($DHFR^{TS}$ "K2") used in Faden et al. (2016) was kindly provided by the Dissmeyer lab and cloned as an *NptII* N-terminal fusion at the Sainsbury Laboratory. Construct pTC147 was purchased from Addgene and previously used in Cermak et al.

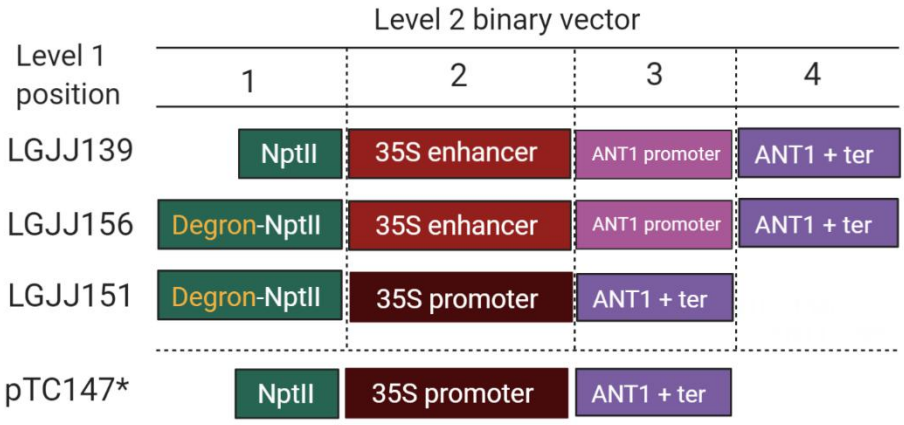


Figure 4.2 Binary vectors assembled using the Golden Gate cloning method to test the performance of the *Degron-NptII* method for reducing non-transformed tissue growth during tomato tissue culture.

4.4 Using an N-terminal degradation signal (degron) to reduce the growth of non-transformed tissues

The tomato explant transformation workflow was revised and adapted to use with the *degron-NptII* construct. After the *Agrobacterium* co-cultivation step, explants were placed at 28 ˚C for three days on non-selective regeneration medium during which transiently expressed degron-NPTII should be degraded. Explants were then moved to kanamycin containing medium and shifted to 17 ˚C, the temperature at which the degron-NPTII proteins are stable (Faden et al., 2016). During this two-week period at 17 ˚C, only cells with a stably integrated T-DNA were expected to survive. As transient NPTII proteins were likely degraded at 28 ˚C, tissues without a T-DNA should not have persisted beyond the co-cultivation step. Explants were then moved to 22 ˚C after the two-week period at 17 ˚C, as development and growth of tomato tissues were noticeably slowed at 17 ˚C.

The performance of the degron-NPTII for reducing the survival of non-transformed cells was visually assessed 6 weeks after transformation by comparing the growth ratio of green and purple tissue when using vectors LGJJ139 (Figure 4.3, 1st panel), and LGJJ156 (Figure 4.3, 2nd panel), LGJJ151 (Figure 4.3, 3rd panel) and pTC147 (Figure 4.3, 4th panel). Explants transformed with LGJJ151 displayed a clear reduction in green tissue growth compared to pTC147 (Figure 4.3, 3rd and 4th panels, top and bottom section), and 90% and 100% of the transformed explants developing purple sectors, respectively (Figure 4.4). Due to the more

variable and ambiguous phenotypes arising from the 35S enhancer:*ANT1* visual assay, the difference between explants transformed with LGJJ139 and LGJJ156 (Figure 4.3, 1st and 2nd panels, bottom section) with regards to green tissue growth is ambiguous. Purple sectors were less frequent as only 35% and 53% of the transformed explants displayed at least one purple sector (Figure 4.4). Nevertheless, low magnification images of these explants indicate an overall reduction in the growth of green tissue (Figure 4.3, top section). These data results suggest that the use of the *degron-NptII* helped reduce the growth of untransformed tissue at the callus stage.

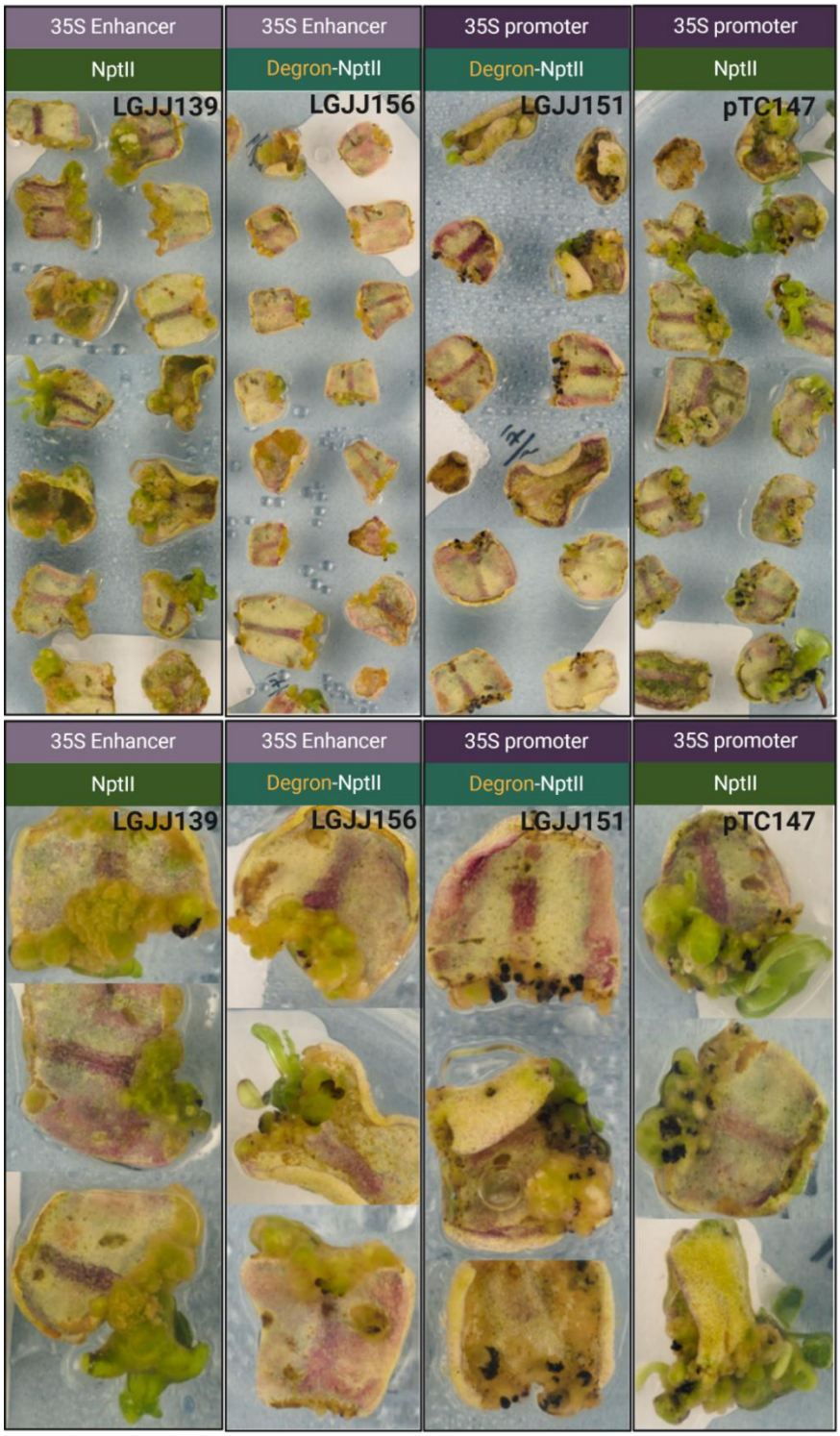


Figure 4.3 Visual estimation of the reduction of non-transformed tissue growth when using a *degron-NptII* or a standard *NptII*. Each column contains pictures of explants transformed with vectors LGJJ139, LGJJ156, LGJJ151 or pTC147. The top panel gives an overview on the growth of green tissues for a selection of explants. The bottom panel provides a higher magnification view of explants to better visualize the ratio of green to purple callus tissue growth. The abaxial side of the explants were photographed six weeks after transformation.

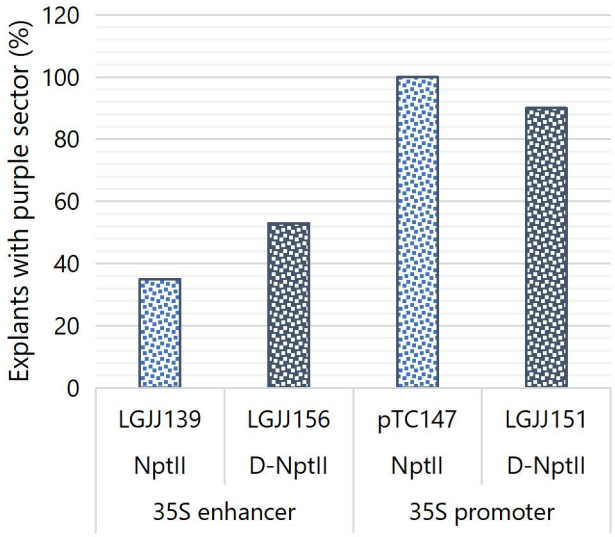


Figure 4.4 Bar chart representing the percentage of explants with at least one purple sector from the total of explants transformed. D-NptII=degron-NptII.

To quantify and corroborate this visual assessment, the image analysis programme Image J was used to measure callus size and purple tissue area within callus. The protocol was as detailed in chapter 2, section 2.2.13. Due to the more ambiguous phenotype arising from 35S enhancer:*ANT1*-mediated visual screen (e.g. LGJJ139 and LGJJ156), I carried out the digital measurement analysis on explants that received a 35S promoter:*ANT1* expression cassette (i.e. pTC147 and LGJJ151), as this cassette allowed for a more pronounced purple phenotype and therefore clearer measurements. Using the raw images of explants taken on the same day six weeks after transformation, using the same camera settings, pictures of each individual explant per plate for each of the four constructs were collated into a document (pictures available in the Appendix B). This document was uploaded onto Image J where the region of growing callus on each cotyledon was defined manually, then measured. The average callus area on an explant (measured in pixels) approached 2,990 pixels (median value = 2938) in the *degron-NptII* group, whereas callus area reached 11,800 pixels (median value = 9858) in the standard *NptII* group (Figure 4.5, left graph). An unpaired student t test analysis revealed that the difference between these two groups is statistically significant ($p < .0001$). Next, the regions of purple sectors within each callus were highlighted and selected as areas of interest using the

thresholding tool of the programme. Purple sector area is represented as a percentage of the highlighted pixels (achieved *via* thresholding of the image) contained in the manually selected callus region. The *degron-NptII* group displayed a mean of 7.9% (median value = 7.3%) of purple callus tissue, while only 3.7% (median value = 2.7%) of callus mass was purple in the standard *NptII* group (Figure 4.5, right). This difference was demonstrated to be statistically significant by an unpaired student t test ($p < .0001$). Using the value of the percentage of purple callus area and multiplying it to the value corresponding to total callus area, then dividing the product by 100, I was able to obtain the number of pixels considered as purple for each of the callus areas. Surprisingly, the mean number of purple pixels per callus area is 250 for the *degron-NptII* group, whereas the mean reaches 481 in the standard *NptII* group. Together, these data corroborate the conclusion of my initial visual assessment that the *degron-NptII* limits the growth of green background tissues.

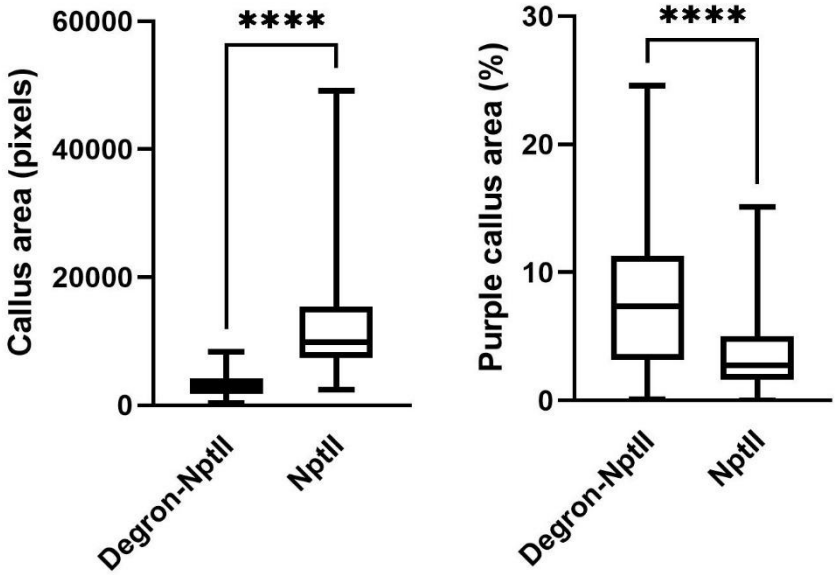


Figure 4.5 Box plot representing the results of the Image J analysis of callus size, measured as pixels within the manually selected regions of the images (left graph). The portion of the callus area constituting of purple cells is represented as a percentage of this area (right graph). Symbols **** indicate the statistical significance of the mean difference of the 2 groups based on an unpaired t test ($p < .0001$). Graphical representation and unpaired t test performed with GraphPad Prism 8.4.3.

Potential benefits of this method were also assessed at later stages in the regeneration of explants. The regenerating material was examined three months after transformation to count

the proportion of green to purple shoots for each of the groups. Detail of the scoring method can be found in chapter 2, section 2.2.7.2 As shown in Figure 4.6, explants transformed with a T-DNA containing a standard *NptII* only had 15% and 16% of their shoots purple (pTC147, LGJJ139), the rest being green putative escape shoots. On the other hand, the proportion of purple shoots to green shoots increased in the groups expressing the *degron-NptII*, where 24% and 48% of shoots were purple (LGJJ156, LGJJ151, Figure 4.6). The number of purple shoots arising from explants transformed with LGJJ151 is surprisingly low (24%) compared to that of those transformed with LGJJ156 (48%), despite both these vectors carrying the *degron-NptII* gene. As shown (Figure 4.6), explant and shoot development from LGJJ151-transformed material showed some retardation compared to the three other groups. Despite vector LGJJ151 underperforming in this experiment, a chi-square test revealed that the counts of green to purple shoots between the *NptII* groups (207:38 respectively, data combined between pTC147 and LGJJ139) and *degron-NptII* groups (103:77, combined between LGJJ151 and 156) was statistically significant, with χ^2 (1, N = 425) = 39.1, p < .0001. As predicted, using the *degron-NptII* strategy increased the proportion of purple shoots to green shoots growth from tomato explants.

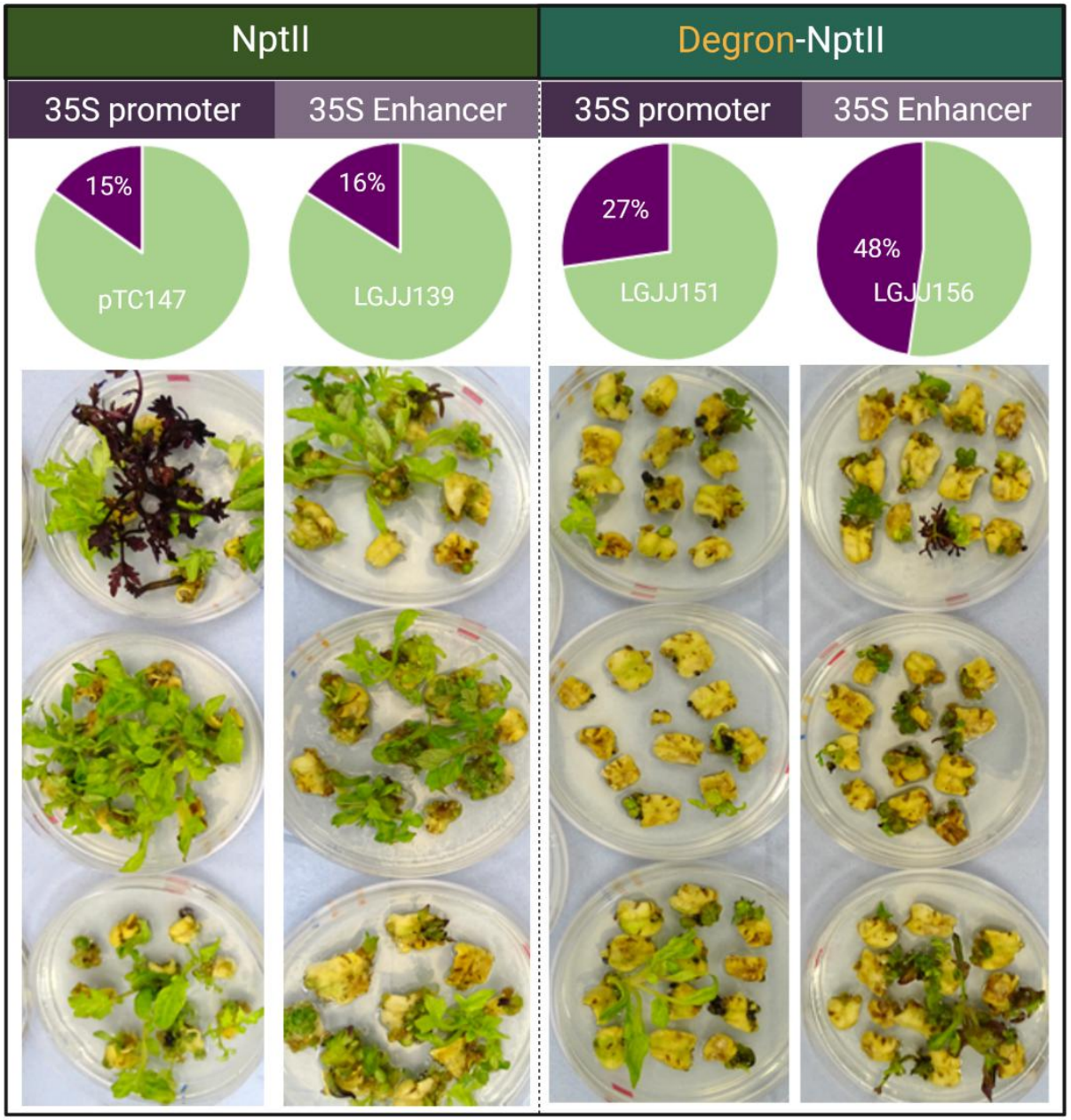


Figure 4.6 Proportion of green to purple shoots between *degron-NptII* expressing explants and the control *NptII* at three months after transformation. The pie charts represent the proportion of purple (purple slice) and green shoots (green slice) calculated from every developed shoot on each explant. The pictures below the pie charts are representative of the state of the regenerating tomato material three month after transformation in terms of progression of shoot development.

The shoot regeneration and shoot production ability of the explants transformed with *degron-NptII* or *NptII* was assessed four months after transformation. This was determined by counting the number of explants developing shoots measuring between 0-1 cm, between 1-3

cm and 4 cm and above (method described in chapter 2, section 2.2.7.2). As shown in Figure 4.7, 40% and 60% of explants had shoots over 4 cm long in the groups that received a standard *NptII*, whereas only 5% and 15% had shoots over 4 cm long in the *degron-NptII* groups. In the *degron-NptII* groups, shoots regenerated slower compared to the standard *NptII* group, which showed stunted shoot growth between 0-1cm for 45% to 80% of the explants. These data indicate that although the *degron-NptII* reduced the number of escapes, growth defects were increased compared to the standard *NptII*.

Overall, the degron strategy helped to reduce the growth of green tissues at the callus stage which in turn translated into an improvement of the green to purple shoot ratio at three months post transformation.

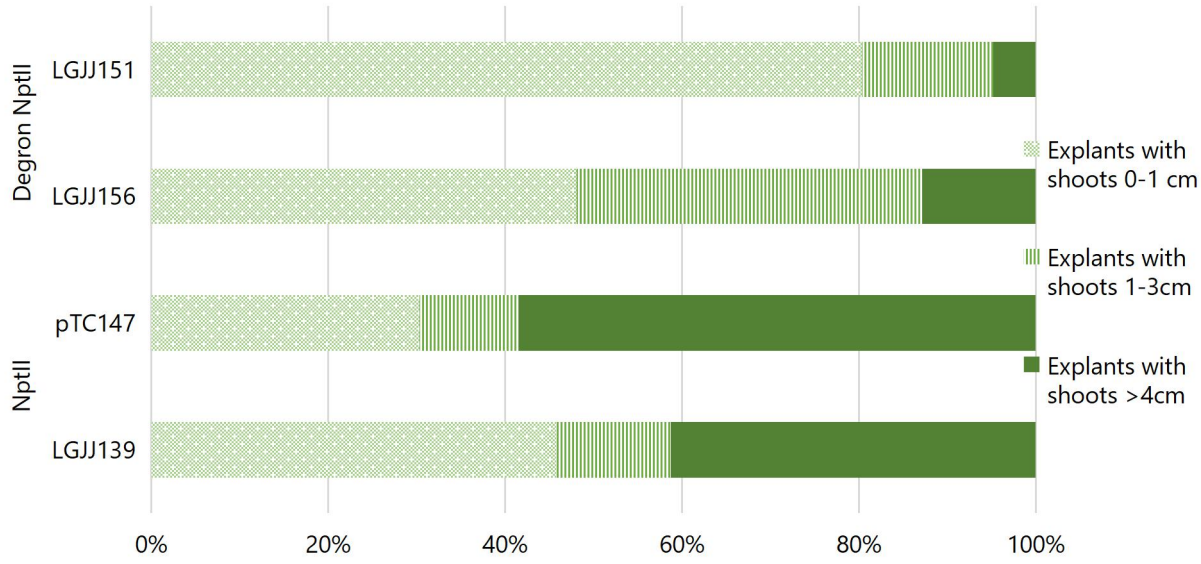


Figure 4.7 Chart showing shoot production and regeneration ability of explants transformed with constructs LGJJ139, LGJJ156, pTC146, LGJJ151. Proportion of explants only producing shoots between 0-1cm in length is shown in bright pale green with dotted pattern. Proportion of explants with intermediate size shoots, between 1-3 cm is depicted in the green section with a stripy pattern. The fraction of explants developing shoots above 4cm in length is shown in the dark green, solid fill sections. Data gathered four months after transformation.

4.5 The use of the *Degron-NptII* increases tomato transformation efficiency

Transformation efficiency can be measured at the end of the tissue culture procedure as a percentage of rooted shoots from the total number of shoots produced. The higher the number

of rooted shoots, the higher the transformation efficiency. Due to the prominent occurrence of escape shoots during tomato tissue culture, these non-rooted shoots bring the transformation efficiency down and create unnecessary additional subculturing work.

The performance of the degron was evaluated at the rooting stage of tomato transformation. Employing the degron strategy was expected to reduce the number of escapes which would bring transformation efficiency up. As seen in Figure 4.8, explant samples that were transformed with a vector containing a *degron-NptII* (LGJJ151, LGJJ156) developed more green rooted shoots (green shoots: 29 and 26) than green non-rooted shoots (24 and 14) for LGJJ156 and LGJJ151, respectively. In contrast, more green non-rooted shoots regenerated (101 and 53) than green rooted shoots (18 and 19, for pTC147 and LGJJ139 respectively) in the standard *NptII* groups. However, the majority of purple shoots surprisingly did not root, both for the standard *NptII* and *degron-NptII* groups (Figure 4.8). Notably, no data is available on the rooting of purple shoots for vector LGJJ151, as the growth of purple material was stunted and no viable purple shoot was available for transfer to rooting. Yet, there seems to be twice as many purple shoots arising from LGJJ156 (a total of 56) than from the standard *NptII* groups (17 and 32 for pTC147 and LGJJ139, respectively). The overall percentage of rooting shoots out of the total number of shoots was established and presented in Figure 4.9. As predicted, the percentage of rooting shoots was higher in the *degron-NptII* groups (44% and 65% for LGJJ156 and LGJJ151), compared to the standard *NptII* groups (17% and 29% for pTC147 and LGJJ139). In summary, between 70% to 80% of the regenerated shoots from the standard *NptII* groups are putative escapes, whereas the number of escapes is between 35% and 45% in the *degron-NptII* groups. These data indicate that the *degron-NptII* reduced the number of escaped and increased the transformation efficiency of tomato explants as more rooted shoots were produced.

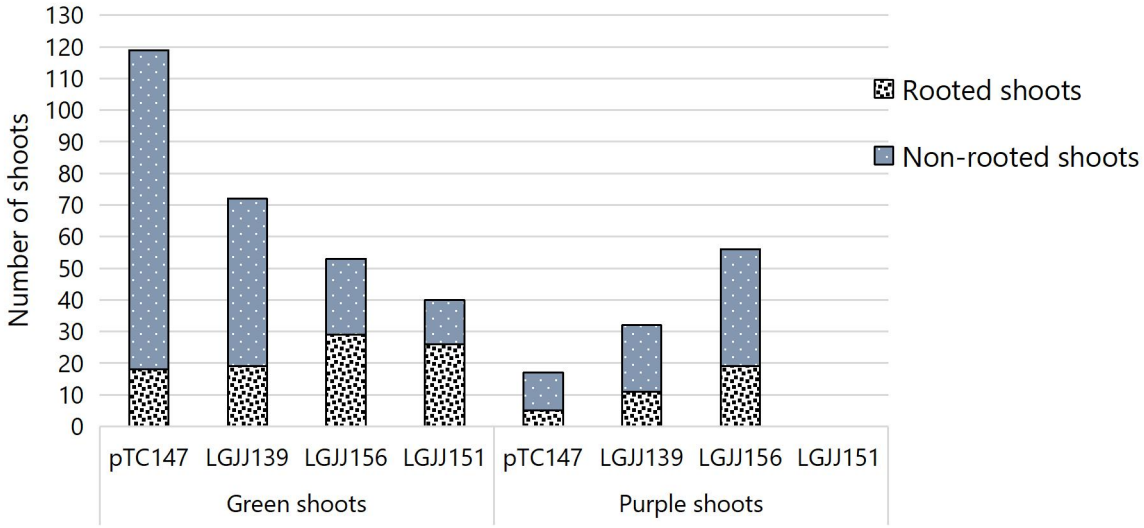


Figure 4.8 Bar chart representing the total number of rooted and non-rooted green shoots and purples shoots arising for each of the constructs tested, pTC147, LGJJ139, LGJJ156 and LGJJ151. Scoring of the rooting was carried out three to four weeks after the shoots were transferred to rooting medium.

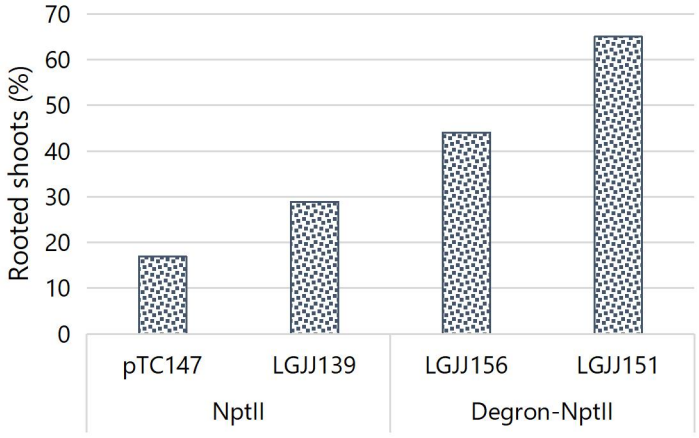


Figure 4.9 Bar chart showing the percentage of rooted shoots from the total pool of shoots produced by each of the transformed constructs.

To corroborate that the rooting phenotype of a shoot in a selective medium is a reliable indicator of whether that shoot is transformed, an *NptII* copy number assay was performed. For each of the transformation groups, ten individual green shoots were selected, five of which with a rooted phenotype (well-developed root system) and five with a non-rooted phenotype. Likewise, ten

purple shoots, containing five rooted and five non-rooted shoots, were selected for each group (except for the LGJJ151 group, as no purple shoot had developed to an adequate stage at the time of the analysis). The result of this *NptII* copy number analysis which was performed by the company IDNA genetics (Norwich, UK) is gathered in table 1. The protocol is described in chapter 2, section 2.2.2.1. As expected, 95% of the rooted shoots (green and purple combined) have at least one copy of the *NptII* cassette, except two green samples from the pTC147 group, which had a rooted phenotype but, surprisingly, *NptII* was not detected. Indeed, PCR-negative rooted shoots have been observed in other tomato transformation experiments (Frary and Aerle, 1996). In the non-rooted shoots category, the results are more complex. Focusing on green non-rooted shoots first, it appears that 80% of shoots from the standard *NptII* groups do not have a copy of the *NptII* cassette, which corroborates the occurrence of escape shoots. On the other hand, 100% of the shoots from the *degron-NptII* groups had at least one copy of the *NptII* cassette. Although initially surprising, this finding shows that escape shoots are reduced when using the *degron-NptII* as every shoot analysed in the *degron-NptII* groups had at least one copy of the *NptII* cassette, despite a proportion of those not rooting despite being transformed. For the purple shoots, 13 out 14 of the non-rooted shoots (*NptII* and *degron-NptII* groups combined) have at least one copy of the *NptII* cassette, which is expected based on their purple phenotype as this could only arise if a T-DNA insertion had occurred. Altogether, this set of experiments showed that use of the *degron-NptII* helped reduce the growth of non-transformed tissue at the callus stage, and therefore eliminated the occurrence of escape shoots during tomato regeneration, increasing the transformation efficiency of tomato explants.

Table 4.1 Copy number assay of the NptII cassette in rooted and non-rooted shoots. For each of the constructs, ten green shoots were selected (five rooted and five non-rooted) and ten purple shoots (five rooted and five non-rooted). No purple shoots were available for LGJJ151. qPCR primers annealed to the Nos terminator of the NptII cassette.

			Green shoots		Purple shoots	
Construct Content	Construct ID	Shoot Rep	Rooted	Non-rooted	Rooted	Non-rooted
NptII *35Senhancer*	LGJJ139	*1*	9	0	2	3
		2	9	0	2	2
		3	1	0	2	2
		4	2	4	1	1

		5	1	2	5	2
		6	2			
NptII 35Spromoter	pTC147	*1*	2	0	9	2
		2	2	0	2	2
		3	0	0	4	2
		4	10	0	9	0
		5	0	0	5	6
Samples with *NptII*			82%	20%	100%	90%
Degron-NptII 35Spromoter	LGJJ151	*1*	4	7		
		2	9	2		
		3	2	5		
		4	2	2		
		5	6	15		
Degron-NptII 35Senhancer	LGJJ156	*1*	2	1	2	3
		2	1	1	2	4
		3	1		1	3
		4	12	14	2	4
		5	2	1	3	2
Samples with *NptII*			100%	100%	100%	100%

4.6 Discussion

In the work presented in this chapter, I attempted to develop a method whereby the growth of untransformed plant tissue is reduced during tissue culture to favour the growth of transformed cells, which express a visual marker transgene, the 35S enhancer/promoter:*ANT1* expression cassette, alongside the *NptII* antibiotic selectable marker.

I expected that, for the *degron-NptII* groups, at the callus stage, purple sectors would be the prevalent cells to regenerate from the callus, with none, or little green tissues growing. At the shoot development stage, I also expected most of these to be purple, due to the insertion of the T-DNA with both the antibiotic selectable marker and the visual marker (35S enhancer/promoter:*ANT1*). At the rooting stage, I expected a majority of purple rooted shoots, and few green non-rooted shoots. The observations made and the data collected align to these expectations, but with some significant deviations, as discussed below.

4.6.1 Major reduction of green tissue growth at the callus stage

Based on a digital assessment of callus growth six weeks after transformation *via* Image J, it appeared that explants which had received a *degron-NptII* selectable marker cassette had a four-fold reduction in growing callus area compared to those which received a standard *NptII*. This supports the idea that the presence of the degron, combined to an adapted regeneration protocol, promotes the reduction of non-transformed, wild-type, green cells, therefore reducing the overall amount of actively growing callus. Furthermore, the proportion of purple cells per callus area was doubled in the *degron-NptII* group compared to the standard *NptII* group, which brings additional evidence that growth of background untransformed tissues is limited under the *degron-NptII* approach. However, the number of purple pixels within selected callus areas showed, on average, fewer purple pixels per callus area in the *degron-NptII* group than in the standard *NptII* group. This could indicate that the impact of the degron occurs mostly by preventing escape tissues from growing, yet does not impact the occurrence of purple sectors, since the purple sector area is smaller in the *degron-NptII* group. Alternatively, the purple sector area may be smaller in the *degron-NptII* group due to the presence of the Omega leader sequence at the 3' end of the 35S promoter on the anthocyanin expression cassette. The Omega leader sequence serves to accentuate anthocyanin overexpression and the associated fitness costs may hinder the development of those purple sectors, compared to purple sectors arising from the transformation of pTC147 which lacks the Omega leader sequence. To clarify this, future experiments should be repeated with a modified version of vector LGJJ151 where the Omega leader sequence has been removed from the 35S promoter module.

4.6.2 Increased ratio of purple to green shoots in the *degron-NptII* groups

Three months after transformation, approximately 25% of the shoots derived from the *NptII* groups were purple, whereas 27% (LGJJ151) to 48% (LGJJ156) of the shoots were purple in the *degron NptII* groups. Surprisingly, the proportion of purple to green shoots was low for the group of explants transformed with LGJJ151. Although these explants displayed numerous purple sectors, none of these sectors had developed into a purple shoot at the time of scoring and therefore could not be counted in the analysis. The growth retardation observed for LGJJ151-derived purple sectors and shoot primordia could be explained by the fitness cost associated with anthocyanin over-expression, for the same reasons mentioned in previous paragraph. Nevertheless, the ratios of purple to green shoots between the *degron-NptII* groups

and the *NptII* groups were statistically significantly different, despite LGJJ151 (containing *degron-NptII*) underperforming. This emphasises the strong effect of the degron at reducing growth of escape shoots. Looking at purple shoots, 77 regenerated in total between the *degron-NptII* groups, against 33 between the standard *NptII* groups. This suggests the action of the degron not only mediates a reduction of escape shoots, but also favours the recovery and growth of transgenic purple tissues. This seems to contradict the assessment of the data collected from the Image J analysis of calli, as the callus area corresponding to purple sectors was smaller in the *degron-NptII* group, suggesting no influence of the degron on recovering purple material. However, as discussed, the Image J analysis was only carried out on explants transformed with LGJJ151 due to the clearer purple phenotype deriving from the 35S promoter:*ANT1* expression cassette. Nevertheless, caution should be taken when drawing conclusions from this set of calli which are likely hindered in their development. The subsequent grouped analysis based on purple shoots number arising from LGJJ156 as well as LGJJ151 is likely a more reliable interpretation of the impact of the degron, and this analysis indicated that the degron enabled more purple shoots to grow. This could have been enabled as a collateral effect of the degron-mediated elimination of escape cells, thereby potentially reducing the competition faced by purple transgenic cells when surrounded by wild type tissues and therefore favouring growth of purple cells.

4.6.3 Reduction in the total number of shoots arising from *degron-NptII* groups and slow regeneration of degron-containing material

Based on the data from Figure 4.7, there is a reduction in the number of shoots regenerating in the *degron-NptII* groups, as only 5% to 15% of explants grew shoots above 4 cm. This observation could have several explanations. A stunted phenotype may indicate presence of non-transformed background tissue which is inhibited on kanamycin regeneration medium. However, based on the timing of the action of the degron, which should eliminate transient NPTIIs during the three-day 28 °C treatment post transformation, growth of non-transformed background tissue should have blocked. Thus, stunted tissue observed four months post-transformation is unlikely to be escape tissue growth. Moreover, results from the *NptII* copy number assay which revealed that all the 29 shoots tested from the *degron-NptII* groups had at least a one copy number of the *NptII* (table 1), implying that the developing tissues are indeed transformed, which further refutes the escape tissue growth theory.

The other potential explanation is that this stunted material is indeed transformed but cannot regenerate. Although the degron-NPTII proteins were stable during the 17 ˚C two-week treatment, explants were subsequently transferred to 22 ˚C in the growth chamber for the remainder of the tissue culture procedure. Based on Faden et al's data (2016), degron-tagged proteins are partially degraded at 22 ˚C, and this partial degradation of degron-NPTII proteins could slow the detoxification process of the kanamycin, affecting the growth of some of the tissues. The observation that this effect is heterogeneous within and between explants may be explained by the chimeric nature of callus mass. Some cells in a region of a callus may have received more copies of the T-DNA than other cells, leading to an increased expression of *degron-NptII* that could compensate for the partial degradation of its protein product and enable full shoot regeneration on selecting medium.

This highlights the bottleneck of the *degron-NptII* approach, as the optimal temperature for degron-NPTII stability is incompatible with rapid tomato *in vitro* tissue regeneration which is hindered at low temperature. Conversely, at the optimal temperature for tissue culture, degron-NPTII proteins are partially degraded, and this partial degradation is suspected to have a detrimental impact on the transformants. In the future, it would be sensible to quantify degron-NPTII proteins *in vivo* and assess by a western blot the extent of its degradation at 22 ˚C. Further experiments could also consider lowering the kanamycin concentration in the regeneration medium during the 22 ˚C phase to see if this improves the overall growth of the degron-containing tissues. Additionally, to establish whether the fusion of the degron at the N-terminus of the NPTII impacted the kanamycin phosphorylation ability of the enzyme, an enzymatic activity assay should be performed in a follow up experiment. This verification step is important as enzymatic activity of NPTII has previously been reported to decrease in an *E. coli* system as the size of the N-terminal fusion increased (Reiss et al., 1984).

4.6.4 Increased transformation efficiency and elimination of escapes

In a final assessment, shoots produced from every construct were scored at the rooting stage to establish transformation efficiency (based on the proportion of rooting shoots) and the impact of the degron on the later. As predicted, a higher percentage of the shoots from the *degron-NptII* groups rooted compared to the control group, thereby increasing the transformation efficiency. Nevertheless, it is surprising that the majority of shoots regenerating from the *degron-NptII* groups are green rooted shoots (Figure 4.8). Upon T-DNA insertion, recipient

cells would both accumulate anthocyanins due to the presence of the 35S enhancer/promoter:*ANT1* cassette, and be kanamycin resistant as a result of the *NptII* cassette. Rooted shoots in kanamycin rooting medium indicates that a T-DNA insertion has occurred, but the fact these are green indicates that the 35S enhancer/promoter:*ANT1* cassette is not expressed in this transgenic material. This could be caused by partial insertion of the T-DNA, where the left border inserted first, followed by the *degron-NptII*, but the insertion was interrupted before the *ANT1* expression cassette integration. Alternatively, the *ANT1* expression cassette was silenced, potentially triggered by repeated copies of the 35S enhancer/promoter:*ANT1*, caused by multiple T-DNA insertions. Data from the *NptII* copy number assay suggests that on average, three or four copies of the transgene are present, but this may not be extrapolated to the *ANT1* cassette copy number as primers were to the left of the cassette, hence we cannot exclude partial T-DNA insertion. Further verifications by PCR would enable detection of presence/absence of the *ANT1* expression cassette and determine the cause of the absence of purple phenotype in this transgenic material.

The *NptII* copy number assay revealed that 100% of the green non-rooted shoots tested in the *degron-NptII* groups were transgenics, whereas only 20% of the green non-rooted shoots were transgenics in the standard *NptII* group. Although the sample size is small for each group (n=10), it indicates the potential of the degron at eliminating escapes. In future experiments, a larger sample size of green non-rooted shoots could be tested for *NptII* copy number to increase the robustness of this finding. Furthermore, a number of *NptII*-positive green and purple non-rooting shoots were reported in both the *degron-NptII* and *NptII* groups. Issues with root formation from *in vitro* shoots are common. Aerial plant tissues have the ability to *de novo* synthesise cytokinins, which production in these tissues peaks in young, developing, tobacco leaves (Nordström et al., 2004). The regenerating tomato shoots could be synthesising cytokinins even several weeks after excision from their callus and transfer to a hormone-free medium. Cytokinins have a known inhibitory effect on root primordia development (Fukaki and Tasaka, 2009; Atkinson et al., 2014). Therefore, it is possible that the phytohormonal composition in the non-rooted transgenic shoots is preventing rooting due to elevated cytokinin levels. Additionally, tomato shoots usually readily form adventitious roots without exogenous application of plant hormones such as auxin (Frary and Aerle, 1996) which is why our protocol contains a hormone-free rooting medium. Nevertheless, auxin is required for the formation of root primordia in tomato shoot cuttings (Guan et al., 2019) and addition of

indole-3-acetic acid (IAA) to the rooting medium was shown to speed up root development in tomato tissue culture (Gupta and Van Eck, 2016). In future experiments, IAA could be added to the rooting medium to test if it triggers rooting in non-rooted transgenic shoots. The lack of root development in some transgenics could also be attributed to the degron-NPTII proteins being partially degraded at 22 ˚C, temperature at which the growth chamber was set also during the rooting phase. Reduced availability of degron-NPTII enzymes could lower kanamycin detoxification and compromise root growth through intermediate protein synthesis inhibition.

To summarise the findings from this chapter, the degron has proven an effective approach to reduce the growth of non-transformed tissues during tomato tissue culture, to eliminate the occurrence of escape shoots and to increase transformation efficiency by producing more shoots that develop roots. Despite increased numbers of recovered rooted shoots, around half of the shoots obtained in the *degron-NptII* groups were green, and not purple as expected. This could be attributed to partial T-DNA insertion, or silencing of the 35S enhancer/promoter:*ANT1* cassette. Inadvertent selection for truncated T-DNA insertions and or silencing of the transgene leading to green shoots may be occurring as green shoots could be more likely to survive as their fitness is not compromised by anthocyanin overproduction. Moreover, a copy number assay revealed that all the shoots produced from the *degron-NptII* groups had at least one copy of the *NptII* gene, even the non-rooted shoots in kanamycin rooting medium which would have been considered as escapes without the copy number assay. This demonstrated that, based on the tested shoots, escape shoots are eliminated when using the degron, although transformed shoots may not root for other reasons. These recalcitrant transformed shoots might conceivably be pushed to root if transferred to an IAA-containing, or a non-selective/or with reduced selective agent rooting medium. Although escapes have been removed, the temperature settings of the degron-adapted transformation pipeline are suboptimal for tissue culture. During the 17 ˚C two-week treatment (required for degron-NPTII stability and activity), tissue development was hindered. At 22 ˚C (promoting plant growth), partial degradation of degron-NPTIIs is suspected to have a detrimental impact on the regenerating material, suggested by the overall slower growth rate of the material compared to the control (standard *NptII*), and by the stunted growth of large proportion of the tissues. In future experiments, modifying the transformation pipeline i.e. kanamycin concentration,

temperature, duration of the cold phase, could fine-tune the protocol to obtain more transformed shoots, faster.

Importantly for my gene targeting experiments, it seems that the degron approach improved the recovery of purple tissues whilst reducing growth of escapes. Because knock-in events occur at low frequency, reducing the competition with WT cells faced by transgenic purple cells could boost the recovery of knock-ins and this will be tested in chapter 5. Testing the degron alongside the use of the 35S enhancer/promoter:*ANT1* visual marker was a sensible approach to establish whether the degron could help recovering more purple material. Nevertheless, because of the associated fitness cost of anthocyanin overexpression, this approach may have confounded the analysis of the effect of the degron by further slowing/stunting the growth of transformed tissues e.g. material from LGJJ151. A simpler experiment comparing *NptII* to *degron-NptII* transformants may provide a more accurate account of the impact of the degron on escape reduction.

Chapter 5. Testing the effect of different variables on rates of targeted knock-ins at the tomato *ANT1* locus

5.1 Introduction

Gene insertion at targeted loci remains the most difficult endeavour of genome editing, despite being almost a decade into the era of CRISPR-Cas nucleases. The vast majority of KI attempts in plants so far were proof of concept experiments, relying on selectable markers, chemical (Danilo et al., 2019; Wolter et al., 2018; Merker et al., 2020; Barone et al., 2020) or visual (Dahan-Meir et al., 2018; Čermák et al., 2015; Danilo et al., 2018), to assay the rates of KIs promoted by a given set of variables. In these experiments, the size of the DNA sequences to be inserted were relatively short, ranging between a few base pairs to 2 kb. The first application of gene targeting for trait improvement was recently attempted to create a KI golden rice variety by inserting a 5.2 kb carotenoid biosynthesis cassette (Dong et al., 2020). In order to realise the full potential of gene targeting for crop trait improvement, larger sequences encompassing several transcription units may need to be inserted e.g. a triple resistance gene stack. However, the efficiency of HR-mediated KI of large inserts has not been tested.

Successful insertion of genetic material at desired genomic locations by HR is hindered by several biological factors. As described in Chapter 1, the HR pathway is only active from late S-phase of the cell cycle until end of G2 phase, whereas only NHEJ repair is active from G1 to G2 phase. As the G1/S phase checkpoint will stop cell cycle progression upon detection of DNA damage until it is repaired, every Cas9-mediated DSBs delivered in the G1 phase can therefore only be repaired by NHEJ. In such cases, the NHEJ repair of the break may result in base pair mutations at the genomic sequence targeted by Cas9, thus excluding a population of cells from subsequent editing opportunities by making the target site unavailable. Studies in mammalian systems have shown that synchronising Cas9 expression and activity to HR-permissive cell cycle phases improves the recovery of HR-mediated insertions. This was

achieved by creating a fusion protein between the C terminus of Cas9 and the degradation signal (D-box) of a protein whose cellular accumulation is regulated and restricted to late S-phase to G2 phase, such as *hs*Geminin and murine Cyclin B2 (Vicente et al., 2019; Gutschner et al., 2016). This strategy employing a 'cell-cycle Cas9 tag' has not been tested in plants and, although there is no direct plant homologue of Geminin, plant mitotic Cyclins—involved in cell cycle control—could provide a functional D-box for a Cas9 fusion. For instance, the Arabidopsis Cyclin B1;1 (Cycb1) D-box fusion successfully mediated post-translational degradation of a heterologous protein in a cell cycle-dependent fashion (Colón-Carmona et al., 1999). Since Cycb1 D-box is a substrate for the Anaphase Promoting Complex ubiquitin ligase and its co-activating factor CCS52A (Fülöp et al., 2005), which are both present during the G1 phase (Heyman and De Veylder, 2012), proteolysis of Cyb1 D-box tagged proteins, such as Cas9, could be mediated through subsequent degradation by the 26S proteasome in the G1 phase.

Despite NHEJ being the prevalent mode of DSB repair in somatic cells, some properties of the DNA break can favour the recruitment of HR over NHEJ. Different types of DNA lesions, whether blunt ends or staggered cuts, promote a differential engagement of the repair pathways in mammalian cells (Bothmer et al., 2017a). Indeed, it was shown that 5' overhangs created from two nicks delivered by a Cas9 nickase was a more potent substrate for homology-directed repair than 3' overhangs (Bothmer et al., 2017a; Vriend et al., 2016), which leads to increased KI rates (Ran et al., 2013; Cermak et al., 2017). Surprisingly, breaks with 3' overhang polarity were poor inducers of HR repair. This finding is a paradox since single stranded 3' overhangs are required intermediates to perform the homology search during HR. The blunt ends produced by Cas9 are preferentially repaired by NHEJ (Bothmer et al., 2017a; Vriend et al., 2016). In light of these findings, application of a different CRISPR nuclease which, unlike Cas9, generates a 5' staggered DBS could bring a significant asset to the gene targeting toolkit. Characterised from the bacterium *Francisella novicida*, the CRISPR nuclease Cpf1 (CRISPR from *Prevotella* and *Francisella* 1), thereafter referred to as Cas12a, is a class 2 type V nuclease which delivers a staggered DSB at its cleavage site, exposing a 5-nt 5' overhang as illustrated in Figure 5.1 (Zetsche et al., 2015) (*c.f.* Chapter 1, Section 1.3.3. for further detail). Using the *Lachnospiraceae bacterium* allele of Cas12a (*Lb*Cas12a), which outperforms other Cas12a alleles (Tang et al., 2017; Bernabé-Orts et al., 2019), the first successful Cas12a-mediated gene targeting experiments in plants were reported in Arabidopsis (Wolter and Puchta, 2019a), rice (Li et al., 2019; Begemann et al., 2017a), tomato (Vu et al.,

2020) and tobacco (Huang et al., 2021). The Cas12a gRNA expression cassettes used in plants also differ from that used with Cas9 (Tang et al., 2017; Wolter and Puchta, 2019a). The gRNA scaffold is flanked by self-cleaving ribozyme sequences to improve the processing of the of crRNA transcript by removing the Pol III termination-derived U-tail at the 3' end of the transcript (Gao et al., 2018a), resulting in improved nuclease activity (Gao et al., 2018b; Tang et al., 2017). Optimisation of *Lb*Cas12a led to the engineering of a temperature-tolerant allele (tt*Lb*Cas12a), displaying mutagenesis efficiency between two to seven times higher than wild type *Lb*Cas12a over 5 different genomic targets at 22 °C (Schindele and Puchta, 2019). This variant results from one amino-acid substitution, D156R, in *Lb*Cas12a CDS. tt*Lb*Cas12a has successfully been used to produce KIs in Arabidopsis (Merker et al., 2020) and tobacco (Huang et al., 2021).

Design of the donor template could also be modified to boost KI efficiency. Placing nuclease target sites at the extremities of the donor template on T-DNA should promote a coordinated cleavage of the genomic target site and release of the donor template from its insertion site in the genome. Additional information about this strategy, also referred to as "*in planta*", or here as "three DSBs", has been detailed in Chapter 1, section 1.5.2. Several hypotheses are possible for why a three DSB strategy could increase the rates of KIs compared to a one DSB approach. First, the excised donor template could be more readily available at the break site to be used as the repair template. Additionally, creation of repair foci both at the genomic and donor template DNA molecules may favour encounter and interaction between the homologous sequences since some DSBs have been shown to be relocated to specialised structures near the nuclear periphery, at least in animals (Marnef and Legube, 2017; Caridi et al., 2018). The three DSBs approach has been used a lot in combination with Cas9 in Arabidopsis (Fauser et al., 2012; Schiml et al., 2014; Wolter et al., 2018) and in rice (Sun et al., 2016b; Dong et al., 2020) and with Cas12a in Arabidopsis (Wolter and Puchta, 2019a; Merker et al., 2020) and tobacco (Huang et al., 2021). Very few studies have compared side by side the rates of KIs derived from the three DSBs or from the simpler one DSB approach (Peterson et al., 2021; Peng et al., 2020).

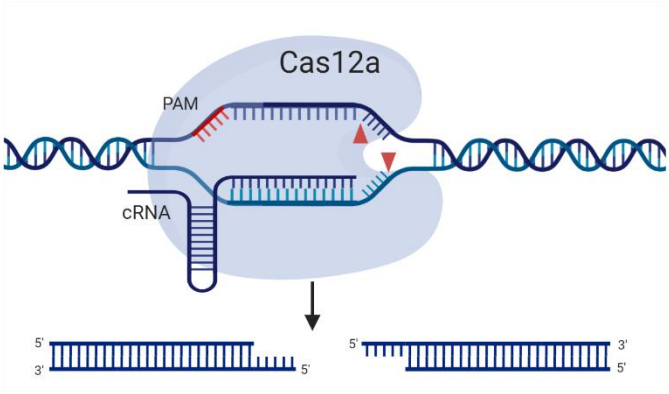


Figure 5.1 Cas12a cleavage of the target sites produces a staggered cut, exposing 5nt in the 5' polarity.

5.2 Aims

The aim of this chapter is to further address the low rates of KI in plants by testing additional variables for their potential to boost gene targeting success. For instance, the CRISPR nuclease tt*Lb*Cas12a will be tested for the first time in tomato to produce KIs and its performance will be compared to that of Cas9. An S/G2-phase-dependent Cas9 will be employed for the first time in plants. Additionally, the KI boosting potential of the three DSBs will be compared to the one DSB approach for Cas9 and for tt*Lb*Cas12a. Importantly, knock-in of the late blight resistance gene *Rpi-vnt1* will be attempted and the rates of KIs obtained when knocking-in a 2.4 kb or a 7.3 kb insert will be compared to assess the impact of insert size on gene targeting success. As performed in Chapter 3, a 35S promoter will be targeted upstream of *ANT1* to score the appearance of purple sectors upon successful insertion. Likewise, the *degron-NptII* strategy will also be tested for its ability to reduce the growth of background tissues in the context of recovering KI events. KI events will be characterised at the sequence level to assess whether the insertion was mediated by HR or NHEJ.

5.3 Construction of golden gate constructs for 35S promoter knock-in at the *ANT1* locus *via* HR

To assess the rates of 35S promoter KI at the *ANT1* locus when using the *degron-NptII* and when removing the geminivirus replicon, two constructs were built, LGJJ180 and LGJJ181, using the Golden Gate cloning method. Vector LGJJ181 harbours Cas9 target sites at the extremities of the two homology regions to *ANT1* (Figure 5.2, red crosses) to mediate cleavage of the donor, which is referred to as the three DSBs strategy. LGJJ180 corresponds to the one DSB approach. A second pair of constructs was built to contain a S/G2-phase dependent Cas9. Mimicking the Cas9 fusion design used by Gutschner et al. (2016), Cas9 was tagged at the C terminus to create an in frame fusion with the DNA sequence encoding the first 160 amino

acids of *Sl CYCLINB1;1 (CYB1;1)* (Solyc10g078330, Zhang et al., 2013) which contains the destruction box motif (RKALGDIGN). The new variant, Cas9-Cycb1, was assembled into a level 1 acceptor vector under the control of the Ubiquitin 10 promoter and E9 terminator used previously. This module was used to assemble vectors LGJJ182, LGJJ183 (Figure 5.2). Vector LGJJ183 carries Cas9-cycb1 target sites for the three DSBs approach. Another set of vectors, LGJJ191 and LGJJ192, carry tt*Lb*Cas12a. The gRNA expression cassette was modified accordingly to function with Cas12a (*c.f.* Chapter 2, section 2.2.11) and contains a new gRNA which targets -50 bp upstream of the *ANT1* CDS. LGJJ192 is designed for the three DSBs approach. Finally, vector LGJJ216 was built to attempt to simultaneously KI the late blight resistance gene *Rpi-vnt1* and the 35S promoter. LGJJ216 carries Cas9 target sites for the three DSBs approach. Vectors LGJJ180, 181, 182, 183 were assembled with the help of Synbio TSL support team member Mark Youles and LGJJ191, LGJJ192, LGJJ216 with the help of research assistant Hsuan Pai.

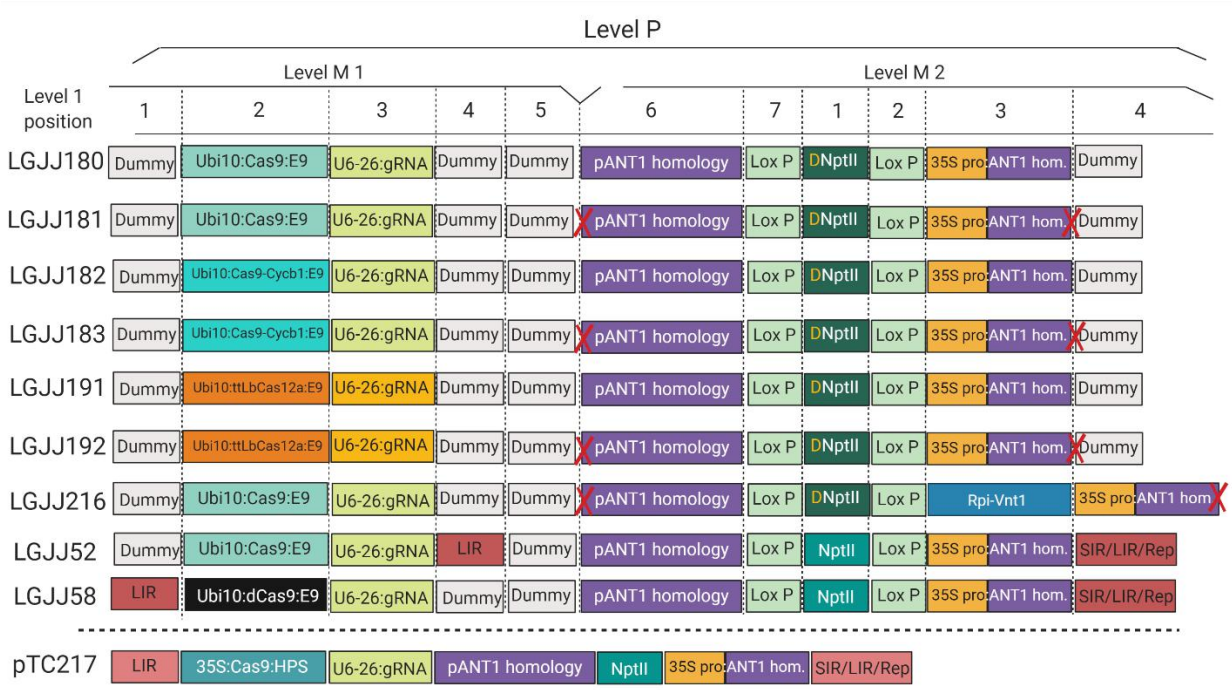


Figure 5.2 Schematic representation of the vectors for 35S promoter or 35S promoter + *Rpi-vnt1* knock-in at the *ANT1* locus. LGJJ180 and LGJJ181 are designed to test KI efficiency with the *degron-NptII* strategy and without the use of the geminivirus replicon. LGJJ182 and LGJJ183 carry a S/G2-phase dependent Cas9. LGJJ191 and LGJJ192 contain a temperature tolerant allele of *Lb*Cas12a (tt*Lb*Cas12a). LGJJ216 aims to knock-in *Rpi-vnt1* and 35S promoter. The red crosses represent Cas9/Cas12a target sites for cleavage of the donor

template in the three DSBs approach. The purple blocks represent the homology regions. pANT1 = *ANT1* promoter.

5.4 High efficiency KI with degron-NptII strategy, cell-cycle dependent Cas9 and ttLbCas12a

In this set of experiments, the seven binary vectors described in section 5.3 were transformed into Moneymaker explants, alongside two control vectors, LGJJ52, the positive control vector (containing the geminivirus replicon and the standard *NptII* cassette as opposed to the *Degron-NptII*) and LGJJ58 as the negative control (expressing a dCas9), as used in experiments conducted earlier (*c.f.* Chapter 3, section 3.3). To reduce the growth of background tissues to benefit the growth of transformed cells, explants were subjected to a heat treatment at 28 ˚C for three days before being placed at 17 ˚C, as per protocol established in Chapter 4. Due to the slow regeneration and growth of callus during the 17 ˚C two-week period, explants were given an additional two weeks before being screened for purple sectors i.e. six weeks after transformation. As described before, the explants displaying at least one purple sector were counted and the resulting KI efficiencies are recorded in Table 5.1.

Table 5.1 Knock-in efficiencies following Moneymaker cotyledons transformation. p35S=35S promoter. Vector IDs are followed by a number, i.e. LGJJ $^{\#N}$, where N refers to the experiment repeat number.

Vector	Nuclease	DSB approach	KI content	Explants used for transformation	Explants with a purple sector	Knock-in efficiency (%)
LGJJ180$^{\#1}$	Cas9	1	p35S	221	28	12.7%
LGJJ181$^{\#1}$	Cas9	3	p35S	185	34	18.4%
LGJJ181$^{\#2}$				281	105	37.4%
LGJJ181$^{\#3}$				316	45	14.2%
LGJJ182$^{\#1}$	Cas9-cycb1	1	p35S	186	21	11.3%
LGJJ183$^{\#1}$	Cas9-cycb1	3	p35S	111	24	21.6%
LGJJ183$^{\#2}$				224	45	20.0%
LGJJ191$^{\#1}$	tt*Lb*Cas12a	1	p35S	210	80	38.1%
LGJJ191$^{\#2}$				243	25	10.3%
LGJJ192$^{\#1}$	tt*Lb*Cas12a	3	p35S	123	48	39.0%

LGJJ192[#2]				166	29	17.5%
LGJJ216[#1]	Cas9	3	p35S +	420	145	34.5%
LGJJ216[#2]			*Rpi-vnt1*	305	64	21.0%
LGJJ52[#1]	Cas9	1	p35S	102	4	3.9%
LGJJ58[#1]	dCas9	-	p35S	88	0	0%

In the initial round of transformation, constructs LGJJ180 and LGJJ181 produced putative knock-ins in 12.7% and 18.4% of the transformed explants, which is four times higher than with the control LGJJ52 which does not contain the *degron-NptII* and carries a geminivirus replicon (Figure 5.3, panel 1). To test the reproducibility of these high rates of KIs, the best performing vector (LGJJ181) was transformed a second and third time into Moneymaker, reaching 37.4% KI efficiency (LGJJ181[#2]) and 14.2% (LGJJ181[#3]) (Table 5.1), with a mean KI efficiency of 23% across the three replicates. Notably, KI efficiencies between replicates were highly variable, increasing and decreasing two-fold. The results obtained with vectors containing a S/G2-phase dependent Cas9 displayed similar KI rates, with 11.3% (LGJJ182) and 21.6% (LGJJ183) of explants exhibiting a putative KI in a first round of transformation. In contrast to LGJJ181, the KI rates from LGJJ183 were high in a second independent transformation round (20.0% KI efficiency). A Chi-square statistical analysis was carried out to establish if a relationship exists between the allele of Cas9 used (Cas9 or Cas9-cycb1) and the rates of KIs observed. For this analysis, the counts of explants with a purple sector (total count: 212) and without purple sectors (total count: 791) obtained from the transformation of LGJJ180 and LGJJ181 were grouped under the categorical variable "Cas9". Similarly, counts of explants with a purple sector (total count: 90) and without (total count: 431) obtained from LGJJ182 and LGJJ183 were grouped under the categorical variable "Cas9-cycb1". The results from the Chi-square test indicated no correlation between the allele of Cas9 used and the rates of KI observed, χ^2 (df=1) = 3.219, p = .073. These data indicate that using the S- to G2 phase-dependent Cas9 did not significantly elevate the efficiency of KIs.

Using tt*Lb*Cas12a, putative KIs were observed in an average of 24.2% and 28.3% of the explants used for transformation for LGJJ191 and LGJJ192 based on two independent replicates of this experiment (Figure 5.3, panel 1). Notably, the second round of transformation produced KI efficiencies four-fold lower than the first round (Table 5.1), suggesting low reproducibility of the KI rates. The average KI efficiency obtained when using

Cas9 reached 20.7% (combining results from LGJJ180 and LGJJ180) as opposed to 26.2% when using tt*Lb*Cas12a (combining results from LGJJ191 and LGJJ192) (Figure 5.3, panel 4). To evaluate whether an association exists between using Cas9 or tt*Lb*Cas12a and the rate of KIs observed, a Chi-square test was performed. Count data used for this analysis included number of explants with (total count: 212) or without a purple sector (total count: 791) combined between LGJJ180 and LGJJ181 under the categorical variable "Cas9", and under "tt*Lb*Cas12a", the total count of explants with (182) and without a purple sector (560) combined between vectors LGJJ191 and LGJJ192. Based on this data, there is insufficient evidence to conclude an association between using Cas9 or tt*Lb*Cas12a and the observed rates of KIs (χ^2 (df=1) = 2.807, p = .094). Utilising tt*Lb*Cas12a appeared to lead to higher KI efficiency compared to Cas9, however, this is not statistically significant.

The potential of the one DSB versus three DSBs approach to boost KI efficiency was assessed when using Cas9 or tt*Lb*Cas12a. Overall, the KI efficiencies obtained when employing the one DSB strategy reached 12.0% (mean from LGJJ180 and LGJJ182) and 22.3% when utilising the three DSBs approach (mean from LGJJ181 and LGJJ183) (Figure 5.3, panel 3a). To search for a relationship between the number of DSB delivered and the observed rate of KIs when using Cas9, count data of explants with (total count: 49) and without purple sectors (total count: 358) obtained from the transformation of LGJJ180 and LGJJ182 was grouped to form the categorical variable "one DSB". Likewise, the number of explants with and without purple sectors from the transformation of LGJJ181 and LGJJ183 were grouped under the categorical variable "three DSB". The Chi-square analysis confirmed an existing relationship between the number of DSB delivered and the rates of KIs observed, χ^2 (df=1) = 21.14, p = .0001. Therefore, the three DSBs approach boosts KI efficiency when employing Cas9 and this observation is statistically substantiated.

With regards to tt*Lb*Cas12a, average KI efficiencies obtained based on two replicates were 24.2% using the one DSB strategy (LGJJ191) and 28.2% when applying the three DSBs approach (LGJJ192) (Figure 5.3, panel 3b). A Chi-square test was subsequently performed as described above. In the case of tt*Lb*Cas12a, there is insufficient evidence to claim an association between the number of DSB generated and the rates of the KIs observed, χ^2 (df=1) = 1.144, p = .285. Therefore, the three DSBs strategy may not lead to increased KI rates when using tt*Lb*Cas12a compared to the one DSB approach.

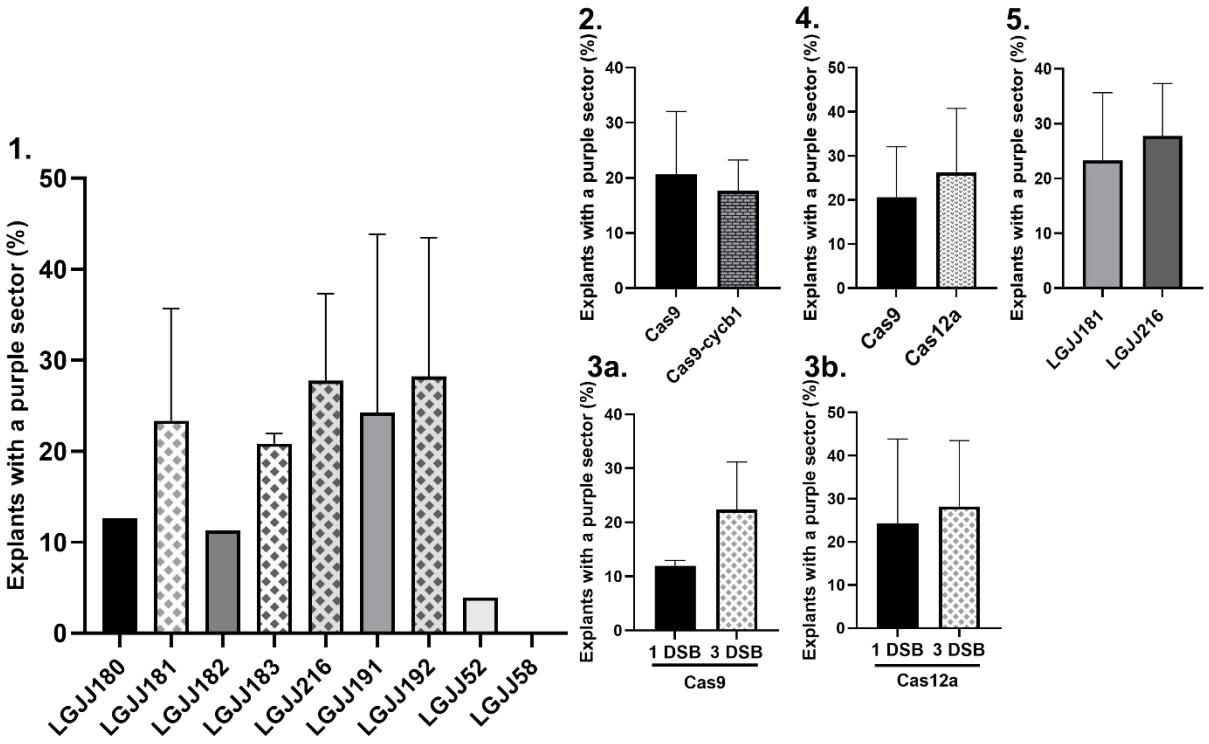


Figure 5.3 Comparison of KI efficiencies obtained between the variables tested (Cas9/Cas9-cycb1, Cas9/tt*Lb*Cas12a, one DSB/three DSB, R gene KI). **1.** Bar chart of the knock-in efficiencies obtained with vectors LGJJ180, LGJJ181, LGJJ182, LGJJ183, LGJJ191, LGJJ192 and LGJJ216. KI efficiency is represented as the percentage of explants that developed at least one purple sector from the total number of explants used for transformation. Bars with solid filled colour: one DSB strategy. Bars with a cross pattern: three DSBs strategy. LGJJ52 is the positive control and LGJJ58 the negative control. Error bars represent the standard deviation from two replicated transformation experiments, except for LGJJ181 which was based on three replicates. Data with no error bars are based on one round of transformation. **2.** Comparison of KI efficiencies obtained when using a standard Cas9 or a cell-cycle dependent Cas9 which is expressed during late S- to G2 phase. Error bars show the standard deviation. **3a**. Differences in KI efficiencies resulting from employing the one DSB or three DSB approach with Cas9. **3b.** Differences in KI efficiencies resulting from employing the one DSB or three DSB approach with tt*Lb*Cas12a. Error bars represent standard deviation from 2 rounds of transformation **4.** Comparison of KI efficiencies obtained when using Cas9 or tt*Lb*Cas12a. Error bars show the standard deviation. **5.** KI efficiencies compared when knocking-in a 2.4 kb DNA molecule (LGJJ181) or a 7.3 kb DNA molecule (LGJJ216). Error bars represent the standard deviation.

5.5 Knock-in of the resistance gene Rpi-vnt1 at the ANT1 locus

In this set of experiments, I attempted to knock in the resistance gene *Rpi-vnt1* and a 35S promoter simultaneously at the *ANT1* locus, making a 7.3 kb insert DNA fragment (*degron-NptII* + *Rpi-vnt1* 35S promoter). Appendix C contains the collated pictures of the purple sectors observed 6 weeks after transformation of vector LGJJ216 into Moneymaker explants. Based on two transformation replicates, 27.8% of transformed explants showed a putative KI (Figure 5.3, panel 5). In comparison, the identical approach (using Cas9 and also applying three DSBs) to knock in a DNA fragment of 2.4 kb with construct LGJJ181 led to an average of 23.3% KI efficiency across three transformation rounds. To assess if a relationship exists between the size of the DNA fragment to be knocked-in and the rates of KIs observed, another Chi-square test was performed. Based on these data, it appears a relationship between the size of the insert and the rate of KIs exists, χ^2 (df=1) = 5.478, p = .019. This finding is surprising since the difference in KI efficiency between the two groups (LGJJ181 and LGJJ216) is seemingly small (23.3% and 27.8%). To analyse the strength of this association, a Cramer's V test was performed and resulted in a co-efficient of 0.06. These data indicate that the association between size of the insert and KI rates is tiny and the small p value resulting from the χ^2 statistic may be an artefact of the large sample size (n=1570).

5.6 Verification and characterisation of knock-in events

5.6.1 Knock-in arising from LGJJ181

After gDNA extraction from purple tissues from explant 181-1, knock-in specific PCRs were performed to amplify the DNA across the left and right junctions of the insert. Three samples were taken from explant 181-1 (181-1_P1, 181-1_2 and 181-1_PG1, Figure 5.4, panel 2), each from a different part of the explant which could potentially reveal independent KI events within the same explant and attest of the chimeric nature of T0 regenerants. Knock-in specific PCR products for the left and right junctions are shown in Figure 5.4, panel 3.

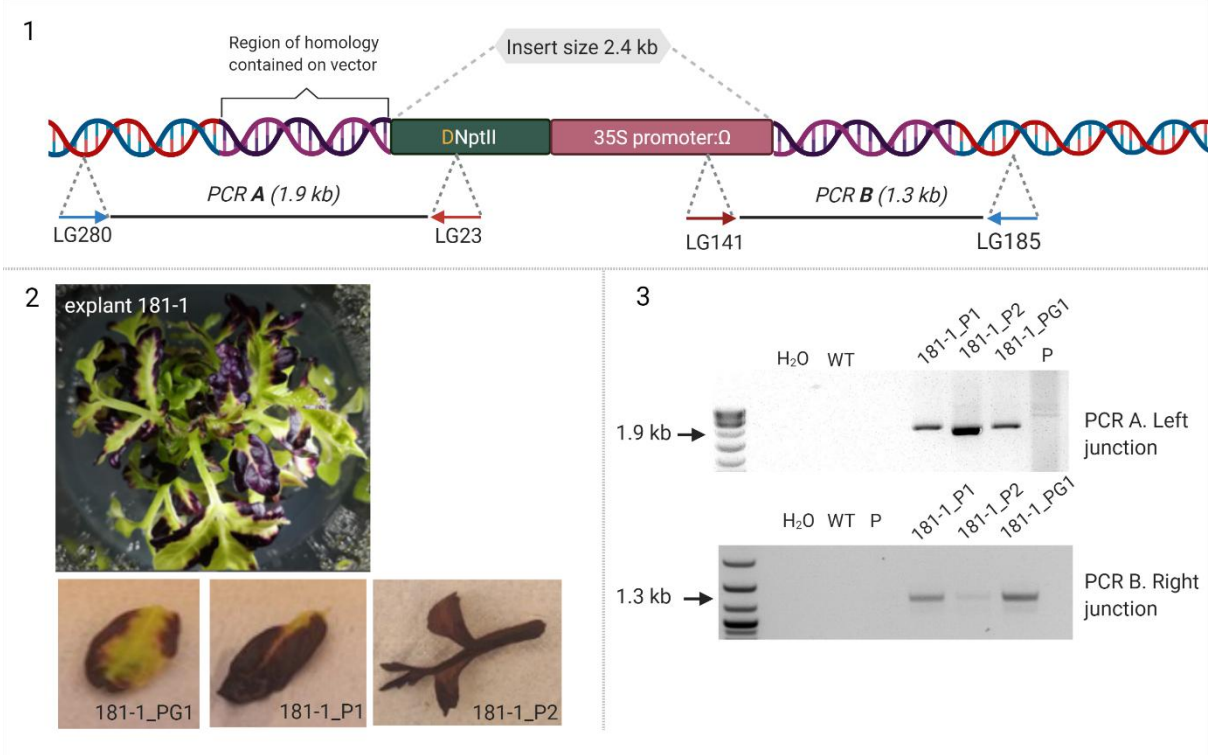


Figure 5.4 Verification of a KI event derived from LGJJ181. **1.** KI specific PCRs performed to characterise the insertion of the KI sequence. The left junction was amplified with LG280 and LG23 (PCR A), the right junction was amplified from the 35S promoter sequence (LG141 and LG185, PCR B). Expected amplicon size is indicated between brackets. Red arrows represent primers that are specific to the insert sequence. Blue arrows represent primers specific to gDNA sequences. **2.** Pictures of the three samples taken for analysis from explant 181-1 are shown (181-1_P1, 181-1_P2 and 181-1_PG1). **3.** Gel images of the KI specific PCRs A and B for samples 181-1_P1, 181-1_P2 and 181-1_PG1. Control PCR reactions included a template-free reaction with water added instead (H_2O), a reaction using wild-type Moneymaker gDNA (WT) and a reaction using plasmid LGJJ181 as template (P).

The PCR B products were subsequently cloned into a vector prior to being sent for Sanger sequencing. Alignments of the sequencing reads for samples 181-1_P1, 181-1_2 and 181-1_PG1 to the reference sequence is shown in Figure 5.5. Full sequencing coverage of the 1.3 kb amplicon is available in Appendix E.3. Looking at the right junction, samples 181-1_PG1 and 181-1_P1 harbour a C to T mutation at the +21 bp after *ANT1* ATG, whereas sample 181-1_P2 contains the same sequence as the reference sequence. Interestingly, the T present in 181-1_PG1 and 181-1_P1 corresponds to the wild-type allele of *ANT1* exon 1 which was purposefully mutated to a C when building the homology region module contained on the

vectors to eliminate a *Bsa*I recognition site which would have interfered with the golden gate cloning procedure. This finding suggests that purple tissues regenerating from explant 181-1 derive from at least two independent KI events.

Figure 5.5 Sequence alignment covering the right and left junctions of samples 181-1_P1, 181-1_P2 and 181-1_PG1 mapped to the reference sequence of the expected KI product. The right junction (top) starts after the 35S promoter sequence (green block annotated on the reference sequence). Dashed black lines at the end of the sequences indicate that only a selection of the sequence covering the junction is shown for ease of representation. These sequences are derived from a single parent explant from which different "sides" of the tissue growth were sampled.

The different outcome of the KI in samples 181-1_P1, 181-1_2 and 181-1_PG1 can be explained by a different mobilisation of the synthesis-dependent strand annealing (SDSA) repair pathway, the principal mechanism leading to KI by HR in plants (Puchta, 1998; Huang and Puchta, 2019). In the event of exonuclease-mediated degradation of the genomic DNA ends before strand invasion, genetic material is lost at the target site (Figure 5.6, B). The deleted sequence at the genomic target will be re-introduced *via* DNA synthesis during repair as this sequence is contained on the donor template, thus, the allele contained on the template will be introduced at the genomic locus. However, in the absence of genomic DNA degradation at the cut site, SDSA DNA synthesis starts at the junctions between the

homology region and the donor template and stops when reaching the next junction to a homology region, thus keeping the genomic sequence intact (Figure 5.6, A). This SDSA-mediated KI and exonuclease-induced genomic degradation can explain why some samples harbour the T>C mutant or the wild-type allele of *ANT1* exon 1. The rest of the sequences did not contain any mutation which suggest the right junction was repaired *via* HR. Several attempts at cloning the 1.9 kb amplicon of the left junction (PCR A) were unsuccessful. Direct sequencing of the left junction between the end of the left homology region and donor DNA was performed on gel purified PCR product. Sequence alignments suggest insertion via HR at this junction but cannot infer on the sequence at the junction between the endogenous *ANT1* promoter and the 5' end of the left homology region.

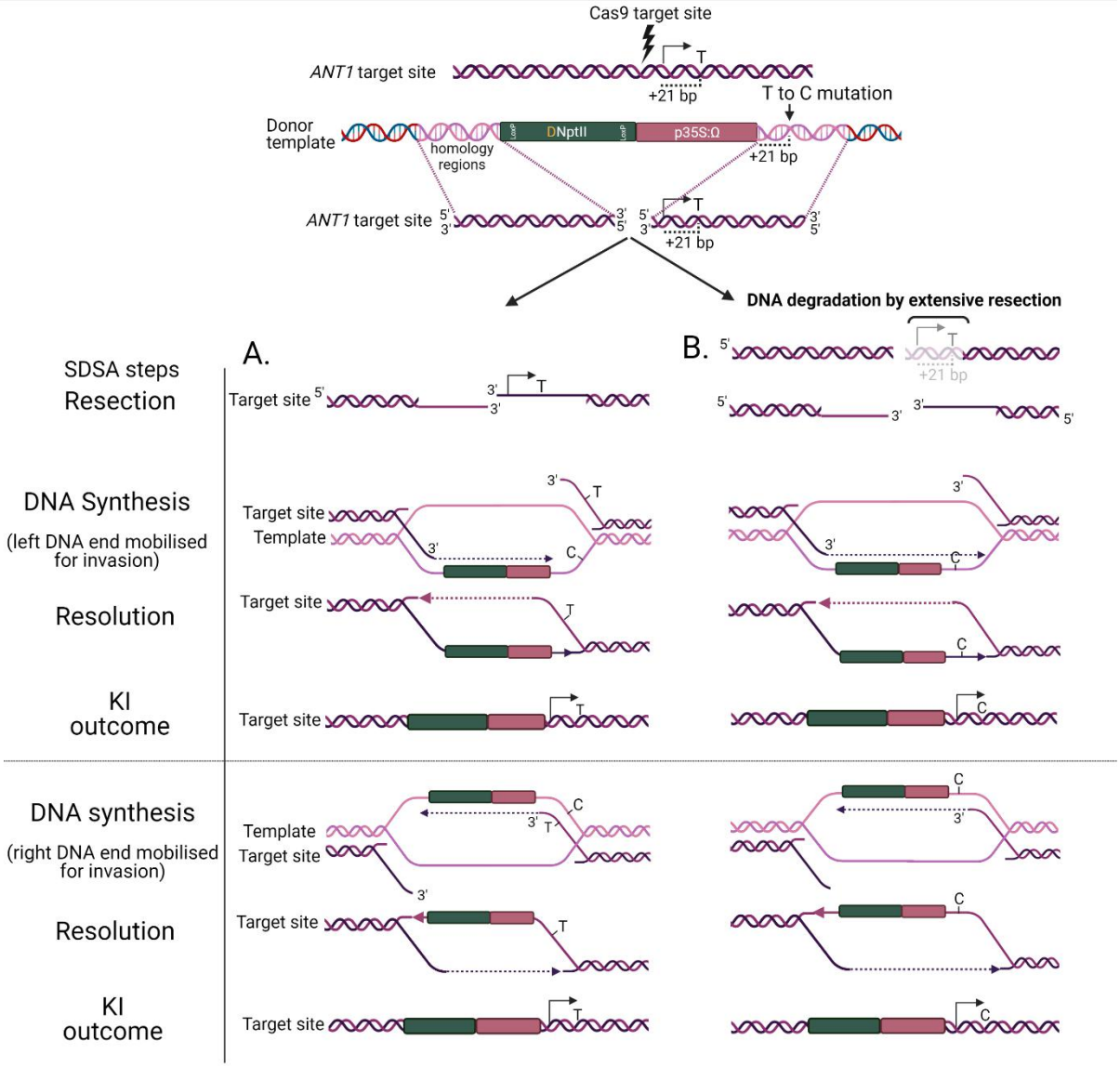


Figure 5.6 Model for DSB repair by SDSA at the *ANT1* locus resulting in two different alleles at the target site post KI. **A**. After DSB formation by Cas9, resection of dsDNA in a 3' to 5'

direction exposes 3' ssDNA to perform homology search in the nucleus. As per SDSA mechanism, either the left or right genomic DNA end is mobilised for strand invasion and D loop formation. Sequences contained between the flanking homology regions are synthesised at the target site. *ANT1* exon 1 is unmodified and retains a T at +21 bp. **B.** After DSB induction, potential degradation of the genomic DNA ends by exonucleases before the strand invasion step results in the loss of genetic material. Therefore, during repair, DNA synthesis starts copying genetic information from further inward to the homologous sequence contained on the template (left genomic DNA end mobilisation), or further down into the homologous sequence (right genomic DNA end mobilisation), inserting a T>C mutation at +21 bp in the exon 1 sequence.

5.6.2 Knock-in arising from LGJJ216

Samples 216_1, 216_2 and 216_20 were chosen for KI characterisation. To confirm the presence of a full-length KI, four PCRs were run on each sample. The left junction of the KI was amplified using primers LG203 (annealing to gDNA) and LG283 (annealing to *Rpi-vnt1* promoter on the insert) (Figure 5.7, panel 1, PCR A). All samples showed a band at the predicted size of 3.7 kb, indicating successful insertion of the *degron-NptII* cassette until at least the promoter of *Rpi-vnt1* (Figure 5.7, panel 3). The right junction was amplified in a PCR using LG297 (annealing to *Rpi-vnt1* CDS on the insert) and LG185 (annealing to gDNA) (Figure 5.7, panel 1, PCR B). A PCR product of the expected size was observed for samples 216_1 and 216_2 but not for 216_20, suggesting full length KI in 216_1 and 216_2, but not in 216_20. Additionally, the right junction was amplified in a second, shorter PCR (PCR C) to expose putative partial insertions. Amplification from every sample displayed a 1.2 kb band, corroborating the outcome of PCR B for samples 216_1 and 216_2 and revealing a putatively partial integration of the template DNA in 216_20. To provide further evidence that a genuine KI event occurred in these samples, a fourth PCR (PCR D) was run with primers amplifying from *ANT1* promoter to the *ANT1* terminator across the whole insert. Remarkably, the 10 kb band expected upon KI was present in samples 216_1 and 216_20 but could not be seen in 216_2 (Figure 5.7, panel 3). This is surprising as results from PCR A and B suggest a full KI in sample 216_2. Inability to amplify the 10 kb fragment from 216_2 could be caused by a lower quality extraction of the gDNA from this sample e.g. shearing of the DNA, presence of contaminants. Lower gDNA quality could prevent successful amplification of more challenging PCRs i.e. 10 kb in length. Other amplicons were also generated from PCR D: the 3 kb band produced in the three samples corresponds to the wild-type allele of

ANT1, indicating the samples are heterozygous for the KI. Moreover, intermediate size amplicons are also visible as faint bands in each sample i.e. a 6 kb band and a 2.5 kb band. Although results from PCR D align with the outcome of PCR B and C for sample 216_1, they are inconsistent with results obtained from PCR B for samples 216_2 and 216_20. For sample 216_20, the absence of a product generated by PCR B indicated a partial insertion of the donor template DNA. Therefore, it is surprising to generate a PCR product size corresponding to a full-length insertion in sample 216_20. Inversely for sample 216_2, outcomes from PCR A and B suggest a full-length insertion, yet a PCR product corresponding to such an event was not seen in this sample when carrying out PCR D. These data indicate that KI-specific PCRs confirmed the presence of a KI in the three samples tested.

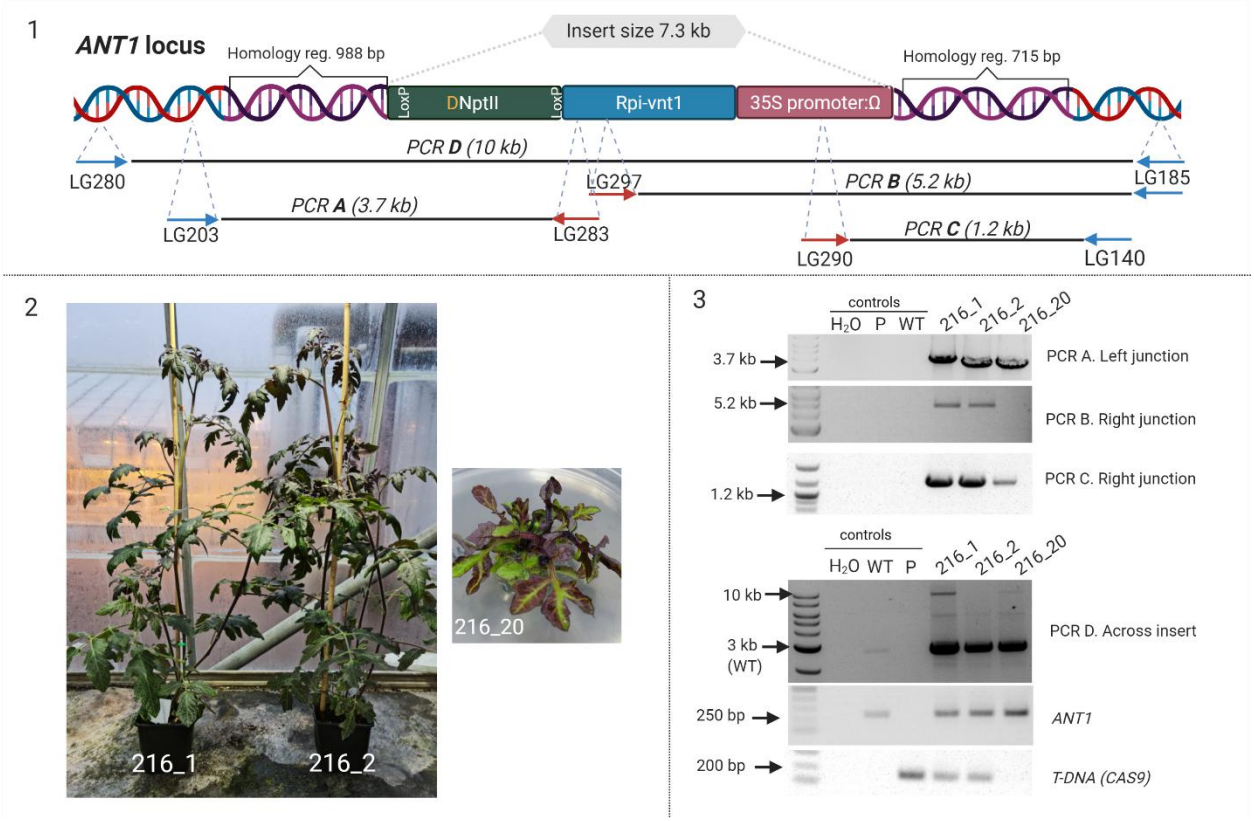


Figure 5.7 Verification of KI events derived from vector LGJJ216. **1.** KI-specific PCRs performed to characterise the insertion of the KI sequence. The left junction was amplified with primers LG203 and LG283 (PCR A), the right junction was amplified either starting from the beginning of *Rpi-vnt1* CDS (LG297 and LG185, PCR B) or starting from the 35S promoter sequence (LG290 and LG140, PCR C). The full insert was amplified with primers LG280 and LG185 (PCR D) **2.** Pictures of the three purple samples selected for characterisation of their KI event, 216_1, 216_2 and 216_20. **3.** Gel images of the KI specific PCRs A, B, C and D for samples 216_1, 261_2 and 216_20. PCRs for gDNA quality and T-DNA presence were also performed (*ANT1* and *T-DNA (Cas9)*). Control PCR reactions

included a template-free reaction with added water (H_2O), a reaction using wild-type Moneymaker gDNA (WT) and a reaction using plasmid LGJJ216 as template (P).

Verification for random T-DNA insertions was done by setting up a PCR to amplify a fragment from the *Cas9* gene, which should only be present if the T-DNA has been inserted. T-DNA insertions were detected in samples 216_1 and 216_2 but not in 216_20 (Figure 5.7, panel 3). Additionally, the copy number of the *NptII* cassette was measured by a qPCR assay *via* IDNA genetics services to assess the number of T-DNA insertions. The *NptII* cassette was present eight, nine and two times in samples 216_1, 216_2 and 216_20, respectively (Table 5.2). Since the *NptII* cassette is part of the DNA integrated at the *ANT1* locus, the number of random T-DNA insertions can be calculated by subtracting one copy from the total copy number. The qPCR data corroborated my PCR results for sample 216_1 and 216_2. PCR results did not detect T-DNA insertions from sample 216_20 (based on *Cas9* sequence) whereas the qPCR assay estimated one T-DNA insertion (based on the Nos terminator of the *NptII* cassette). These data show that between one to nine T-DNA molecules have been integrated in the samples tested, which is unusually high for a for a regular transformation procedure. These data could suggest that indeed having a higher copy number of the donor DNA (here, in the form of multiple T-DNA inserts) facilitates the occurrence of KIs, as suggested by the elevated rates of KI observed elsewhere when using a mild replicon approach (Čermák et al., 2015; Vu et al., 2020c).

Table 5.2 *NptII* (Nos terminator) copy number assay in LGJJ216-derived transgenics.

Sample	*NptII* copy number	Detected by PCR
216_1	8	Yes
216_2	9	Yes
216_20	2	No

To verify the nature of the insertion at the DNA level, PCR products generated by PCR A and C from samples 216_1, 216_2 and 216_20 were cloned into vectors and transformed into *E. coli*. Vectors from one to four positive *E. coli* clones were purified then sent for Sanger sequencing. Analysis of the sequencing data revealed a perfect alignment between the sample sequences and the reference sequence across the entire amplified region at the left and right

junction in all of the samples, suggesting KIs occurred *via* HR in each case (Figure 5.8). As observed in samples 181-1_PG1 and 181-1_P1 (Figure 5.5), samples 216_1 and 216_20 harbour the wild type allele of *ANT1* exon 1, whereas 216_2 carries the domesticated allele (Figure 5.8, A). Available sequences covering left and right junctions for samples 216_1, 216_2 and 216_10 are displayed in Appendix E.4 to E.9.

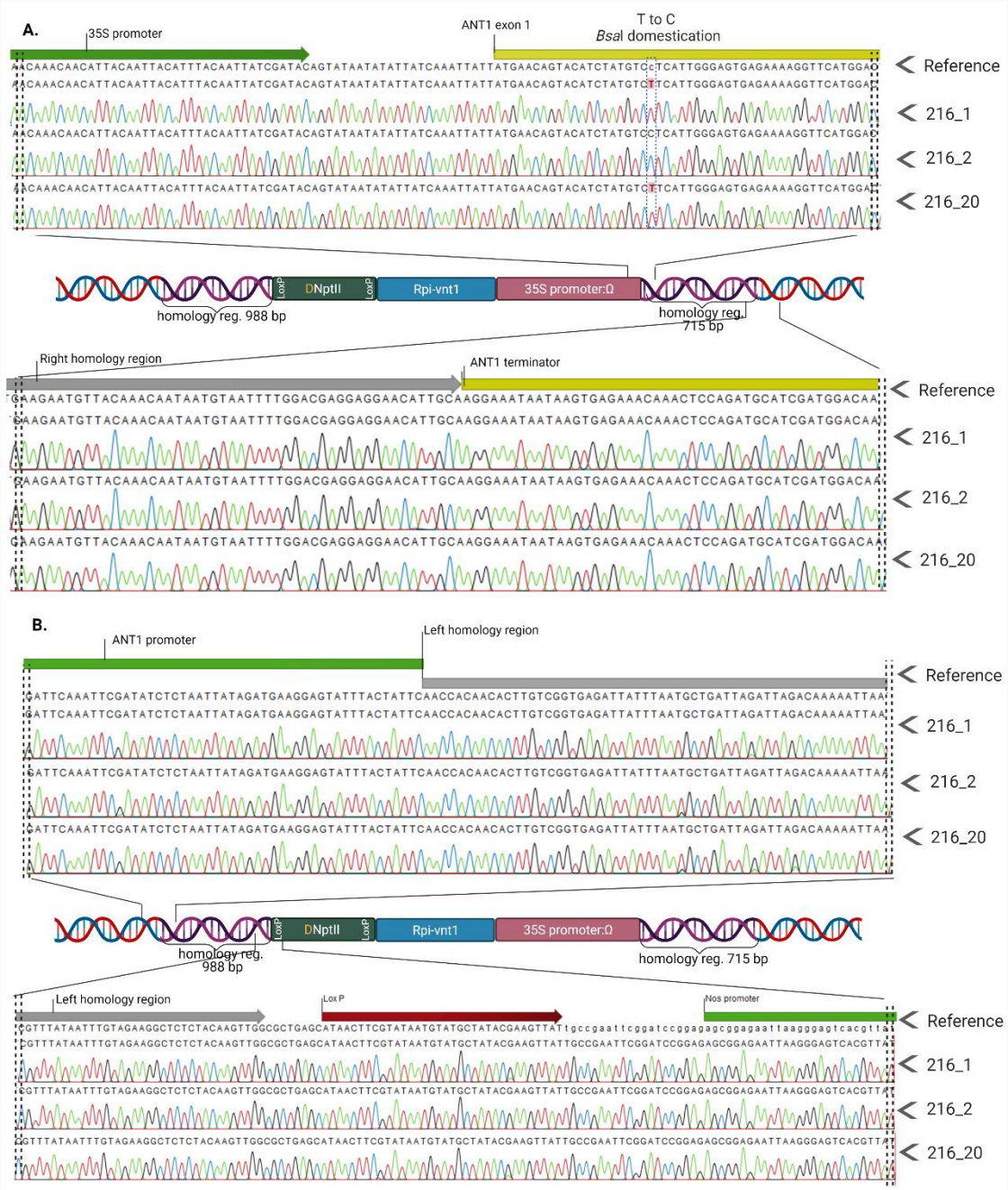


Figure 5.8 Sequence alignments covering the right and left junctions of the KIs in 216_1, 216_2 and 216_20. **A.** Coverage of the right junction at the section between the insert (35S promoter) and the 5' end of the right homology region (top alignments) and at the section between the 3' end of the right homology region and the genomic *ANT1* terminator sequence

(lower alignments). **B.** Coverage of the left junction at the section between genomic *ANT1* promoter sequence and the 5' of the left homology region (top alignment) and at the section between 3' end of the left homology region and the insert (LoxP, *Degron-NptII*) (lower alignment). Dashed black lines at the end of the sequences indicate that only a selection of the sequencing covering the whole junction is shown for ease of representation.

Presence of the T>C base pair mutation in the donor template enabled tracking of independent KI events arising in each KI line by assessing plant material from different parts of the originating explant, or by sampling material at different times during explant development. For the KI line 216_1 (i.e. material generated by explant 216_1), two samples were harvested from a shoot cluster during tissue culture (216_1 and 216_1_w), a third sample was harvested later from one rooted shoot from the initial shoot cluster (216_1_1). Looking at the Sanger sequencing data covering the border of the inserted 35S promoter and the *ANT1* promoter, I detected the occurrence of at least two independent KI events from KI line 216_1. The glasshouse sample 216_1_1 harboured the domesticated allele of ANT1 while 216_1 carried the wild type allele (Figure 5.9). This indicates that these two samples, despite emerging from the same shoot have a different cellular origin, which highlights the chimeric nature of T0 transformant plants. The third sample deriving from line 216_1, 216_1_w contained the wild-type allele of *ANT1*. Intriguingly, out of the four sequenced clones containing the right junction, three of them contained the wild-type allele of *ANT1* whereas one out of the four contained the domesticated allele. This could indicate that the population of cells within sample 216_1 was chimeric and contained both genotypes, but it cannot be excluded that this finding could result from PCR-derived chimeric amplification. In KI line 216_2, samples collected from tissue culture and from the glasshouse both harboured the domesticated allele of *ANT1*, and the two samples collected from tissue culture for line 216_20 displayed the wild-type allele of *ANT1*. No chimerism has been detected in these two KI lines and it can be assumed that all purple material arose from the same KI event.

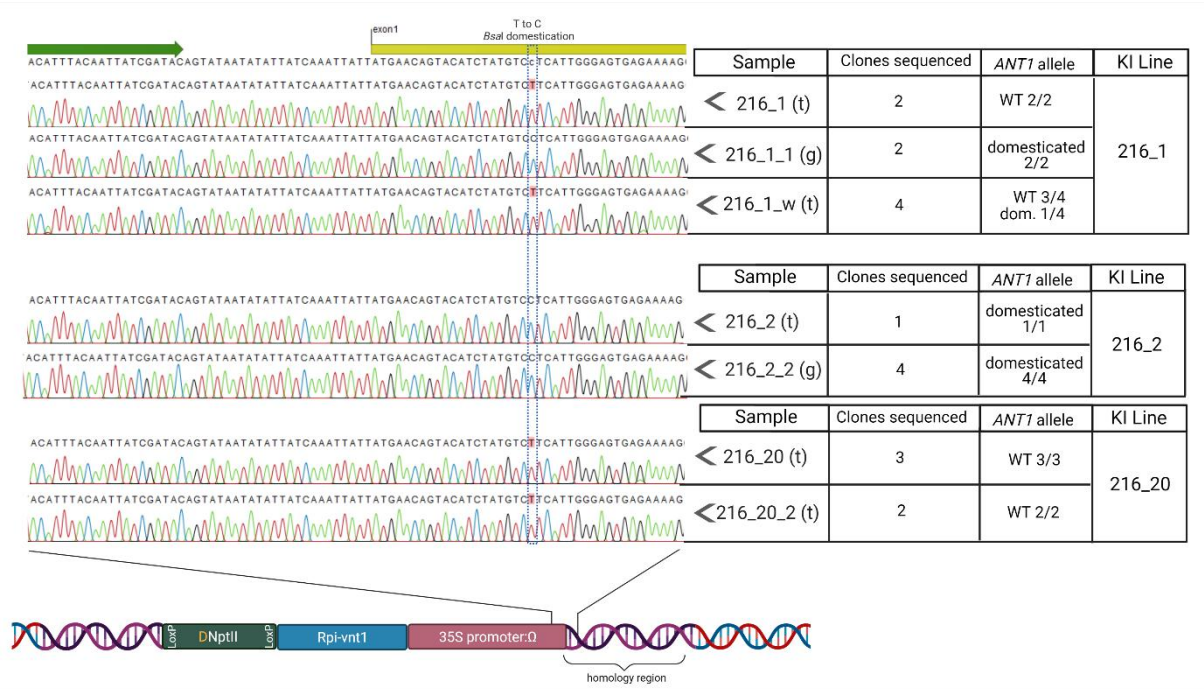


Figure 5.9 Tracking the occurrence of independent knock-in events in LGJJ216-derived samples. (t= sampled at the tissue culture stage) (g= sampled from mature plants growing in the glasshouse).

5.6.3 Knock-in arising from LGJJ191

The purple sample 191_2 deriving from the transformation of LGJJ191 (1 DSB), where the genomic DSB was produced by tt*Lb*Cas12a, was chosen for molecular verification (Figure 5.10, panel 2). The left junction of the KI was amplified using primers LG203 (annealing to the *ANT1* promoter) and LG23 (annealing to the *NptII* on the inserted DNA) (Figure 5.10, panel 1, PCR A). Intriguingly, the PCR product generated from PCR A on sample 191_2 resulted in a band of approximately 2.8 kb, where a 1.4 kb product was expected, suggesting a larger DNA fragment was inserted at the *ANT1* locus (Figure 5.10, panel 3). Moreover, an amplicon of 4 kb was obtained from the negative control reaction using plasmid LGJJ191 as template. Follow-up verifications by Sanger sequencing revealed this band contained spurious amplification from vector LGJJ191, where primer LG203, supposedly not binding to any sequence on the vector, still enabled amplification from the vector despite only 8 bp identity between the primer and the vector. The right junction was amplified with LG290 (annealing to the 35S promoter) and LG140 (annealing to *ANT1* terminator) (PCR B). As expected, a 1.2 kb PCR product was produced from the reaction. A third PCR was devised to amplify across the whole insertion to provide additional evidence for the KI (PCR C). Although the

expected product of 5.2 kb could not be seen, the PCR product resulting from the amplification of the wild type *ANT1* allele was present, alongside an unexpected, shorter, 1.5 kb fragment which is absent in the wild-type gDNA control reaction. Based on PCR C, it seemed that the two occurring alleles at the *ANT1* locus in the cell population were an unedited wild-type allele and another allele harbouring a deletion of approximately 1 kb. These results are inconsistent with the outcomes from PCR A and B which suggested that a KI had occurred (positive PCR B results) albeit with an unexpected insertion arrangement at the left border leading to a larger fragment amplified through PCR A.

Figure 5.10 Verification of a KI event derived from LGJJ191. **1.** KI specific PCRs performed to characterise the insertion of the KI sequence. The left junction was amplified with LG203 and LG23 (PCR A), the right junction was amplified from the 35S promoter sequence (LG290 and LG140, PCR B). Amplification covering the full insert at the genomic target was attempted with primers LG203 and LG140 (PCR C). Expected amplicon size is indicated between brackets. Red arrows represent primers that are specific to the insert sequence. Blue arrows represent primers specific to gDNA sequences. **2.** Picture of the sample selected for characterisation of a KI event, 191_2. **3.** Gel images of the KI specific PCRs A, B and C for samples 191_2. PCRs for gDNA quality was also performed *(ANT1)*. Control PCR reactions

included a template-free reaction with water added instead (H_2O), a reaction using wild-type Moneymaker gDNA (WT) and a reaction using plasmid LGJJ191 as template (P).

To clarify the nature of the insertion, unpurified PCR products from PCR A and PCR B were cloned into a vector and transformed into *E. coli* and vectors from positive clones sent for Sanger sequencing. Looking at the sequences covering the junctions of the KI, it appears the insertion took place *via* HR, as perfect sequence alignment to the reference is observed both at the left and right junction (Figure 5.11) and the analysis revealed that sample 191_2 harbours the domesticated allele of *ANT1*. Available sequences covering left and right junctions are displayed in Appendix E.10 and E.11.

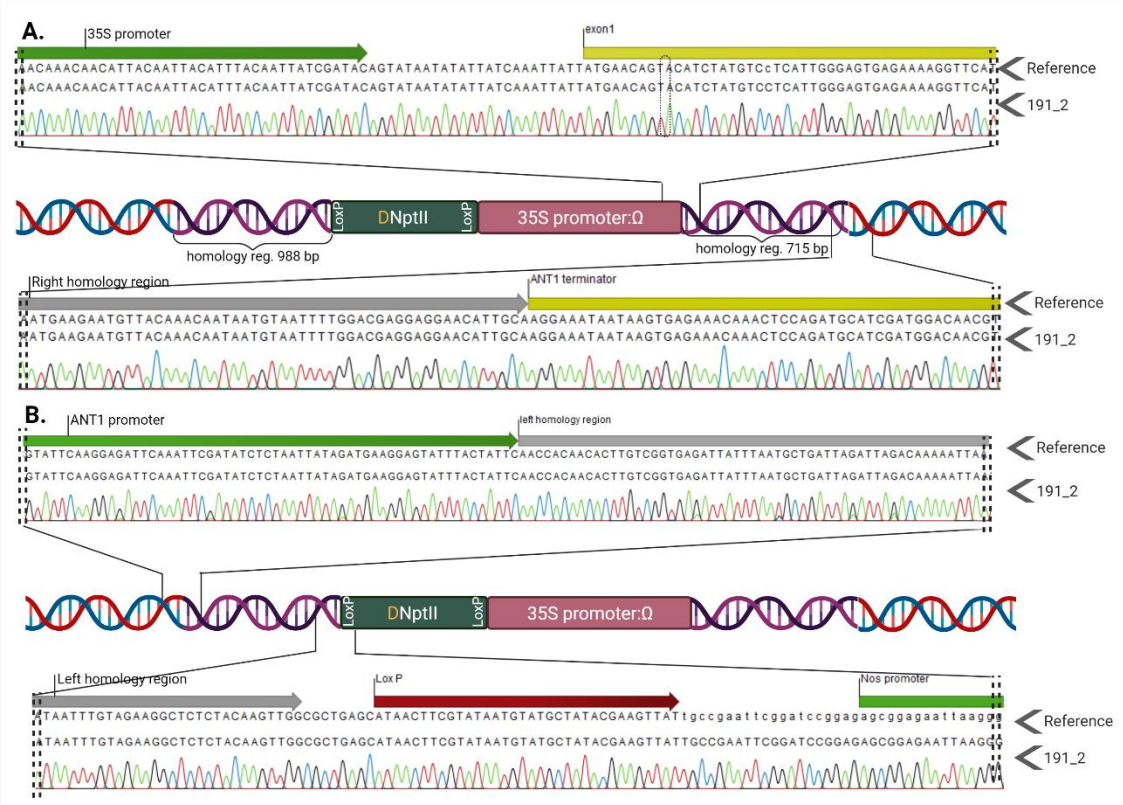


Figure 5.11 Sequence alignments covering the right and left junctions of the KI (*degron-NptII* + 35S promoter) in 191_2. Dashed black lines at the end of the sequences indicate that only a selection of the sequencing covering the whole junction is shown for ease of representation.

Nevertheless, results from PCR A indicated a longer fragment than expected was inserted at the left junction (Figure 5.12, panel 1). To determine the DNA content of the unexpected product, individual Sanger sequencing reactions were performed using four different primers (A, B, C and D) binding at regular intervals along the length of the cloned PCR product to

obtain full coverage of the sequence (Figure 5.12, panel 3). Surprisingly, the sequencing results showed that the DNA sequence cloned into the vector is identical to sequence expected from a perfect KI. Perfect alignment was observed from the start of the amplicon at -1,170 bp from ATG in the *ANT1* promoter (which is outside the homology region contained on the KI template) until the end of the amplicon containing insert-specific sequences (i.e. LoxP site, *Degron-NptII*) (Figure 5.12, panel 4). These sequencing data were observed across the four positive *E. coli* transformants selected for the analysis. These data did not reveal the presence of unknown DNA at the left border of the KI which would have explained the increased size of the amplicon produced by PCR A. Moreover, the vectors containing the cloned PCR A product were subjected to an enzyme restriction digestion to further investigate their content. If containing the expected 1.4 kb product, the restriction digestion should generate three fragments, one of 2,500 bp, one of 1,130 bp and another of 240 bp (Figure 5.12, panel 5). Each of these fragments were indeed present, but with the addition of one unexpected band of 1,000 bp. These results are in disagreement with the Sanger sequencing results, but they do corroborate the outcome of PCR A and add further evidence that the left junction contains additional DNA sequence and is not a perfect insertion. Although the accurate insertion of the KI has been validated at the right junction (Figure 5.10, panel 3, PCR B and Figure 5.11), the data gathered here does not enable confirmation of the nature of the insertion at the left junction due to inconsistencies in the results obtained from PCR, sequencing, and restriction digestion analyses.

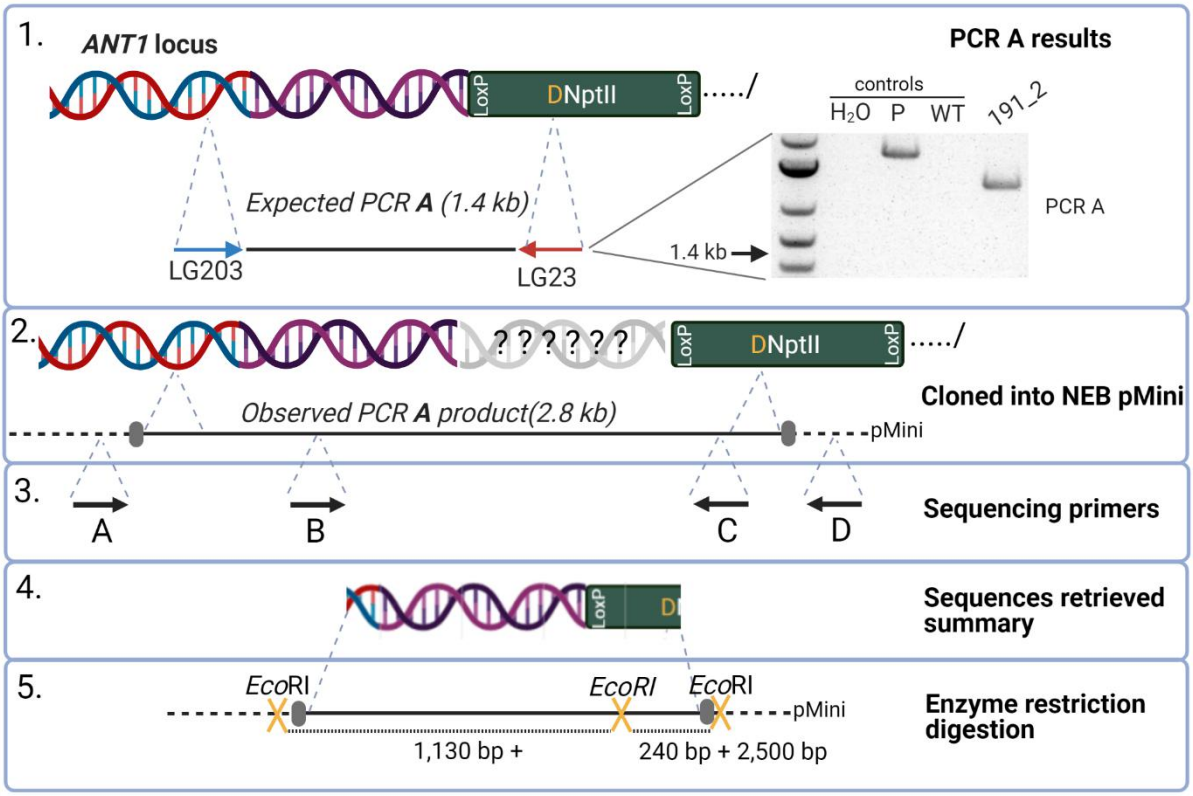


Figure 5.12 Verification of the content of PCR A (left KI junction) from sample 191_2. **1.** PCR A, expected to produce a 1.4 kb fragment when using primers LG203 and LG23, generated a 2.8 kb DNA fragment. **2.** PCR A product (2.8 kb) was cloned into the NEB cloning vector. **3.** To cover the whole insert, sequencing primers A (annealing to the backbone of the vector), B (annealing to the *ANT1* promoter region), C (annealing to the *NptII*) and D (annealing to the backbone of the vector) were used on each of the four clones analysed by Sanger sequencing. **4.** Sequencing data retrieved from the four individual clones contains the following sequence is cloned into NEB pMini: *ANT1* promoter followed by LoxP-*NptII*. **5.** Restriction digestion with *Eco*RI of the vectors containing PCR A product. Expected band pattern: 2,500 bp, 1,130 bp and 240 bp.

5.7 Discussion

The work presented in this chapter investigated new ways to boost KI efficiency in tomato. To help with the regeneration of purple sectors, the *degron-NptII* approach was applied here as well. Importantly, the methods developed in this project were utilised to attempt knocking in a gene.

5.7.1 Elevated rates of KIs with a geminivirus replicon-free, *degron-NptII* approach

The rates of KIs obtained following the transformation of my re-designed binary vectors (to express the *degron-NptII* and remove the modules generating the geminivirus replicon) were approximately 20-fold higher than the rates initially obtained (*c.f.* Section 3.4). Across the variables I tested here, KI efficiencies averaged just above 22%. Such rates of KIs can indeed be regarded as 'high efficiency' when compared to other gene targeting studies performed in tomato, and in other plant species. In other attempts in tomato, KI efficiencies obtained using a geminivirus replicon reached 25% for a 281 bp KI (Dahan-Meir et al., 2018) and 10% for a 1,938 bp KI (Čermák et al., 2015). Vu et al. (2020) obtained approximately 5% KI efficiency for a 1,938 bp insertion, however their scoring method for KIs is different to the one employed here and in the studies cited above; therefore, they are not strictly comparable. Without the use of the geminivirus replicon, rates of KIs of a 1,013 bp fragment attained 1.29% in tomato explants (Danilo et al., 2018).

With regards to the alleged KI-enhancing benefits provided by geminivirus replicon, data gathered in this chapter conflicts with what others have found. In solanaceous species, three times more KIs were observed when using the replicon than without (Vu et al., 2020b), others even reported a ten-fold and a hundred-fold increase when using the replicon (Čermák et al., 2015; Baltes et al., 2014). This trend was also reported in wheat cells (Gil-Humanes et al., 2017). However, in the data presented here, the rates of KI where over five times higher once the replicon had been removed. It is worth noting that, as detailed in Chapter 3, the replicon used here on vector LGJJ52 was based on a different BeYDV strain than the one used to create the replicon in the above-mentioned studies. Based on an acute strain of the BeYDV which replicates to reach a copy number of replicons in the nucleus at least six times higher than the mild BeYDV replicon (chapter 3, 3.10). Elevated presence of the replicon and of viral proteins such as Rep/RepA is suspected to have a significant detrimental effect on the plant regeneration ability and growth. This is likely the reason why my KI attempts using the acute replicon underperformed and explains the discrepancy with other studies performed in tomato. Nevertheless, a recent publication also demonstrates higher rates of heritable KIs in their replicon-free experiment compared to when using the Geminivirus replicon approach in barley (Lawrenson et al., 2021). Using my set of constructs built without the replicon (LGJJ180-LGJJ216), the elevated rates of observed KIs could be a result of the improved capacity of the plant material to grow and regenerate without the pleiotropic effects of the geminivirus replicon. Nevertheless, gaining a 20-fold increase in KI rates just by removing the replicon seems unlikely, as previous gene targeting attempts in tomato without the use of the replicon

only reached around 1% KI efficiency (Čermák et al., 2015; Danilo et al., 2018), indicating that another variable included in this set of experiments impacted, and boosted, KI efficiency.

In order to make the tomato cotyledon regeneration process more efficient by reducing the growth of escape tissue, the *Degron-NptII* approach was applied in this set of KI experiments. This approach may also have helped boost the KI efficiency by removing the competition faced by purple edited cells against wild type, escape cells early on during callus formation and shoot development. Based on my observations detailed in chapter 4, the use of the *degron-NptII* has certainly increased the number of regenerating purple shoots based on T-DNA insertion, hence it is reasonable to assume that a similar effect is observed here. Besides, another difference in the experimental set up compared to that of chapter 1 is the two-week period at 17 ˚C. Anthocyanin production can be triggered as a stress response mechanism against cold temperatures (Olsen et al., 2009; Qiu et al., 2016). In tomato, anthocyanin content was shown to gradually increase as temperature decreased from 28 ˚C to 18 ˚C and 12 ˚C (Løvdal et al., 2010). The transcription factor AN2 (*Solyc*10g086250) was determined a positive regulator of the cold stress response (Kiferle et al., 2015). Being a neighbouring gene of *ANT1* (*Solyc*10g086260) on chromosome 10, it is possible that lower temperatures led to the upregulation of *AN2* and induced an opening of the chromatin structure and configuration at this genomic location, which could improve nuclease accessibility to the target locus, but also more generally to HR machinery and template. Moreover, it was shown that DSB repair by HR occurs more efficiently near actively transcribed genes (Aymard et al., 2014), therefore, upregulation of *AN2* at cooler temperatures also may favour KI by HR at *ANT1*.

Based on the experiments presented in this chapter, we are unable to distinguish how much of the increase in KI efficiency can be accounted for the absence of the acute geminivirus replicon or by the presence of the *degron-NptII*. To uncouple the effect on KI rates of each these variables, future experiments could compare KI efficiency following the transformation of, for example, vector LGJJ181 alongside two control vectors. These control vectors would have an identical content to LGJJ181 except for one control expressing a standard *NptII* instead of *Degron-NptII* and the second control vector containing the *degron-NptII* as well as the modules to express the mild allele of the replicon. The first control would enable to assess the direct contribution of the *degron-NptII* strategy to the recovery of KIs, and the second control would test whether a cumulative effect can be seen if combining the use of the *degron-NptII* and a more physiologically tolerable replicon - as successfully used in several tomato gene targeting experiments (Cermak et al., 2017; Vu et al., 2020b; Dahan-Meir et al., 2018) – to

elevate KI efficiency higher still. Although the average of KIs rates obtained were high across the vectors tested here, KI efficiency varied significantly between experiments, with sometimes as much as a four-fold difference for the same vector between two experiments. This variability has been seen elsewhere (Huang et al., 2021), but few other comparisons can be made as other gene targeting experiments are often reported based on one experiment (Merker et al., 2020; Dahan-Meir et al., 2018; Danilo et al., 2018), or the rates reported are the average of several experiments and details of individual rates are not specified in the publication (Čermák et al., 2015). Nevertheless, variability in explant transformation success which can be linked to the quality of the starting plant material or other uncontrolled variables could also impact the overall outcome of the KI experiment and may in part be responsible for KI rates variability between repeat experiments.

5.7.2 Utilisation of tt*Lb*Cas12a boosts KI rates

Looking for variables with a significant potential to increase the rates of KIs, the type of DNA lesion created at the target site has come as a strong candidate. In an endeavour to create 5' overhangs at the genomic target site to preferentially recruit the HR machinery (Bothmer et al., 2017b; Vriend et al., 2016) to promote the KI, the nuclease tt*Lb*Cas12a was employed and its performance compared to that of Cas9. Consistent with others work, the average rate of KIs of a 2.4 kb insert was higher when tt*Lb*Cas12a was expressed (26.2%) compared to Cas9 (20.7%). Nonetheless, the difference in proportions of the explant count "with purple sector"/ "without purple sector" was not statistically significant between the tt*Lb*Cas12a and Cas9 groups. Nevertheless, this seems significant since this increase in KI rates translates into additional material bearing the desired mutation, with the potential to be regenerated to a full-size plant, self-fertilised, and the progeny taken forward. Using the *Lb*Cas12a allele in tomato, the KI-mediated purple spot formation rate was observed to increase to 81% from 54% when using *Lb*Cas12a instead of *Sp*Cas9 (Vu et al., 2020b). In Arabidopsis, *Lb*Cas12a was also found to increase the rates of KIs (Wolter and Puchta, 2019a). With regards to the tt*Lb*Cas12a allele, its use for gene targeting applications has been reported in Arabidopsis (Merker et al., 2020), and recently in another solanaceous species, tobacco (Huang et al., 2021). In tobacco, tt*Lb*Cas12a-mediated KI efficiencies ranged between 13% and 32% (average 20%), doubling the KI efficiency obtained with Cas9 (average 9%). Although the rates obtained from tt*Lb*Cas12a are comparable to mine, the Cas9-mediatd KI efficiencies were much lower than mine. This could be accounted to the fact they used the *Staphylococcus aureus* orthologue of Cas9 rather than SpCas9, or simply reflects the variability in KI efficiency between

experiments. Interestingly, an elevated ability for Cas12a to potentiate KIs has been reported in zebrafish too (Moreno-Mateos et al., 2017).

The reason why Cas12a outperforms Cas9 in gene targeting experiments has not been demonstrated, but I favour the interpretation that the production of 5' overhangs upon Cas12a cleavage could preferentially recruit HR machinery to the break to favour KI, since these types of DNA breaks, i.e. harbouring 5' overhangs, have been shown as HR-inducive in mammalian (Bothmer et al., 2017b; Vriend et al., 2016; Ran et al., 2013) and plant cells (Cermak et al., 2017). Nevertheless, these studies were performed using paired nickase to obtain a staggered DSB where the length of the overhangs would range between 20 to 100 bp, as opposed to the 5 bp overhangs created by Cas12a (Vriend et al., 2016; Bothmer et al., 2017a; Cermak et al., 2017). However, it is unclear whether the same enhancement for HR is promoted with the shorter 5 bp overhangs created by Cas12a, as 50 bp seemed to be optimal overhang length, with HR rates decreasing with shorter (20 bp) and longer (100 bp) overhangs (Vriend et al., 2016). The other common speculation on Cas12a enhanced performance at promoting KIs lies with its cleavage site being distal to the PAM. The DSB occurs about 20 bp downstream of the PAM (Zetsche et al., 2015), theoretically preserving both the PAM and seed sequence of the gRNA from mutations post NHEJ-repair. Unaltered, secondary Cas12a binding and cleavage of the genomic target sequence followed by KI remains possible. Additionally, the different expression system for Cas12a gRNAs (i.e. processed by ribozymes post RNA pol III transcription) could influence the efficiency of mutagenesis and thus KI rates in the Cas12a expressing group. However, it has been reported that Cas12a rates of mutagenesis display a non-linear relationship with the rates of KIs (Wolter and Puchta, 2019a) so the different gRNA expression system might not be the cause for Cas12a-promoted rates of KIs.

An unknown aspect of my experiment is the impact of the 17 ˚C two-week period on nuclease activity. Derived from the bacterium *Streptococcus pyogenes*, with optimal growing conditions nearing 40 ˚C (Panos and Cohen, 1964), Cas enzymes (Cas9 and Cas12a) have been shown to perform mutagenesis at higher frequency at elevated temperatures in several plants: Arabidopsis (Le Blanc et al., 2017; Malzahn et al., 2019), rice and maize (Malzahn et al., 2019), citrus (Le Blanc et al., 2017), wheat (Milner et al., 2020). Hence, it is reasonable to expect that enzymatic activity of wild type Cas9 will be low during the two-week 17 ˚C period. On the other hand, tt*Lb*Cas12a has been specifically engineered to retain enzymatic activity even at lower temperatures. Therefore, tt*Lb*Cas12a might continue cleaving the genomic target site, priming it for HR and KI of the insert, conferring an advantage over using the potentially

inactivated Cas9 during the 17 ˚C two-week period. Although tt*Lb*Cas12a has been studied for its ability to surpass its wild type counterpart in mutagenesis at lower ambient temperature, these studies only monitored the nuclease activity between 22 ˚C and 28 ˚C (Schindele and Puchta, 2019; Merker et al., 2020), therefore we can only speculate about tt*Lb*Cas12a efficiency at 17 ˚C. Future experiments could test tt*Lb*Cas12a mutagenic activity at 17 ˚C to assess the impact of the *Degron-NptII*-compatible tomato transformation pipeline on genomic target cleavage frequency. *Lb*Cas12a has repeatedly been reported to promote higher rates of KIs compared to Cas9 and in several plant species, and in other systems like zebrafish. Moreover, the tt*Lb*Cas12a variant demonstrated superior performance in boosting KIs compared to Cas9 (present study), and also compared to *Lb*Cas12a (Merker et al., 2020), making it a compelling nuclease choice for future gene targeting endeavours.

5.7.3 No improvement in knock-in efficiency was observed with a cell-cycle dependent Cas9

Another aim was to investigate whether utilising a cell cycle dependant Cas9, whereby enzymatic activity would be restricted to the cell cycle phase when the HR machinery is present i.e. from late S-phase until the end of G2 phase, would increase KIs. Here, no improvement in KI efficiency was observed when employing a chimeric Cas9 fused to the degradation signal of the tomato Cyclin B1 at the C terminus (Cas9-cycb1), with an average of 17.6% KI efficiency, compared to 20.7% KI efficiency when using Cas9. In a human cell line system where Cas9 C-terminus was tagged with the degradation domain of the murine Cyclin B2, rates of HR-mediated insertions were decreased compared to when using wild type Cas9. Interestingly, rates were increased when an N-terminus fusion of Cas9 to the Cyclin B2 D-box was used compared to wild type Cas9 (Vicente et al., 2019). Nevertheless, another study in human cell lines achieved elevated rates of KIs when employing a C-terminus fusion of Cas9 to a different cell-cycle tag i.e *hs*Geminin D-box (Gutschner et al., 2016), indicating the fusion positioning at the C-terminus of Cas9 is not *per se* responsible for the lack of improvement of KI rates in my case. Although the DSB induction ability of the chimeric Cas9-cycb1 was not assessed, we do not anticipate that cleavage efficiency has been drastically reduced by the Cycb1 D-box tag. Based on the KI rates as a proxy for DSB induction (which is a prerequisite for successful KI), Cas9-cycb1 gave a similar performance than Cas9. Although the Cas9-cycb1 was not beneficial in the present study, it is worth noting that these data are relatively preliminary, and that additional verification steps are needed before ruling out that a cell-cycle dependent Cas9 can promote enhancement of KIs. It is unclear whether the post-translational degradation of Cas9-cycb1 is indeed occurring and could be verified in

future experiments through western blots to detect differential protein accumulation between Cas9 and Cas9-cycb1 expressing cells. Additionally, an N-terminal fusion of Cas9 with Cycb1 could be tested since a similar approach has worked elsewhere (Vicente et al., 2019).

5.7.4 The three DSBs approach increases rates of KI in Cas9-conducted experiments but not Cas12a

Another aspect of the experimental set up which may promote high KI frequency is the simultaneous cleavage of the genomic target site and of the donor template, essentially excising it from the T-DNA. The data gathered here showed that a significant increase in KI efficiency is promoted when using the three DSB strategy with Cas9. This is in line with the other study having compared side by side the one and three DSBs approach in plants (Peterson et al., 2021), and also in human cell lines (Zhang et al., 2017a). In some cases, the three DSBs approach did not lead to high efficiency KI in tomato (Danilo et al., 2018). In this study, each of the three DBS were produced using a different gRNA, as opposed to one identical gRNA for each of the breaks in my case. Because gRNA efficiency is variable, it is possible that in their case, a gRNA with reduced activity compromised the outcome. This suggests that to realise the full potential of the three DSBs approach, using one gRNA with verified mutagenic activity is preferable. In another study in Arabidopsis (Peng et al., 2020), the three DSB failed to recapitulate the high rates of KI observed by another group (Wolter et al., 2018). This discrepancy was attributed to the utilisation of *Sp*Cas9 to conduct the experiment, rather than *Staphylococcus aureus* homologue of Cas9. Again, this highlights the importance of choosing the best adapted tools for a particular system to achieve high efficiency KIs.

Surprisingly, the enhancement of KI efficiency *via* the three DSBs strategy was not observed when carrying out the experiments with tt*Lb*Cas12a, and both the one and three DSBs strategies conferred high rates of KIs. It is unclear why the three DSBs approach failed to increase the rates of KIs in this case, but it suggests that the repair procedure of DSBs differs according to whether the DSBs were generated by Cas9 or tt*Lb*Cas12a. The tt*Lb*Cas12a methodology may not benefit from the three DSBs strategy because one cut harbouring a 5' overhang already has a strong enough potential to induce HR. These findings are based on two independently repeated experiments. Additional repeats of the experiment and attempts in a different plant species would increase the robustness of the data.

5.7.5 Efficient knock-in of a 7.3 kb insert containing the R gene *Rpi-vnt1*

In this chapter, I also demonstrated the feasibility of Cas9-mediated targeted insertions of novel large DNA molecules with the KI of the blight disease resistance gene *Rpi-vnt1*.

Remarkably, the 7.3 kb donor template was knocked-in at the *ANT1* locus with an average efficiency of 27.8%. This is the first time a resistance gene is inserted at a chosen locus by HR. Previous published attempts achieved a 5.2 kb KI containing a carotenoid biosynthesis cassette in rice (Dong et al., 2020) with a lower KI efficiency of 6.25%. Moreover, the researchers reported that none of these insertions occurred by HR, but instead were integrated by NHEJ in a reverse orientation. Here, all of the three purple samples selected for KI event characterisation had undergone HR at the left and right junction to realise the KI. Strikingly, a side-by-side comparison knocking in a DNA fragment of a medium size (2.3 kb) and of a larger size (7.3 kb) revealed no difference in the resulting KI efficiency. The impact of the insert size on KI efficiency has not been well studied in plants, but a KI study carried out in tobacco protoplasts using ZFNs showed that large DNA molecules of 5 kb, 10 kb and 20 kb could be successfully inserted at 28%, 19% and 18% efficiency, respectively (Schiermeyer et al., 2019), corroborating my findings. Likewise, a handful of studies in mammalian systems reported that no significant difference in KI efficiency were observed between a 500 bp and 1794 kb insert size (Shy et al., 2016; Gu et al., 2018). This seems to suggest that the main hurdles to successful KI occur prior to HR engagement for insertion. Once launched, the synthesis of the novel DNA sequence at the targeted locus is seemingly capable of inserting small and large DNA sequences with the same efficiency.

Considering the high performance of tt*Lb*Cas12a at mediating KIs, it would be interesting to repeat this *Rpi-vnt1* KI experiment using ttLbCas12a to see if this elevates the rates of KIs even further. Future experiments should analyse the progeny of the KI samples to assess the segregation ratio of the edit conferred by the KI. Moreover, a pathogen assay should be carried out in T-DNA-free progeny to verify that the transgene *Rpi-vnt1* is indeed expressed after KI at the *ANT1* locus and that its expression confers resistance to blight.

5.7.6 Characterisation of knock-ins by PCR and Sanger sequencing

Characterisation of KI events enables us to assess the accuracy of the insertion depending on whether it occurred by HR or NHEJ at the left and right junctions of the KI. Here, I established the nature of KI events in five T0 KI lines. Insertions took place by HR at both junctions in at least six out of the seven the samples analysed, with one unresolved junction for sample 191_2. Overall, these data indicate that KIs took place with high precision, with HR mobilised to repair both junctions. This is in line with other gene targeting experiments in tomato, where 65% or above of the KI events derived from HR at both junctions (Čermák et al., 2015; Dahan-Meir et al., 2018; Danilo et al., 2018, 2019; Vu et al., 2020b). Similar high

rates of perfect KI events were observed in tobacco (Huang et al., 2021), wheat (Gil-Humanes et al., 2017), and rice (Lu et al., 2020), although others only recovered NHEJ-mediated targeted insertions as inversions in this species (Dong et al., 2020). Interestingly, occurrence of perfect KIs is much lower in Arabidopsis, where about only half of the KIs are fully HR-mediated (Merker et al., 2020; Wolter and Puchta, 2019a; Wolter et al., 2018). It is not known whether this difference between species is significant, and if it is, whether it is caused by inherent biological properties or experimental design.

Due to inconsistencies between PCR, restriction digestion and Sanger sequencing results, I was unable to resolve the nature of the insertion at the left junction of the KI in sample 191_2. Both the PCR and restriction enzyme digest indicate the presence of a ~2.8 kb product but the full sequencing of the product indicates a 1.4 kb molecule. Presence of duplicated sequences at the left junction could be a potential explanation for the discrepancy in the sequencing result. Alignments of Sanger sequencing files covering between 500 bp to 800 bp per read may fail to show the presence of duplicated sequences at the junction once assembled onto the reference sequence. Relying on long read (>2 kb) sequencing technology to would address this limitation. Additional restriction digestion with other enzymes could also help decipher the content of the left border of the KI in 191_2.

Additional steps to confirm the presence of a full-length KI include performing a PCR encompassing the whole insert. This PCR is useful to establish homozygosity or heterozygosity of the edit. Samples derived from vector LGJJ216 produced the two bands expected from a heterozygous mutant containing a KI (3 kb and 10 kb bands). Nevertheless, other bands were also observed (at ~6 kb and ~2.8 kb) in the three samples. These bands may either be unspecific amplifications or represent additional alleles occurring in the cell population constituting the selected purple leaf. Indeed, the CRISPR reagents being present in these cells, subsequent edits may be created at the wild type *ANT1* locus of heterozygous cells, generating deletions or potential KIs. When the full-length KI PCR was performed on sample 191_2, two bands, one representing the wild-type allele of *ANT1*, and, unexpectedly, a second band of 1.5 kb were produced, indicating an *ANT1* allele harbouring a deletion of several hundred of base pairs. The expected 5.1 kb product containing the whole insert was not seen. The *Taq* polymerase extension time for this PCR was adapted to the amplification of a 5.1 kb product. According to the outcome of PCR A, a larger fragment than expected has been knocked-in (section 5.7.3), therefore, the extension time may have been too short to enable amplification of the full-length KI product. Gradually increasing extension time in future

PCRs would clarify this possibility. Alternatively, the amplification of the full-length KI may have been quickly outcompeted through the preferential amplification of the two smaller products generated during PCR C, and thus, not detected despite a full-length KI actually residing at the *ANT1* locus of sample 191_2.

The current methodology most widely available to characterise KI events relies heavily on PCR and Sanger sequencing which have some inherent limitations. PCRs amplification failures can require laborious troubleshooting which may prevent timely resolution of some KI events. Moreover, PCRs can create artefacts and false positive results arising from *in vitro* recombinant molecules (Meyerhans et al., 1990) which are observed as in KI junction-specific PCRs and reported in zebrafish (Won and Dawid, 2017) and barley gene targeting experiments (Lawrenson et al., 2021). PCR artefact formation can be limited by using specific polymerases i.e. such as the Kapa polymerase used in my experiments, and by limiting PCR cycle number to ≤ 30 (Sze et al., 2019). This highlights the importance of relying on a robust visual screen for successful KI, such as anthocyanin overexpression, as this provides confidence that the PCR products generated whilst characterising KI events are indeed genuine. Southern blotting has been used in several studies to corroborate PCR results (Schiml et al., 2014; Dahan-Meir et al., 2018; Fauser et al., 2012; Čermák et al., 2015), and with additional time, southern blots could have been performed on my samples to further confirm true KIs. The field of gene targeting will benefit from the advances of cutting edge molecular techniques such as Samplix Xdrop® (Madsen et al., 2020; Blondal et al., 2021). With this technique, targeted enrichment of the locus of interest followed by multiple displacement amplification in droplets ensures unbiased amplification of single high molecular weight DNA molecules (>50 kb), preventing inter-template chimera formation. These long range amplicons of the target site combined with long-read sequencing technology enables a coverage spanning up to 100 kb around the locus of interest that can detect larger scale genomic rearrangements, which would go undetected with commonly used PCR assays and Sanger sequencing (Kosicki et al., 2018; Burgio and Teboul, 2020).

Chapter 6. General discussion and Future outlook

6.1 Tackling the low rates of targeted gene knock-ins

This research project aimed to test and identify variables that could boost the rates of KI in plants. Inserting novel DNA material at predefined genomic loci is a biotechnological endeavour with the potential for key crop trait improvements. Although conventional plant genetic engineering methods, such as T-DNA-based transgene insertion, can achieve transgenesis to confer novel traits, targeted insertion of novel DNA sequences would have many advantages. These include: control over insertion site, absence of T-DNA sequences at chosen locus and reduced screening effort associated with transgenic events at different genomic loci. Moreover, gene targeting offers the possibility of complex trait stacking to create varieties quicker and easier than with conventional breeding methods. Shorter timelines are needed for generation of new cultivars with enhanced resilience against rapidly evolving biotic and abiotic threats. Nevertheless, the current rates of KI in plants are low, which inhibits development of targeted KI as a viable tool for genome editing of crops. By testing and combining different approaches, KI efficiency was raised in my experiments from 3% to above 20%, occasionally reaching almost 40% in experiments in the tomato variety Moneymaker. KI efficiency here is defined as the percentage of explants displaying a KI out of the total number of explants transformed in the experiment.

6.1.1 Reducing the barriers to high efficiency knock-ins

Consistent with the premise set out for this study, higher efficiency KIs were achieved by combining engineering solutions which improved both the frequency of KI events and the recovery of edited plant tissue over background non-transformed tissues. Based on the data gathered in this thesis, several conclusions can be drawn with regards to the optimal strategies to adopt in future KI efforts.

Replicon-based amplification of the donor template is not essential for high efficiency knock-ins

We established that high efficiency KIs, i.e. >20%, can be achieved without the use of a replicon for the first time in tomato (*c.f.* Chapter 5), starkly contrasting with the 1% previously achieved elsewhere without a replicon (Danilo et al., 2018; Čermák et al., 2015) and closely rivalling the 25% KI efficiency reported by others when using a replicon (Dahan-Meir et al., 2018). Augmenting the total copy number of the donor template with an acute form of the replicon did not further increase the rates of KIs and, on the contrary, reduced KI rates in comparison to the use of a mild form of the replicon (Čermák et al., 2015). Indeed, use of the acute strain of the BeYDV to create a replicon for high copy number replication of the donor template turned out to be unsuitable for high efficiency KI (*c.f.* Chapter 3). There are several potential benefits to using replicon-free KI strategies. First, it simplifies plasmids design and reduces plasmid size to accommodate other contents. Moreover, it prevents undesirable side-effects of the replicon on plant regeneration (Atkins and Voytas, 2020) and linear donor template are supposedly preferred to circular ones for the repair of DNA by HR (Song and Stieger, 2017). Therefore, the use of replicon-based amplification of donor templates is not required for generation of high-efficiency KIs, and may in fact be detrimental in my experiments.

Fostering favourable conditions to knock-ins

The potential to produce KIs for the nucleases tt*Lb*Cas12a and *Sp*Cas9 was compared for the first time, and utilisation of tt*Lb*Cas12a led to more KIs than Cas9 (*c.f.* Chapter 5). This is in concordance with previous studies which reported the superior performance of tt*Lb*Cas12a over *Sa*Cas9 (Huang et al., 2021) and *Lb*Cas12a over *Sp*Cas9, indicating that KI-favouring features of Cas12a are conserved in the engineered temperature tolerant allele. Some evidence suggests that tt*Lb*Cas12a is superior even to *Lb*Cas12a with regards to generating KIs (Merker et al., 2020).

Simultaneous induction of DSBs at the genomic target site and the extremities of the donor template was tested in a side-by-side comparison. My data bring additional evidence that cleaving both the target site and the donor fragment in the 3 DSBs approach significantly elevates KI success when employing Cas9. The same comparison experiment was conducted for tt*Lb*Cas12a. Surprisingly, no significant difference in KI rates was observed between the 1 DSB and 3 DSBs approaches (24.2% vs. 28.2% respectively, *c.f.* Chapter 5). This indicates that the 3 DSBs strategy may not be necessary when employing tt*Lb*Cas12a. Additional replicates of this experiment and tests in different plant species would help substantiating this preliminary inference.

6.1.2 Fine-tuning the optimal knock-in protocol

It is unclear how well the *degron-NptII* promoted the high rates of KIs observed in this study, or whether it should be used as a KI-enhancing tool in future KI experiments in tomato. Based on other gene targeting experiments carried out in tomato, attempts undertaken with a standard *NptII* reached 1% KI efficiency (Čermák et al., 2015; Danilo et al., 2018), suggesting the *degron-NptII* may have positively influenced the recovery of observable purple sectors in this study. Nevertheless, a recent publication demonstrated the feasibility of achieving high efficiency KIs in tobacco (average 20% with tt*Lb*Cas12a or 8.3% with *Sa*Cas9) without using the *degron-NptII* (Huang et al., 2021). Notably, kanamycin selection was only applied for two weeks to select for transformants, then replaced by herbicide resistance to recover KI events. The different mechanisms underlying resistance to kanamycin by NPTII and resistance to the herbicide (imidazolinone) by SuRB in Huang et al (2021) could explain their high rates of KIs without using the *degron-NptII*. As discussed in Chapter 4, growth of non-transformed, escape tissues from callus may in part occur due to the formation of concentration gradients of active kanamycin and inactive, phosphorylated kanamycin molecules across the regenerating callus mass resulting from NPTII enzymatic activity. Thus, non-transformed tissues may be able to grow in neighbouring localised pockets of tissues containing deactivated kanamycin. Other modes of action providing chemical resistance *in planta*, i.e. herbicide resistance from modified ALS, the mechanismof which does not rely on inactivation of the active compound, may reduce this effect. Indeed, transgenics selection *via* SuRB and herbicide resistance was shown to promote fewer escapes in Chinese cabbage (Konagaya et al., 2013). By utilising this more efficient selection mechanism, the growth of edited tissues containing a KI is likely enhanced due to the absence of escape tissues (eradication of competition between the two tissue types), elevating successful recovery of KIs in a degron-independent manner. Additionally, we hypothesised that escapes formation could be triggered by T-DNA-based transient expression of the selectable marker shortly after transformation. Such transient expression could not have occurred in Huang et at (2021) as the two base pair mutations conferring herbicide resistance are being knocked in at the *SuRB* locus, thus herbicide resistance could not be provided by transient T-DNA expression, and thus limits the formation of escape tissue. Overall, this suggests that a degron-based approach might be helpful to boost the recovery of KIs in selection systems with high occurrence of escape tissues and shoots, such as the NPTII system, along with methodologies where, due to the selectable marker being contained on the donor template, transient expression of the latter can occur and sustain the development of escape tissues. This emphasises the importance of efficient and precise selection of cells containing a

KI from the rest of the callus mass and that some gene targeting methodologies may benefit from optimising the transgenic selection step to enhance the recovery of KIs.

Using *ANT1* as the landing pad in conjunction with anthocyanin overproduction as a visual maker was both successful, efficient, and convenient with regards to enabling a screen for desired KI events, as reported elsewhere (Čermák et al., 2015; Vu et al., 2020c). This approach conferred a clear and robust phenotype upon DNA integration and enabled an early assessment (i.e. four to six weeks post *Agro* transformation of the explants) of KI efficiency induced by a particular variable. Thanks to purple pigmentation, edited callus cells could be specifically isolated and regenerated into shoots harbouring a KI. Nevertheless, significant drawbacks were encountered with regards to the tissue culture regeneration of anthocyanin-overexpressing tissues containing a KI. As previously discussed (*c.f.* Chapter 3, section 3.13), *ANT1* over-expression may cause a metabolic drain on the cells. As such, the KI protocol developed here may not enable viable and reproducible production of genome edited cultivars due to the low proportion of purple sectors that will develop into a viable plant. This phenomenon was also observed by Vu et al. (2020), although attributed to the effect of the geminivirus replicon rather than anthocyanin: "the regeneration of the purple calli into plants was not completely proportional probably due to pleiotropic impacts of the new replicon systems". Improving purple sector regeneration into shoots and plants is the next key step for the betterment of our KI protocol. To be noted, selectable-marker free KIs at the *ANT1* locus could not be observed, likely caused by out-competition of edited cells by wild-type cells (Čermák et al., 2015). Thus, unless efficiencies are substantially elevated, KI strategies that target *ANT1* are likely to be always constrained to also rely on chemical selection of transformed cells, which is undesirable in genetically engineered products (Yau and Stewart, 2013), and will subsequent steps to remove the selectable marker gene from the edited plant with site-specific nucleases. In conclusion, the *ANT1* locus has the potential to provide an efficient landing pad for KI that facilitates comparing the efficiency of various KI design, but further fine-tuning of the expression of *ANT1* will be required to prevent edited tissues being compromised during development. This will be discussed in section 6.1.4.

Other visual phenotypes for successful KI indication included the *carotenoid isomerase* locus, restoring red fruit colour of tomatoes (Dahan-Meir et al., 2018). Because the phenotype corresponding to successful KI is restricted to the fruits, phenotype-based assessment of the KI efficiency can only be undertaken several months after transformation, as opposed to several weeks with *ANT1*. Strategies relying on herbicide resistant callus to select and detect KIs are

designed to replace (*via* the KI) a few base pairs of an endogenous gene which will confer resistance to a given active compound e.g. *ALS* locus, (Danilo et al., 2019; Huang et al., 2021) OR *protoporphyrinogen oxidase,* (*PPO)* (De Pater et al., 2018). As the required mutations for chemical resistance occurs within the coding sequence of the gene, it does not easily enable the use a donor template which would contain both the chemical resistance-conferring sequence alongside a novel gene of interest to be knocked in. Such a donor template would therefore lead to the insertion of the gene of interest in the coding sequence of the endogenous selectable marker gene, disrupting the expression of the latter and disabling its selectable marker capacity. Although a useful proof of concept approach for gene allele replacement, this endogenously based selectable marker gene strategy to select for KI events does not appear compatible with inserting novel genes. Therefore, *ANT1* selection for KI detection were deemed most appropriate as they allowed screening of young calli and visual inspection for KI.

HR has been widely acknowledged as challenging to harness for gene targeting purposes and substantial trouble-shooting and fine-tuning of methodologies have been necessary to recently reach KI efficiencies ranging between 6 to 30% (Čermák et al., 2015; Dahan-Meir et al., 2018; Barone et al., 2020; Huang et al., 2021; Vu et al., 2020; this thesis). While HR has been considered the optimal repair pathway for KI, alternative DNA repair pathways, such as NHEJ and MMEJ (introduced in Chapter 1), could be suitable to support site-specific DNA insertions. For instance, glyphosate-resistant rice calli were produced employing an NHEJ-based sequence targeted insertion at the *EPSPS* locus with 2.2% efficiency for a 1.6 kb insert (Li et al., 2016). In another study in rice, targeted insertions of DNA fragments ranging from 26 bp to 2,049 bp were obtained *via* NHEJ at frequencies ranging from 3.9% and 47.2% (Lu et al., 2020). In human cells, NHEJ enabled the targeted insertion of large DNA inserts, 12 kb and 34 kb, at the efficiency of 7.49% and 1.18% respectively (He et al., 2016). Nevertheless, it is worth noting that NHEJ does not ensure directional insertion of the fragment, which can be problematic if performing a gene replacement, or if requiring an in-frame insertion and therefore may not be suitable for all applications.

MMEJ-based KI approaches have shown promising results in mammalian systems (Nakade et al., 2018; Aida et al., 2016) and preliminary experiments in our lab have demonstrated the feasibility of this approach in plants. Although MMEJ-based KIs in plants have not been published yet, several publications have revealed the potential of MMEJ for targeted genomic deletion and translocations in plants (Beying et al., 2020; Weiss et al., 2020; Tan et al., 2020). Relying on MMEJ for targeted sequence insertion has the advantages of simplifying plasmid

design and cloning as the microhomologies of ~25 bp can be added by PCR to existing modules if using the Golden Gate cloning approach. Reducing the size of the homologies also enables greater size allocation on the plasmid for more DNA sequences to be knocked in. Besides, PCR characterisation of KI events may also be simpler as shorter PCRs will be required to amplify the KI junctions (as opposed to PCRs spanning across ~900 bp of homology sequences). Success rate of generating shorter amplicons containing the KI junctions should be enhanced and subsequent analysis thus made simpler.

Taken together, the findings gathered in this thesis help the construction of a consolidated protocol to achieve high KI efficiency of larger DNA fragments, such as genes, in tomato.

6.1.3 Resistance genes can be knocked in with high efficiency at the *ANT1* locus

I have shown that large DNA inserts (7.3 kb) can be knocked in with precision, and at high efficiency, at the tomato *ANT1* locus. Based on these data, it seems that insert length need not negatively impact the success rate of KIs as 2.4 kb and 7.3 kb donor templates were inserted with equivalent efficiency (Chapter 5). Additional experiments are necessary to further validate that claim, for instance, rates of KI of a 7.3 kb DNA fragment should be tested at a different locus to assess reproducibility. To the best of our knowledge, the present study is the second which has used CRISPR-Cas HR-induced KIs to insert an entire gene to confer an agronomically relevant trait into a crop species (*c.f.* Dong et al., 2020). These encouraging findings should support the transition from proof-of-concept studies to translational applications in the plant gene targeting field as it offers unprecedented advantages compared to other available methods to breed disease resistant varieties. Deploying durable disease resistance in crops is challenging with conventional breeding methods as several resistance genes and quantitative trait loci are usually required to achieve this goal. This incurs complex and laborious crossbreeding over several years. Moreover, due to linkage drag, introgression of new sources of resistance into elite cultivars can have undesirable repercussions, such as yield loss (Esse et al., 2020). Using classical genetic engineering methods, one or more genes conferring desirable traits can be introduced into a cultivar at once, but the lack of control over the genomic insertion sites creates other issues. Transgene expression is affected by the genomic context and therefore is variable depending on the insertion site (Butaye et al., 2005), and reciprocally, exogenous DNA has the potential to alter the expression of endogenous genes if inserted within a coding sequence (Bouché and Bouchez, 2001). Extensive testing is therefore necessary to ensure correct expression of both the transgene and the endogenous

genes neighbouring the insertion site. The drawbacks above-mentioned are largely revoked with the use of targeted gene KIs as targeted insertion of one or more genes could be focused to pre-established genomic "safe harbours"(Dong et al., 2020). The research pipeline could therefore be speeded up as fewer screening and verification steps may be necessary when using an optimised targeted knock-in protocol. With a standard GM approach, transgenes insertions will be scattered across the genome, requiring extensive downstream screening of multiple transgenic lines and transgenic events in order to establish location of transgene (ensure no insertion in an endogenous gene or regulatory element) and expression level of the transgene.

Besides, targeted KIs offer the possibility to stack several genes at one locus through sequential transformations. This should facilitate breeding of multi-genes by simplifying the segregation and tracking of the traits-conferring stack in the progeny of the parental line. Again, this facilitates the downstream screening process to breed an elite cultivar as these transgenes will be co-segregating as one locus in the progeny of the T0 edited line. With a standard GM approach, genes of interest can also segregate as one locus as long as the transgenes are delivered on the same plasmid, as one unit. This approach is therefore limited by the number of genes than be contained on a vector, with a likely trade-off between plasmid size and transformation efficiency. Due to the lack of control over insertion site, standard GM techniques do not enable subsequent addition of beneficial genes at a previously engineered locus. Transgenes added subsequently will follow their own segregation pattern which will require more complex genetic analyses to ensure loci containing transgenes are present in progenies of that cultivar.

Targeted KIs can also promote the creation of new alleles by performing an allele switch, whereby only few base pairs are being altered in an endogenous gene based on the sequence provided on a repair template. Such outcome is not feasible using standard GM techniques. At present, plant products obtained via site-directed nucleases are to be regulated under the EU GMO legislation (discussed in more detail in section 6.2.2.2) and would therefore require lengthy and costly safety assessments before being released on the market, preventing faster delivery of improved crops to growers in Europe. Nevertheless, allele replacements consisting of few base pair changes, not introducing foreign DNA into the genome, may become exempt from the GMO legislation which would help promoting their development in Europe (EFSA, 2020). Such products are also not to be regulated in the U.S. (USDA APHIS, 2020) on the

basis that such genetic outcomes could have also arisen from mutation breeding or conventional breeding, which are considered safe breeding techniques.

Another advantage over a standard GM plant product is that KI lines can be made T-DNA-free by removing any *Agrobacterium*-derived sequences i.e. left/right border, or other transformation-associated DNA i.e. selectable markers by crossing them out (provided no T-DNA inserts are linked to the targeted locus). This would not be feasible in a GM product as the transgene is contained on the T-DNA. This may mostly be favourable in a regulatory context for commercialisation of a novel product. For instance, a product free from foreign DNA sequences e.g. *Agrobacterium* should be deregulated in the U.S. (USDA APHIS, 2020). A plant product resulting from a targeted KI which inserted a gene which already occurs in the gene pool of that plant (i.e. within the species and/or sexually compatible species), provided no foreign DNA remains, would be exempted from regulation - also on the basis that such an outcome could have occurred via traditional breeding.

To be noted, the KI tomato line produced in the present study would indeed qualify as a GMO and would be regulated as such in most countries due to the presence of foreign DNA: (1) *Rpi-vnt1*, although isolated from a solanaceous, does not occur in the gene pool of *Solanum lycopersicum*, (2) the 35S promoter knocked in upstream of *ANT1* derives from the CaMV virus, (3) random T-DNA inserts are yet to be crossed out to produce T-DNA and CRISPR-Cas free lines. This shows that from a regulatory standpoint, many targeted KI derived-plant products are no different from plant products based on standard GM technology and offer little advantage over the latter. Nevertheless, countries around the world are gradually updating their GMO legislation to take into account products derived from targeted genome editing and suitably deregulate products where the genetic outcome could have occurred by conventional mutagenesis or breeding (Turnball et al., 2021). This is where targeted KIs have a higher potential than standard GMOs to be bred and commercially released quicker to meet the rapidly evolving demand for improved crop varieties.

Other methods exist to create site-specific gene insertions and rely on recombinases such as the Cre-*Lox*, flipase-flipase (FLP-*FRT*) and R-*Rs* systems, which recombine DNA sequences situated between the recombinase i.e. Cre, FLP, respective recognition sequences i.e. *Lox*, *FRT*, *Rs*. Efficiencies of site-specific integration of single transgenes and transgene stacks range between 2% and 7% in plants (Nanto et al., 2005; Anand et al., 2019; D'Halluin et al., 2013; Hou et al., 2014). The limitation of this approach is the unavoidable presence of the

recombinase recognition sequences in the genome of the recipient cultivar, which restricts genomic locations suitable for site-directed insertion.

6.1.4 Future work

The immediate next steps of this research project include verification of the heritability of the KI containing *Rpi-vnt1* at the *ANT1* locus and assessment of Mendelian segregation in the T1 progeny, phenotypically and by PCR. In T-DNA-free progeny, expression of the transgene *Rpi-vnt1* should be assessed, followed by a disease resistance assay by inoculating leaves with a *Phytophthora* isolate resisted by *Rpi-vnt1* to verify its function.

In future experiments, the Omega leader sequence could be removed from the 35S promoter to reduce *ANT1* overexpression, reducing the selective disadvantage of cells carrying events of interest. Moreover, it would be interesting to test the suitability of other promoters to trigger *ANT1* expression and purple tissue pigmentation without compromising plant fitness, as was demonstrated from a *DWARF4 (DWF4) promoter::ANT1* visual selectable marker cassette in tomato (Jin et al., 2012). Due to the tissue-specific expression pattern of *DWF4*, restricted to cotyledons and actively growing tissue (Kim et al., 2006), the *DWF4* promoter could be knocked in alongside a gene of interest upstream on *ANT1*. Purple pigmentation would be retained for scoring and isolating purple sectors but the lower levels of anthocyanin would not cause any adverse effect on the plant. Moreover, once reaching maturity, *ANT1* expression would cease in tissues, potentially restoring a green phenotype to the plant.

Additional variables to test in the future include the use of MMEJ to perform the KI for the reasons explained above. Moreover, successful KI strategies developed in mammalian systems have the potential to be translated to plants to further elevate rates of KI. An example of that is the application of a tethering approach by linking the Cas9 nuclease to the donor template to ensure physical proximity between the DSB and the repair template containing the desirable sequence. Initially developed in human cells (Savic et al., 2018; Gu et al., 2018), this approach led to 20% KI efficiency when adapted into rice (Ali et al., 2020b). However, this attempt focused on an allele exchange at the *ALS* locus which involved a short sequence insertion. It would be interesting to test the feasibility of this approach with long DNA donor templates.

Continuation of this research project would initially involve the removal of the *NptII* selectable marker gene from the *ANT1* locus in the edited purple T1 samples. Flanked by LoxP sites, *NptII* excision can be mediated by crossing the edited purple T1 line to a transgenic Moneymaker line which carries the associated Cre recombinase to recover *NptII*-free offspring.

Subsequent KIs of *R* genes could then be attempted at the *ANT1* locus to produce an *R* gene stack.

Remaining and arising questions relating to the targeted KI methodology developed in this study include the following. How much did the use of the degron impact the rates of KI observed in Chapter 5? Could relying on a more stringent selectable marker gene than *NptII* eliminate the need to use the degron? Would the degron still be required if exploiting another landing pad conferring, upon KI, a phenotype bearing no cost to the fitness of edited cells? Could the degron be a suitable strategy to improve recovery of KIs in other crop species? With regards to using anthocyanin accumulation as a visual marker, how much growth impairment results from *ANT1* overexpression in edited cells? What is the exact cause of the low proportion of purple sectors regenerating into purple shoots? How could we increase successful shoot regeneration from purple sectors? Ultimately, could the KI rates observed at the *ANT1* locus be recapitulated at a different locus?

6.2 Potential and limitations facing the Gene Targeting field

Recent improvement in gene targeting efficiency in plants - now reaching KI rates above 20% in tobacco, rice and tomato - are representative of the efforts directed at the improvement of this technology which are starting to come to fruition. But what can we hope to gain from gene targeting for crop trait improvements, what should be its technical improvement roadmap, and what are the main hurdles to developing this technology?

6.2.1 The potential

Although the technological advancement and refinement of the genome editing toolkit already offers a wealth of opportunities for crop trait improvement *via* gene knock-outs, base-pair editing, and prime editing (*c.f.* Chapter 1, section 1.3.3.), some traits may only be gained from the insertion of entire genes, and potentially gene networks. Gene targeting has the potential to induce a leap forward in harnessing plant synthetic biology to benefit agriculture in several ways. Development of 'smart plants', displaying engineered resilience to adverse environmental factors, could help curb crop losses. Specifically, drought stress is thought to be the most prevalent abiotic stress with potentially catastrophic outcomes on crop production (Seleiman et al., 2021). Biotechnological solutions could include, for example, the use of transgenic expression of synthetic receptors, to artificially activate the plant's abscisic acid-controlled drought tolerance pathway to tightly control water use, enabling plant survival during a drought (Park et al., 2015). Additionally, sustainability of agricultural practices could be improved. Application of synthetic nitrogen fertiliser can be damaging to the environment.

Over half of the nitrogen used in cereal production escapes into the environment (Raun and Johnson, 1999), causing atmospheric pollution in the form of potent greenhouse gases like nitrous oxides or causing waterway pollution after leaching as soluble nitrates (Glendining et al., 2009). Furthermore, the industrial production of nitrogen fertiliser is one of agriculture's largest uses of fossil fuel, and a significant source of greenhouse gases emission. Reducing reliance on nitrogen fertilisers is therefore of paramount importance. An ambitious biotechnological approach is aiming to transfer the legume-rhizobium symbiosis pathway into cereal crops to enable biological nitrogen fixation (Rogers and Oldroyd, 2014). Gene targeting could pave the way for reliable and controlled implementation of these biotechnological solutions into elite cultivars.

Aspirations for the gene targeting field include the development of methodologies that do not rely on a selectable marker gene to selected for targeted gene insertions. Although considered safe for environmental and human health by the European Food Safety Authority ((EFSA), 2004, 2009), the presence of selectable marker genes such as *NptII* in genetically modified crops should be avoided as it raises concerns from regulatory bodies, such as the European Commission, and from the public. Moreover, some selectable marker genes are not permitted in commercialised genetically modified crops. KI attempts free of selectable marker gene at the targeted locus have had variable success: 0.66% (Vu et al., 2020c) and 25% (Dahan-Meir et al., 2018) KI efficiency in tomato, 6.25% in rice (Dong et al., 2020), 8% in Arabidopsis (Miki et al., 2018) and 4.7% in maize (Barone et al., 2020). Such endeavours will require a high baseline KI efficiency in order to recover KI events without chemical selection. Achieving this may require additional efforts to unravel the HR-mediated repair mechanism in plants, as a more in depth understanding of the pathway could bring about more innovations to tackle the system's limiting factors to high efficiency gene targeting. For instance, the ultimate rewarding discovery would be the identification of mitotic recombination hotspots in plant somatic cells, where occurrence of HR is prevalent upon DSB induction. Such sites could become ideal landing pads.

6.2.2 Limitations

6.2.2.1 Transformation and tissue culture methods are a bottleneck to gene editing

Another significant bottleneck to the effective deployment of genome editing by gene targeting are plant transformation and regeneration of edited plants by *in vitro* tissue culture. Despite 40 years of research into transformation technologies, transformation and regeneration remains challenging for most crops (Altpeter et al., 2016). Utilisation of *Agrobacterium* for

transformation is limited by the recalcitrance of many plants to this method, as cell susceptibility to this bacterium is a complex and variable trait (Lacroix and Citovsky, 2019). Therefore, transformation success can be very genotype-dependent within a single plant species (Nam et al., 1997) and require extensive protocol elaboration to obtain satisfactory transformation rates. This excludes many crop species from the potential of genome editing for improvements. Employment of biolistic bombardment to introduce the CRISPR-Cas reagents inside cells can overcome this constraint as bombardment is amenable to a wider range of species (Altpeter et al., 2005), but the bottleneck of low *in vitro* regeneration ability displayed by many plants remains. Besides, tissue culture practices introduce DNA mutations in the genome (Phillips et al., 1994), and as well as epigenetic mutations (Miguel and Marum, 2011), which may be linked to unintended phenotypes regularly observed in transgenic plants (Latham et al., 2005).

Thus, there is a need to develop transformation protocols that are genotype-independent and that do not rely on *in vitro* tissue culture regeneration to broaden the range of crop species that can benefit from gene editing technology. To this end, researchers have developed strategies involving plant developmental regulators to bypass the need for tissue culture. Remarkably, by co-delivering the CRISPR reagents with a mix of developmental regulators to *N. benthamiana* seedlings i.e. WUSCHEL, BABYBOOM, SHOOT MERISTEMLESS, isopentenyl transferase, researchers were able to trigger *de novo* formation of meristems on cotyledons, which developed into gene edited shoots containing a heritable mutation (Maher et al., 2020). This strategy was also demonstrated on soil-grown plants and on several crops e.g. tomato, potato and grape. Although this study relied on *Agrobacterium* delivery of the components, other delivery methods could be used to broaden the range of plants where this can be applied. For instance, advancements in nanotechnology have recently shown the feasibility of plasmid DNA delivery mediated by carbon nanotube nanoparticles (Demirer et al., 2019). Interestingly, this strategy resulted in high efficiency transient transgene expression without causing its stable integration and was successfully applied in three unrelated plants, arugula, wheat and cotton.

6.2.2.2 The regulatory landscape of gene edited crops is a major hurdle to technology development in Europe

Gene editing has the potential to revolutionise plant breeding and agriculture, but achieving this goal is ultimately dependent upon appropriate regulation of gene edited crops and the regulatory landscape for biotech crops is a complex one. Indeed, different countries have built

different regulatory frameworks around genetic engineering, imparting different legal definitions of genetically modified organisms (GMOs) and different regulations. Currently, such regulatory frameworks are either "product-based" where the trigger for regulation is based on the safety risks posed by the product (e.g. in United States, Argentina, Canada), or "process-based", where the trigger for regulation is focused on the techniques employed to create a new variety (e.g. in Japan, New Zealand, China, European Union, Australia). These regulations were implemented with the aim to protect both the environment and human health from unintended consequences linked to the release and consumption of GMOs.

In the European Union (EU), the regulations subjected to GMOs are set out in the Directive 2001/18/EC and Directive 2009/41/EC, but dating from pre-2010, these directives do not explicitly accommodate gene edited crops, casting legal uncertainty over the regulation of products resulting from new plant breeding techniques (NPBTs) that involve use of site-directed nucleases (SDNs), such as CRISPR-Cas. Half a decade ago, several authorities expressed a favourable opinion for the adoption of gene edited crops that do not contain exogenous DNA, either considering them similar to those derived by classical breeding methods, or proposing adapted regulations (Sprink et al., 2016). Risks associated with varieties developed using NPBTs were found to be equivalent to that of conventional breeding (EASAC, 2013) and conventional transgenesis and mutagenesis (EFSA, 2012). Nevertheless, following the ruling from the European Court of Justice (ECJ) in July 2018, gene edited crop varieties produced with NPBT fall within the scope of the Directive 2001/18/EC, and therefore "constitute GMOs within the meaning of that provision" as stated in the Judgment of the Court, Case C-528/16, paragraph 54. This decision creates a considerable disadvantage to breeding and biotech companies. In a recent survey of private breeding companies, the top factor seen as the main hurdle to using new plant breeding techniques was the regulation of products generated with such techniques under the Directive 2001/18/EC since the 2018 ECJ ruling (Jorasch, 2020). Negative impacts prompted by the ruling included: prohibitive costs, lengthy timelines (assessment, approval) and labelling requirements, which incited companies to reduce R&D activities and investments, at least for products destined to the EU market. Besides, an overwhelming majority of the breeding companies reported that, should gene edited plants not fall under GMO regulations, they would increase their R&D investments into NPBTs, highlighting the important potential associated with these technologies (Jorasch, 2020).

Although there has been disagreement between competent European authorities and non-governmental organisations over the legal interpretation of the Directive 2001/18/EC on whether the regulation is product-based or process based (Sprink et al., 2016; Meer et al., 2021), the interpretation of the Directive 2001/18/EC has mostly been interpreted as strictly process-based (Krämer, 2015), which explains the outcome of the 2018 ECJ ruling with regards to crops produced with NPBTs. A process-based approach is not fit for purpose when dealing with gene edited crops. Indeed, genomic outcomes generated by NPBTs are classed into three types: type SDN-1, which contains few base pair mutations, insertion/deletion without the use of a homologous template; type SDN-2, which displays a few base pair mutations or insertion (<20 bp) provided on a repair template; and the last category, SDN-3, for events which involve targeted insertion of larger fragments of exogenous DNA *via* HR or NHEJ (Podevin et al., 2013). The nature of the genetic modifications incurred by SDN-1 and SDN-2 are equivalent to those occurring naturally or induced by chemical or radiation mutagenesis, making gene edited products arising from these techniques indistinguishable from varieties generated by conventional breeding or mutagenesis. This raises the problem of traceability, labelling, and subsequent international trade, as technically impossible to discriminate conventional products from gene edited ones, placing such products in a regulatory impasse (Purnhagen et al., 2018). On these grounds, it appears that an exclusively process-based interpretation of the GMO legislation is rationally flawed and incoherent from a scientific viewpoint. From a legislative perspective, the decision of the 2018 ECJ ruling very much aligns with the Precautionary Principle followed by the EU with regards to agricultural biotechnology (Vives-Vallés and Collonnier, 2020). Nevertheless, it has been argued that the Precautionary Principle has been used inadequately to guide current GMO legislation, and that the precaution to risk appraisal of NPBTs is disproportionate and largely exaggerated (Jouanin et al., 2018; Zhang et al., 2020a; Aerni, 2019; Jones, 2019), based on the accumulating evidence over the safety of genetically modified crops over the last 20 years (Pellegrino et al., 2018; Klümper and Qaim, 2014; ICGEB).

Fortunately, several countries, some with a process-based approach to GMO regulatory triggers, have already exempted gene edited crops resulting from SDN-1 and SDN-2 from the scope of GMO regulation (e.g. US, Japan, Argentina, Brazil, Chile, Colombia, Japan, Israel). A glimmer of hope for a more logical and science-based regulation of gene edited crops in the EU has recently emerged. In a study commissioned by the ECJ on the status of NPBTs in relation to the ECJ judgment in case C-528/16, the European Commission (EC) produced a report which highlight the shortcomings of the current legislation and provides a proposal for

suitable measures to be implemented (European Commission, 2021). Importantly, the EC will pioneer the initiative to draw up a targeted policy action for gene edited crops, with "proportionate regulatory oversight" and adaptation of the risk assessment, authorisation, labelling procedures. Interestingly, the EC included products resulting from SDN-1 and SDN-2 to qualify for exemption, but also included some SDN-3 events, on the condition that they result in cisgenesis i.e. insertion of a gene, unaltered, from a sexually compatible species to the recipient organism, with absence of plasmid DNA insertion. Regulation of cisgenesis products would follow on case-by-case basis, based on the potential risks associated with the gene, or genes, being integrated. Amending current EU GMO regulations based on these recommendations could pave the way for the deployment of gene targeting to produce improved varieties by cisgenesis without excessively prohibitive regulations.

Additionally, the role of NPBTs in improving the sustainability of agricultural systems (key objective of the EU Green Deal and Farm to Fork strategy) was emphasised in the EC study, which is a crucial factor to be conveyed during public outreach activities as it may assist public acceptance and endorsement of gene edited crops. Furthermore, the Department for Environment and Rural affairs launched earlier this year a public consultation regarding the regulation of gene edited products to assess the feasibility of their deregulation in the UK in a post-Brexit context (DEFRA, 2021). Also earlier this year, France took a stance with the Agriculture Minister announcing that crops developed with NPBTs should not be regulated as GMOs (Reuters, 2021), which is a first for the country which has a strong anti-GMO history (Seifert, 2020). Overall, an optimistic outlook for the future of gene edited crops future in the EU is appearing, but nothing is set in stone yet. Concerted efforts should be made to ensure the harmonisation and clarification of GMO/NPBTs products regulations amongst Member States of the EU.

6.2.3 Outstanding questions to the gene targeting field

To conclude, here are some the remaining questions to the gene targeting field. To what extent can we further boost the rates of KIs in plants? Which aspect of HR is the major barrier to unlocking high frequency KI? Different plant species seem to have disparate abilities to perform high efficiency KIs; the moss *Physcomitrella patens* can reach KI efficiency up to 100% (Collonnier et al., 2017), whereas Arabidopsis gene targeting experiments never exceed 9% KI efficiency. What causes these apparent intrinsic differences in plant species capability to facilitate KIs? Unravelling differences in the HR pathway between *P. patens* and other plants may shed light on new ways to improve KI efficiencies in other plant species. Are some regions

of the genome more amenable to high efficiency KI due to being more prone to being repaired by HR? What is the largest DNA fragment that can be knocked in with high efficiency? These important questions remain to be answered in order to truly unlock the potential of gene targeting as a robust and powerful crop trait improvement tool.

Bibliography

(EFSA), E.F.S.A. (2009). Consolidated presentation of the joint Scientific Opinion of the GMO and BIOHAZ Panels on the "Use of Antibiotic Resistance Genes as Marker Genes in Genetically Modified Plants" and the Scientific Opinion of the GMO Panel on "Consequences of the Opinion o. EFSA J. **7**: 1108.

(EFSA), E.F.S.A. (2004). Opinion of the Scientific Panel on Genetically Modified Organisms on the use of antibiotic resistance genes as marker genes in genetically modified plants. EFSA J. **2**: 1–48.

Acquaah, G. (2012). Principles of Plant Genetics and Breeding.

Adegbeye, M.J., Kanth, P.R., Obaisi, A.I., Elghandour, M.M.M.Y., Oyebamiji, K.J., Salem, A.Z.M., Morakinyo-fasipe, O.T., and Cipriano-salazar, M. (2020). Sustainable agriculture options for production, greenhouse gasses and pollution alleviation, and nutrient recycling in emerging and transitional nations - An overview. J. Clean. Prod. **242**: 118319.

Aerni, P. (2019). Politicizing the Precautionary Principle: Why Disregarding Facts Should Not Pass for Farsightedness. Front. Plant Sci. **10**: 26–29.

Aida, T., Nakade, S., Sakuma, T., Izu, Y., Oishi, A., Mochida, K., Ishikubo, H., Usami, T., Aizawa, H., Yamamoto, T., and Tanaka, K. (2016). Gene cassette knock-in in mammalian cells and zygotes by enhanced MMEJ. BMC Genomics **17**: 1–18.

Aird, E.J., Lovendahl, K.N., Martin, A.S., Harris, R.S., and Gordon, W.R. (2018). Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. Commun. Biol.

Albert, N.W., Davies, K.M., Lewis, D.H., Zhang, H., Montefiori, M., Brendolise, C., Boase, M.R., Ngo, H., Jameson, P.E., and Schwinn, K.E. (2014). A Conserved Network of Transcriptional Activators and Repressors Regulates Anthocyanin Pigmentation in Eudicots. Plant Cell **26**: 962–980.

Alfenito, M.R., Souer, E., Goodman, C.D., Buell, R., Mol, J., Koes, R., and Walbot, V. (1998). Functional Complementation of Anthocyanin Sequestration in the Vacuole by Widely Divergent Glutathione S -Transferases. **10**: 1135–1149.

Ali, Z., Shami, A., Sedeek, K., Kamel, R., Alhabsi, A., Tehseen, M., Hassan, N., Butt, H., Kababji, A., Hamdan, S.M., and Mahfouz, M.M. (2020a). Fusion of the Cas9

endonuclease and the VirD2 relaxase facilitates homology-directed repair for precise genome engineering in rice. Commun. Biol. **3**.

Ali, Z., Shami, A., Sedeek, K., Kamel, R., Alhabsi, A., Tehseen, M., Hassan, N., Butt, H., Kababji, A., Hamdan, S.M., and Mahfouz, M.M. (2020b). Fusion of the Cas9 endonuclease and the VirD2 relaxase facilitates homology-directed repair for precise genome engineering in rice. Commun. Biol. **3**.

Altpeter, F. et al. (2016). Advancing Crop Transformation in the Era of Genome Editing. Plant Cell **28**: 1511–1520.

Altpeter, F. et al. (2005). Particle bombardment and the genetic enhancement of crops : myths and realities. Mol. Breed. **15**: 305–327.

Anand, A., Wu, E., Li, Z., TeRonde, S., Arling, M., Lenderts, B., Mutti, J.S., Gordon-Kamm, W., Jones, T.J., and Chilcoat, N.D. (2019). High efficiency Agrobacterium-mediated site-specific gene integration in maize utilizing the FLP-FRT recombination system. Plant Biotechnol. J. **17**: 1636–1645.

Angstenberger, M., Signori, F. De, Vecchi, V., Osto, L.D., and Bassi, R. (2020). Cell Synchronization Enhances Nuclear Transformation and Genome Editing via Cas9 Enabling Homologous Recombination in Chlamydomonas reinhardtii. ACS Synth. Biol. **9**: 2840–2850.

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., & A.R., and Liu, D.R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature **566**: 235–238.

Atkins, P.A. and Voytas, D.F. (2020). Overcoming bottlenecks in plant gene editing. Curr. Opin. Plant Biol. **54**: 79–84.

Atkinson, J.A., Rasmussen, A., Traini, R., Voß, U., Sturrock, C., Mooney, S.J., Wells, D.M., and Bennett, M.J. (2014). Branching out in roots: Uncovering form, function, and regulation. Plant Physiol. **166**: 538–550.

Audsley, E., Stacey, K., Parsons, D.J., and Williams, A.G. (2009). Estimation of the greenhouse gas emissions from agricultural pesticide manufacture and use.

Aymard, F., Bugler, B., Schmidt, C.K., Guillou, E., Caron, P., Briois, S., Iacovoni, J.S., Daburon, V., Miller, K.M., Jackson, S.P., and Legube, G. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nat. Struct. Mol. Biol. **21**: 366–374.

Bachmair, A., Finley, D., and Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. Science (80-.). **234**: 179–186.

Baltes, N.J., Gil-Humanes, J., Cermak, T., Atkins, P.A., and Voytas, D.F. (2014). DNA Replicons for Plant Genome Engineering. Plant Cell **26**: 151–163.

Baltes, N.J., Gil-Humanes, J., and Voytas, D.F. (2017). Chapter One - Genome Engineering and Agriculture: Opportunities and Challenges. In Gene Editing in Plants, D.P. Weeks and B. Yang, eds, Progress in Molecular Biology and Translational Science. (Academic Press), pp. 1–26.

Baltes, N.J. and Voytas, D.F. (2015). Enabling plant synthetic biology through genome engineering. Trends Biotechnol.

Barakate, A., Keir, E., Oakey, H., and Halpin, C. (2020). Stimulation of homologous recombination in plants expressing heterologous recombinases.: 1–10.

Barkal, A.A., Srinivasan, S., Hashimoto, T., Gifford, D.K., and Sherwood, R.I. (2016). Cas9 functionally opens chromatin. PLoS One **11**.

Barone, P., Wu, E., Lenderts, B., Anand, A., Gordon-Kamm, W., Svitashev, S., and Kumar, S. (2020). Efficient Gene Targeting in Maize Using Inducible CRISPR-Cas9 and Marker-free Donor Template. Mol. Plant **13**: 1219–1227.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science (80-.). **315**: 1709–1712.

Bartlett, J.G., Alves, S.C., Smedley, M., Snape, J.W., and Harwood, W.A. (2008). High-throughput Agrobacterium-mediated barley transformation. Plant Methods **4**: 1–12.

Bastet, A., Zafirov, D., Giovinazzo, N., Guyon-debast, A., and Nogu, F. (2019). Mimicking natural polymorphism in eIF4E by CRISPR-Cas9 base editing is associated with resistance to potyviruses. Plant Biotechnol. J.: 1736–1750.

Beck, E., Ludwig, G., Aucxswald, E.A., Reiss, B., and Schaller, H. (1982). Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. Gene **19**: 327–336.

Begemann, M.B., Gray, B.N., January, E., Gordon, G.C., He, Y., Liu, H., Wu, X., Brutnell, T.P., Mockler, T.C., and Oufattole, M. (2017a). Precise insertion and guided editing of higher plant genomes using Cpf1 CRISPR nucleases. Sci. Rep. **7**: 1–6.

Begemann, M.B., Gray, B.N., January, E., Singer, A., Kesler, D.C., He, Y., Liu, H., Guo,

H., Jordan, A., Brutnell, T.P., Mockler, T.C., and Oufattole, M. (2017b). Characterization and Validation of a Novel Group of Type V, Class 2 Nucleases for in vivo Genome Editing. bioRxiv.

Behrens, C.E., Smith, K.E., Iancu, C. V., Choe, J. yong, and Dean, J. V. (2019). Transport of Anthocyanins and other Flavonoids by the Arabidopsis ATP-Binding Cassette Transporter AtABCC2. Sci. Rep. **9**: 1–15.

Bernabé-Orts, J.M., Casas-Rodrigo, I., Minguet, E.G., Landolfi, V., Garcia-Carpintero, V., Gianoglio, S., Vázquez-Vilar, M., Granell, A., and Orzaez, D. (2019). Assessment of Cas12a-mediated gene editing efficiency in plants. Plant Biotechnol. J.: 1–14.

Bernd Zetsche, Jonathan S. Gootenberg, Omar O. Abudayyeh, Ian M. Slaymaker, Kira S. Makarova, Patrick Essletzbichler, Sara E. Volz, Julia Joung, John van der Oost, A.R. and Eugene V. Koonin, and F.Z. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Mol. Cell **9**: 901–913.

Beying, N., Schmidt, C., Pacher, M., Houben, A., and Puchta, H. (2020). CRISPR–Cas9-mediated induction of heritable chromosomal translocations in Arabidopsis. Nat. Plants **6**: 638–645.

Bharat, S.S., Li, S., Li, J., Yan, L., and Xia, L. (2020). Base editing in plants: Current status and challenges. Crop J. **8**: 384–395.

Bhat, M.A., Bhat, M.A., Kumar, V., Wani, I.A., Bashir, H., Shah, A.A., Rahman, S., and Jan, A.T. (2020). The era of editing plant genomes using CRISPR/Cas: A critical appraisal. J. Biotechnol. **324**: 34–60.

Le Blanc, C., Zhang, F., Mendez, J., Lozano, Y., Chatpar, K., Irish, V., and Jacob, Y. (2017). Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. Plant J. **9**: 1–10.

Blondal, T., Gamba, C., Møller Jagd, L., Su, L., Demirov, D., Guo, S., Johnston, C.M., Riising, E.M., Wu, X., Mikkelsen, M.J., Szabova, L., and Mouritzen, P. (2021). Verification of CRISPR editing and finding transgenic inserts by Xdrop indirect sequence capture followed by short- and long-read sequencing. Methods.

Boch, J. and Bonas, U. (2010). *Xanthomonas* AvrBs3 Family-Type III Effectors: Discovery and Function. Annu. Rev. Phytopathol. **48**: 419–436.

Bothmer, A., Phadke, T., Barrera, L.A., Margulies, C.M., Lee, C.S., Buquicchio, F., Moss, S., Abdulkerim, H.S., Selleck, W., Jayaram, H., Myer, V.E., and Cotta-Ramusino, C.

(2017a). Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. Nat. Commun. **8**: 1–12.

Bothmer, A., Phadke, T., Barrera, L.A., Margulies, C.M., Lee, C.S., Buquicchio, F., Moss, S., Abdulkerim, H.S., Selleck, W., Jayaram, H., Myer, V.E., and Cotta-Ramusino, C. (2017b). Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. Nat. Commun. **8**.

Bouché, N. and Bouchez, D. (2001). Arabidopsis gene knockout: phenotypes wanted. Curr. Opin. Plant Biol. **4**: 111–117.

Budhagatapalli, N., Rutten, T., Gurushidze, M., Kumlehn, J., and Hensel, G. (2015). Targeted Modification of Gene Function Exploiting Homology-Directed Repair of TALEN-Mediated Double-Strand Breaks in Barley. G3: Genes|Genomes|Genetics **5**: 1857–1863.

Burgio, G. and Teboul, L. (2020). Anticipating and Identifying Collateral Damage in Genome Editing. Trends Genet. **36**: 905–914.

Butaye, K.M.J., Cammue, B.P.A., Delaure, S.L., and Bolle, M.F.C. De (2005). Approaches to minimize variation of transgene expression in plants. Mol. Breed. **16**: 79–91.

Butler, N.M., Baltes, N.J., Voytas, D.F., and Douches, D.S. (2016a). Geminivirus-Mediated Genome Editing in Potato (Solanum tuberosum L.) Using Sequence-Specific Nucleases. Front. Plant Sci. **7**.

Butler, N.M., Baltes, N.J., Voytas, D.F., and Douches, D.S. (2016b). Geminivirus-Mediated Genome Editing in Potato (Solanum tuberosum L.) Using Sequence-Specific Nucleases. Front. Plant Sci. **7**.

Butt, H., Rao, G.S., Sedeek, K., Aman, R., Kamel, R., and Mahfouz, M. (2020). Engineering herbicide resistance via prime editing in rice. Plant Biotechnol. J. **18**: 2370–2372.

Caridi, C.P., D'agostino, C., Ryu, T., Zapotoczny, G., Delabaere, L., Li, X., Khodaverdian, V.Y., Amaral, N., Lin, E., Rau, A.R., and Chiolo, I. (2018). Nuclear F-actin and myosins drive relocalization of heterochromatic breaks. Nature **559**: 54–60.

Carroll, D. (2011). Genome engineering with zinc-finger nucleases. Genetics **188**: 773–782.

Catalá, R., Medina, J., and Salinas, J. (2011). Integration of low temperature and light signaling during cold acclimation response in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. **108**: 16475–16480.

Cebrian-Serrano, A. and Davies, B. (2017). CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. Mamm. Genome **28**: 247–261.

Čermák, T., Baltes, N.J., Čegan, R., Zhang, Y., and Voytas, D.F. (2015). High-frequency, precise modification of the tomato genome. Genome Biol. **16**: 2–3.

Cermak, T., Curtin, S.J., Gil-Humanes, J., Čegan, R., Kono, T.J.Y., Konečná, E., Belanto, J.J., Starker, C.G., Mathre, J.W., Greenstein, R.L., and Voytas, D.F. (2017). A multi-purpose toolkit to enable advanced genome engineering in plants. Plant Cell.

Cervera, M., Pina, J.A., Juárez, J., Navarro, L., and Peña, L. (1998). Agrobacterium-mediated transformation of citrange: Factors affecting transformation and regeneration. Plant Cell Rep. **18**: 271–278.

Chalfun-Junior, A., Mes, J.J., Mlynárová, L., Aarts, M.G.M., and Angenent, G.C. (2003). Low frequency of T-DNA based activation tagging in Arabidopsis is correlated with methylation of CaMV 35S enhancer sequences. FEBS Lett. **555**: 459–463.

Chalker-Scott, L. (1999). Environmental significance of anthocyanins in plant stress responses. Photochem. Photobiol. **70**: 1–9.

Charpentier, M. et al. (2018). CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. Nat. Commun. **9**.

Chawla, H.S., Cass, L.A., and Simmonds, J.A. (1999). Developmental and environmental regulation of anthocyanin pigmentation in wheat tissues transformed with anthocyanin regulatory genes. Vitr. Cell. Dev. Biol. - Plant **35**: 403–408.

Chein, Y.-H. and Davidson, N. (1978). RNA:DNA hybrids are more stable than DNA:DNA duplexes in concentrated perchlorate and trichloroacetate solutions. Nucleic Acids Res. **5**: 1627–1637.

Chen, X., Janssen, J.M., Liu, J., Maggio, I., T'Jong, A.E.J., Mikkers, H.M.M., and Gonçalves, M.A.F.V. (2017). In trans paired nicking triggers seamless genome editing without double-stranded DNA cutting. Nat. Commun. **8**.

Collonnier, C., Epert, A., Mara, K., Maclot, F., Guyon-Debast, A., Charlot, F., White, C., Schaefer, D.G., and Nogué, F. (2017). CRISPR-Cas9-mediated efficient directed mutagenesis and RAD51-dependent and RAD51-independent gene targeting in the moss Physcomitrella patens. Plant Biotechnol. J.

Colón-Carmona, A., You, R., Haimovitch-Gal, T., and Doerner, P. (1999). Spatio-temporal

analysis of mitotic activity with a labile cyclin-GUS fusion protein. Plant J. **20**: 503–508.

Le Cong, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, Xuebing Wu, Wenyan Jiang, Luciano A. Marraffini, F.Z. (2009). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (80-.).: 819–821.

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., and Zhang, F. (2017). RNA editing with CRISPR-Cas13. Science (80-.). **358**: 1019–1027.

Cruz-becerra, G. and Kadonaga, J.T. (2020). Enhancement of homology-directed repair with chromatin donor templates in cells. Elife **9**: 1–12.

Curtis, I.S., Power, J.B., and Davey, M.R. (1995). NPTII assays for measuring gene expression and enzyme activity in transgenic plants. In Methods in molecular biology (Clifton, N.J.), pp. 149–159.

Czako, M. and Marton, L. (1994). The Herpes Simplex Virus Thymidine Kinase Gene as a Conditional Negative-selection Market gene in Arabidopsis Thaliana. Plant Physiol.: 1067–1071.

D'Halluin, K., Vanderstraeten, C., Van Hulle, J., Rosolowska, J., Van Den Brande, I., Pennewaert, A., D'Hont, K., Bossut, M., Jantz, D., Ruiter, R., and Broadhvest, J. (2013). Targeted molecular trait stacking in cotton through targeted double-strand break induction. Plant Biotechnol. J. **11**: 933–941.

Dahan-Meir, T., Filler-Hayut, S., Melamed-Bessudo, C., Bocobza, S., Czosnek, H., Aharoni, A., and Levy, A.A. (2018). Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. Plant J. **95**: 5–16.

Dan, Y., Armstrong, C.L., Dong, J., Feng, X., Fry, J.E., Keithly, G.E., Martinell, B.J., Roberts, G.A., Smith, L.A., Tan, L.J., and Duncan, D.R. (2009). Lipoic acid-an unique plant transformation enhancer. Vitr. Cell. Dev. Biol. - Plant **45**: 630–638.

Danilo, B., Perrot, L., Botton, E., Nogué, F., and Mazier, M. (2018). The DFR locus: A smart landing pad for targeted transgene insertion in tomato. PLoS One **13**: e0208395.

Danilo, B., Perrot, L., Mara, K., Botton, E., Nogué, F., and Mazier, M. (2019). Efficient and transgene-free gene targeting using Agrobacterium-mediated delivery of the CRISPR/Cas9 system in tomato. Plant Cell Rep. **38**: 459–462.

Daugaard, M. et al. (2012). LEDGF (p75) promotes DNA-end resection and homologous recombination. Nat. Struct. Mol. Biol. **19**: 803–810.

Davis, L. and Maizels, N. (2014). Homology-directed repair of DNA nicks via pathways

distinct from canonical double-strand break repair. Proc. Natl. Acad. Sci. **111**: E924–E932.

DEFRA (2021). The regulation of genetic technologies.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature **471**: 602–607.

Demirer, G.S., Zhang, H., Goh, N.S., González-grandío, E., and Landry, M.P. (2019). Carbon nanotube – mediated DNA delivery without transgene integration in intact plants. Nat. Protoc. **14**.

Diamos, A.G. and Mason, H.S. (2019). Modifying the replication of geminiviral vectors reduces cell death and enhances expression of biopharmaceutical proteins in nicotiana benthamiana leaves. Front. Plant Sci. **9**.

DIssmeyer, N. (2019). Conditional Protein Function via N-Degron Pathway-Mediated Proteostasis in Stress Physiology. Annu. Rev. Plant Biol. **70**: 83–117.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature **330**: 576–578.

Dohmen, R.J., Wu, P., and Varshavskyt, A. (1994). Heat-inducible Degron : A Method for Constructing Temperature-Sensitive Mutants. Science (80-.). **263**.

Dong, O.X. et al. (2020). Marker-free carotenoid-enriched rice generated through targeted gene insertion using CRISPR-Cas9. Nat. Commun. **11**: 1–10.

Dong, O.X. and Ronald, P.C. (2021). Targeted DNA insertion in plants. PNAS **118**: 1–9.

EASAC (2013). Planting the future : opportunities and challenges for using crop genetic improvement technologies for sustainable agriculture.

EC (2020). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A Farm to Fork Strategy for a fair, healthy and environmentally-friendly food system COM/2020/381 final.

EC (2019). Communication from the Commission to the European Parliement, the Council, the European Council, the European Economic and social Committee and the Committee of the Regions. The European Green Deal. COM/2019/640 final.

EFSA (2012). Scientific opinion addressing the safety assessment of plants developed using Zinc Finger Nuclease 3 and other Site-Directed Nucleases with similar function. EFSA J. **10**: 31.

Enciso-Rodriguez, F., Manrique-Carpintero, N.C., Nadakuduti, S.S., Buell, C.R., Zarka, D., and Douches, D. (2019). Overcoming self-incompatibility in diploid potato using CRISPR-cas9. Front. Plant Sci. **10**: 1–12.

Endo, A., Masafumi, M., Kaya, H., and Toki, S. (2016). Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from Francisella novicida. Sci. Rep. **6**: 38169.

Endo, M., Ishikawa, Y., Osakabe, K., Nakayama, S., Kaya, H., Araki, T., Shibahara, K.I., Abe, K., Ichikawa, H., Valentine, L., Hohn, B., and Toki, S. (2006). Increased frequency of homologous recombination and T-DNA integration in Arabidopsis CAF-1 mutants. EMBO J. **25**: 5579–5590.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.M., Werner, S., Jones, J.D.G., Patron, N.J., and Marillonnet, S. (2014). A Golden Gate modular cloning toolbox for plants. ACS Synth. Biol.

Espley, R. V., Hellens, R.P., Putterill, J., Stevenson, D.E., Kutty-Amma, S., and Allan, A.C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. Plant J. **49**: 414–427.

Esse, H.P. Van, Esse, H.P. Van, and Reuber, T.L. (2020). Genetic modification to improve disease resistance in crops. New Phytol. **225**: 70–86.

European Commission (2021). Study on the status of new genomic techniques under Union law and in light of the Court of Justice ruling in Case C-528/16.

Even-Faitelson, L., Samach, A., Melamed-Bessudo, C., Avivi-Ragolsky, N., and Levy, A.A. (2011). Localized egg-cell expression of effector proteins for targeted modification of the Arabidopsis genome. Plant J. **68**: 929–937.

Faden, F., Ramezani, T., Mielke, S., Almudi, I., Nairz, K., Froehlich, M.S., Höckendorff, J., Brandt, W., Hoehenwarter, W., Dohmen, R.J., Schnittger, A., and Dissmeyer, N. (2016). Phenotypes on demand via switchable target protein degradation in multicellular organisms. Nat. Commun. **7**: 73–78.

Fauser, F., Roth, N., Pacher, M., Ilg, G., Sanchez-Fernandez, R., Biesgen, C., and Puchta, H. (2012). In planta gene targeting. Proc. Natl. Acad. Sci. **109**: 7535–7540.

Feng, Z. et al. (2014). Multigeneration analysis reveals the inheritance, specificity, and

patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. Proc. Natl. Acad. Sci. **111**: 4632–4637.

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature: 1–19.

Fraley, R.T. et al. (1983). Expression of bacterial genes in plant cells. PNAS **80**: 4803–4807.

Frary, A. and Aerle, E. (1996). An examination of factors affecting the efficiency of Agrobacterium mediated transformation of tomato. Plant Cell Rep. **16**: 235–240.

Fukaki, H. and Tasaka, M. (2009). Hormone interactions during lateral root formation. Plant Mol. Biol. **69**: 437–449.

Fukushima, T., Takata, M., Morrisonţ, C., Araki, R., Fujimori, A., Abe, M., Tatsumi, K., Jasin, M., Dhar, P.K., Sonoda, E., Chiba, T., and Takeda, S. (2001). Genetic Analysis of the DNA-dependent Protein Kinase Reveals an Inhibitory Role of Ku in Late S–G 2 Phase DNA Double-strand Break Repair*.

Fülöp, K., Tarayre, S., Kelemen, Z., Horváth, G., Kevei, Z., Nikovics, K., Bakó, L., Brown, S., Kondorosi, A., and Kondorosi, E. (2005). Arabidopsis anaphase-promoting complexes: Multiple activators and wide range of substrates might keep APC perpetually busy. Cell Cycle **4**: 4084–4092.

Gallego-Bartolomé, J., Gardiner, J., Liu, W., Papikian, A., Ghoshal, B., Kuo, H.Y., Zhao, J.M.C., Segal, D.J., and Jacobsen, S.E. (2018). Targeted DNA demethylation of the arabidopsis genome using the human TET1 catalytic domain. Proc. Natl. Acad. Sci. U. S. A. **115**: E2125–E2134.

Gallie, D.R. (2002). The 5-leader of tobacco mosaic virus promotes translation through enhanced recruitment of eIF4F. Nucleic Acids Res. **30**: 3401–3411.

Gao, H. et al. (2020). Complex Trait Loci in Maize Enabled by CRISPR-Cas9 Mediated Gene Insertion. Front. Plant Sci. **11**: 1–14.

Gao, L., Cox, D.B.T., Yan, W.X., Manteiga, J.C., Schneider, M.W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N., and Zhang, F. (2017). Engineered Cpf1 variants with altered PAM specificities. Nat. Biotechnol. **35**: 789–792.

Gao, Z., Herrera-Carrillo, E., and Berkhout, B. (2018a). Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III. Mol. Ther. - Nucleic Acids **10**: 36–44.

Gao, Z., Herrera-Carrillo, E., and Berkhout, B. (2018b). Improvement of the CRISPR-Cpf1 system with ribozyme-processed crRNA. RNA Biol. **15**: 1458–1467.

Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature **468**: 67–71.

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

Glendining, M.J., Dailey, A.G., Williams, A.G., Evert, F.K. Van, Goulding, K.W.T., and Whitmore, A.P. (2009). Is it possible to increase the sustainability of arable and ruminant agriculture by reducing inputs ? Agric. Syst. **99**: 117–125.

Gu, B., Posfai, E., and Rossant, J. (2018). Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. Nat. Biotechnol. **36**: 632–637.

Guan, L., Tayengwa, R., Cheng, Z.M., Peer, W.A., Murphy, A.S., and Zhao, M. (2019). Auxin regulates adventitious root formation in tomato cuttings. BMC Plant Biol. **19**: 1–16.

Gupta, S. and Van Eck, J. (2016). Modification of plant regeneration medium decreases the time for recovery of Solanum lycopersicum cultivar M82 stable transgenic lines. Plant Cell. Tissue Organ Cult. **127**: 417–423.

Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G.F., and Chin, L. (2016). Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. Cell Rep.

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A.P.M. (2018a). Homology-Directed Repair of a Defective Glabrous Gene in Arabidopsis With Cas9-Based Gene Targeting. Front. Plant Sci. **9**: 1–20.

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A.P.M. (2018b). Homology-Directed Repair of a Defective Glabrous Gene in Arabidopsis With Cas9-Based Gene Targeting. Front. Plant Sci. **9**: 1–13.

Halley-Stott, R.P., Tanzer, F., Martin, D.P., and Rybicki, E.P. (2007). The complete nucleotide sequence of a mild strain of Bean yellow dwarf virus. Arch. Virol.

Hanley-Bowdoin, L., Bejarano, E.R., Robertson, D., and Mansoor, S. (2013). Geminiviruses:

Masters at redirecting and reprogramming plant processes. Nat. Rev. Microbiol. **11**: 777–788.

Hanley-Bowdoin, L., Settlage, S.B., and Robertson, D. (2004). Reprogramming plant gene expression: A prerequisite to geminivirus DNA replication. Mol. Plant Pathol. **5**: 149–156.

He, X., Tan, C., Wang, F., Wang, Y., Zhou, R., Cui, D., You, W., Zhao, H., Ren, J., and Feng, B. (2016). Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. Nucleic Acids Res. **44**.

Heyman, J. and De Veylder, L. (2012). The anaphase-promoting complex/cyclosome in control of plant development. Mol. Plant **5**: 1182–1194.

Hickey, L.T., Hafeez, A.N., Robinson, H., Jackson, S.A., Leal-bertioli, S.C.M., Tester, M., Gao, C., Godwin, I.D., Hayes, B.J., and Wulff, B.B.H. (2019). Breeding crops to feed 10 billion. Nat. Biotechnol. **37**.

Horlbeck, M.A., Witkowsky, L.B., Guglielmi, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., Torigoe, S.E., Tjian, R., and Weissman, J.S. (2016). Nucleosomes impede Cas9 access to DNA in vivo and in vitro.

Horvath, P., Barrangou, R., Horvath1, P., and Barrangou2, R. (2010). CRISPR/Cas, the Immune System of Bacteria and Archaea. Science (80-.). **327**: 167–170.

Hou, L., Yau, Y.Y., Wei, J., Han, Z., Dong, Z., and Ow, D.W. (2014). An Open-Source System for in Planta Gene Stacking by Bxb1 and Cre Recombinases. Mol. Plant **7**: 1756–1765.

Howden, S.E., McColl, B., Glaser, A., Vadolas, J., Petrou, S., Little, M.H., Elefanty, A.G., and Stanley, E.G. (2016). A Cas9 Variant for Efficient Generation of Indel-Free Knockin or Gene-Corrected Human Pluripotent Stem Cells. Stem Cell Reports **7**: 508–517.

Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z., and Liu, D.R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature.

Huang, T.-K. and Puchta, H. (2019). CRISPR/Cas-mediated gene targeting in plants: finally a turn for the better for homologous recombination. Plant Cell Rep. **0**: 0.

Huang, T.K., Armstrong, B., Schindele, P., and Puchta, H. (2021). Efficient gene targeting in Nicotiana tabacum using CRISPR/SaCas9 and temperature tolerant LbCas12a. Plant

Biotechnol. J.: 1–11.

Hummel, A.W., Chauhan, R.D., Cermak, T., Mutka, A.M., Vijayaraghavan, A., Boyher, A., Starker, C.G., Bart, R., Voytas, D.F., and Taylor, N.J. (2018). Allele exchange at the EPSPS locus confers glyphosate tolerance in cassava. Plant Biotechnol. J.: 1–8.

Hustedt, N. and Durocher, D. (2017). The control of DNA repair by the cell cycle. Nat. Publ. Gr. **19**: 1–9.

ICGEB Collection of biosafety reviews.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J. Bacteriol. **169**: 5429–5433.

James, D.J., Passey, A.J., Derek, J., and Bevan, M. (1989). Genetic transformation of apple (Malus pumila Mill.) using a disarmed Ti-binary vector. Plant Cell Rep. **7**: 658–661.

Jansen, R., Embden, J.D.A. van, Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. **43**: 1565–1575.

Janssen, B.J. and Gardner, R.C. (1990). Localized transient expression of GUS in leaf discs following cocultivation with Agrobacterium. Plant Mol. Biol. **14**: 61–72.

Jasper, F., Koncz, C., Schell, J., and Steinbiss, H.H. (1994). Agrobacterium T-strand production in vitro: sequence-specific cleavage and 5{\textquoteright} protection of single-stranded DNA templates by purified VirD2 protein. Proc. Natl. Acad. Sci. **91**: 694–698.

Jensen, K.T., Fløe, L., Petersen, T.S., Huang, J., Xu, F., Bolund, L., Luo, Y., and Lin, L. (2017). Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. FEBS Lett. **591**: 1892–1901.

Jin, F., Li, S., Dang, L., Chai, W., Li, P., and Wang, N.N. (2012). PL1 fusion gene: A novel visual selectable marker gene that confers tolerance to multiple abiotic stresses in transgenic tomato. Transgenic Res. **21**: 1057–1070.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (80-.). **337**: 816–821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife **2013**: 1–9.

Johnson, E.S., Gonda, D.K., and Varshavsky, A. (1990). Cis-trans recognition and subunit-specific degradation of short-lived proteins. Nature **346**: 287–291.

Jones, H.D. (2019). Future-proofing regulation for rapidly changing biotechnologies. Transgenic Res. **28**: 107–110.

Jorasch, P. (2020). Potential, Challenges, and Threats for the Application of New Breeding Techniques by the Private Plant Breeding Sector in the EU. Front. Plant Sci. **11**: 1–13.

Jordan, M.C. and McHughen, A. (1988). Transformed callus does not necessarily regenerate transformed shoots. Plant Cell Rep. **7**: 285–287.

Jouanin, A., Boyd, L., Visser, R.G.F., and Smulders, M.J.M. (2018). Development of Wheat With Hypoimmunogenic Gluten Obstructed by the Gene Editing Policy in Europe. Front. Plant Sci. **9**: 1–8.

Kallimasioti-Pazi, E.M., Thelakkad Chathoth, K., Taylor, G.C., Meynert, A., Ballinger, T., Kelder, M.J.E., Lalevée, S., Sanli, I., Feil, R., and Wood, A.J. (2018). Heterochromatin delays CRISPR-Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair. PLoS Biol. **16**: 1–22.

Kass, E.M. and Jasin, M. (2010). Collaboration and competition between DNA double-strand break repair pathways. FEBS Lett. **584**: 3703–3708.

Keskin, H., Shen, Y., Huang, F., Patel, M., Yang, T., Ashley, K., Mazin, A. V, and Storici, F. (2014). Transcript-RNA-templated DNA recombination and repair. Nature **515**: 436–439.

Kiferle, C., Fantini, E., Bassolino, L., Povero, G., Spelt, C., Buti, S., Giuliano, G., Quattrocchio, F., Koes, R., Perata, P., and Gonzali, S. (2015). Tomato R2R3-MYB proteins SlANT1 and SlAN2: Same protein activity, different roles. PLoS One **10**.

Kim, C.Y., Ahn, Y.O., Kim, S.H., Kim, Y.H., Lee, H.S., Catanach, A.S., Jacobs, J.M.E., Conner, A.J., and Kwak, S.S. (2010). The sweet potato IbMYB1 gene as a potential visible marker for sweet potato intragenic vector system. Physiol. Plant. **139**: 229–240.

Kim, H.B., Kwon, M., Ryu, H., Fujioka, S., Takatsuto, S., Yoshida, S., An, C.S., Lee, I., Hwang, I., and Choe, S. (2006). The regulation of DWARF4 expression is likely a critical mechanism in maintaining the homeostasis of bioactive brassinosteroids in arabidopsis. Plant Physiol. **140**: 548–557.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc. Natl. Acad. Sci. **93**: 1156–1160.

Kirik, A., Pecinka, A., Wendeler, E., and Reiss, B. (2006). The chromatin assembly factor subunit FASCIATA1 is involved in homologous recombination in plants. Plant Cell **18**: 2431–2442.

Klümper, W. and Qaim, M. (2014). A meta-analysis of the impacts of genetically modified crops. PLoS One **9**.

Knoll, A., Fauser, F., and Puchta, H. (2014). DNA recombination in somatic plant cells: Mechanisms and evolutionary consequences. Chromosom. Res. **22**: 191–201.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature.

Konagaya, K.I., Tsuda, M., Okuzaki, A., Ando, S., and Tabei, Y. (2013). Application of the acetolactate synthase gene as a cisgenic selectable marker for Agrobacterium-mediated transformation in Chinese cabbage (Brassica rapa ssp. pekinensis). Plant Biotechnol. **30**: 125–133.

Kortstee, A.J., Khan, S.A., Helderman, C., Trindade, L.M., Wu, Y., Visser, R.G.F., Brendolise, C., Allan, A., Schouten, H.J., and Jacobsen, E. (2011). Anthocyanin production as a potential visual selection marker during plant transformation. Transgenic Res. **20**: 1253–1264.

Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nat. Biotechnol. **36**.

Krämer, L. (2015). Legal questions concerning new methods for changing the genetic conditions in plants commissioned by Arbeitsgemeinschaft bäuerliche Landwirtschaft (AbL), Bund für Umwelt und Naturschutz (BUND), Bund Ökologische Lebensmittelwirtschaft (BÖLW), etc.: 23.

Ku, H.K. and Ha, S.H. (2020). Improving Nutritional and Functional Quality by Genome Editing of Crops: Status and Perspectives. Front. Plant Sci. **11**.

Kummu, M., Heino, M., Taka, M., and Olli, V. (2021). Climate change risks to push one-third of global food production outside Safe Climatic Space. One Earth: 1–10.

Lacroix, B. and Citovsky, V. (2019). Pathways of DNA transfer to plants from agrobacterium tumefaciens and related bacterial species. Annu. Rev. Phytopathol. **57**: 231–251.

Latham, J.R., Wilson, A.K., and Steinbrecher, R.A. (2005). The Mutational Consequences of Plant Transformation. J. Biomed. Biotechnol. **2006**: 1–7.

Lawrenson, T., Hinchliffe, A., Clarke, M., Morgan, Y., and Harwood, W. (2021). In-planta Gene Targeting in Barley Using Cas9 With and Without Geminiviral Replicons. Front. Genome Ed. **3**: 1–12.

Ledford, H. and Callaway, E. (2020). Pioneers of CRISPR gene editing win Chemistry Nobel. Nature **586**: 346–347.

Lee, C.S., Wang, R.W., Chang, H.H., Capurso, D., Segal, M.R., and Haber, J.E. (2016). Chromosome position determines the success of double-strand break repair. Proc. Natl. Acad. Sci. U. S. A. **113**: E146–E154.

Lee, D.W. (2002). Anthocyanins in leaves: Distribution, phylogeny and development. Adv. Bot. Res. **37**: 37–53.

Lévy, F., Johnston, J.A., and Varshavsky, A. (1999). Analysis of a conditional degradation signal in yeast and mammalian cells. Eur. J. Biochem. **259**: 244–252.

Li, J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., Li, J., and Gao, C. (2016). Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. Nat. Plants **2**: 1–6.

Li, J., Sun, Y., Du, J., Zhao, Y., and Xia, L. (2017). Generation of Targeted Point Mutations in Rice by a Modified CRISPR/Cas9 System. Mol. Plant.

Li, S., Li, J., He, Y., Xu, M., Zhang, J., Du, W., Zhao, Y., and Xia, L. (2019). Precise gene replacement in rice by RNA transcript-templated homologous recombination. Nat. Biotechnol. **37**: 445–450.

Li, X., Fan, J., Gruber, J., Guan, R., Frentzen, M., and Zhu, L.-H. (2013). Efficient selection and evaluation of transgenic lines of Crambe abyssinica. Front. Plant Sci. **4**: 162.

Li, Z., Liu, Z.-B., Xing, A., Moon, B.P., Koellhoffer, J.P., Huang, L., Ward, R.T., Clifton, E., Falco, S.C., and Cigan, A.M. (2015). Cas9-Guide RNA Directed Genome Editing in Soybean. Plant Physiol.

Lieberman-Lazarovich, M. and Levy, A.A. (2011). Homologous recombination in plants: an antireview. Methods Mol. Biol. **701**: 51–65.

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., and Gao, C. (2020). Prime genome editing in rice and wheat. Nat. Biotechnol. **38**: 582–585.

Liobikas, J., Skemiene, K., Trumbeckaite, S., and Borutaite, V. (2016). Anthocyanins in cardioprotection: A path through mitochondria. Pharmacol. Res. **113**: 808–815.

Liu, C. et al. (2020a). Extension of the in vivo haploid induction system from diploid maize to hexaploid wheat. Plant Biotechnol. J.: 316–318.

Liu, G., Yin, K., Zhang, Q., Gao, C., and Qiu, J.-L. (2019). Modulating chromatin accessibility by transactivation and targeting proximal dsgRNAs enhances Cas9 editing efficiency in vivo. Genome Biol. **20**.

Liu, G., Zhang, Y., and Zhang, T. (2020b). Computational approaches for effective CRISPR guide RNA design and evaluation. Comput. Struct. Biotechnol. J. **18**: 35–44.

Liu, H.-J. et al. (2020c). High-Throughput CRISPR/Cas9 Mutagenesis Streamlines Trait Gene Identification in Maize. Plant Cell **32**: 1397–1413.

Liu, L., Saunders, K., Thomas, C.L., Davies, J.W., and Stanley, J. (1999). Bean yellow dwarf virus RepA, but not Rep, binds to maize retinoblastoma protein and the virus tolerates mutations in the consensus binding motif. Virology **256**: 270–279.

Liu, L., Van Tonder, T., Pietersen, G., Davies, J.W., and Stanley, J. (1997). Molecular characterization of a subgroup I geminivirus from a legume in South Africa. J. Gen. Virol. **78**: 2113–2117.

Løvdal, T., Olsen, K.M., Slimestad, R., Verheul, M., and Lillo, C. (2010). Synergetic effects of nitrogen depletion, temperature, and light on the content of phenolic compounds and gene expression in leaves of tomato. Phytochemistry **71**: 605–613.

Lowder, L.G., Zhang, D., Baltes, N.J., Paul, J.W., Tang, X., Zheng, X., Voytas, D.F., Hsieh, T.F., Zhang, Y., and Qi, Y. (2015). A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. Plant Physiol. **169**: 971–985.

Lu, Y., Tian, Y., Shen, R., Yao, Q., Wang, M., Chen, M., Dong, J., Zhang, T., Li, F., Lei, M., and Zhu, J.K. (2020). Targeted, efficient sequence insertion and replacement in rice. Nat. Biotechnol. **38**: 1402–1407.

Lu, Y. and Zhu, J.K. (2017). Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System. Mol. Plant.

Ma, L., Ruan, J., Song, J., Wen, L., Yang, D., Zhao, J., Xia, X., Chen, Y.E., Zhang, J., and Xu, J. (2020). MiCas9 increases large size gene knock-in rates and reduces undesirable on-target and off-target indel edits. Nat. Commun. **11**.

Ma, M., Zhuang, F., Hu, X., Wang, B., Wen, X.-Z., Ji, J.-F., and Xi, J.J. (2017). Efficient generation of mice carrying homozygous double-floxp alleles using the Cas9-Avidin/Biotin-donor DNA system. Cell Res. **27**: 578–581.

Madsen, E.B., Höijer, I., Kvist, T., Ameur, A., and Mikkelsen, M.J. (2020). Xdrop: Targeted sequencing of long DNA molecules from low input samples using droplet sorting. Hum. Mutat. **41**: 1671–1679.

Maher, M.F., Nasti, R.A., Vollbrecht, M., Starker, C.G., Clark, M.D., and Voytas, D.F. (2020). Plant genome editing through de novo induction of meristems. Nat. Biotechnol. **38**: 84–89.

Maio, F., Helderman, T.A., Arroyo-Mateos, M., van der Wolf, M., Boeren, S., Prins, M., and van den Burg, H.A. (2020). Identification of Tomato Proteins That Interact With Replication Initiator Protein (Rep) of the Geminivirus TYLCV. Front. Plant Sci. **11**: 1–15.

Makarova, K.S. et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. **18**: 67–83.

Malzahn, A.A. et al. (2019). Application of CRISPR-Cas12a temperature sensitivity for improved genome editing in rice, maize, and Arabidopsis. BMC Biol. **17**: 1–14.

Manova, V. and Gruszka, D. (2015). DNA damage and repair in plants – from models to crops. Front. Plant Sci. **6**.

Marnef, A. and Legube, G. (2017). Organizing DNA repair in the nucleus: DSBs hit the road. Curr. Opin. Cell Biol.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat. Biotechnol.

Mathews, H. (2003). Activation Tagging in Tomato Identifies a Transcriptional Regulator of Anthocyanin Biosynthesis, Modification, and Transport. Plant Cell Online **15**: 1689–1703.

Mathur, J. and Koncz, C. (1998). PEG-Mediated Protoplast Transformation with Naked DNA. In Arabidopsis Protocols, J.M. Martinez-Zapater and J. Salinas, eds (Humana Press: Totowa, NJ), pp. 267–276.

Mbow, C., Rosenzweig, C., Barioni, L.G., Benton, T.G., Herrero, M., Krishnapillai, M., Liwenga, E., Pradhan, P., Rivera-Ferre, M.G., Sapkota, T., Tubiello, F.N., and Xu, Y. (2019). Food Security. In Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystem, J.S. Shukla et al., eds, pp. 439–550.

McHughen, A. and Jordan, M.C. (1989). Recovery of transgenic plants from "escape" shoots. Plant Cell Rep. **7**: 611–614.

Meer, P. VAN DER et al. (2021). The Status under EU Law of Organisms Developed through Novel Genomic Techniques. Eur. J. Risk Regul.: 1–20.

Merker, L., Schindele, P., Huang, T.K., Wolter, F., and Puchta, H. (2020). Enhancing in planta gene targeting efficiencies in Arabidopsis using temperature-tolerant CRISPR/LbCas12a. Plant Biotechnol. J. **18**: 2382–2384.

Meyerhans, A., Vartanian, J., and Wain-hobson, S. (1990). DNA recombination during PCR. **18**: 1687–1691.

Miguel, C. and Marum, L. (2011). An epigenetic view of plant cells cultured in vitro: Somaclonal variation and beyond. J. Exp. Bot. **62**: 3713–3725.

Miki, D., Zhang, W., Zeng, W., Feng, Z., and Zhu, J.K. (2018). CRISPR/Cas9-mediated gene targeting in Arabidopsis using sequential transformation. Nat. Commun. **9**: 1–9.

Milner, M.J., Craze, M., Hope, M.S., and Wallington, E.J. (2020). Turning Up the Temperature on CRISPR: Increased Temperature Can Improve the Editing Efficiency of Wheat Using CRISPR/Cas9. Front. Plant Sci. **11**: 1–9.

Mishra, R., Joshi, R.K., and Zhao, K. (2020). Base editing in crops: current advances, limitations and future implications. Plant Biotechnol. J. **18**: 20–31.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol.

Molinier, J., Oakeley, E.J., Niederhauser, O., Kovalchuk, I., and Hohn, B. (2005). Dynamic response of plant genome to ultraviolet radiation and other genotoxic stresses. Mutat. Res. - Fundam. Mol. Mech. Mutagen. **571**: 235–247.

Montefiori, M., Brendolise, C., Dare, A.P., Kui, L.W., Davies, K.M., Hellens, R.P., and Allan, A.C. (2015). In the Solanaceae, a hierarchy of bHLHs confer distinct target specificity to the anthocyanin regulatory complex. J. Exp. Bot. **66**: 1427–1436.

Moreno-Mateos, M.A., Fernandez, J.P., Rouet, R., Vejnar, C.E., Lane, M.A., Mis, E., Khokha, M.K., Doudna, J.A., and Giraldez, A.J. (2017). CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. Nat. Commun. **679**.

Moscou, M.J. and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by

TAL effectors. Science (80-.). **326**: 1501.

Movahedi, A. et al. (2021). Enhancement of CRISPR-mediated homologous recombination-based gene knock-in efficiency by simultaneous HDR overexpression and NHEJ suppression in poplar. bioRxiv: 2020.07.04.188219.

Naim, F., Shand, K., Hayashi, S., O'Brien, M., McGree, J., Johnson, A.A.T., Dugdale, B., and Waterhouse, P.M. (2020). Are the current gRNA ranking prediction algorithms useful for genome editing in plants? PLoS One **15**: 1–12.

Nakabayashi, R. et al. (2014). Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. Plant J. **77**: 367–379.

Nakade, S., Mochida, K., Kunii, A., Nakamae, K., Aida, T., Tanaka, K., Sakamoto, N., Sakuma, T., and Yamamoto, T. (2018). Biased genome editing using the local accumulation of DSB repair molecules system. Nat. Commun. **9**: 1–5.

Nam, J., Matthysse, A.G., and Gelvinayl, S.B. (1997). Differences in susceptibility of Arabidopsis ecotypes to crown gall disease may result from a deficiency in T-DNA integration. Plant Cell **9**: 317–333.

Nanto, K., Yamada-Watababe, K., and Ebinuma, H. (2005). Agrobacterium -mediated RMCE approach for gene replacement. Plant Biotechnol. J. **3**: 203–214.

Nicole M. Gaudelli, Alexis C. Komor, Holly A. Rees, Michael S. Packer, Ahmed H. Badran, D.I.B.& D.R.L. (2017). Commentary: Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. Nature.

Nishizawa-Yokoi, A., Mikami, M., and Toki, S. (2020). A Universal System of CRISPR/Cas9-Mediated Gene Targeting Using All-in-One Vector in Plants. Front. Genome Ed. **2**: 1–12.

Nordström, A., Tarkowski, P., Tarkowska, D., Norbaek, R., Åstot, C., Dolezal, K., and Sandberg, G. (2004). Auxin regulation of cytokinin biosynthesis in Arabidopsis thaliana: A factor of potential importance for auxin-cytokinin-regulated development. Proc. Natl. Acad. Sci. U. S. A. **101**: 8039–8044.

Offringa, R., de Groot, M.J., Haagsman, H.J., Does, M.P., van den Elzen, P.J., and Hooykaas, P.J. (1990). Extrachromosomal homologous recombination and gene targeting in plant cells after Agrobacterium mediated transformation. EMBO J. **9**: 3077–3084.

Olsen, K.M., Slimestad, R., Lea, U.S., Brede, C., Løvdal, T., Ruoff, P., Verheul, M., and

Lillo, C. (2009). Temperature and nitrogen effects on regulators and products of the flavonoid pathway: Experimental and kinetic model studies. Plant, Cell Environ. **32**: 286–299.

Orel, N., Kyryk, A., and Puchta, H. (2003). Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. Plant J.

Ow, D.W., Jacobst, J.D., and Howell, S.H. (1987). Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by use of the firefly luciferase gene as a reporter of promoter activity. Biochemistry **84**: 4870–4874.

Pabo, C.O., Peisach, E., and Grant, R.A. (2001). Design and selection of novel Cys2 His2 Zinc Finger Proteins. Annu. Rev. Biochem **70**: 313–40.

Padilla, I.M.G., Webb, K., and Scorza, R. (2003). Early antibiotic selection and efficient rooting and acclimatization improve the production of transgenic plum plants (Prunus domestica L .). Plant Cell Rep **22**: 38–45.

Panos, C. and Cohen, H. (1964). Growth rates of Streptococcus pyogenes and derived L form at various temperatures. J. Bacteriol. **87**: 1242–1243.

Park, S.-Y., Peterson, F.C., Mosquana, A., Yao, J., Volkman, B.F., and Cutler, S.R. (2015). Agrochemical control of plant water use using engineered abscisic acid receptors. Nature **250**: 545–548.

Paszkowski, J., Baur, M., Bogucki, A., and Potrykus, I. (1988). Gene targeting in plants. EMBO J. **7**: 4021–6.

De Pater, S., Klemann, B.J.P.M., and Hooykaas, P.J.J. (2018). True gene-targeting events by CRISPR/Cas-induced DSB repair of the PPO locus with an ectopically integrated repair template. Sci. Rep.

Pellegrino, E., Bedini, S., Nuti, M., and Ercoli, L. (2018). Impact of genetically engineered maize on agronomic, environmental and toxicological traits: A meta-analysis of 21 years of field data. Sci. Rep. **8**: 1–12.

Peña, L., Pérez, R.M., Cervera, M., Juárez, J.A., and Navarro, L. (2004). Early events in Agrobacterium-mediated genetic transformation of citrus explants. Ann. Bot. **94**: 67–74.

Peng, F., Zhang, W., Zeng, W., Zhu, J.K., and Miki, D. (2020). Gene targeting in Arabidopsis via an all-in-one strategy that uses a translational enhancer to aid Cas9 expression. Plant Biotechnol. J. **18**: 892–894.

Peterson, D., Barone, P., Lenderts, B., Schwartz, C., Feigenbutz, L., Clair, G.S., Jones, S., and Svitashev, S. (2021). Advances in Agrobacterium transformation and vector design result in high frequency targeted gene insertion in maize. Plant Biotechnol. J.

Phillips, R.L., Kaeppler, S.M., and Olhoft, P. (1994). Genetic instability of plant tissue cultures: Breakdown of normal controls. PNAS **91**: 5222–5226.

Pireyre, M. and Burow, M. (2015). Regulation of MYB and bHLH transcription factors: A glance at the protein level. Mol. Plant **8**: 378–388.

Podevin, N., Davies, H. V., Hartung, F., Nogué, F., and Casacuberta, J.M. (2013). Site-directed nucleases: A paradigm shift in predictable, knowledge-based plant breeding. Trends Biotechnol. **31**: 375–383.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology.

Prado, J.R., Segers, G., Voelker, T., Carson, D., Dobert, R., Phillips, J., Cook, K., Cornejo, C., Monken, J., Grapes, L., Reynolds, T., and Martino-Catt, S. (2014). Genetically engineered crops: From idea to product. Annu. Rev. Plant Biol. **65**: 769–790.

Puchta, H. (2016). Breaking DNA in plants: How I almost missed my personal breakthrough. Plant Biotechnol. J. **14**: 437–440.

Puchta, H. et al. (1999). Double-strand break-induced recombination between ectopic homologous sequences in somatic plant cells. Genetics **152**: 1173–81.

Puchta, H. (2002). Gene replacement by homologous recombination in plants. Plant Mol. Biol. **48**: 173–182.

Puchta, H. (1998). Repair of genomic double-strand breaks in somatic plant cells by one-sided invasion of homologous sequences. Plant J.

Puchta, H., Dujon, B., and Hohn, B. (1993). Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. Nucleic Acids Res. **21**: 5034–5040.

Puchta, H. and Fauser, F. (2013). Gene targeting in plants: 25 years later. Int. J. Dev. Biol. **57**: 629–637.

Purnhagen, K.P., Kok, E., Kleter, G., Schebesta, H., Visser, R.G.F., and Wesseler, J. (2018). EU court casts new plant breeding techniques into regulatory limbo. Nat. Publ. Gr. **36**: 799–800.

Qi, Y., Zhang, Y., Zhang, F., Baller, J.A., Cleland, S.C., Ryu, Y., Starker, C.G., and Voytas, D.F. (2013). Increasing frequencies of site-specific mutagenesis and gene targeting in *Arabidopsis* by manipulating DNA repair pathways. Genome Res. **23**: 547–554.

Qiu, Z., Wang, X., Gao, J., Guo, Y., Huang, Z., and Du, Y. (2016). The tomato Hoffman's anthocyaninless gene encodes a bHLH transcription factor involved in anthocyanin biosynthesis that is developmentally regulated and induced by low temperatures. PLoS One **11**: 1–22.

Rahavi, S.M.R. and Kovalchuk, I. (2013). Changes in homologous recombination frequency in Arabidopsis thaliana plants exposed to stress depend on time of exposure during development and on duration of stress exposure. Physiol. Mol. Biol. Plants **19**: 479–488.

Raman, R. (2017). The impact of Genetically Modified (GM) crops in modern agriculture: a review. GM Crops Food **5698**: 00–00.

Ramsay, N.A. and Glover, B.J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. **10**: 63–70.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013). Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. Cell.

Raun, W.R. and Johnson, G. V (1999). Improving Nitrogen Use Efficiency for Cereal Production. **91**: 357–363.

Reiss, B., Sprengel, R., and Schaller, H. (1984). Protein fusions with the kanamycin resistance gene from transposon Tn5. EMBO J. **3**: 3317–3322.

Remy, S. et al. (2014). Efficient gene targeting by homology-directed repair in rat zygotes using TALE nucleases. Genome Res. **24**: 1371–1383.

Ren, B., Yan, F., Kuang, Y., Li, N., Zhang, D., Zhou, X., Lin, H., and Zhou, H. (2018). Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-guided hyperactive hAID mutant. Mol. Plant.

Ren, Q. et al. (2021). PAM-less plant genome editing using a CRISPR–SpRY toolbox. Nat. Plants **7**.

Renckens, S., De Greve, H., Van Montagu, M., and Hernalsteens, J.P. (1992). Petunia plants escape from negative selection against a transgene by silencing the foreign DNA via methylation. MGG Mol. Gen. Genet. **233**: 53–64.

Reuters (2021). France backs non-GMO regulation for crop gene-editing in EU.

Richter, K.S., Kleinow, T., and Jeske, H. (2014). Somatic homologous recombination in plants is promoted by a geminivirus in a tissue-selective manner. Virology **452–453**: 287–296.

Rogers, C. and Oldroyd, G.E.D. (2014). Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. J. Exp. Bot. **65**: 1939–1946.

Roldan, M.V.G., Engel, B., de Vos, R.C.H., Vereijken, P., Astola, L., Groenenboom, M., van de Geest, H., Bovy, A., Molenaar, J., van Eeuwijk, F., and Hall, R.D. (2014). Metabolomics reveals organ-specific metabolic rearrangements during early tomato seedling development. Metabolomics **10**: 958–974.

Rommens, C.M., Richael, C.M., Yan, H., Navarre, D.A., Ye, J., Krucker, M., and Swords, K. (2008). Engineered native pathways for high kaempferol and caffeoylquinate production in potato. Plant Biotechnol. J. **6**: 870–886.

Rosellini, D., Capomaccio, S., Ferradini, N., Savo Sardaro, M.L., Nicolia, A., and Veronesi, F. (2007). Non-antibiotic, efficient selection for alfalfa genetic engineering. Plant Cell Rep. **26**: 1035–1044.

Sato, S. et al. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature.

Sauer, N.J. et al. (2016). Oligonucleotide-mediated genome editing provides precision and function to engineered nucleases and antibiotics in plants. Plant Physiol. **170**: pp.01696.2015.

Savic, N., Ringnalda, F., Bargsten, K., Li, Y., Berk, C., Hall, J., Neri, D., Jinek, M., and Schwank, G. (2018). Covalent linkage of the DNA repair template to the CRISPR/Cas9 complex enhances homology-directed repair. Elife: 1–18.

Scheben, A., Wolter, F., Batley, J., Puchta, H., and Edwards, D. (2017). Towards CRISPR/Cas crops - bringing together genomics and genome editing. New Phytol.

Schiermeyer, A. et al. (2019). Targeted insertion of large DNA sequences by - homology - directed repair or non - homologous end joining in engineered tobacco BY - 2 cells using designed zinc finger nucleases. Plant Direct **3**: 1–14.

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

Schindele, P. and Puchta, H. (2019). Engineering CRISPR/LbCas12a for highly efficient,

temperature-tolerant plant gene editing. Plant Biotechnol. J.: 0–3.

Schreiber, G., Reuveni, M., Evenor, D., Oren-Shamir, M., Ovadia, R., Sapir-Mir, M., Bootbool-Man, A., Nahon, S., Shlomo, H., Chen, L., and Levin, I. (2012). ANTHOCYANIN1 from Solanum chilense is more efficient in accumulating anthocyanin metabolites than its Solanum lycopersicum counterpart in association with the ANTHOCYANIN FRUIT phenotype of tomato. Theor. Appl. Genet. **124**: 295–307.

Seifert, F. (2020). National specificity and convergence in the European anti-GM movement: the cases of Austria, Germany, France and the UK. Innov. Eur. J. Soc. Sci. Res.: 1–22.

Seleiman, M.F., Al-suhaibani, N., Ali, N., Akmal, M., Alotaibi, M., Refay, Y., Dindaroglu, T., Abdul-wajid, H.H., and Battaglia, M.L. (2021). Drought stress impact on plants and different approaches to alleviate its adverse effects. Plants **10**: 1–25.

Shaked, H., Melamed-Bessudo, C., and Levy, A.A. (2005). High-frequency gene targeting in Arabidopsis plants expressing the yeast RAD54 gene. PNAS **102**: 12265–12269.

Sharma, A., Shahzad, B., Rehman, A., Bhardwaj, R., Landi, M., and Zheng, B. (2019). Response of phenylpropanoid pathway and the role of polyphenols in plants under abiotic stress. Molecules **24**: 1–22.

Shi, J., Gao, H., Wang, H., Lafitte, H.R., Archibald, R.L., Yang, M., Hakimi, S.M., Mo, H., and Habben, J.E. (2017). ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. Plant Biotechnol. J.

Shibata, A., Jeggo, P., and Löbrich, M. (2018). The pendulum of the Ku-Ku clock. DNA Repair (Amst). **71**: 164–171.

Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., Zhang, F., and Koonin, E. V. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol. Cell.

Shukla, V.K. et al. (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature **459**: 437–441.

Shy, B.R., Macdougall, M.S., Clarke, R., and Merrill, B.J. (2016). Co-incident insertion enables high efficiency genome engineering in mouse embryonic stem cells. Nucleic Acids Res. **44**: 7997–8010.

Sikora, D. and Rzymski, P. (2021). Chapter 13 - Public Acceptance of GM Foods: A Global

Perspective (1999–2019). In Policy Issues in Genetically Modified Crops, P. Singh, A. Borthakur, A.A. Singh, A. Kumar, and K.K. Singh, eds (Academic Press), pp. 293–315.

Smith, E.F. and Townsend, C.O. (1907). A plant-tumor of bacterial origins. Science (80-.). **25**: 671–673.

Smith, J. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res.

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., and Kucherlapati, R.S. (1985). Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. Nature **317**: 230–234.

Solleti, S.K., Bakshi, S., and Sahoo, L. (2008). Additional virulence genes in conjunction with efficient selection scheme, and compatible culture regime enhance recovery of stable transgenic plants in cowpea via Agrobacterium tumefaciens-mediated transformation. J. Biotechnol. **135**: 97–104.

Song, F. and Stieger, K. (2017). Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. Mol. Ther. - Nucleic Acids **7**: 53–60.

Spampinato, C.P. (2017). Protecting DNA from errors and damage: an overview of DNA repair mechanisms in plants compared to mammals. Cell. Mol. Life Sci. **74**: 1693–1709.

Sprink, T., Eriksson, D., Schiemann, J., and Hartung, F. (2016). Regulatory hurdles for genome editing: process- vs. product-based approaches in different regulatory contexts. Plant Cell Rep. **35**: 1493–1506.

Stadler, L.J. (1928). Genetic effects of X-rays in maize. Proc. Natl. Acad. Sci. **14**.

Stan J. J. Brouns, 1* Matthijs M. Jore, 1* Magnus Lundgren, 1 Edze R. Westra, 1 Rik J. H. Slijkhuis, 1 Ambrosius P. L. Snijders, 2 Mark J. Dickman, 2 Kira S. Makarova, 3 Eugene V. Koonin, 3 John van der Oost1† (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. Science (80-.).: 960–963.

Steinert, J., Schiml, S., and Puchta, H. (2016). Homology-based double-strand break-induced genome engineering in plants. Plant Cell Rep. **35**: 1429–1438.

Steyn, W.J., Wand, S.J.E., Holcroft, D.M., and Jacobs, G. (2002). in vegetative tissues : Anthocyanins in unified function a proposed photoprotection. New Phytol. **155**: 349–361.

Suárez-López, P. and Gutiérrez, C. (1997). DNA replication of wheat dwarf geminivirus vectors: Effects of origin structure and size. Virology **227**: 389–399.

Sun, Y., Li, J., and Xia, L. (2016a). Precise Genome Modification via Sequence-Specific Nucleases-Mediated Gene Targeting for Crop Improvement. Front. Plant Sci.

Sun, Y., Zhang, X., Wu, C., He, Y., Ma, Y., Hou, H., Guo, X., Du, W., Zhao, Y., and Xia, L. (2016b). Engineering Herbicide-Resistant Rice Plants through CRISPR/Cas9-Mediated Homologous Recombination of Acetolactate Synthase. Mol. Plant **9**: 628–631.

Svitashev, S., Young, J.K., Schwartz, C., Gao, H., Falco, S.C., and Cigan, A.M. (2015). Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA. Plant Physiol.

Sze, M.A., Schloss, P.D., and Arbor, A. (2019). The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data.

Taagen, E., Bogdanove, A.J., and Sorrells, M.E. (2020). Counting on Crossovers: Controlled Recombination for Plant Breeding. Trends Plant Sci. **25**: 455–465.

Tan, J. et al. (2020). Efficient CRISPR/Cas9-based plant genomic fragment deletions by microhomology-mediated end joining. Plant Biotechnol. J. **18**: 2161–2163.

Tang, J., Cho, N.W., Cui, G., Manion, E.M., Shanbhag, N.M., Botuyan, M.V., Mer, G., and Greenberg, R.A. (2013). Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination. Nat. Struct. Mol. Biol. **20**: 317–325.

Tang, X. et al. (2017). A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants. Nat. Plants.

Thomas, K.R. and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell **51**: 503–512.

Tilman, D., Balzer, C., Hill, J., and Befort, B.L. (2011). Global food demand and the sustainable intensi fi cation of agriculture. PNAS **108**: 20260–20264.

Tudi, M., Ruan, H.D., Wang, L., Lyu, J., Sadler, R., Connell, D., Chu, C., and Phung, D. tri (2021). Agriculture Development , Pesticide Application and Its Impact on the Environment. Int. J. Environ. Res. Public Health **18**: 1–23.

Turnbull C., Lillemo M. and Hvoslef-Eide TAK. (2021) Global Regulation of Genetically Modified Crops Amid the Gene Edited Crop Boom – A Review. *Front. Plant Sci.* 12:630396. doi: 10.3389/fpls.2021.630396

Uanschou, C., Siwiec, T., Pedrosa-Harand, A., Kerzendorfer, C., Sanchez-Moran, E., Novatchkova, M., Akimcheva, S., Woglar, A., Klein, F., and Schlögelhofer, P. (2007). A novel plant gene essential for meiosis is related to the human CtIP and the yeast

COM1/SAE2 gene. EMBO J. **26**: 5061–5070.

USDA APHIS, (2020), Federal register: SECURE rule, movement of certain genetically engineered organisms, 7 CFR Parts 330, 340, and 372. Available online at: https://www.aphis.usda.gov/aphis/ourfocus/biotechnology/biotech-rule-revision/secure-rule/secure-about/340_2017_perdue_biotechreg (accessed Jan 18, 2022)

Vanitharani, R., Chellappan, P., and Fauquet, C.M. (2005). Geminiviruses and RNA silencing. Trends Plant Sci. **10**: 144–151.

Varshavsky, A. (2011). The N-end rule pathway and regulation by proteolysis. Protein Sci. **20**: 1298–1345.

Veillet, F., Durand, M., Kroj, T., Cesari, S., and Gallois, J.-L. (2020). Precision Breeding Made Real with CRISPR: Illustration through Genetic Resistance to Pathogens. Plant Commun. **1**: 100102.

Vicente, M.M., Mendes, A., Cruz, M., Vicente, J.R., and Barreto, V.M. (2019). A CyclinB2-Cas9 fusion promotes the homology-directed repair of double-strand breaks. bioRxiv.

Vives-Vallés, J.A. and Collonnier, C. (2020). The Judgment of the CJEU of 25 July 2018 on Mutagenesis: Interpretation and Interim Legislative Proposal. Front. Plant Sci. **10**: 1–15.

Vriend, L.E.M., Prakash, R., Chen, C.C., Vanoli, F., Cavallo, F., Zhang, Y., Jasin, M., and Krawczyk, P.M. (2016). Distinct genetic control of homologous recombination repair of Cas9-induced double-strand breaks, nicks and paired nicks. Nucleic Acids Res. **44**: 5204–5217.

Vu, G.T.H., Cao, H.X., Watanabe, K., Hensel, G., Blattner, F.R., Kumlehn, J., and Schubert, I. (2014). Repair of Site-Specific DNA Double-Strand Breaks in Barley Occurs via Diverse Pathways Primarily Involving the Sister Chromatid. Plant Cell **26**: 2156–2167.

Vu, T. Van, Das, S., Tran, M.T., Hong, J.C., and Kim, J.-Y. (2020a). Precision Genome Engineering for the Breeding of Tomatoes: Recent Progress and Future Perspectives. Front. Genome Ed. **2**: 1–16.

Vu, T. Van, Sivankalyani, V., Kim, E.J., Doan, D.T.H., Tran, M.T., Kim, J., Sung, Y.W., Park, M., Kang, Y.J., and Kim, J.Y. (2020b). Highly efficient homology-directed repair using CRISPR/Cpf1-geminiviral replicon in tomato. Plant Biotechnol. J.: 2133–2143.

Vu, T. Van, Sivankalyani, V., Kim, E.J., Doan, D.T.H., Tran, M.T., Kim, J., Sung, Y.W.,

Park, M., Kang, Y.J., and Kim, J.Y. (2020c). Highly efficient homology-directed repair using CRISPR/Cpf1-geminiviral replicon in tomato. Plant Biotechnol. J. **18**: 2133–2143.

Van Vu, T., Sung, Y.W., Kim, J., Doan, D.T.H., Tran, M.T., and Kim, J.Y. (2019). Challenges and Perspectives in Homology-Directed Gene Targeting in Monocot Plants. Rice **12**: 1–29.

van der Walt, E., Rybicki, E.P., Varsani, A., Polston, J.E., Billharz, R., Donaldson, L., Monjane, A.L., and Martin, D.P. (2009). Rapid host adaptation by extensive recombination. J. Gen. Virol. **90**: 734–746.

Wang, M., Lu, Y., Botella, J., Mao, Y., Hua, K., and Zhu, J. (2017). Gene Targeting by Homology-directed Repair in Rice using a Geminivirus-based CRISPR/Cas9 System. Mol. Plant.

Weigel, D. et al. (2000). Activation tagging in Arabidopsis. Plant Physiol. **122**: 1003–1013.

Weiss, T., Wang, C., Kang, X., Zhao, H., Elena Gamo, M., Starker, C.G., Crisp, P.A., Zhou, P., Springer, N.M., Voytas, D.F., and Zhang, F. (2020). Optimization of multiplexed CRISPR/Cas9 system for highly efficient genome editing in Setaria viridis. Plant J. **104**: 828–838.

Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012). Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. Bioeng. Bugs.

Werr, W. and Lörz, H. (1986). Transient gene expression in a Gramineae cell line. Mol. Gen. Genet. MGG **202**: 471–475.

Wilmink, A. and Dons, J.J.M. (1993). Selective agents and marker genes for use in transformation of monocotyledonous plants. Plant Mol. Biol. **11**: 165–185.

Wolt, J.D. and Wolf, C. (2018). Policy and governance perspectives for regulation of genome edited crops in the United States. Front. Plant Sci. **871**: 1–12.

Wolter, F., Klemm, J., and Puchta, H. (2018). Efficient in planta gene targeting in Arabidopsis using egg-cell specific expression of the Cas9 nuclease of *S. aureus*. Plant J.

Wolter, F. and Puchta, H. (2019a). In planta gene targeting can be enhanced by the use of CRISPR/Cas12a . Plant J.

Wolter, F. and Puchta, H. (2019b). In planta gene targeting can be enhanced by the use of CRISPR/Cas12a. Plant J. **100**: 1083–1094.

Won, M. and Dawid, I.B. (2017). PCR artifact in testing for homologous recombination in genomic editing in zebrafish. PLoS One: 1–10.

Wright, D.A., Townsend, J.A., Winfrey, R.J., Irwin, P.A., Rajagopal, J., Lonosky, P.M., Hall, B.D., Jondle, M.D., and Voytas, D.F. (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J. **44**: 693–705.

Xu, W., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends Plant Sci. **20**: 176–185.

Yan, S., Chen, N., Huang, Z., Li, D., Zhi, J., Yu, B., Liu, X., Cao, B., and Qiu, Z. (2020). Anthocyanin Fruit encodes an R2R3-MYB transcription factor, SlAN2-like, activating the transcription of SlMYBATV to fine-tune anthocyanin content in tomato fruit. New Phytol. **225**: 2048–2063.

Yang, D., Scavuzzo, M.A., Chmielowiec, J., Sharp, R., Bajic, A., and Borowiak, M. (2016). Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. Sci. Rep.

Yao, L., Zhang, Y., Liu, C., Liu, Y., Wang, Y., Liang, D., Liu, J., Sahoo, G., and Kelliher, T. (2018). OsMATL mutation induces haploid seed formation in indica rice. Nat. Plants **4**.

Yao, Y. and Kovalchuk, I. (2011). Abiotic stress leads to somatic and heritable changes in homologous recombination frequency, point mutation frequency and microsatellite stability in Arabidopsis plants. Mutat. Res. - Fundam. Mol. Mech. Mutagen. **707**: 61–66.

Yarrington, R.M., Verma, S., Schwartz, S., Trautman, J.K., and Carroll, D. (2018). Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. Pnas **115**: 9351–9358.

Yau, Y.Y. and Stewart, C.N. (2013). Less is more: Strategies to remove marker genes from transgenic plants. BMC Biotechnol. **13**.

Yoshioka, Y., Takahashi, Y., Matsuoka, K., Nakamura, K., Koizumi, J., Kojima, M., and Machida, Y. (1996). Transient gene expression in plant cells mediated by Agrobacterium tumefaciens: Application for the analysis of virulence loci. Plant Cell Physiol. **37**: 782–789.

Yu, Q.H., Wang, B., Li, N., Tang, Y., Yang, S., Yang, T., Xu, J., Guo, C., Yan, P., Wang, Q., and Asmutola, P. (2017). CRISPR/Cas9-induced Targeted Mutagenesis and Gene Replacement to Generate Long-shelf Life Tomato Lines. Sci. Rep.

Zaidi, S.S. e. A., Mahas, A., Vanderschuren, H., and Mahfouz, M.M. (2020). Engineering crops of the future: CRISPR approaches to develop climate-resilient and disease-resistant

plants. Genome Biol. **21**: 1–19.

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., and Zhang, F. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell **163**: 759–771.

Zhang, J.P. et al. (2017a). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. Genome Biol. **18**: 1–18.

Zhang, K., Raboanatahiry, N., Zhu, B., and Li, M. (2017b). Progress in Genome Editing Technology and Its Application in Plants. Front. Plant Sci. **8**.

Zhang, P., Potrykus, I., and Puonti-Kaerlas, J. (2000). Efficient production of transgenic cassava using negative and positive selection. Transgenic Res. **9**: 405–415.

Zhang, T., Wang, X., Lu, Y., Cai, X., Ye, Z., and Zhang, J. (2013a). Genome-wide analysis of the cyclin gene family in tomato. Int. J. Mol. Sci. **15**: 120–140.

Zhang, Y., Pribil, M., Palmgren, M., and Gao, C. (2020a). A CRISPR way for accelerating improvement of food crops. Nat. Food **1**: 200–205.

Zhang, Y., Zhang, F., Li, X., Baller, J.A., Qi, Y., Starker, C.G., Bogdanove, A.J., and Voytas, D.F. (2013b). Transcription Activator-Like Effector Nucleases Enable Efficient Plant Genome Engineering. Plant Physiol. **161**: 20–27.

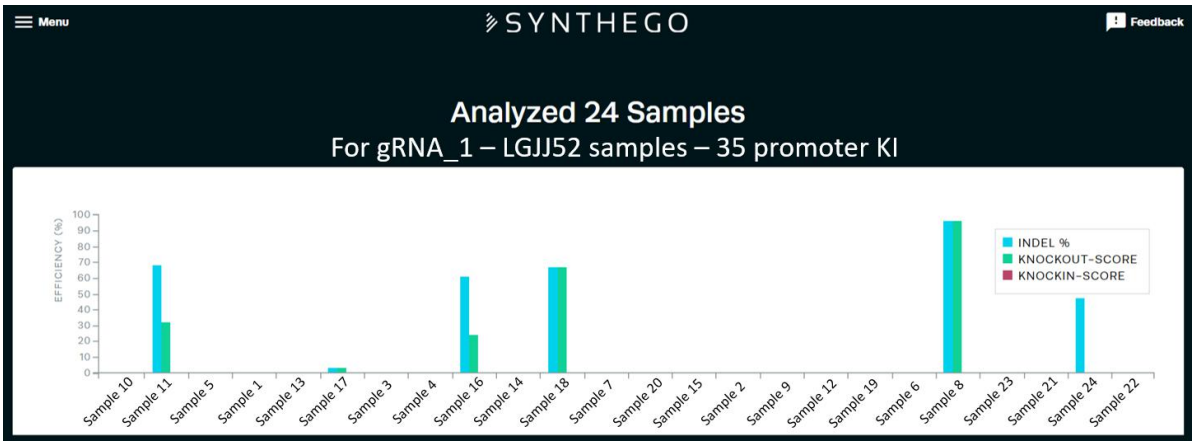
Zhang, Z. et al. (2020b). The strawberry transcription factor FaRAV1 positively regulates anthocyanin accumulation by activation of FaMYB10 and anthocyanin pathway genes. Plant Biotechnol. J. **18**: 2267–2279.

Zhu, H., Li, C., and Gao, C. (2020). Applications of CRISPR–Cas in agriculture and plant biotechnology. Nat. Rev. Mol. Cell Biol. **21**: 661–677.

Appendices

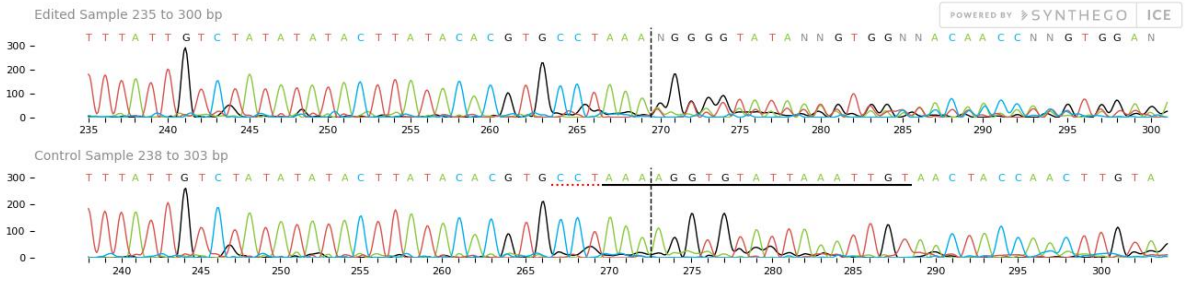
Appendix A.

ICE results of the trace decomposition comparison. Top chromatograms are the Sanger sequencing traces for the edited samples, bottom chromatograms correspond to the wild-type control Sanger sequencing traces.

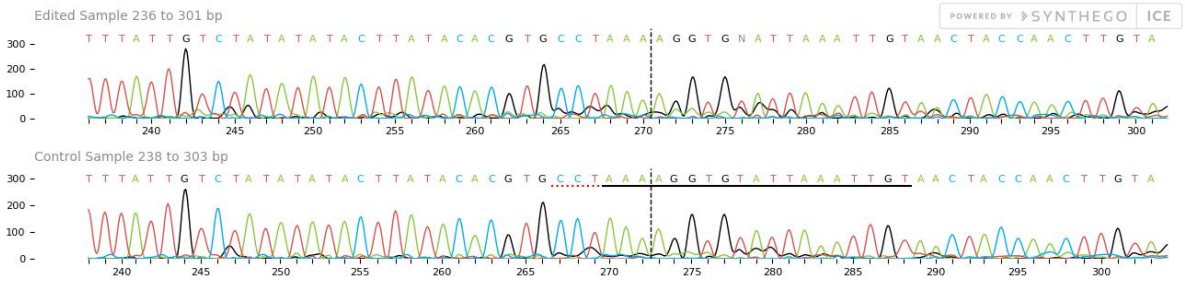


Appendix A.1 LGJJ52.....

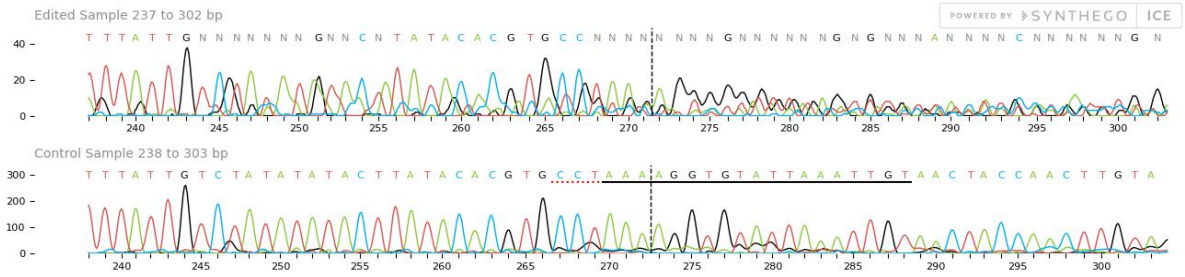
Sample 11



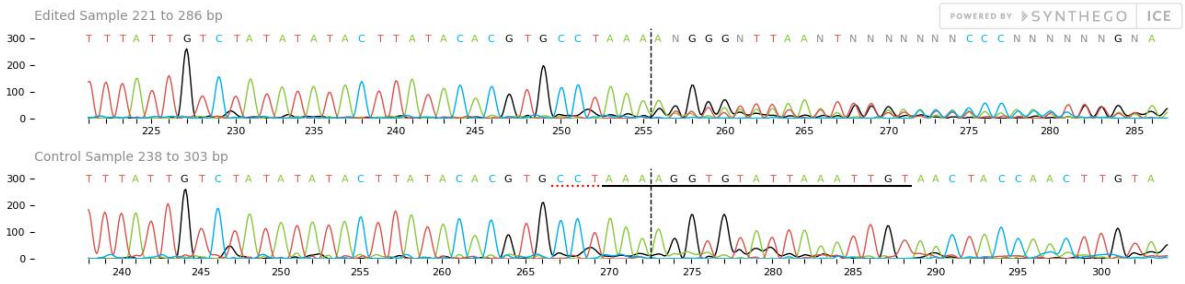
Sample 17



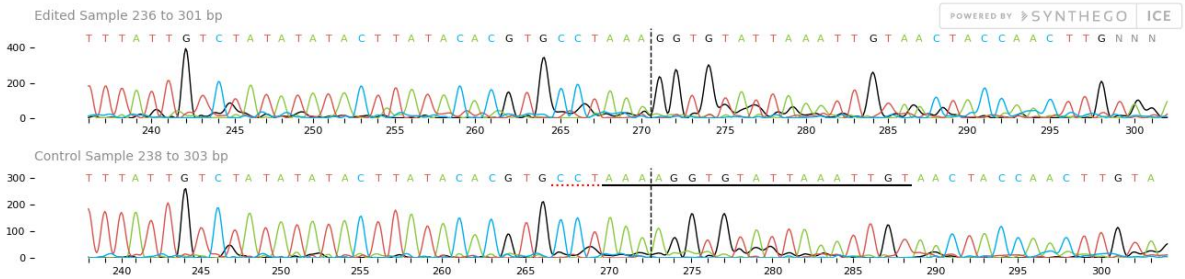
Sample 16



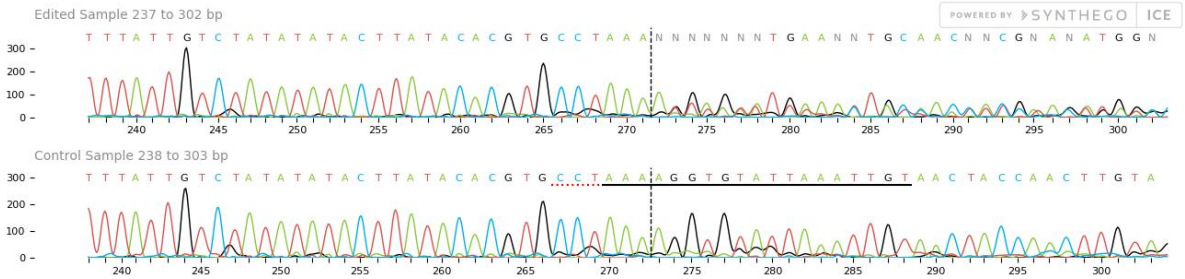
Sample 18



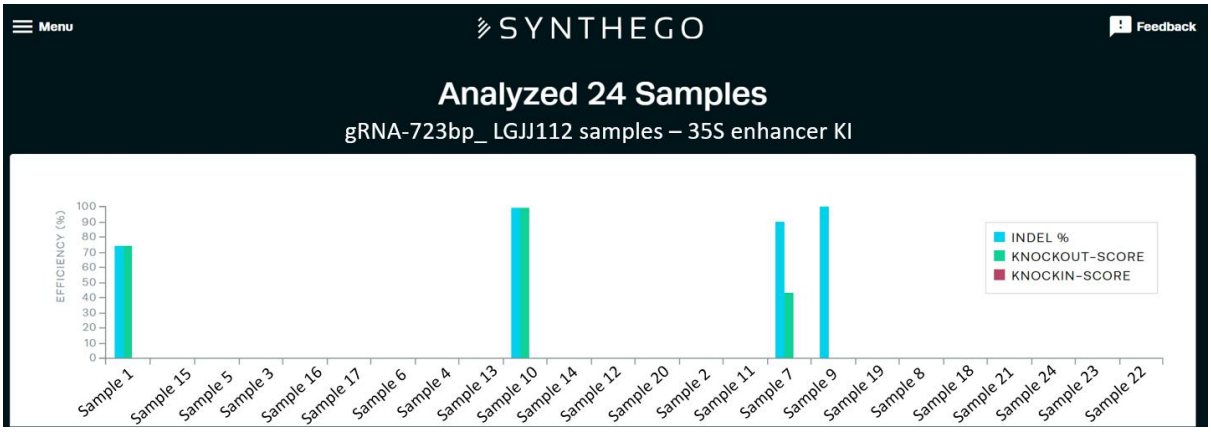
Sample 8



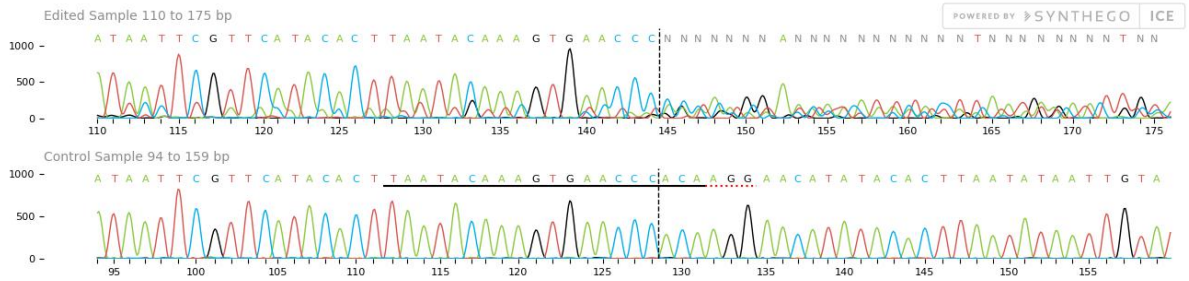
Sample 24



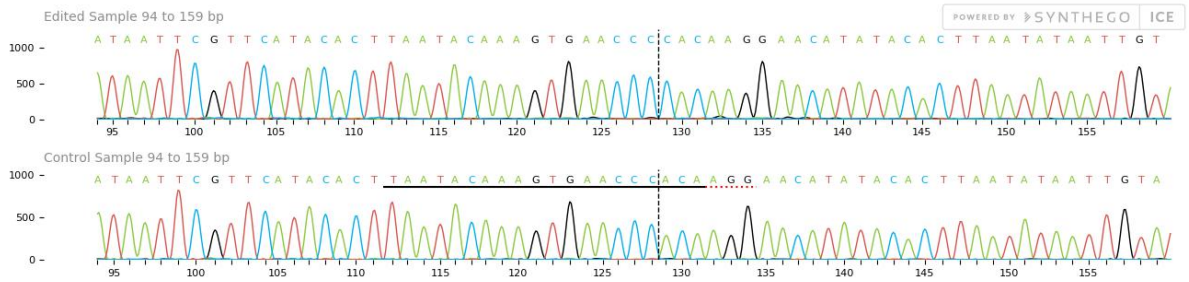
Appendix A.2. LGJJ112 samples



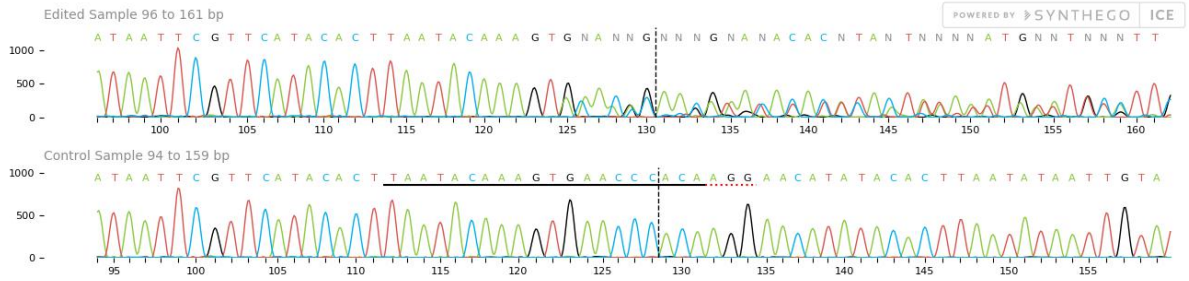
Sample 1



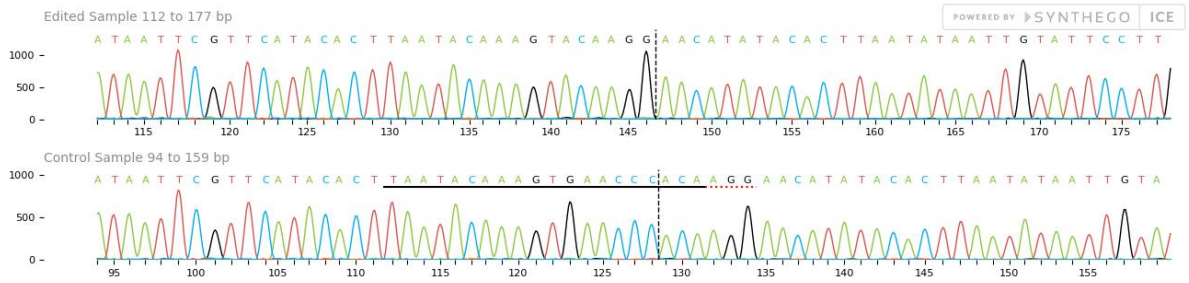
Sample 10



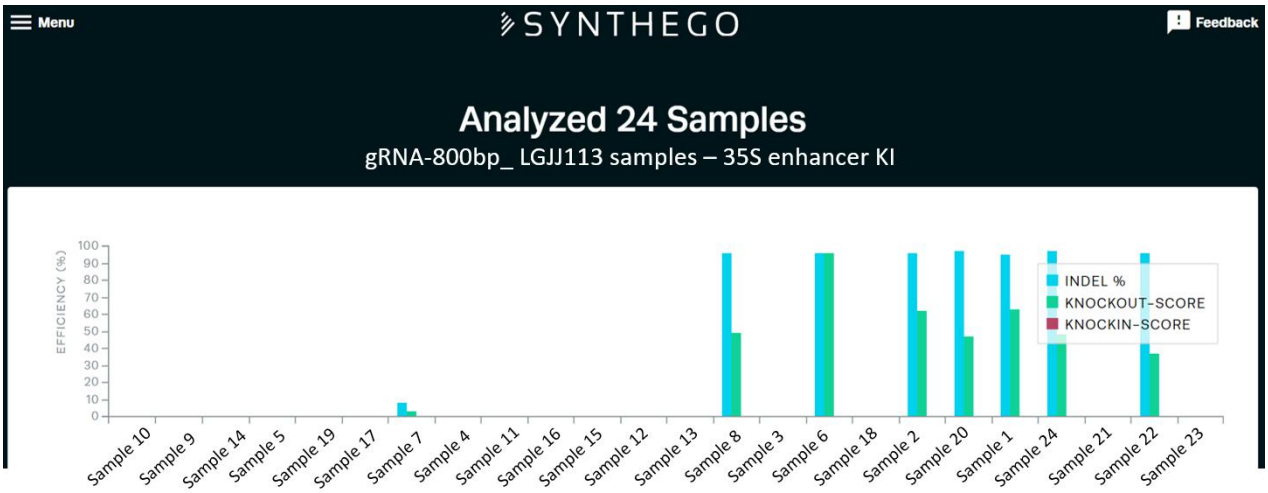
Sample 7



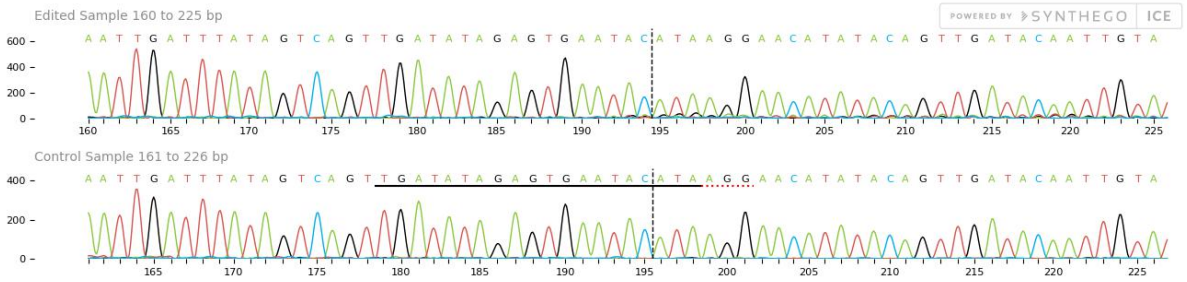
Sample 9



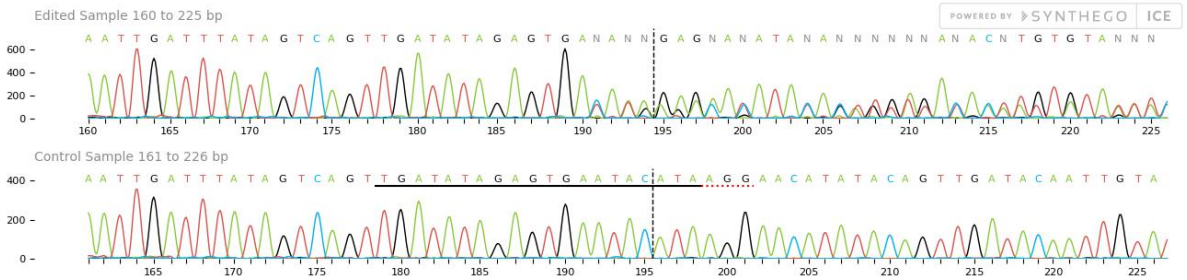
Appendix A.3. LGJJ113 samples



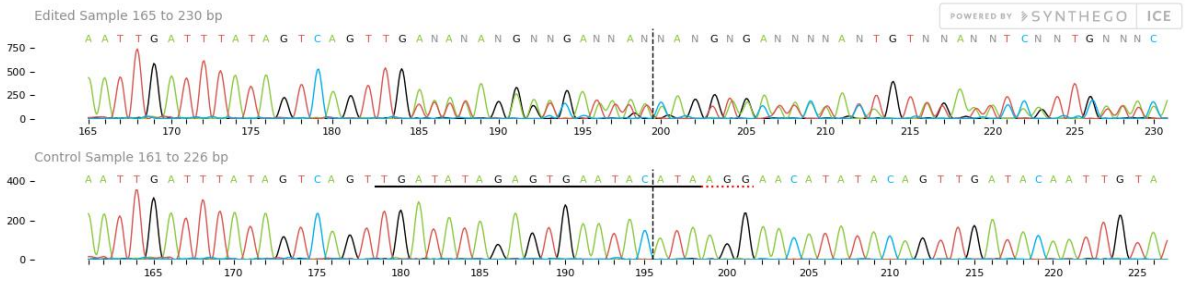
Sample 7



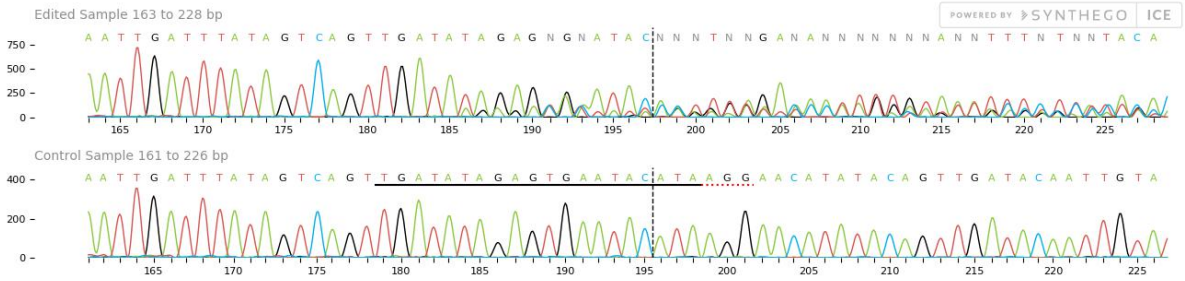
Sample 8



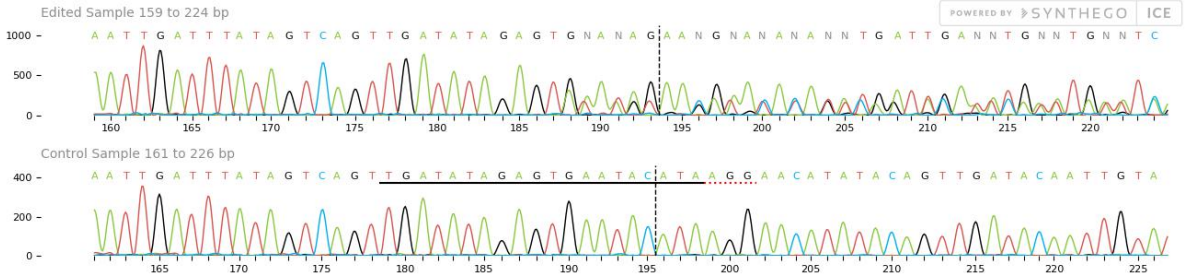
Sample 6



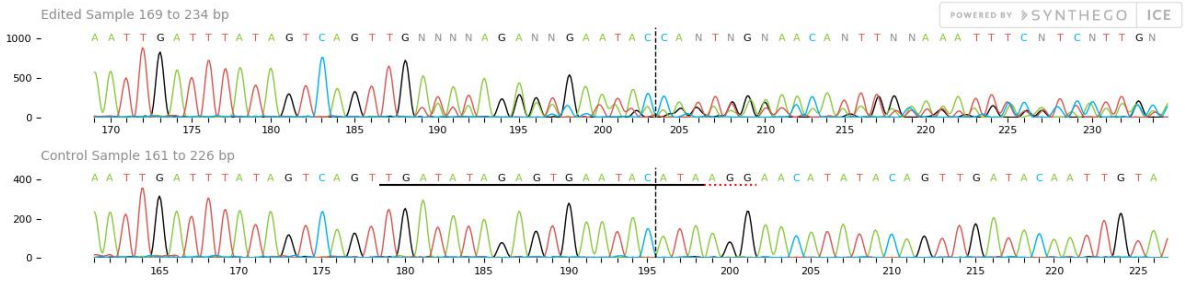
Sample 2



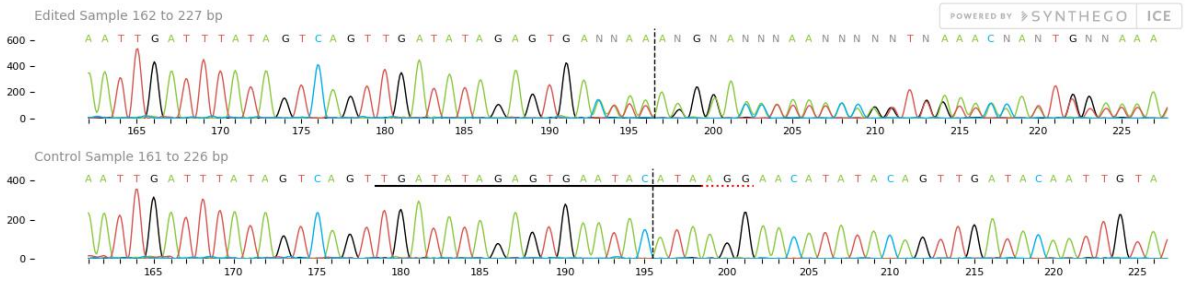
Sample 20



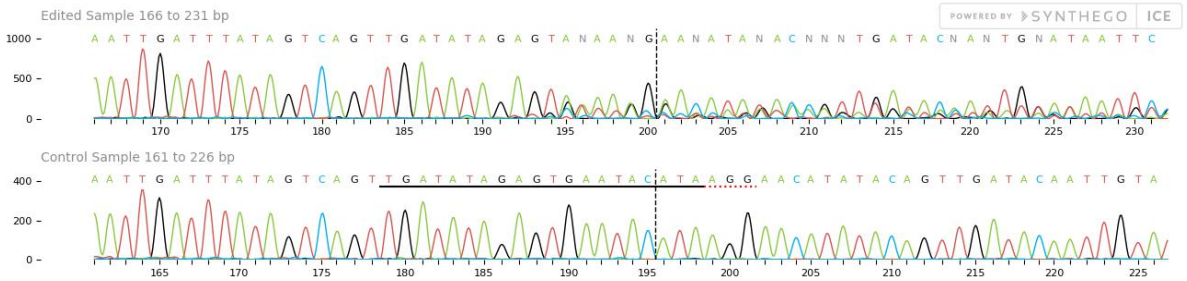
Sample 1



Sample 24



Sample 22

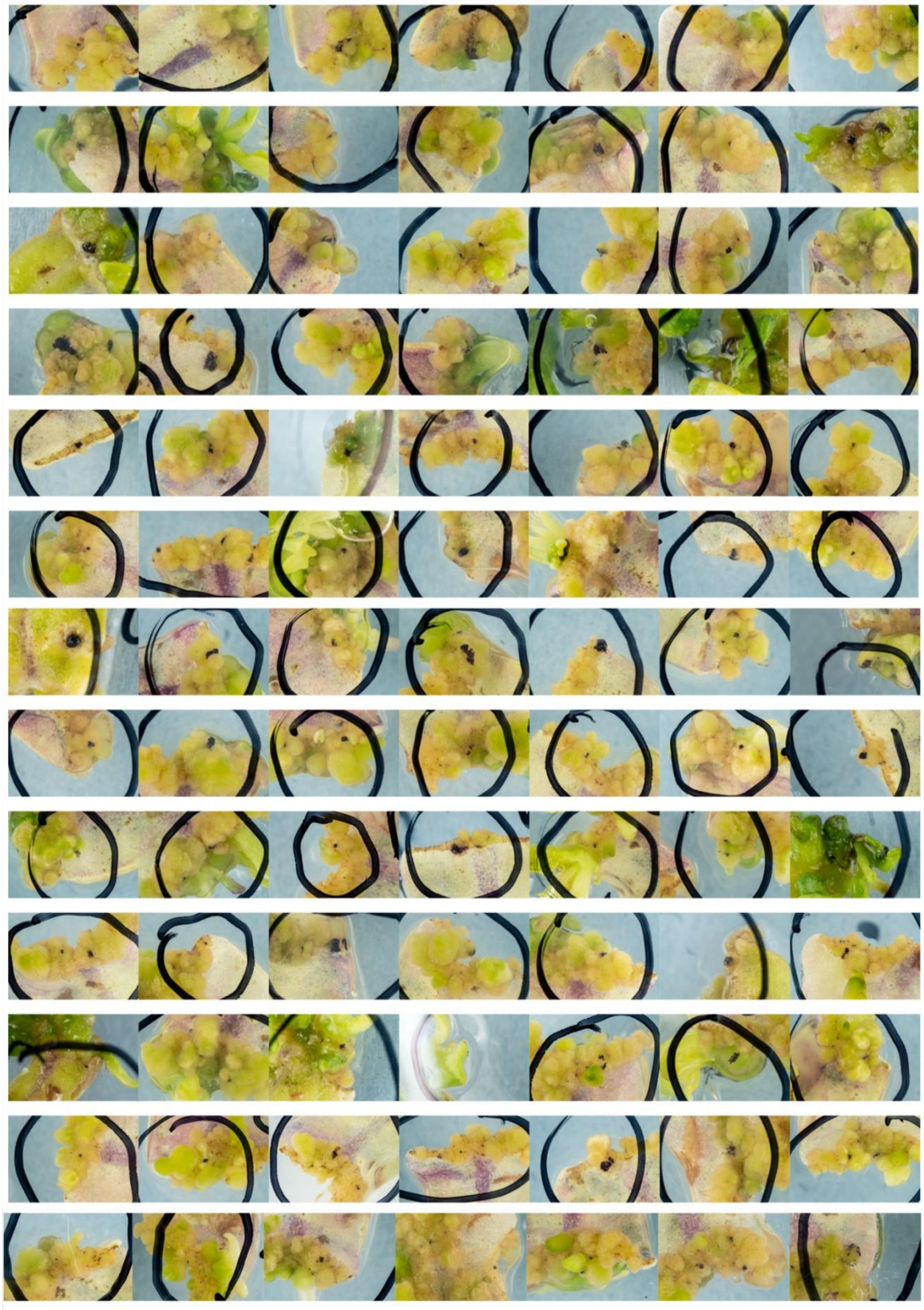

Edited Sample 166 to 231 bp
POWERED BY SYNTHEGO ICE
Control Sample 161 to 226 bp

Appendix B. Raw images used for Image J analysis

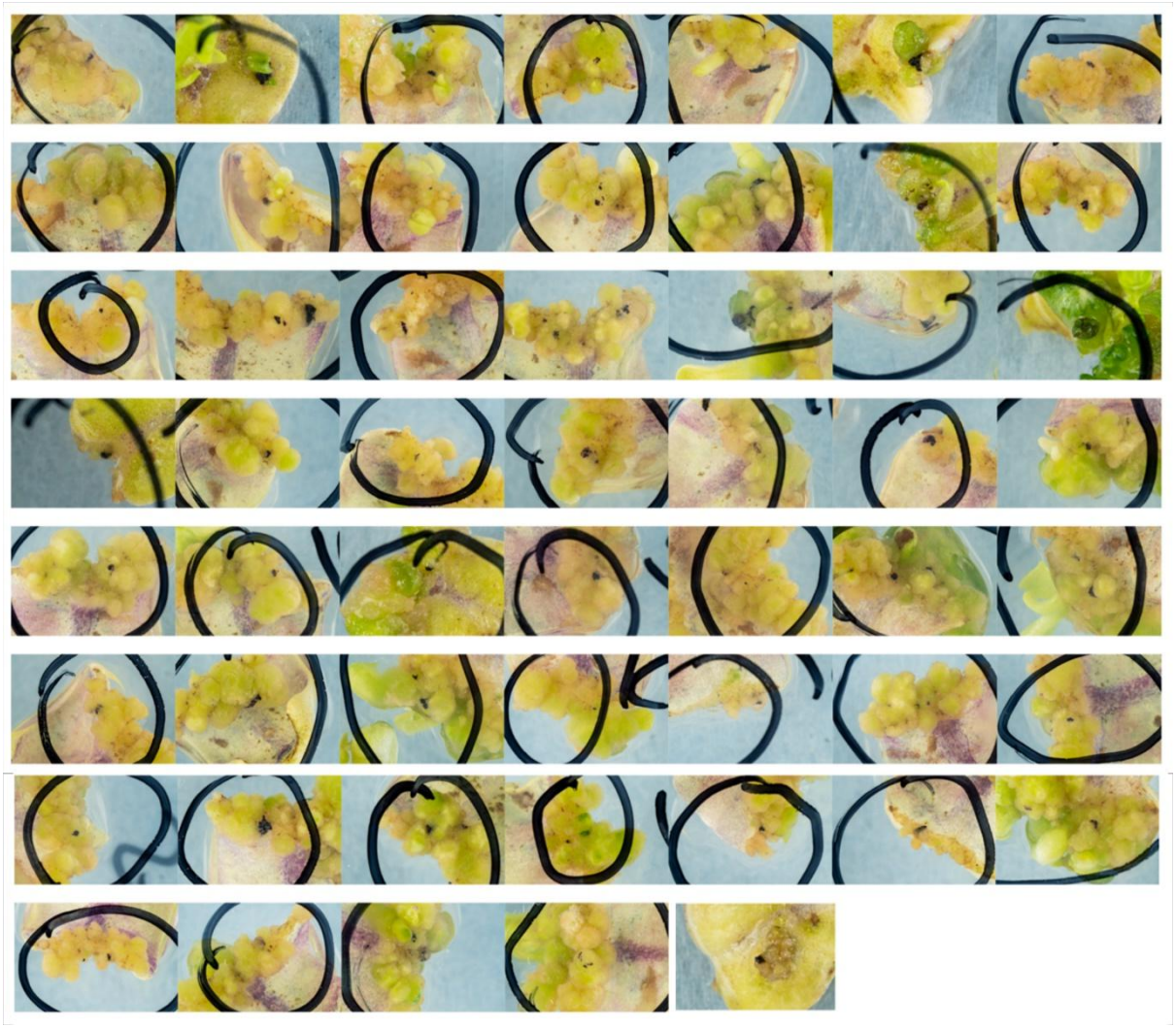
Appendix B.1. Collated pictures of explants from the raw NEF file used for the Image J analysis (LGJJ151)

Appendix B.2. Collated pictures of explants from the raw NEF file used for the Image J analysis (pTC147)

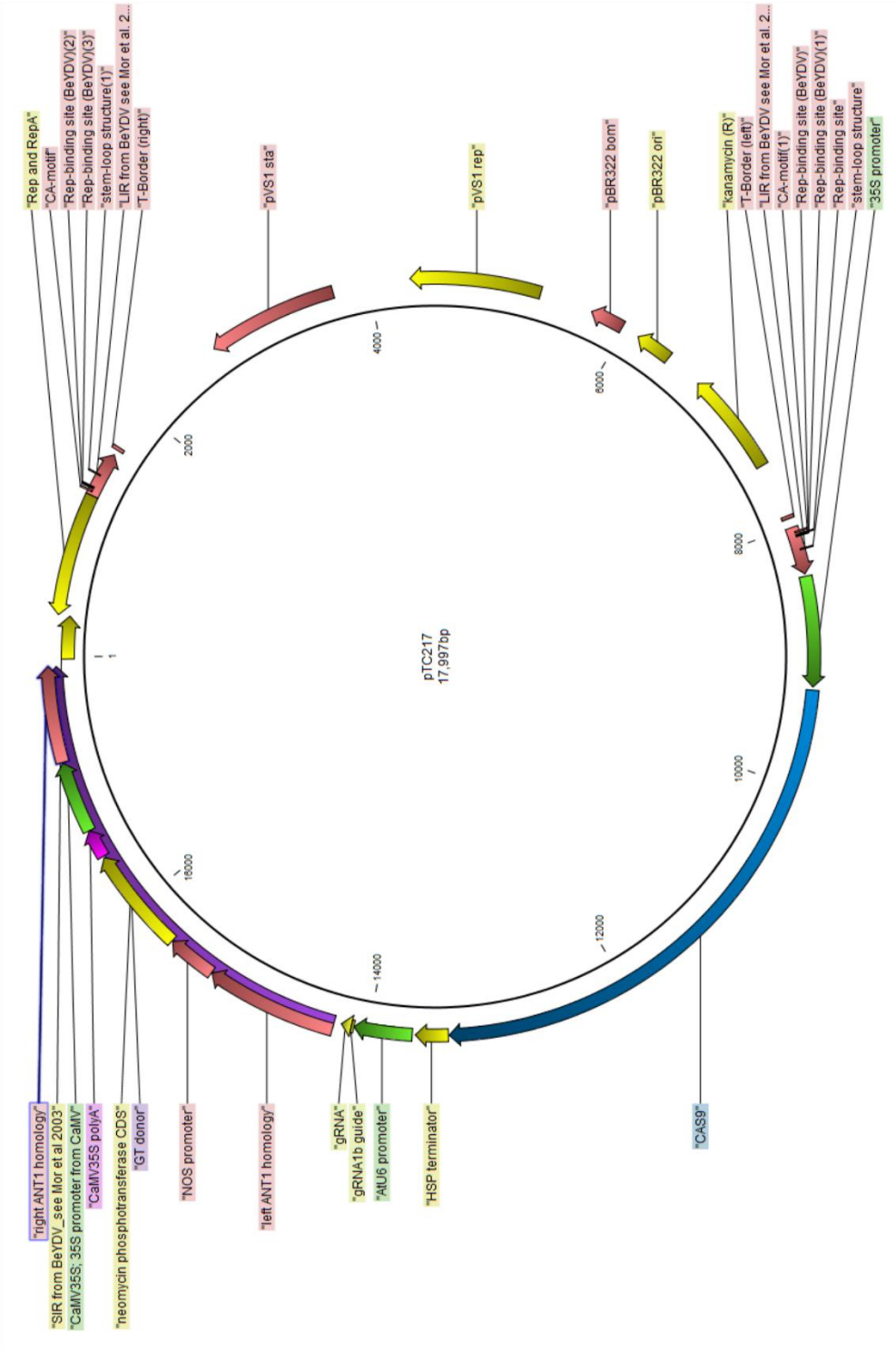
Appendix C. Collated pictures of the 145 purples sectors observed following the first transformation of vector LGJJ216 into Moneymaker cotyledons.



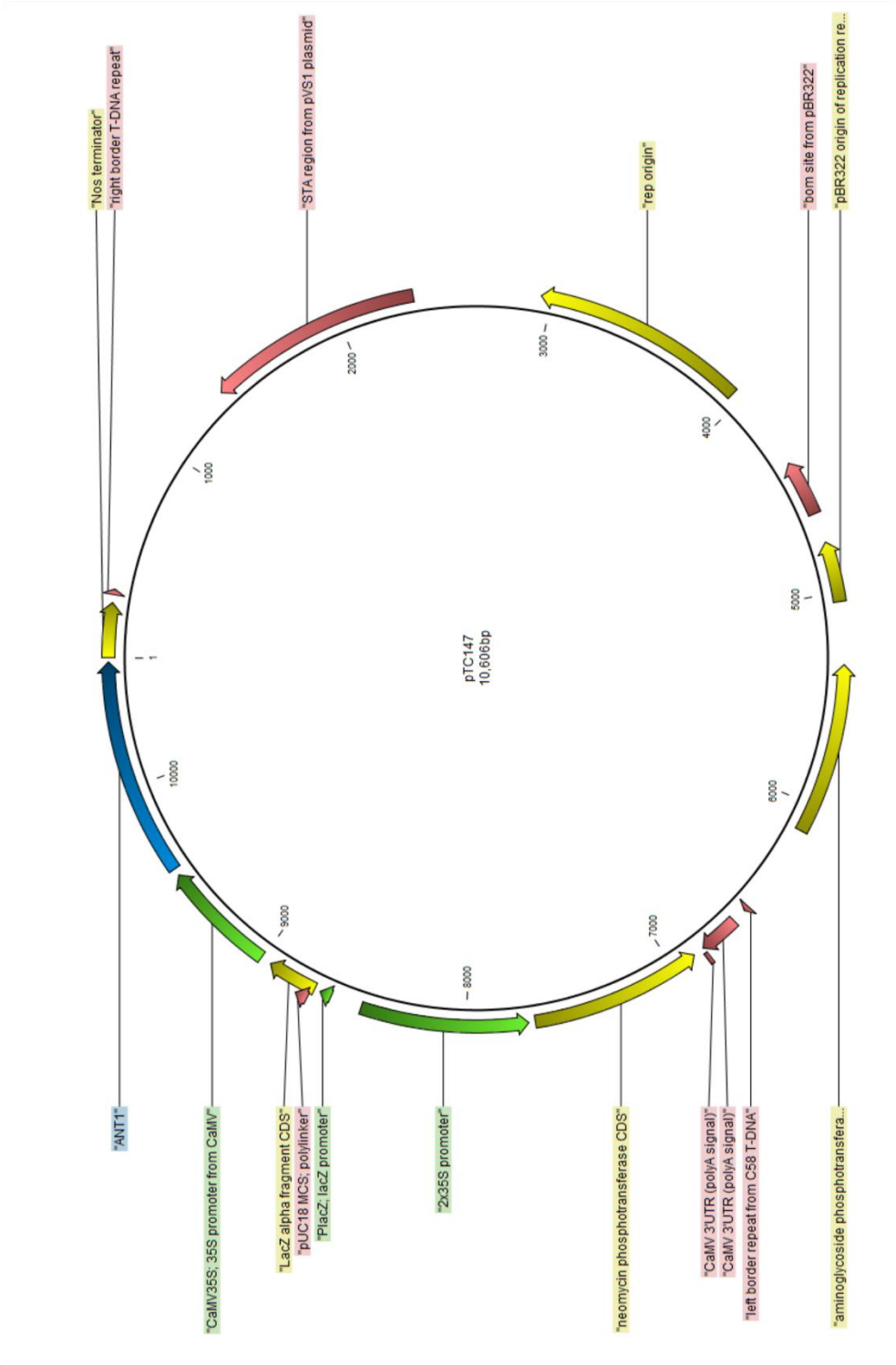
Appendix C continued.



Appendix D. Maps of constructs from the Cermak et al's publication (2015) used in the present study



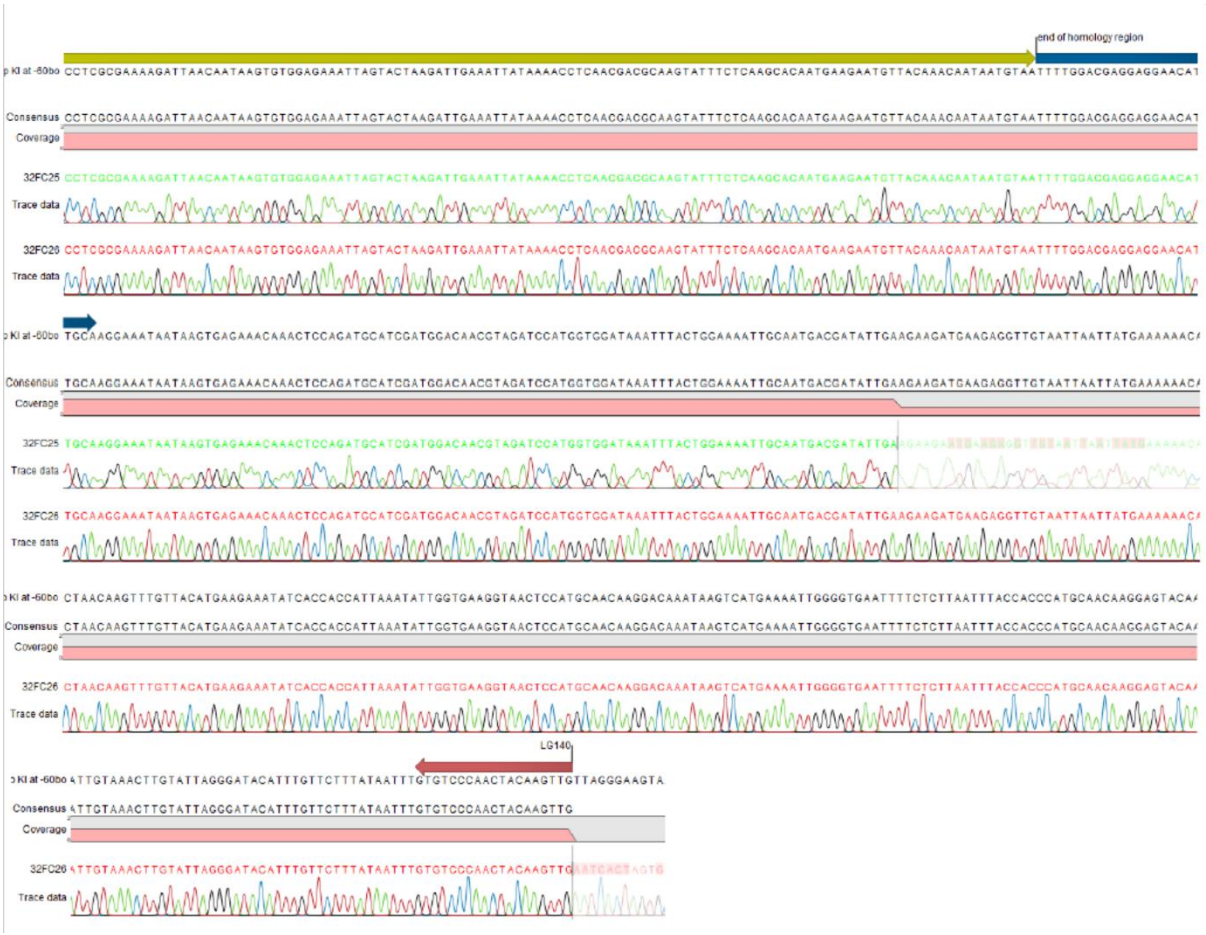
Appendix D.1 Map of vector pTC217.



Appendix D.2 Map of vector pTC147.

Appendix E. Sequencing reads covering the junctions between the KI DNA and genomic DNA at the target locus.

… continued



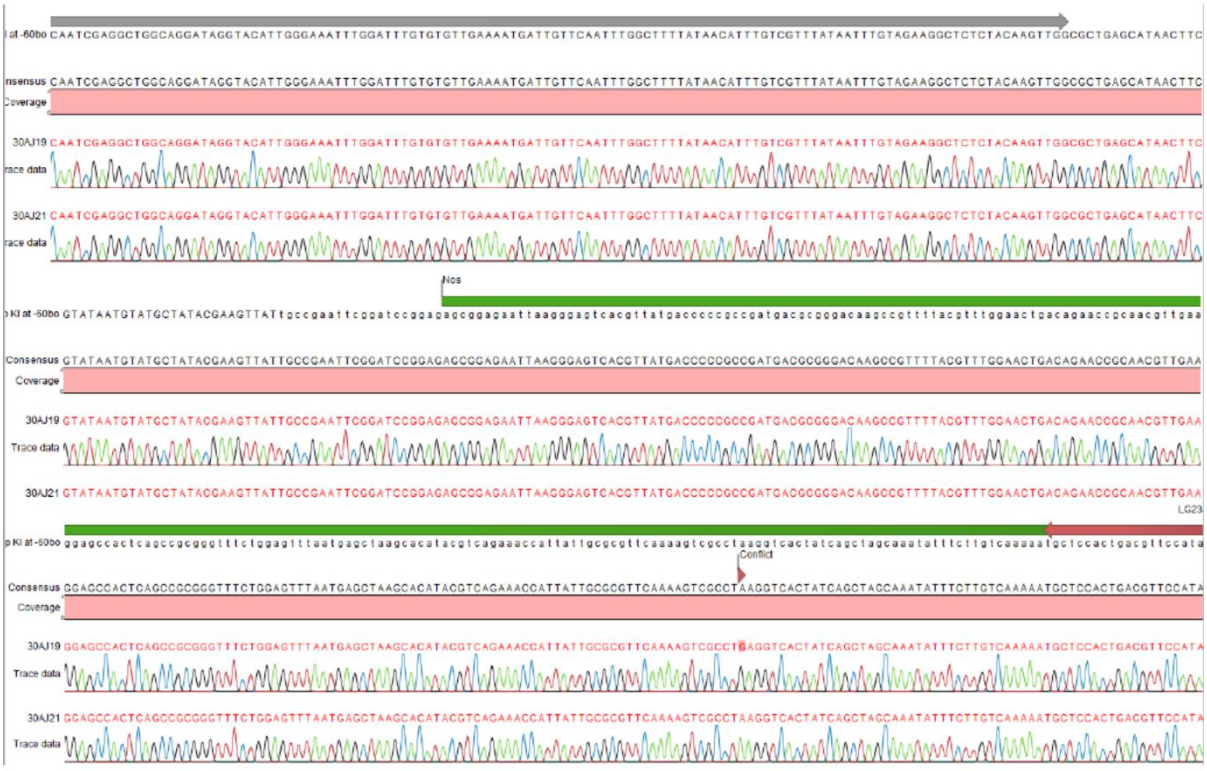
Appendix E.1 Full sequencing coverage of the cloned 1.3 kb PCR amplicon spanning the right junction of the 35S promoter KI of sample arising from vector LGJJ52 obtained with primers LG141 and LG140. Green sequences: forward reads of 2 clones sequenced. Red sequence: reverse read of 1 out of 2 clones sequenced. 2nd clone sequencing reaction failed so no included. Mutation in one clone +90 bp downstream of primer LG141 binding site is likely a sequencing error as only present in one of the two clones sequenced.


LG203
ANT1 promoter
left homology region
Consensus
Coverage
Trace data
30AJ20_
30AJ18
30AJ19
30AJ21

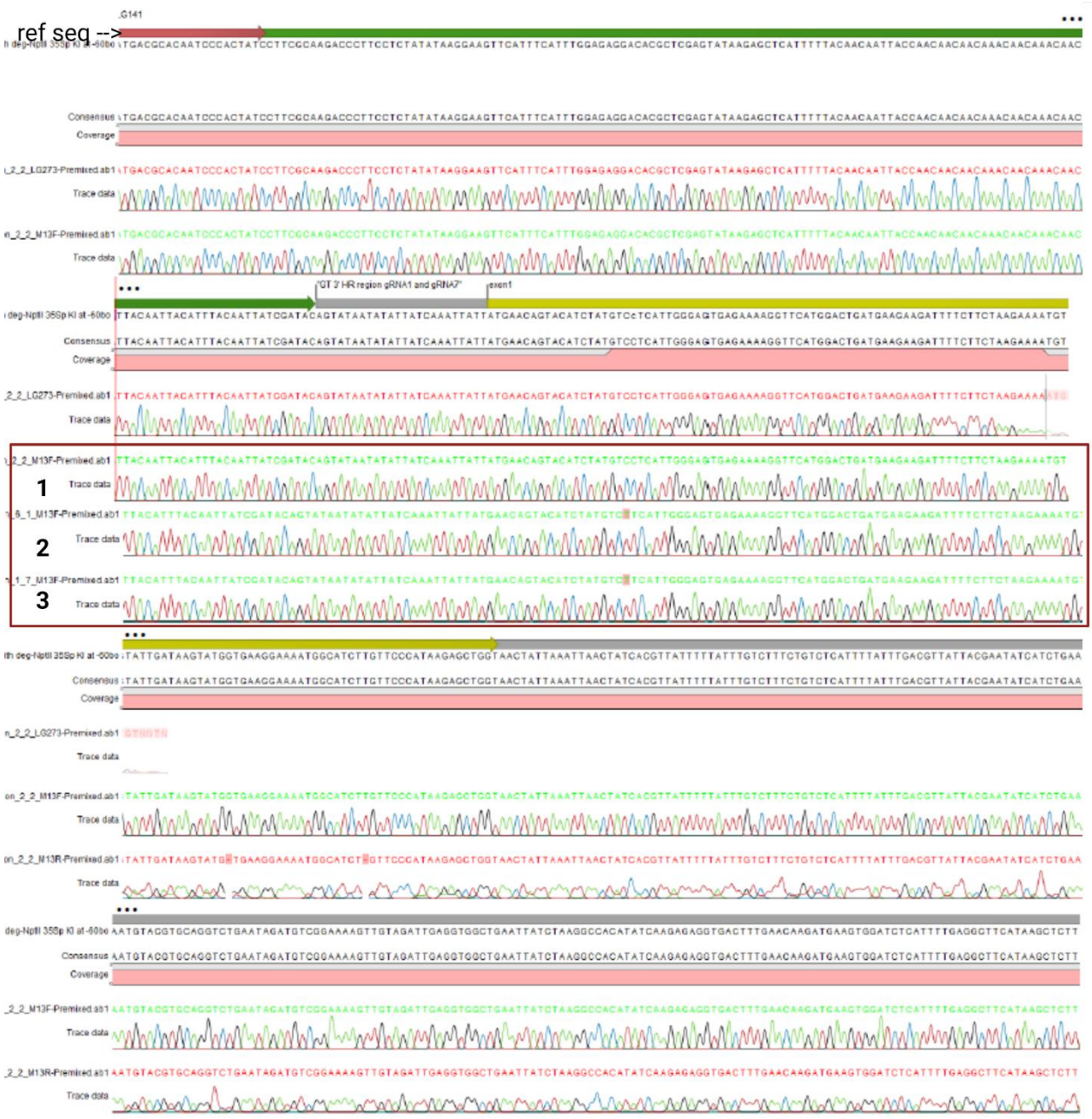
…continued


Kl at -60bo
Consensus
Coverage
30AJ20_
Trace data
30AJ18
30AJ19
30AJ21

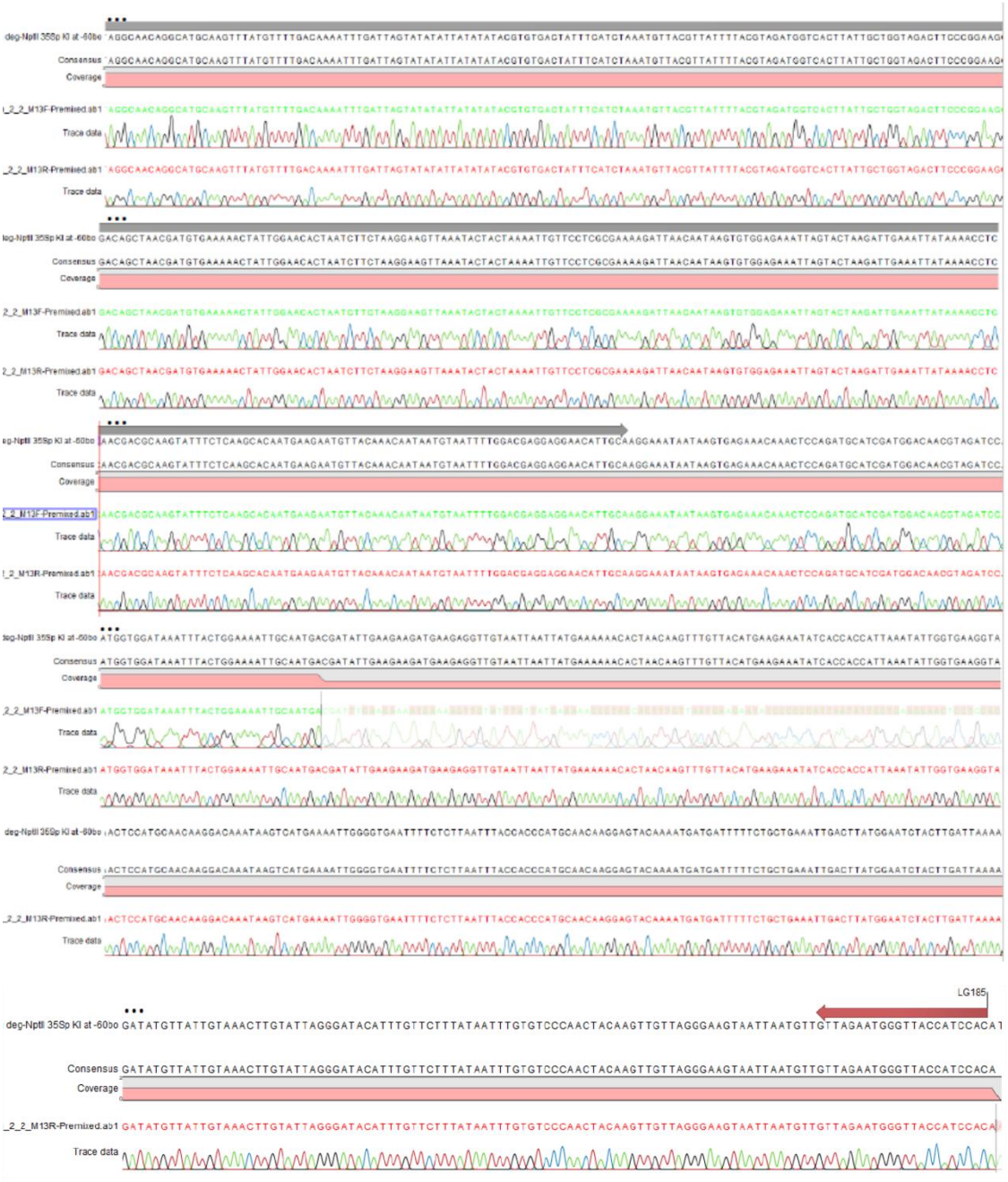
… continued



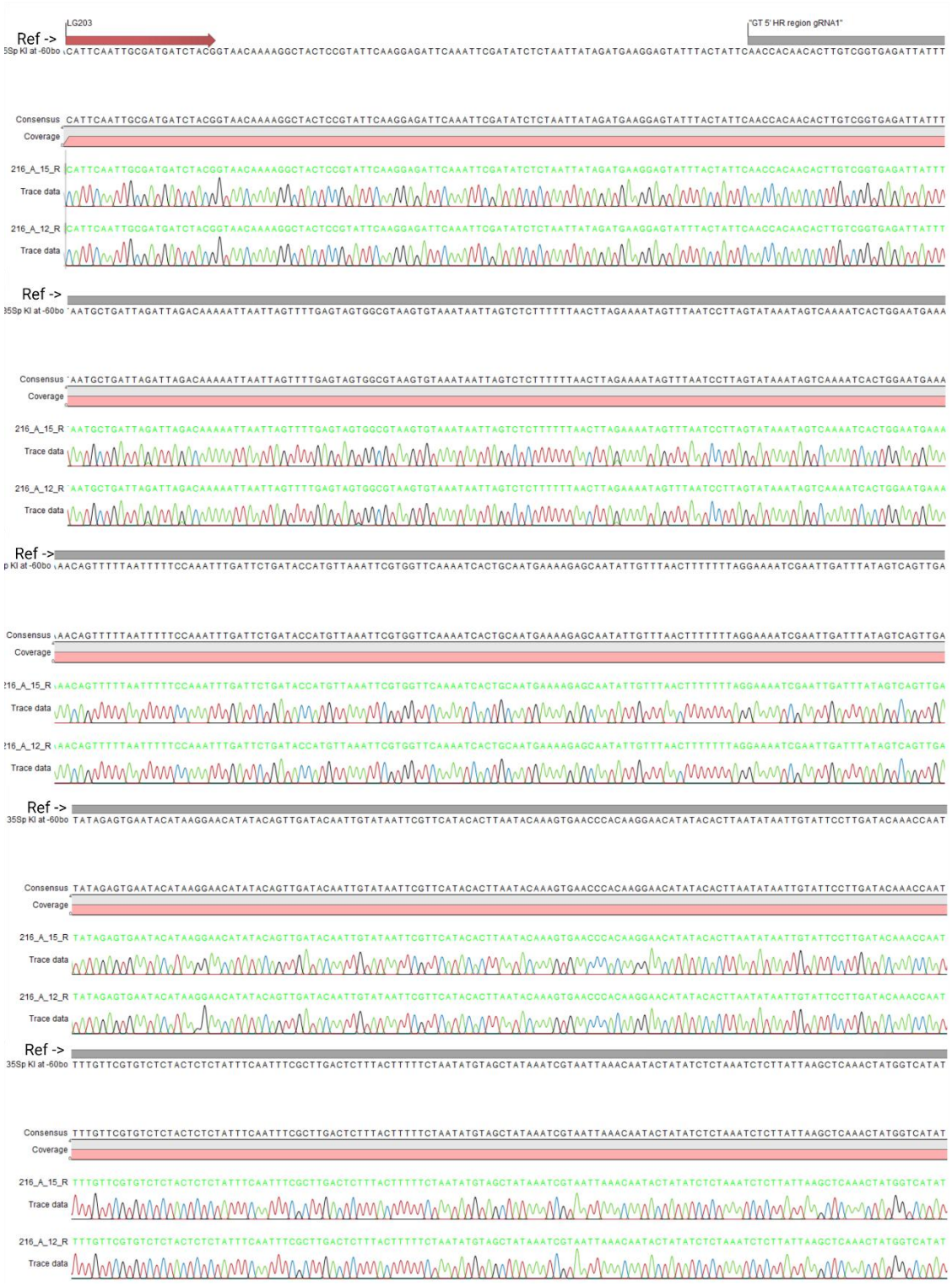
Appendix E.2 Full sequencing coverage of the cloned 1.3 kb PCR amplicon spanning the left junction of the 35S promoter KI of sample arising from vector LGJJ52 obtained with primers LG203 and LG23. Green sequences: forward reads of 2 clones sequenced. Red sequence: reverse reads of 2 clones sequenced. The one mutation detected in the sequencing +60 bp upstream of primer LG23 binding site is likely a sequencing reaction error as only one clone out of 2 harbours this mutation. The other clone contains the expected sequence.


ref seq -->
Consensus
Coverage
Trace data
"GT 3' HR region gRNA1 and gRNA2"
exon1
1
2
3

Continued…



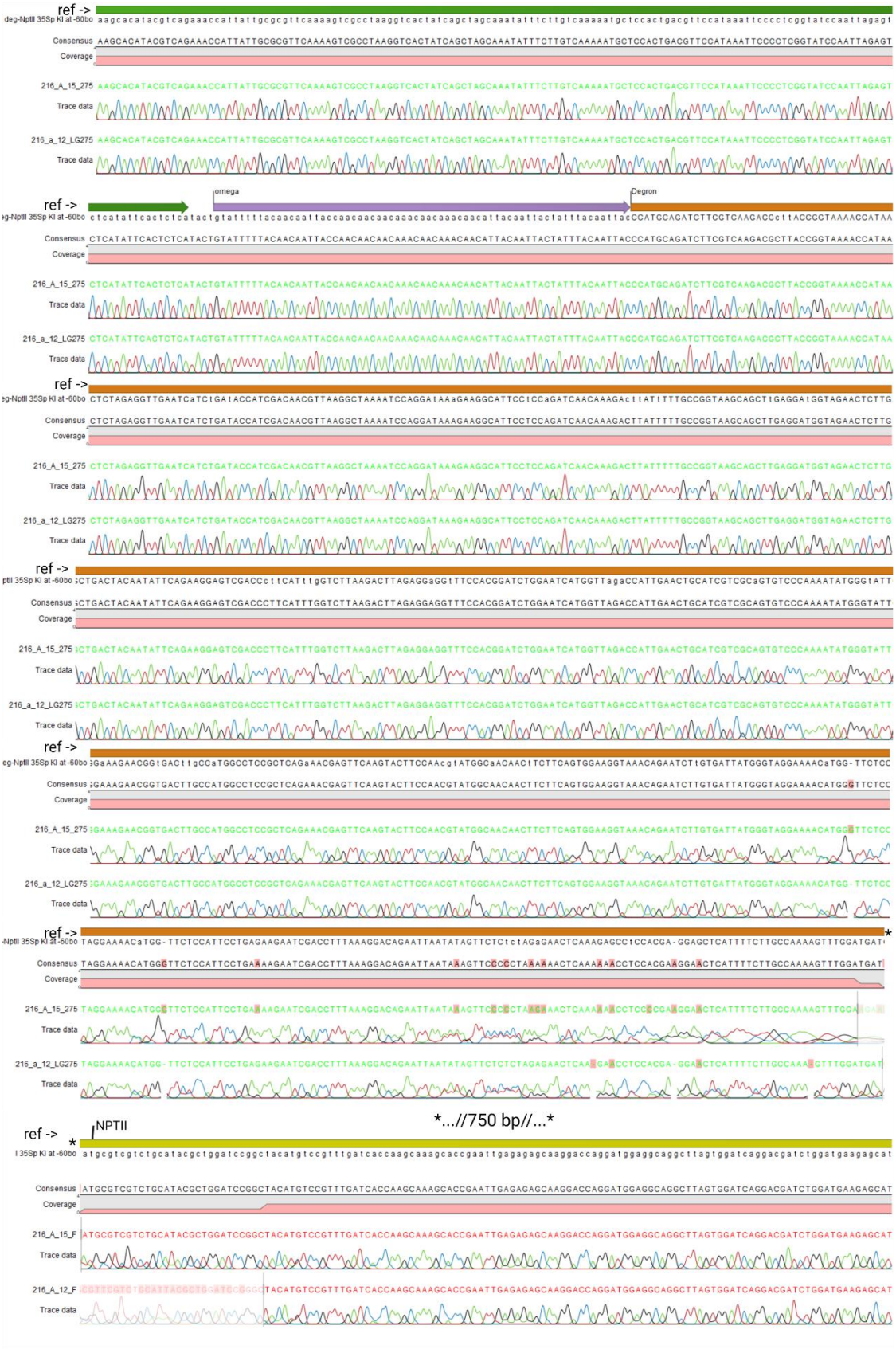
Appendix E.3. Full sequencing coverage of the cloned 1.3 kb PCR amplicon spanning the right junction of the 35S promoter KI of samples arising from vector LGJJ181 obtained with primers LG141 and LG185. Sequencing for on clone representing the sequence contained in sample 181_P2 is shown in its entirety (green sequence: forward reads, red sequence: reverse reads). Sequences retrieved for samples 181_P1 and 181_PG1 (not shown here) were identical to the sequences of 181_P2 along across the 1.3 kb fragment, except the region spanning the exon 1 of *ANT1* (shown here), highlighted by a red square in the first part of this figure. In this square, line labelled "1" represent sequencing from sample 181_P2, "2" from 181_P1 and "3" from 181_PG1.


LG203
"GT 5' HR region gRNA1"
Ref ->
Consensus
Coverage
216_A_15_R
Trace data
216_A_12_R
Trace data

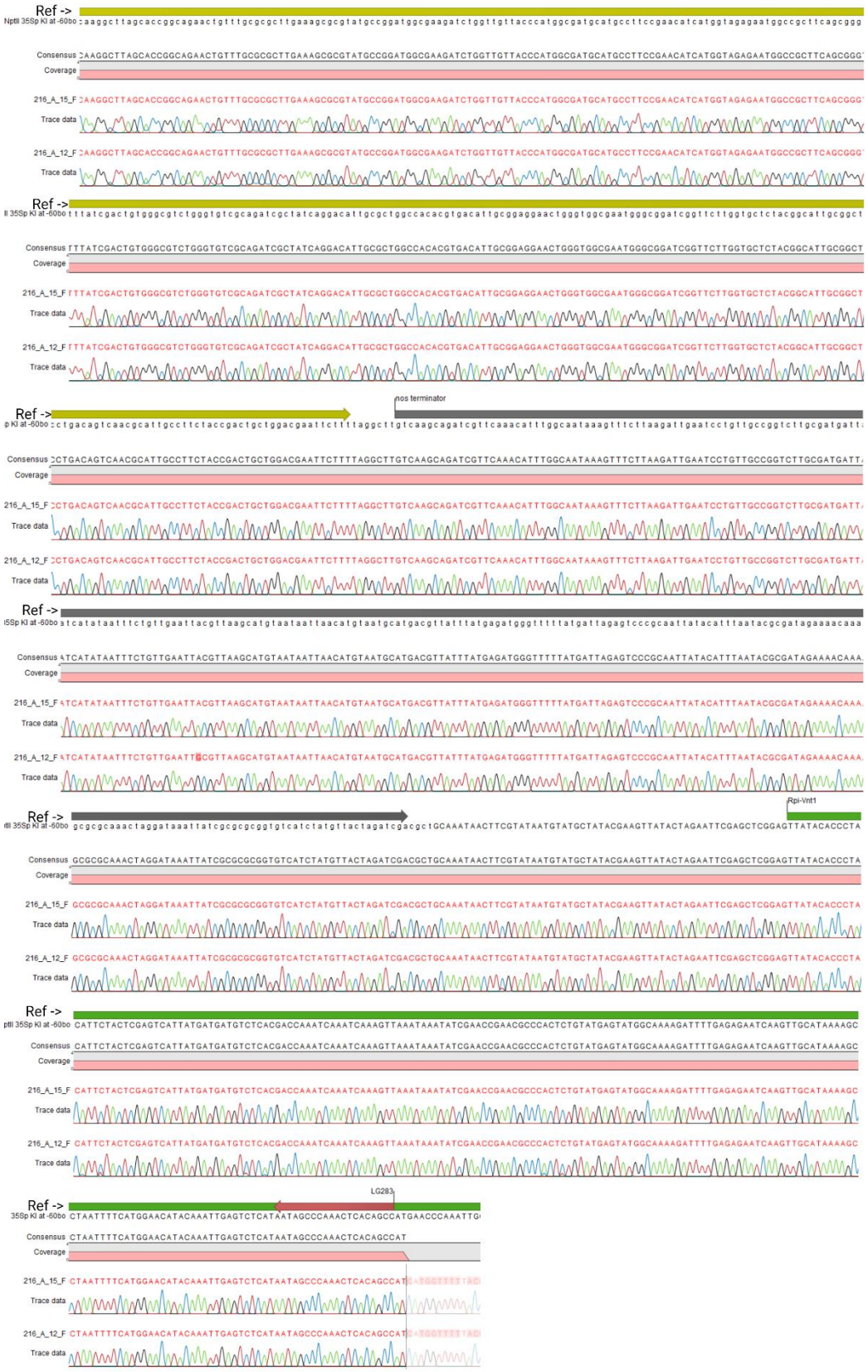
…continued


ref ->
Consensus
Coverage
216_A_15_R
216_A_12_R
216_A_15_275
216_a_12_LG275
Trace data

…continued

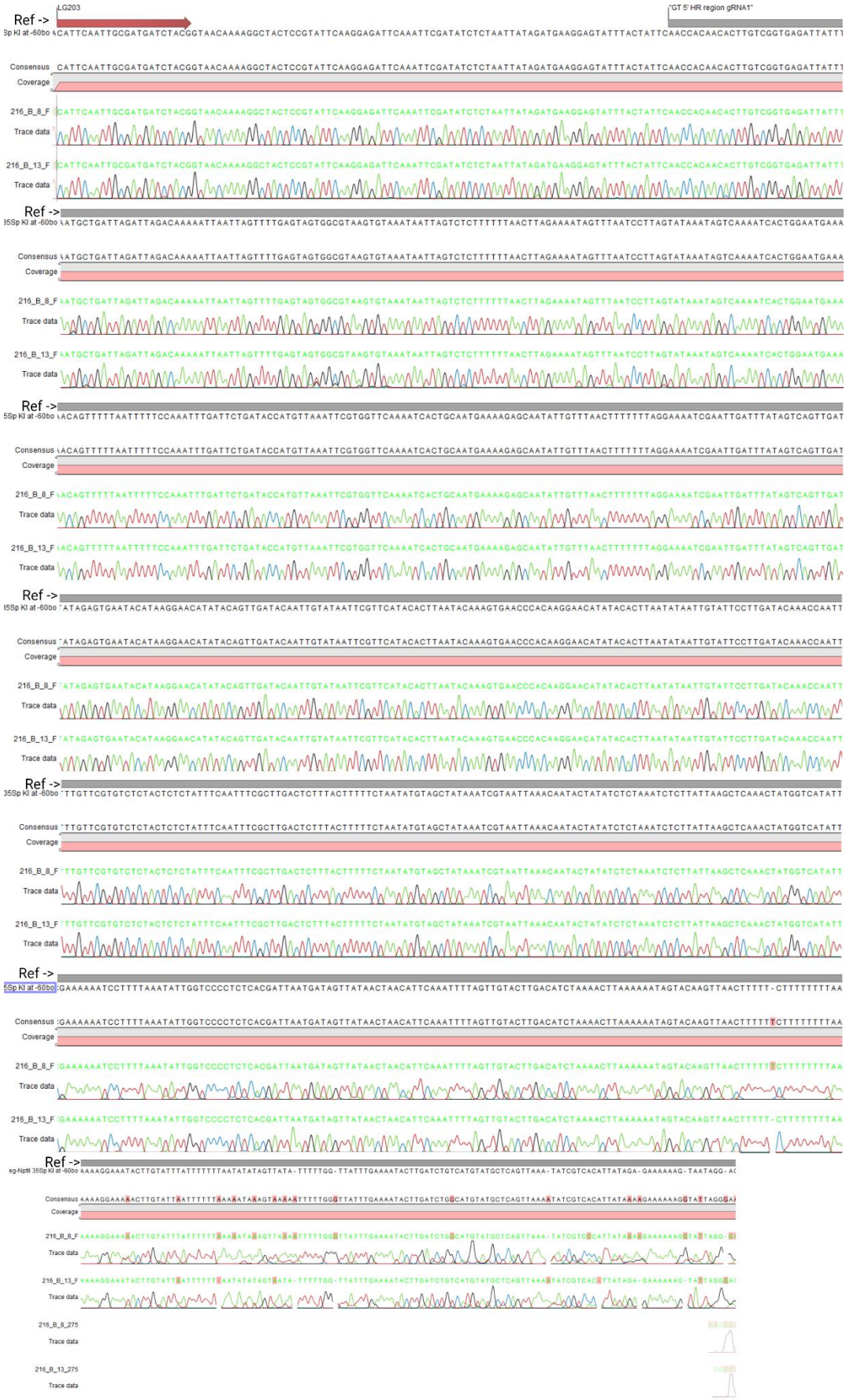

ref ->
Consensus
Coverage
216_A_15_275
Trace data
216_a_12_LG275
omega
Degron
...//750 bp//...
NPTII

…continued

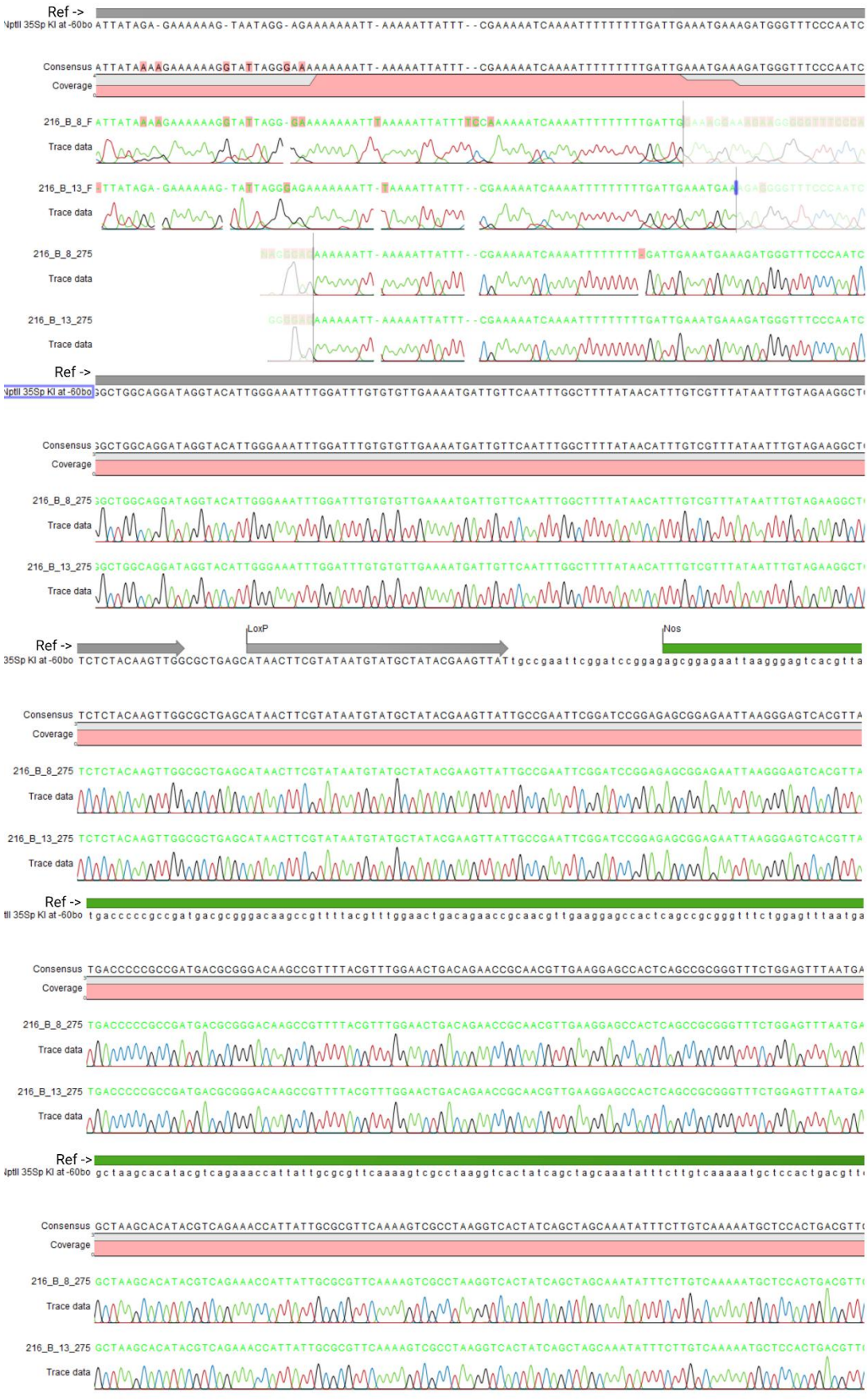


Appendix E.4. Sequencing coverage of the cloned 3.7 kb PCR amplicon spanning the left junction of the KI of sample 216_1 arising from vector LGJJ216 obtained with primers

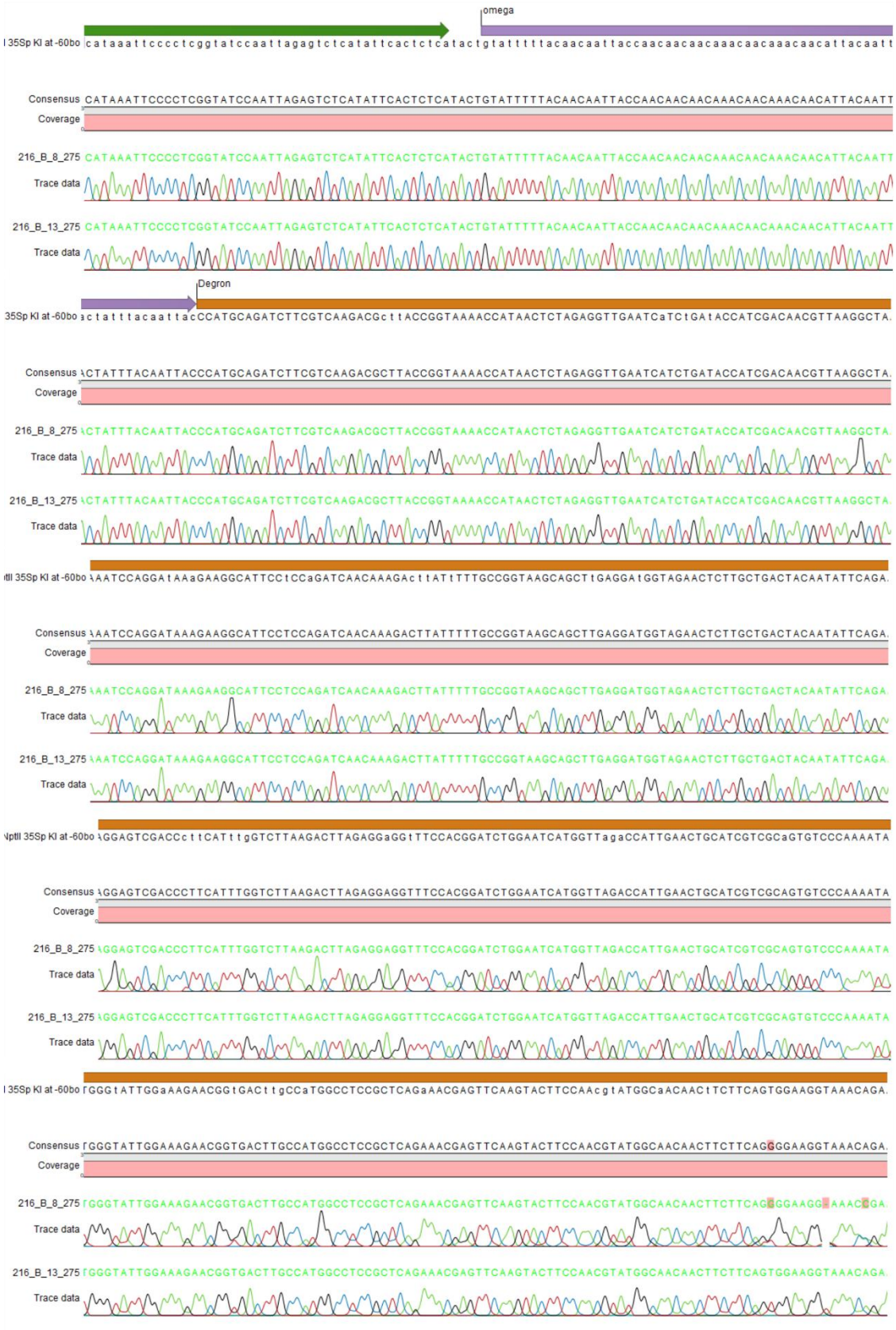
LG203 and LG283.). Gap in sequencing coverage spanning 750 bp (demonstrated as *//..//* on the sequence alignment) within the donor DNA, from mid-degron (orange bar annotation on reference ("ref") sequence) to the 5' region of the *NPTII* CDS (yellow bar annotation). Sequencing from two clones shown here. Green sequences: forward sequencing reads. Red sequences: reverse sequencing reads. Base pair mutations observed in one out of the two clones analysed are considered as sequencing errors.


LG203
"GT 5' HR region gRNA1"
Ref ->
Consensus
Coverage
216_B_8_F
Trace data
216_B_13_F
Trace data

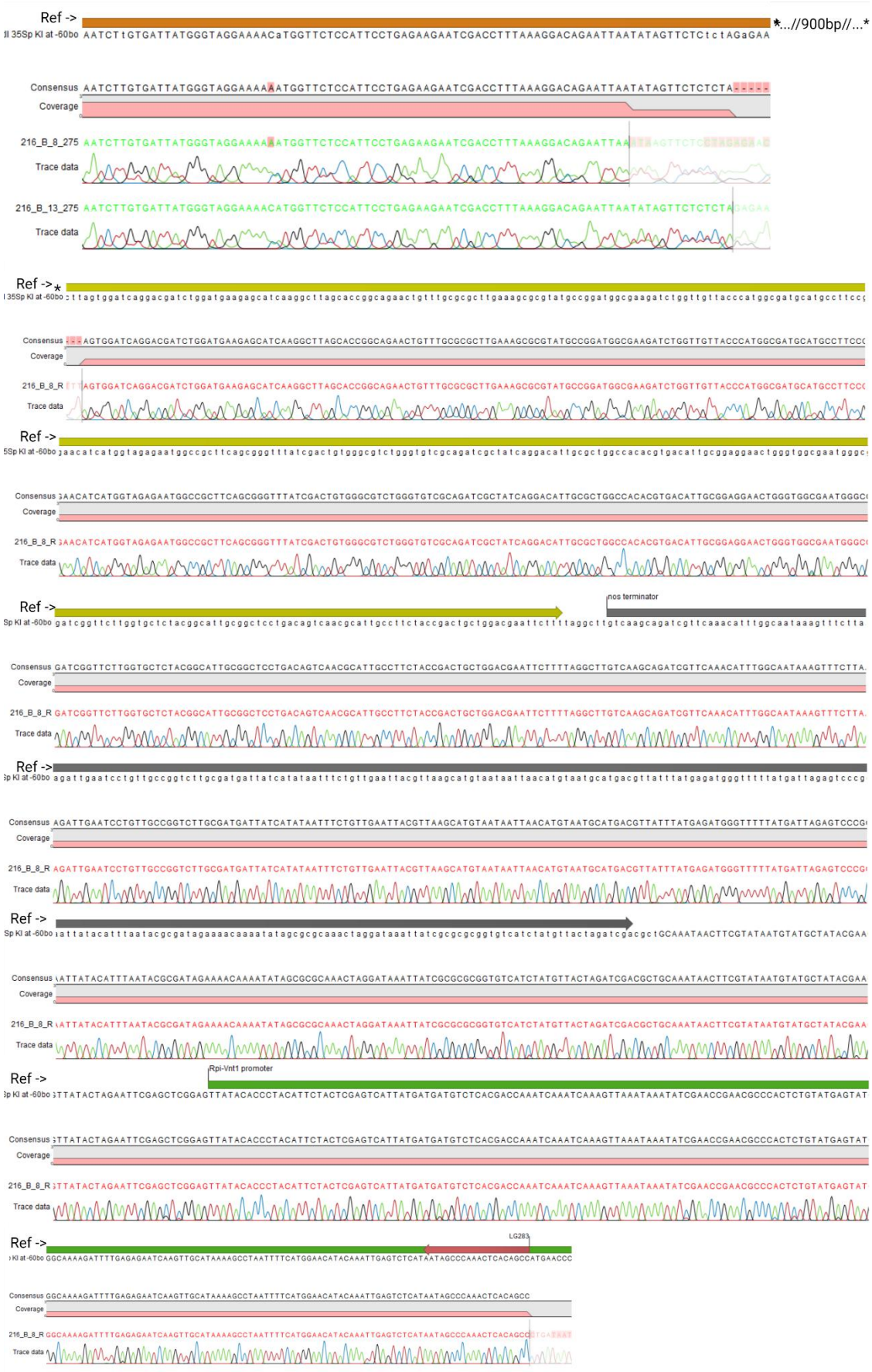
…continued


Ref ->
Consensus
Coverage
216_B_8_F
Trace data
216_B_13_F
216_B_8_275
216_B_13_275
LoxP
Nos

…continued

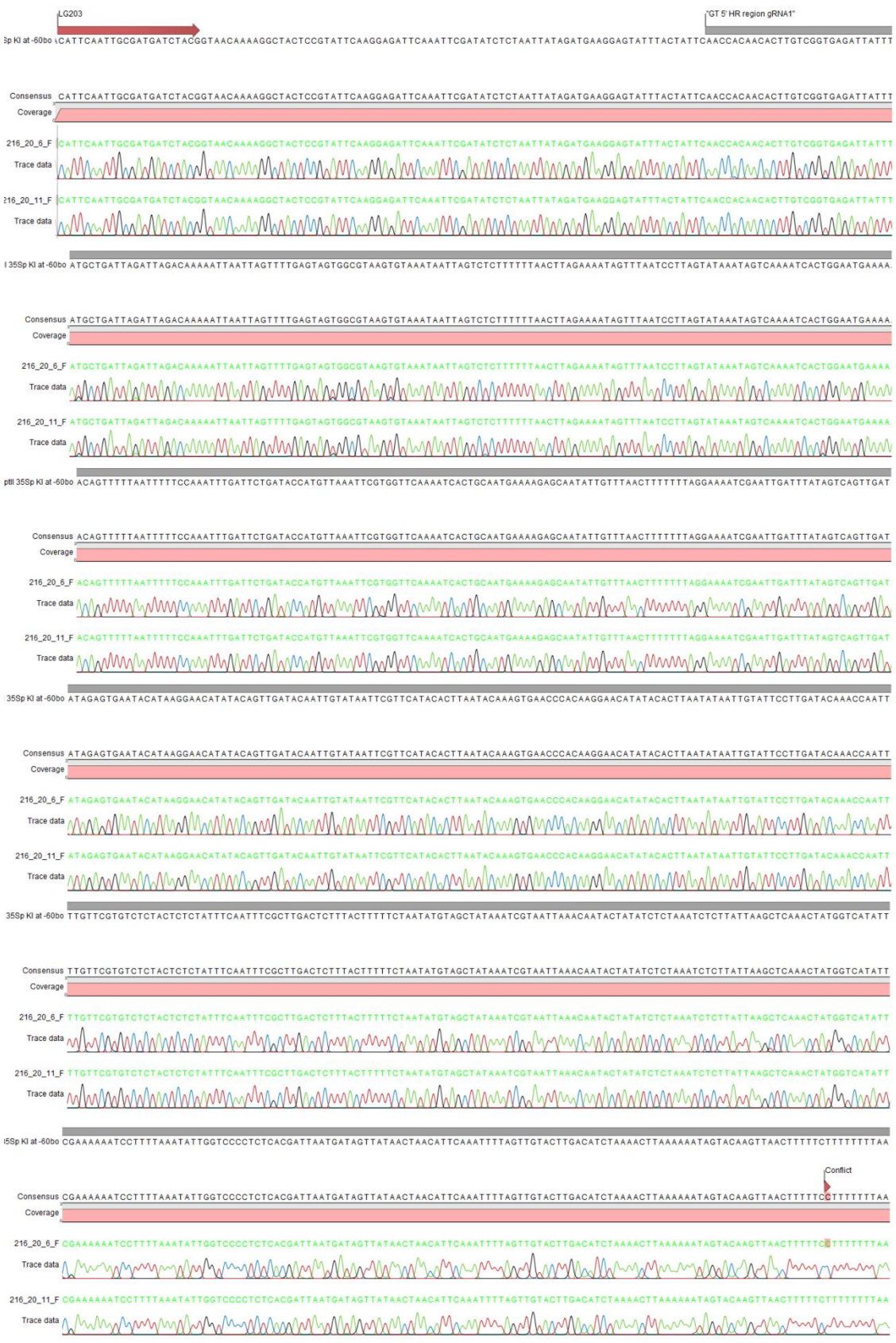

omega
Consensus
Coverage
216_B_8_275
Trace data
216_B_13_275
Trace data
Degron

…continued



Appendix E.5. Sequencing coverage of the cloned 3.7 kb PCR amplicon spanning the left

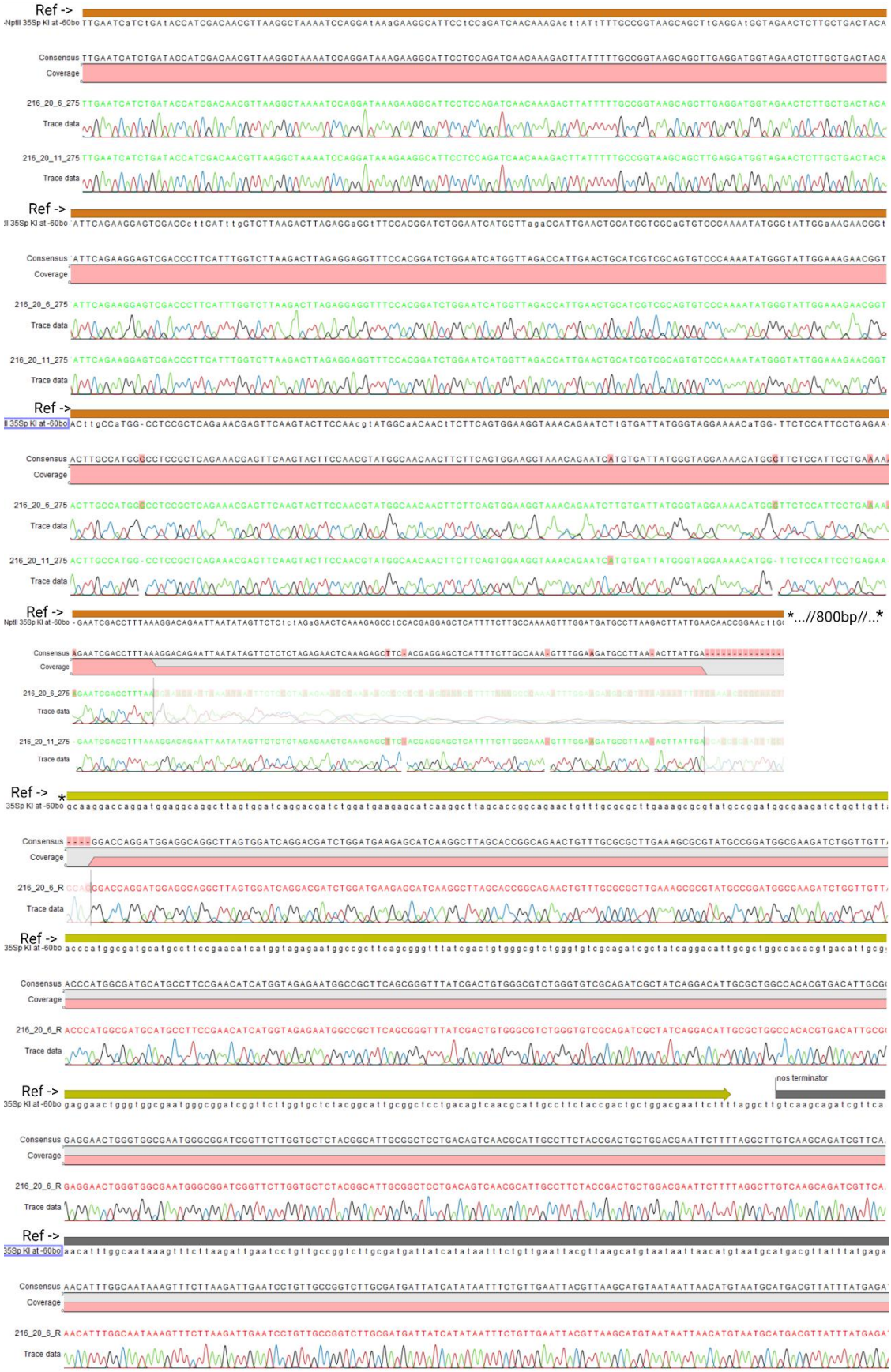
junction of the KI of sample 216_2 arising from vector LGJJ216 obtained with primers LG203 and LG283.). Gap in sequencing coverage spanning 900 bp (demonstrated as *//..//* on the sequence alignment) within the donor DNA, from mid-degron (orange bar annotation on reference ("ref") sequence) to the 5' region of the *NPTII* CDS (yellow bar annotation). Green sequences: forward sequencing reads from two clones. Base pair mutations observed in one out of the two clones analysed are considered as sequencing errors. Red sequences: reverse sequencing reads from one out of two clones sequenced. Sequencing from second clone failed so not shown here.


LG203
"GT 5' HR region gRNA1"
Consensus
Coverage
216_20_6_F
Trace data
216_20_11_F
Trace data
Conflict

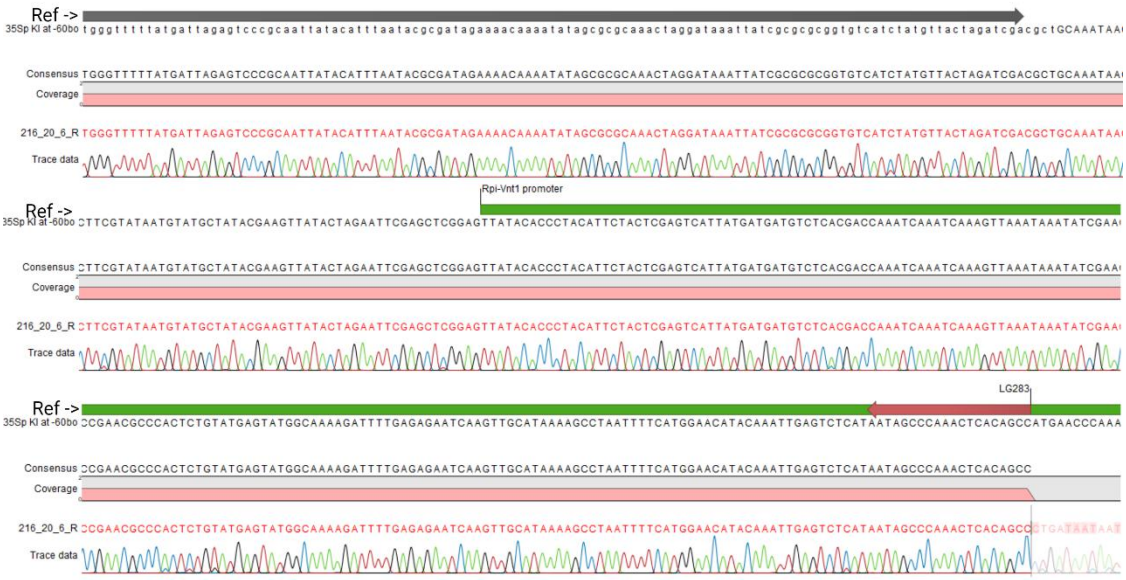
…continued


Ref ->
Consensus
Coverage
216_20_6_F
216_20_11_F
Trace data
216_20_6_275
216_20_11_275
LoxP
Nos
omega
Degron

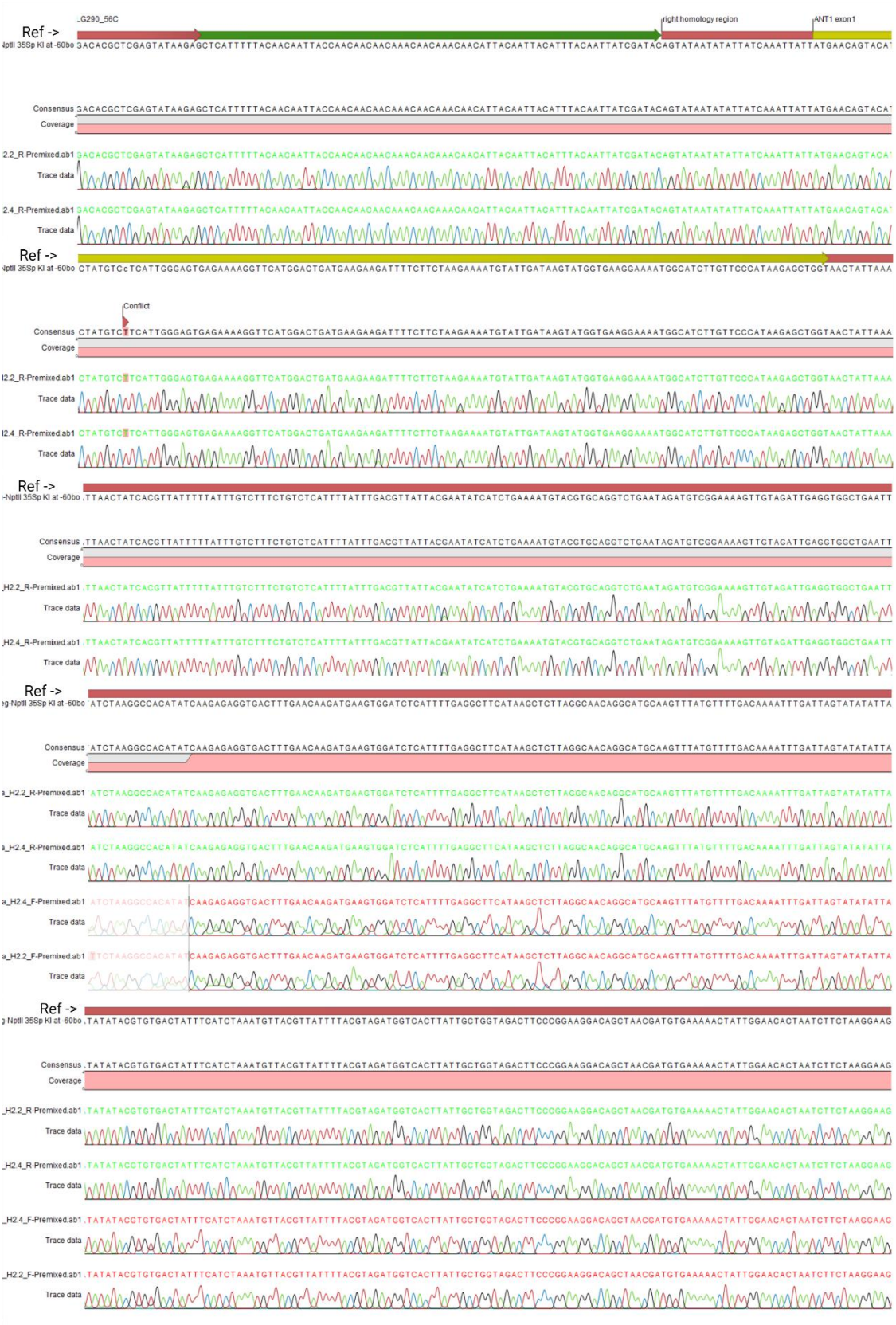
…continued


Ref ->
Consensus
Coverage
216_20_6_275
Trace data
216_20_11_275
Trace data
Ref ->
Consensus
Coverage
216_20_6_275
Trace data
216_20_11_275
Trace data
Ref ->
Consensus
Coverage
216_20_6_275
Trace data
216_20_11_275
Trace data
Ref ->
...//800bp//...
Consensus
Coverage
216_20_6_275
Trace data
216_20_11_275
Trace data
Ref ->
Consensus
Coverage
216_20_6_R
Trace data
Ref ->
Consensus
Coverage
216_20_6_R
Trace data
nos terminator
Ref ->
Consensus
Coverage
216_20_6_R
Trace data
Ref ->
Consensus
Coverage
216_20_6_R
Trace data

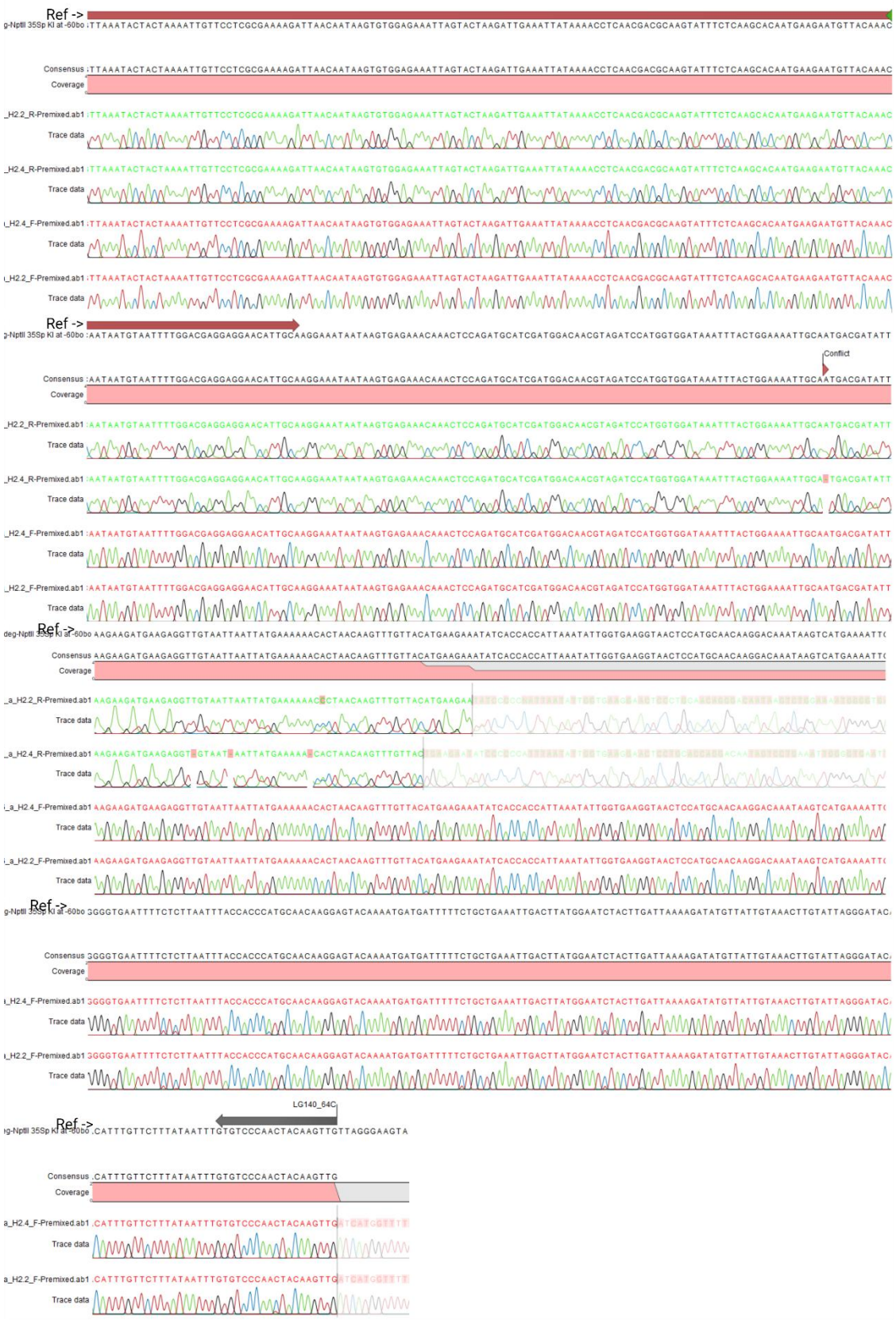
…continued



Appendix E.6. Sequencing coverage of the cloned 3.7 kb PCR amplicon spanning the left junction of the KI of sample 216_20 arising from vector LGJJ216 obtained with primers LG203 and LG283.). Gap in sequencing coverage spanning 800 bp (demonstrated as *//..//* on the sequence alignment) within the donor DNA, from mid-degron (orange bar annotation on reference ("ref") sequence) to the 5' region of the *NPTII* CDS (yellow bar annotation). Green sequences: forward sequencing reads from two clones. Base pair mutations observed in one out of the two clones analysed are considered as sequencing errors. Red sequences: reverse sequencing reads from one out of two clones sequenced. Sequencing from second clone failed so not shown here.

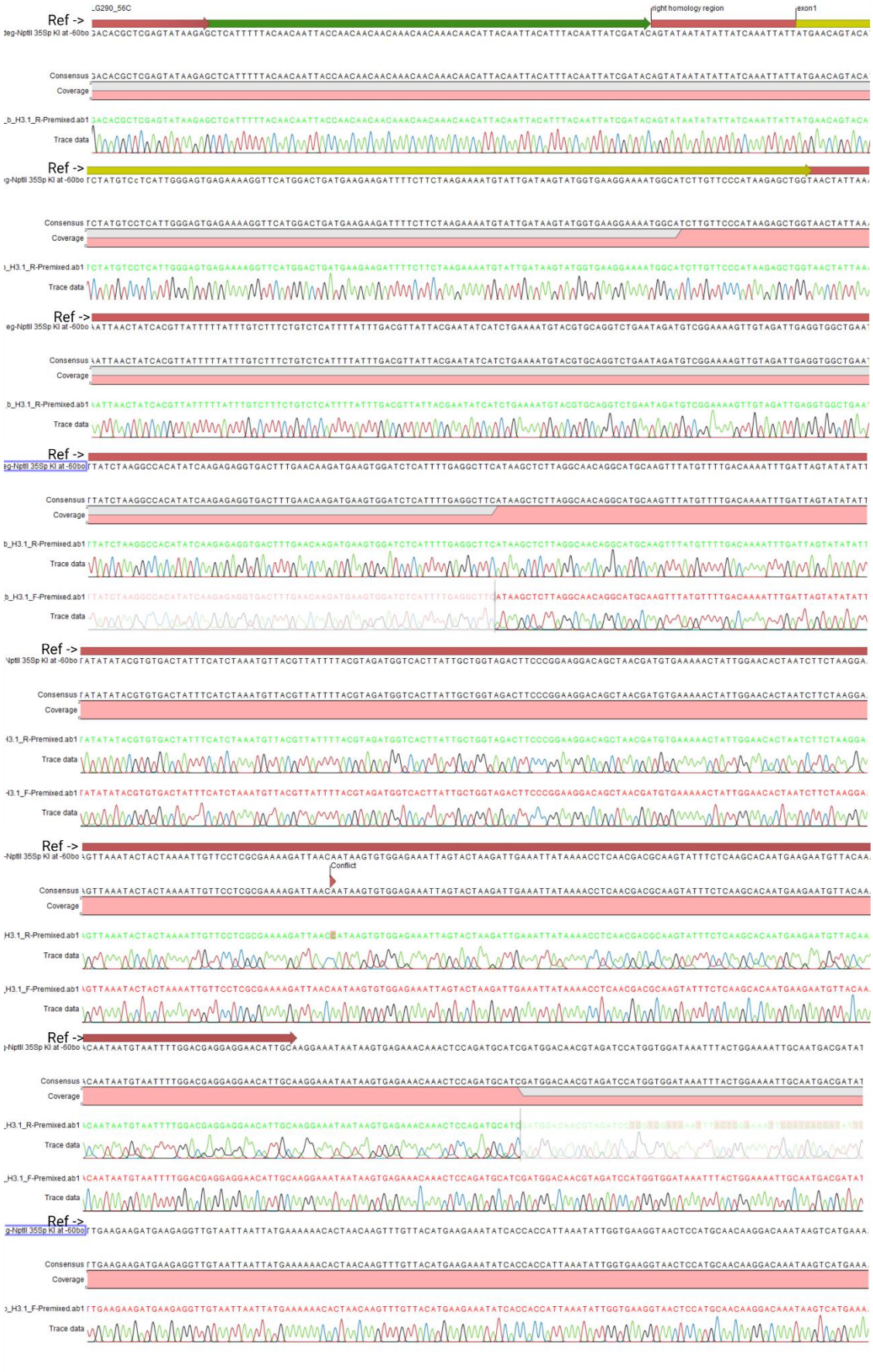

Ref ->
right homology region
ANT1 exon1
Consensus
Coverage
Trace data
Conflict

…continued



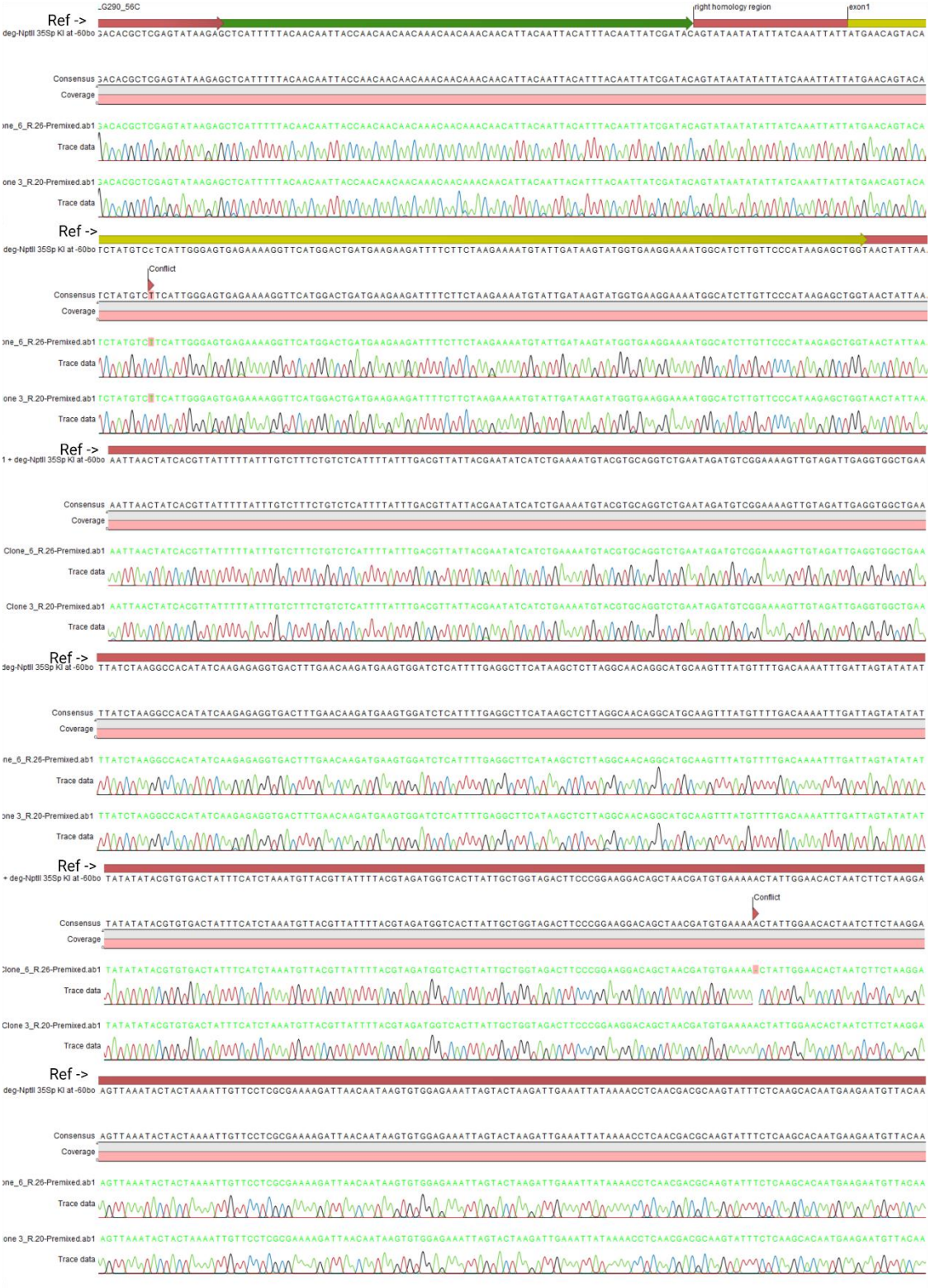
Appendix E.7. Sequencing coverage of the cloned 1.2 kb PCR amplicon spanning the right junction of the KI of sample 216_1 arising from vector LGJJ216 obtained with primers LG290 and LG140.). Green sequences: forward sequencing reads from two clones analysed.

Red sequences: reverse sequencing reads from two clones analysed. Base pair mutations observed in one out of the two clones analysed are considered as sequencing errors.

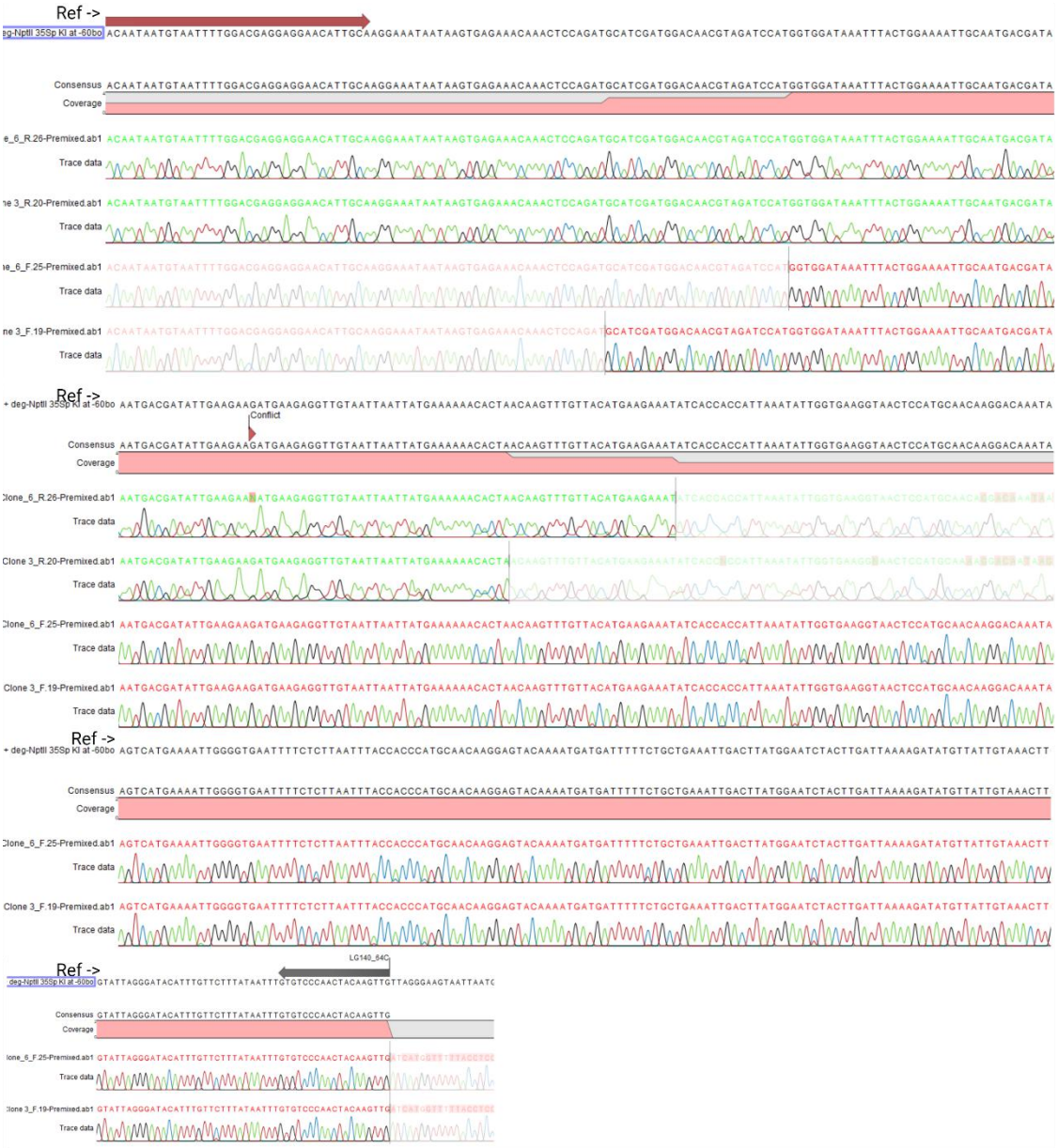

_G290_56C
right homology region
exon1
Ref ->
Consensus
Coverage
Trace data
Conflict

..continued

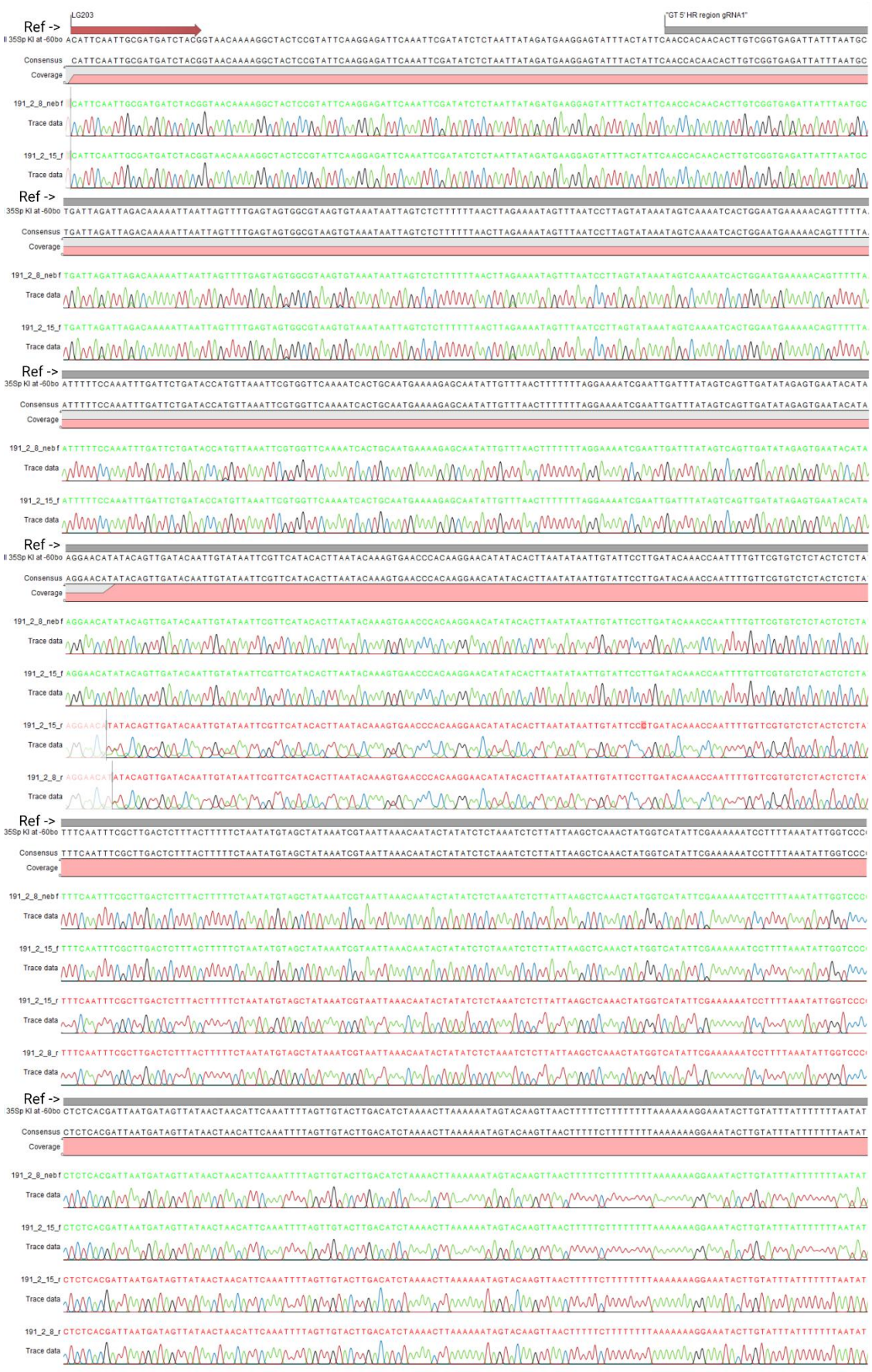
Appendix E.8. Sequencing coverage of the cloned 1.2 kb PCR amplicon spanning the right junction of the KI of sample 216_2 arising from vector LGJJ216 obtained with primers LG290 and LG140.). Green sequences: forward sequencing reads from one clone analysed. Red sequences: reverse sequencing reads from one clone analysed. Base pair mutation observed in the forward or reverse reads but not in its counterpart are considered as sequencing mutations.


Ref ->
right homology region
exon1
Consensus
Coverage
Trace data
Conflict
Clone_6_R.26-Premixed.ab1
Clone 3_R.20-Premixed.ab1

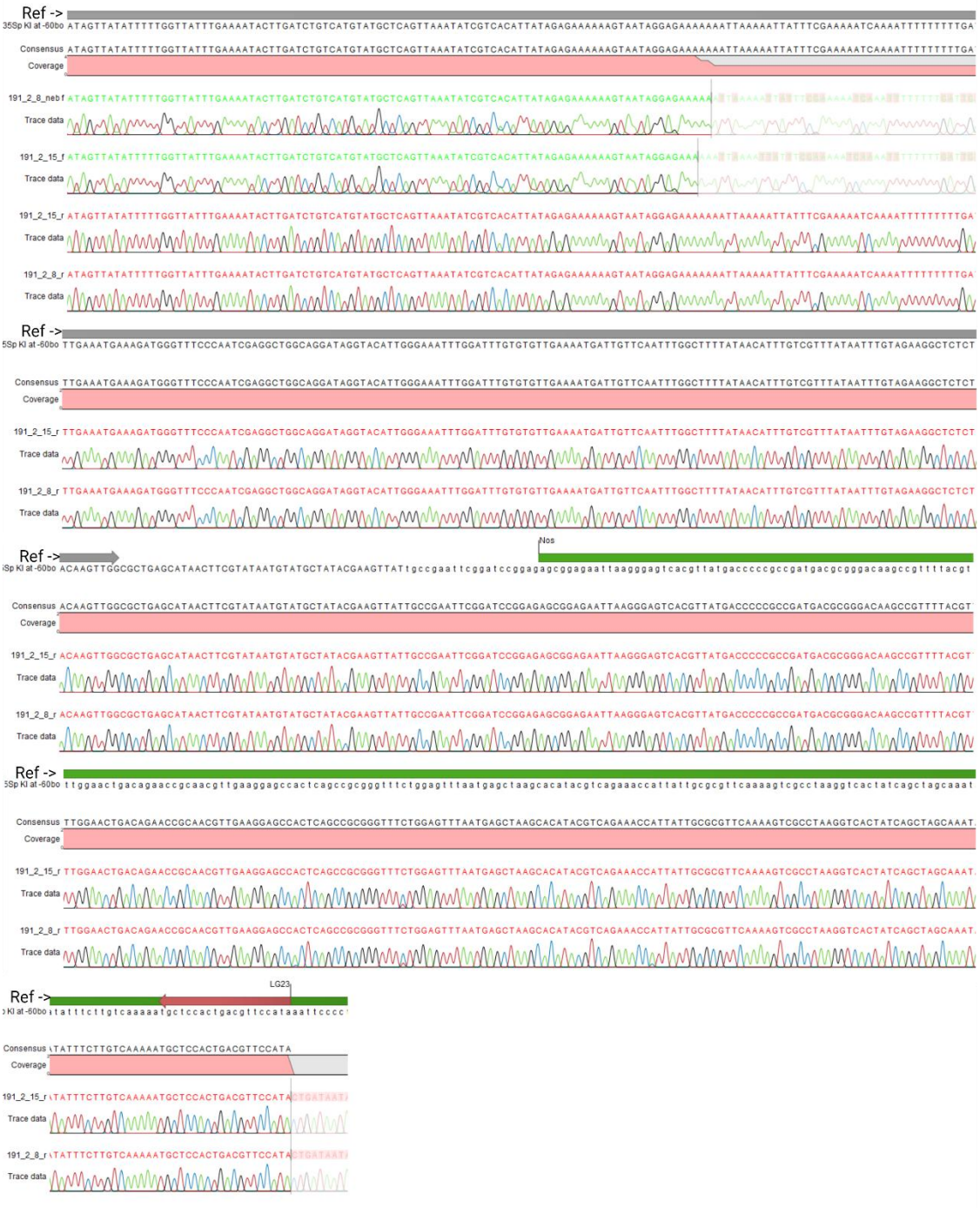
…continued



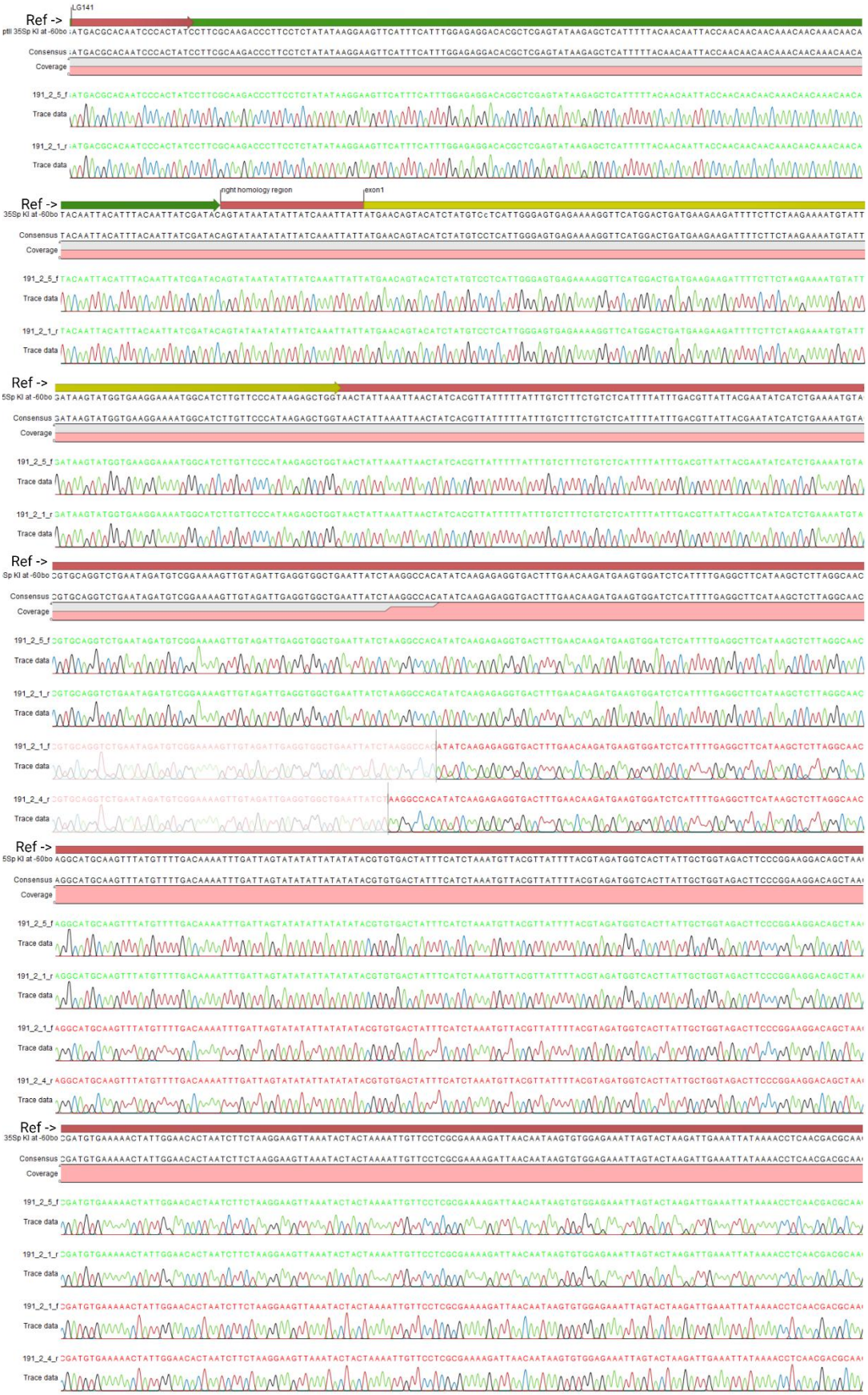
Appendix E.9. Sequencing coverage of the cloned 1.2 kb PCR amplicon spanning the right junction of the KI of sample 216_20 arising from vector LGJJ216 obtained with primers LG290 and LG140.). Green sequences: forward sequencing reads from two clones analysed. Red sequences: reverse sequencing reads from two clones analysed. Base pair mutations observed in one out of the two clones analysed are considered as sequencing errors.


LG203
"GT 5' HR region gRNA1"
Ref ->
Consensus
Coverage
Trace data
191_2_8_nebf
191_2_15_f
191_2_15_r
191_2_8_r

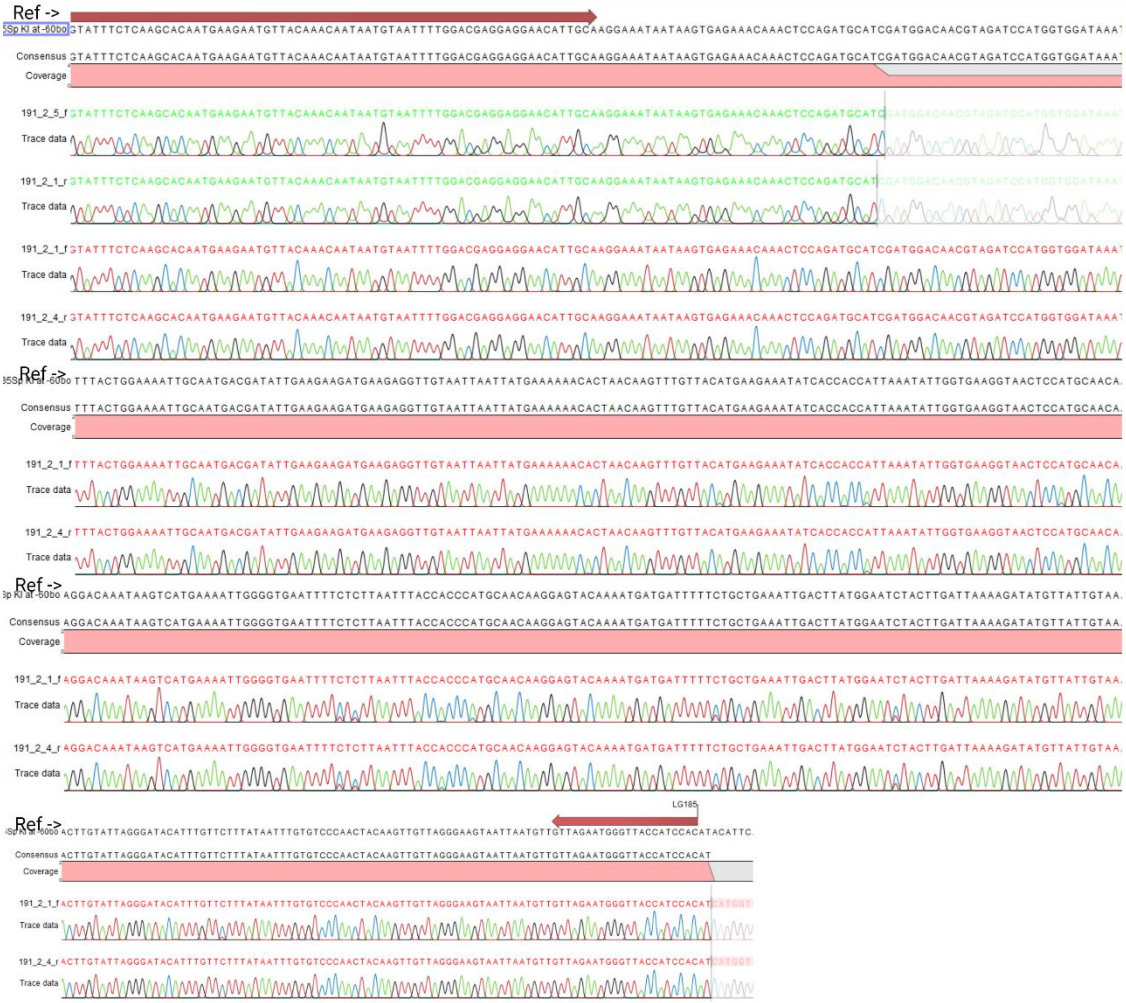
…continued



Appendix E.10. Sequencing coverage of the cloned 1.4 kb PCR amplicon spanning the left junction of the KI of sample 191_2 arising from vector LGJJ191 obtained with primers LG203 and LG23.). Green sequences: forward sequencing reads from two clones analysed. Red sequences: reverse sequencing reads from two clones analysed. Base pair mutations observed in one out of the two clones analysed are considered as sequencing errors. Surprisingly, the sequencing result does not reflect the larger than expected size of PCR A in Fig. 5.10.


LG141
Ref ->
ptII 35Sp KI at -60bo
Consensus
Coverage
191_2_5_f
Trace data
191_2_1_r
right homology region
exon1
35Sp KI at -60bo
5Sp KI at -60bo
Sp KI at -60bo
191_2_1_f
191_2_4_r

…continued



Appendix E.11. Sequencing coverage of the cloned 1.2 kb PCR amplicon spanning the right junction of the KI of sample 191_2 arising from vector LGJJ191 obtained with primers LG141 and LG185). Green sequences: forward sequencing reads from two clones analysed. Red sequences: reverse sequencing reads from two clones analysed.