CRISPR specificity in tomato

Ellen Slaman

Propositions

1. CRISPR-Cas mediated genome editing in tomato is specific.
   (this thesis)
2. In crops, the specificity of CRISPR-mediated genome editing is irrelevant for consumer safety.
   (this thesis)
3. Soon, the field of molecular biology will become a subfield of bioinformatics.
4. Epigenetics is an underexplored area of both animal and plant breeding.
5. Sustainability is an unsuitable criterium to use in the regulation of genome-edited crops.
6. Gender inequality is a self-perpetuating problem.

Propositions belonging to the thesis entitled:
“CRISPR specificity in tomato”

Ellen Slaman
Wageningen, 30 August 2023

CRISPR specificity in tomato

Ellen Slaman

Thesis Committee

Promotor

Prof. Dr. G. C. Angenent

Personal chair, Laboratory of Molecular Biology

Wageningen University & Research

Co-promotor

Dr. R. A. de Maagd

Senior Researcher, Business Unit Bioscience

Wageningen University & Research

Other members

Prof. Dr. J. van der Oost, Wageningen University & Research

Dr. R. Wilhelm, Julius-Kühn Institut, Quedlinburg, Germany

Dr. B. S. de Pater, Leiden Universiteit

Dr. J. M. Kooter, Vrije Universiteit Amsterdam

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences.

CRISPR specificity in tomato

Ellen Slaman

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr. A. P. J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 30 August 2023
at 4:00 p.m. in the Omnia Auditorium.

Ellen Slaman

CRISPR specificity in tomato

228 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2023)

With references, with summary in English and Dutch

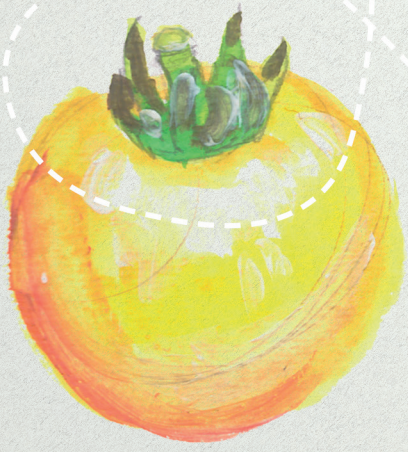
ISBN: 978-94-6447-720-7

DOI: https://doi.org/10.18174/631465

Table of Contents

Chapter 1

General Introduction



Genome editing

Bringing reverse genetics forward

In all fields of molecular biology research, the ability to modify DNA at will is highly sought-after. Scientists have always been interested in how the expression of specific genes affects the characteristics and appearance (phenotype) of their organism of interest. Forward and reverse genetics are the two main approaches to studying the relationship between genome (genotype) and phenotype. In forward genetics, the genetic basis of a particular phenotype is identified. In reverse genetics, one determines if and how a gene relates to a specific phenotype. In the current era of widely available genomics data and increasingly affordable next-generation sequencing techniques, reverse genetics plays an increasingly important role, as available data can often be used to make predictions about gene functions. However, these predictions are hypotheses and need to be tested experimentally. To do this, it is often necessary to interfere with or block the function of a gene, for example, by making a knockout mutant.

Several ways are frequently employed to disrupt genes, both non-specific, targeting the entire genome, or specific, targeting only a region of interest. Non-specific methods may rely on the abilities of transposons or Transfer-DNA (T-DNA) to integrate into the genome, where they can interrupt open reading frames (J. S. Jeon et al., 2000; Kawakami et al., 2017; Koncz et al., 1992; Martienssen, 1998). Additionally, ionizing radiation or chemicals may be used to induce random mutations (Jankowicz-Cieslak & Till, 2015; Mullins et al., 1994). Mutations cannot be generated at too high frequencies, as this causes toxicity or is even lethal. Moreover, such mutations are spread randomly over the genome. Therefore, the major downside of these non-specific methods is that large populations need to be screened to find individuals with the desired mutation. Additionally, especially in the case of transposons and chemical or radiation treatment, the desired mutation will be accompanied by many others that may affect the phenotype, obscuring the role of the gene of interest.

Break it down: DNA breaks as the starting point for specific genome editing

The more elegant, targeted methods for inducing mutations rely on generating double-strand breaks (DSBs) in the genome. Such breaks endanger any organism's survival, and therefore organisms in all branches of the tree of life have developed methods to repair this type of DNA damage. The two most frequently employed methods for DNA repair in eukaryotic cells are non-homologous end joining (NHEJ) and homology-directed repair (HDR). In NHEJ, the two DNA strands

are re-ligated without the use of a repair template. In HDR, the sister chromatid or another stretch of homologous DNA is used as a template to repair the damage (Chatterjee & Walker, 2017; Manova & Gruszka, 2015) (Figure 1). Especially NHEJ is not always perfect - although far more perfect than often portrayed - and during repair, mutations may arise (Ben-Tov et al., 2023; Bétermier et al., 2014). These mutations frequently take the form of small insertions or deletions (indels) that can disrupt open reading frames or other genetic elements. DNA repair through the HDR pathways is generally perfect, leaves no scars in the form of indels, and can be used to modify the targeted locus precisely or to introduce a piece of genetic material by providing a DNA template. Thus, DSB creation is the first step in producing mutations or specific modifications at a pre-defined genomic region of interest.

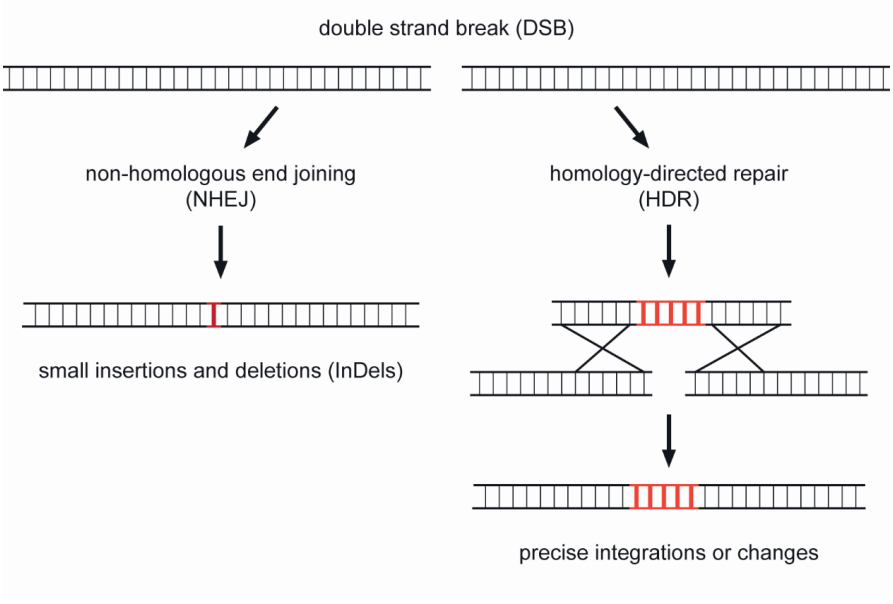


Figure 1 - Repair of double-strand breaks. The two most frequently employed repair pathways are non-homologous end joining (NHEJ, left) and homology-directed repair (HDR, right). NHEJ may lead to gene disruption by resulting in small insertions and deletions. HDR is generally perfect.

Methods for the targeted induction of DSBs

Meganucleases

Prokaryotes employ sequence-specific DNA cleavage to protect themselves against invading DNA, such as that of bacteriophages. Many prokaryotic genomes encode restriction endonucleases: enzymes that recognize and bind to a specific DNA sequence, which they subsequently cleave. These recognition sites are frequently between 4 and 8 bases and often palindromic (Loenen et al.,

2014). As the recognition sites for most restriction enzymes are so short, they occur frequently and are unsuitable for inducing DSBs in larger eukaryotic genomes. A six-basepair motif will occur by chance once every 4^6 (=4,096) bases. If we consider that the tomato genome encodes approximately 950 megabases (=950,000,000 bases), it is clear that using such a restriction enzyme would completely destroy the genome instead of specifically targeting a site of interest. Meganucleases are a potential solution to this problem. Members of this group of endonucleases are not only found in prokaryotes, but also in eukaryote nuclear, mitochondrial, and chloroplast genomes and recognize stretches of DNA more than 12 bp in size (Figure 2). In tomato, a recognition site from 15 bp onwards can be expected to occur only once in the genome. Several hundred of these rare-cutting endonucleases have been described (Chevalier & Stoddard, 2001). Unfortunately, their major advantage – specificity – is also their major disadvantage: only a limited number of sequences can be targeted with the existing suite of enzymes. To overcome this problem, the DNA recognition domains of meganucleases have been modified by varying their amino acid sequences or by fusing domains from different meganucleases (Arnould et al., 2007; Epinat et al., 2003; H. Gao et al., 2010). Although these endeavours have had some success, protein engineering remains complex, and the meganucleases remain unsuitable for targeting most genetic loci.

Zinc Finger Nucleases

A more versatile methodology for the site-specific induction of DSBs is based on zinc finger proteins. Zinc fingers are structural protein motifs that bind DNA and are frequently found in transcription factors - a class of proteins that regulate gene transcription by binding to specific DNA motifs. Every zinc finger recognizes a particular stretch of approximately three nucleotides, and multiple fingers can be combined to recognize longer stretches of DNA. This multiplexing gives the protein the desired sequence specificity, but zinc fingers cannot cleave DNA by themselves. Therefore, the protein containing the zinc finger array is fused to the catalytic domain of a nuclease, typically *FokI*. *FokI* needs to dimerize to be able to catalyze the formation of a DSB. To achieve dimerization, for every target, two zinc finger proteins are designed – one targeting the top strand of DNA, and the other the bottom strand (Figure 2) (Durai et al., 2005; Urnov et al., 2010). These zinc finger nucleases (ZFN) have been successfully used to achieve targeted mutations in a wide range of animal and plant species (Urnov et al., 2010), but they also have significant drawbacks. Although combining zinc fingers that each recognize different nucleotide triplets to produce a protein that can recognize a longer DNA sequence is relatively simple in theory, the practical execution is less so (Ramirez et al., 2008). The binding specificities of individual zinc fingers may overlap, and adjacent zinc fingers may influence each other's binding specificity, causing the failure to bind (Ramirez et al., 2008). These

problems can be partially overcome by utilizing zinc fingers that bind "GNN" motifs, but this, in turn, severely limits the number of genomic loci that can be targeted.

Transcription-activator-like effector nucleases

Transcription activator-like effectors (TALEs) are proteins encoded by the plant pathogen *Xanthomonas*. This bacterium injects TALEs into plant cells, where they bind promoters of specific genes and induce transcription, hijacking the cellular machinery of the plant to the benefit of the bacterium (Boch & Bonas, 2010). Similar to zinc fingers recognizing three bases, each TALE repeat recognizes one specific base. These repeats can be combined in a modular way to produce proteins that bind the desired genomic locus. Like ZFN, TALEs can be fused to a *FokI* endonuclease domain, forming TALE nucleases (TALENs). A pair of TALENs can form the *FokI* dimer that causes the DSB (Figure 2). A significant advantage of TALENs over ZFNs is that the binding specificity of a TALE repeat is not affected by a neighbouring repeat, making TALEN design more straightforward than ZFN design. Like ZFNs, TALENs have been applied for targeted genome editing in the animal and plant kingdoms (Joung & Sander, 2013). Yet, for every target, a new TALEN must be designed and constructed, which can be challenging due to their repetitive nature.

CRISPR-Cas systems

With the advent of ZFNs and TALENs, introducing DSBs at desired positions in genomes has certainly become feasible but not yet routine. This routine came about with the introduction of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) – Cas (CRISPR-associated protein) systems. In these systems, an endonuclease – for example, Cas9 – is guided to a specific position in the genome by a guide RNA or gRNA. Upon gRNA-mediated binding of the nuclease to the target site, the nuclease will induce a DSB, which can then be utilized for genome editing purposes as described above (Figure 2) (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013). The major advantage of this system over ZFN and TALEN technology is the gRNA: it is far simpler to synthesize or express a new gRNA for a new target than to construct a new protein for every target. In the following paragraph, the discovery of CRISPR systems, the adaptation of these systems for genome editing applications of unparalleled versatility and simplicity, and the mode of action of these systems is reviewed more in-depth.

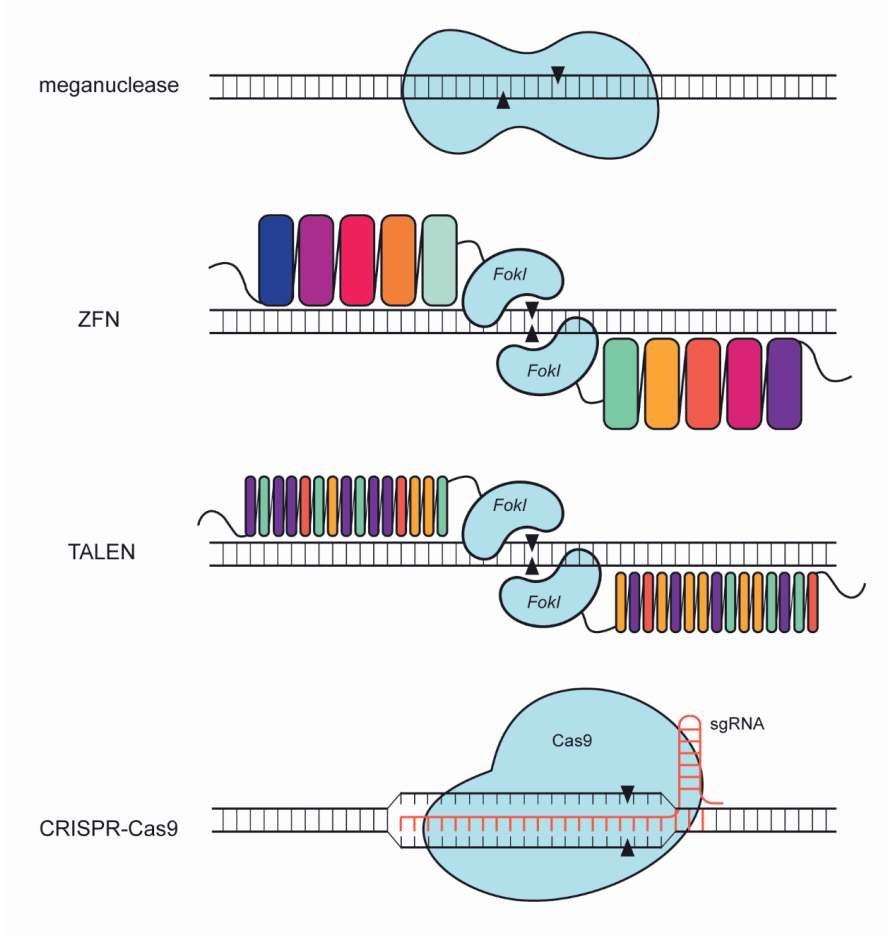


Figure 2 – Schematic overview of several methods for creating DSBs in genomic DNA. DSBs are indicated with black triangles. Meganucleases are a particular class of restriction endonucleases that recognize longer (>12 bp) stretches of DNA. ZFNs consist of several connected zinc fingers, each recognizing three bp, fused to a FokI endonuclease domain. Dimerization results in the activation of this domain and the subsequent formation of a DSB. TALENs work similarly, but specificity is conferred by TALE repeats – each TALE recognizes one base. In the CRISPR-Cas9 system, the Cas9-gRNA complex binds a specific 20 bp stretch of DNA through Watson and Crick base-pairing of the guide RNA with the target DNA strand. The Cas9 endonuclease then cleaves the DNA.

Discovery and development of CRISPR technology

CRISPRs – more than just an "unusual structure"

The first CRISPRs – clustered regularly interspaced short palindromic repeats – were identified long before the name "CRISPR" was coined. In 1987, a Japanese group studying the *iap* gene in *Escherichia coli* reported "an unusual structure" in the 3' flanking region of the gene (Ishino et al.,

1987). They identified five highly homologous, almost perfectly repeated sequences of 29 nucleotides each, interspaced with variable sequences of 32 nucleotides in length (Figure 3). At the time, no homologous sequences in other prokaryotes were found, and the function of this structure remained unknown.

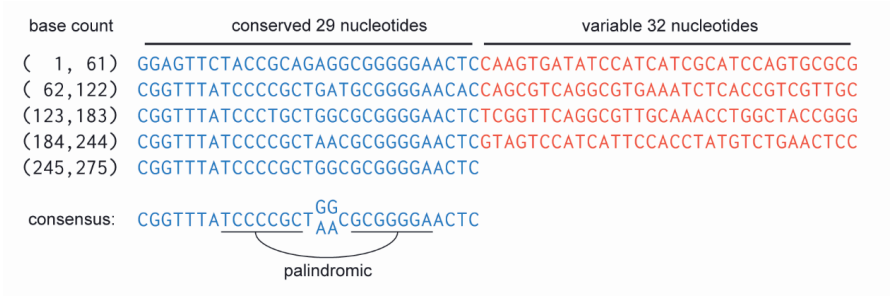


Figure 3 – Overview of the first discovered CRISPR array in E. coli, in the 3' end flanking region of the iap gene. Five highly conserved repeats of a 29 bp, partially palindromic sequence (blue) were found to be regularly interspaced by variable sequences of 32 bp (orange). Adapted from Ishino et al., 1987.

Over time, similar structures were identified in other bacteria as well as in archaea, but not until 2000 were these peculiar repeat structures recognized to be *"an actual family of repeats; their own thing, conserved over all prokaryotes"* (Mojica et al., 2000). As these structures were so conserved and widespread, it was assumed that they must have some biological function, but what exactly this function was, was unknown – as was the origin of these structures. Of course, hypotheses were posed: the size and palindromic nature of the repeats were reminiscent of recognition sites for DNA-binding proteins. The position of some of these repeat structures on prokaryote genomes indicated that they might play a role in replicon partitioning (Mojica et al., 1995).

The first milestone in the study of these structures was achieved in 2002, when after corresponding with Mojica et al., Jansen et al. gave these structures the name they are now known by: CRISPRs (Jansen et al., 2002). Additionally, they discovered that these clusters of repeats were always accompanied by at least one, and often more, CRISPR-associated genes (abbreviated to *cas* genes). In this study, they identified four such genes and named them *cas1-4*. They noted that the *cas3* homologs contained nucleic acid helicase motifs, indicating that the Cas3 protein is capable of unwinding the DNA helix. Additionally, the *cas4* homologues possessed exonuclease activity, meaning that the Cas4 protein can cleave nucleotide bonds. Finally, the *cas1* homologues had a high (basic) isoelectric point, which is a common characteristic of DNA-binding proteins. These clues indicated that the *cas* gene ensemble had the potential for DNA manipulation and might play a role

in the formation of these peculiar CRISPR arrays – although they did not immediately provide insight into the function of the arrays themselves.

The repetitive, palindromic sequences were the most remarkable part of the CRISPR array. The spacers – fixed numbers of nucleotides separating the repeats – were of unknown function and origin. Mojica et al. were among the first to report that these spacers are not unique or random but derived from the DNA sequences of "foreign genetic elements", such as phages and conjugative plasmids (Mojica et al., 2005). This finding was corroborated by a study from Pourcel et al., which showed that the bacterium *Yersinia pestis* had acquired spacers that originated from a bacteriophage, and Bolotin et al., which showed that many spacers from the bacterium *Streptococcus thermophilus* were either derived from phages or plasmids (Bolotin et al., 2005; Pourcel et al., 2005). These studies raised new hypotheses about the function of CRISPR arrays and their associated genes. Mojica et al. observed that the presence of a spacer derived from a specific phage seemed to inhibit the invasion of the bacterium by that phage. Likewise, Bolotin et al. found a negative correlation between the number of spacers a particular strain of *S. thermophilus* had and the sensitivity of that strain to phage infection. Both findings indicated that specific spacers offered resistance against the infection of the bacterium by bacteriophages, or by invasive genetic elements in general, perhaps in an RNAi-like fashion.

CRISPR systems are adaptive prokaryotic immune systems

The hypothesis that CRISPR systems act as immune systems, protecting against invasion by mobile genetic elements, was confirmed not much later. In 2007, Barrangou et al. proved the one-to-one correlation between the presence of specific spacers and phage resistance. They showed that *S. thermophilus* could gain resistance against a strain of virulent bacteriophages by incorporating a spacer originating from this bacteriophage in its CRISPR array. Additionally, they proved that phage resistance was either lost or gained by deleting spacers from or adding spacers to the CRISPR locus (Barrangou et al., 2007).

Exactly how new spacers were acquired or immunity was achieved was unknown. As the products of *cas* genes close to CRISPR arrays were predicted to be able to manipulate DNA, they were prime candidates. At first, only four *cas* families were identified (Jansen et al., 2002), but it quickly became apparent that CRISPR systems were much more complex and versatile. This complexity was first studied *in silico* by Haft et al. and Makarova et al. in 2005 and 2006 (Haft et al., 2005; Makarova et al., 2006).

Barrangou et al. provided the first evidence that the *cas* genes were indeed needed for adaptive immunity as well (Barrangou et al., 2007). They knocked out two of the *cas* genes in *S. thermophilus*: *cas5* and *cas7*. The loss of Cas5, a hypothesized nuclease, resensitized the bacterium to phage infection. The loss of Cas7 did not affect existing phage resistance but rendered the bacterium unable to gain resistance against a new phage, indicating that this gene was involved in the acquisition of new spacers.

It was previously shown that CRISPR arrays were transcribed (Lillestøl et al., 2006; T. H. Tang et al., 2005; T.-H. Tang et al., 2002). Brouns et al. showed that the long transcript from the CRISPR array was processed into shorter CRISPR RNAs (crRNAs), each consisting of both a repeat and a spacer, by a complex of several Cas-proteins they named "Cascade". This processing was necessary to achieve phage resistance (Brouns et al., 2008). Additionally, an *E. coli* strain without a CRISPR system could be rendered resistant to phage Lambda infection by integrating all *cas* genes encoding Cascade, an additional *cas* gene, and an artificial CRISPR array into the genome. Moreover, spacers targeting the coding strand as well as the non-coding strand of the phage provided resistance against viral infection. This was a strong indication that CRISPR systems target viral DNA and not RNA. Previously, it was hypothesized that CRISPR systems might function in an RNAi-like manner (Makarova et al., 2006).

Indeed, several groups showed that CRISPR-mediated immunity was achieved by cleaving DNA rather than RNA (Garneau et al., 2010; Marraffini & Sontheimer, 2008). Cleavage occurs at the position of the protospacer on the plasmid or bacteriophage DNA, indicating that the crRNA guides the nuclease to this position. In a demonstration of the wide variety of CRISPR-Cas systems, another paper published around the same time reported that a CRISPR system in *Pyrococcus furiosus* did target RNA in an RNAi-like manner (Hale et al., 2009). Later, more such systems were found (Burmistrz et al., 2020).

Cleavage of DNA at the site of the protospacer raises the matter of self and non-self differentiation. How can the *cas* nuclease distinguish the protospacer from the incorporated spacer? Research on *S. thermophilus* CRISPR arrays showed that all protospacers are followed by an "NNAGAAW" motif and that its presence is required for CRISPR interference (Deveau et al., 2008). Such a motif was called a protospacer-adjacent motif, or PAM. A more comprehensive study encompassing multiple types of CRISPR systems from different species of prokaryotes was later performed by Mojica et al, who identified a wide range of motifs (Mojica et al., 2009).

A large variety of CRISPR systems exist

We can now divide the adaptive immunity effectuated by CRISPR systems into three stages: (1) adaptation or spacer acquisition, (2) processing of the primary transcript of the CRISPR array into individual, mature crRNAs and (3) CRISPR interference – either through cleavage and degradation of invading DNA, or through RNA interference. All CRISPR systems can be divided into two classes based on their effector module: Class I CRISPR systems have multi-subunit effectors, whereas Class II CRISPR systems rely on large multidomain proteins for interference (Makarova et al., 2015). Class I CRISPR systems are much more abundant and occur in both bacteria and archaea. Class II systems constitute only about 10% of known CRISPR systems and only occur in bacteria. These two classes are further divided into different types and subtypes as reviewed in (Koonin & Makarova, 2019; Makarova et al., 2020).

From adaptive bacterial immunity to precision genome editing

CRISPR systems were initially seen as promising systems to protect industrial bacterial strains against devastating phage infections (Kamerbeek et al., 1997; Mills et al., 2010; Mokrousov et al., 2007; Vergnaud et al., 2007). In 2012 however, CRISPR left the confines of the prokaryotic world and emerged as what might arguably be the most influential scientific discovery of the 21st century thus far: this bacterial immune system could be used as a tool for precision genome editing – not only in prokaryotes, but also in eukaryotes.

A precision gene editing system needs three main features: it should be able to modify DNA, it should be able to do this at a specific, desired location, and it should be as easy as possible to direct this system to different locations. CRISPR-based adaptive immune systems meet all these criteria: a large number of Cas-effectors can cleave DNA, specificity to a target site is conferred by the crRNA, and almost any position can be targeted as long as there is a PAM present. Jinek et al. were the first to prove that a Class II CRISPR-Cas9 system from *Streptococcus pyogenes* could be used as a guided endonuclease to specifically cleave target DNA *in vitro* (Jinek et al., 2012). An advantage of this Class II, Type 2 system is that only three components are necessary for this cleavage reaction: the Cas9 effector nuclease, the crRNA encoding the 20 bp spacer and repeat sequence, and a transactivating RNA or tracrRNA. In the natural *S. pyogenes* system, the tracrRNA hybridizes with the crRNA during the crRNA maturation process to create a specific RNA structure necessary for the formation of the Cas9-RNA complex and subsequent DNA binding and cleavage (Figure 4) (Deltcheva et al., 2011; Jinek et al., 2012). Already seeing the potential for genome editing, Jinek et al. simplified the system

even further by engineering a crRNA-tracrRNA fusion, called a single guide RNA or sgRNA, bringing the number of necessary components down to two.

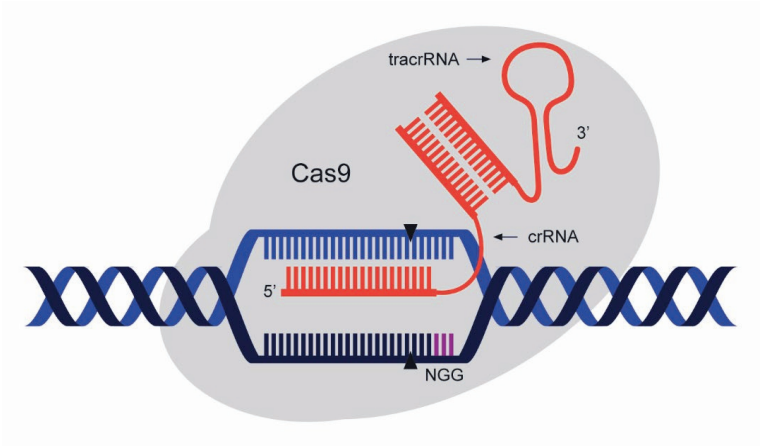


Figure 4 - The CRISPR-Cas9 system for genome editing. The Cas9 endonuclease is guided to the target DNA by the hybridization of a crRNA together with a tracrRNA. This Cas9-RNA complex binds the target site using Watson-Crick base pairing. An "NGG" PAM at the 3' end of the target site is necessary for cleavage. Spacers are 20 bp in length.

Shortly after, by adding a nuclear localization signal, the CRISPR-Cas9 system was shown to cleave endogenous targets in human cells, where it could induce both NHEJ and HDR (Cong et al., 2013; Mali et al., 2013). From this point onward, research on and the application of CRISPR systems for genome editing exploded. Soon, algorithms were developed to predict gRNAs, their efficiencies for genome editing, and their specificities (Bae et al., 2014; Concordet & Haeussler, 2018; Lei et al., 2014). The genome editing system was applied to the plant, animal, and fungal kingdoms and was found to function efficiently in all (Belhaj et al., 2015; Doudna & Charpentier, 2014). Nowadays, multiple clinical trials that use CRISPR-mediated genome editing to combat diseases such as sickle cell disease, diabetes and cancer are in progress (NIH, 2023).

Cas9 Structure and Mechanism

The main actor in CRISPR-Cas9 mediated genome editing is the large, 160 kDa Cas9 enzyme (for reference, the average human protein is about 38 kDa). The main features of this protein are the two endonuclease domains with which it induces DSBs: an HNH-like domain, which cuts the DNA strand complementary to the gRNA (target strand or TS), and a RuvC-like domain, which cuts the DNA strand opposite the complementary strand (non-target strand or NTS) (Figure 5a,b,c) (Jinek et al., 2012).

In its unbound apo state, Cas9 is catalytically inactive and assumes a bilobed structure (Jinek et al., 2014). These two lobes have been designated the recognition or REC lobe and the nuclease or NUC lobe (Figure 5a,b). The recognition lobe contains three alpha-helical domains (REC1-3), which are involved in binding the target DNA. The nuclease lobe contains both the HNH and tripartite RuvC domain, as well as a C-terminal domain (CTD), which is involved in PAM recognition. In the apo state, the CTD is unordered, meaning that PAM recognition cannot take place in the absence of a bound gRNA (Figure 5b) (Sternberg et al., 2014). Upon binding of a gRNA, a significant conformational change occurs in the REC lobe – most notably in the REC3 domain, which moves much closer to the HNH domain (Jiang et al., 2015; Jinek et al., 2014). Additionally, the CTD becomes ordered, which positions two crucial arginines (Arg1333 and Arg1335) so that they can bind a 5'-NGG-3' PAM. Furthermore, the 10 nucleotides of the gRNA that will bind directly upstream of the PAM are pre-ordered, positioning them in such a way as to allow interrogating of the flanking DNA for potential gRNA complementarity. Thus, gRNA binding to the Cas9 nuclease primes the enzyme for PAM and subsequent target site recognition.

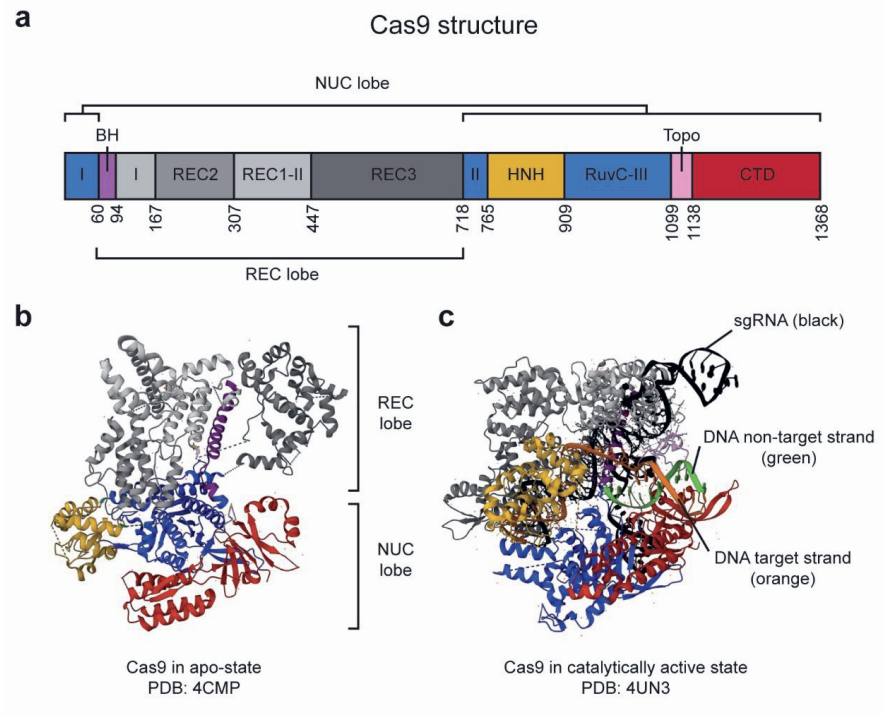


Figure 5 - Structure of Cas9. (a) Schematic overview of the different protein domains. Domains in the 3D structures as shown in (b) and (c) are coloured in agreement with (a). (b) 3D structure of Cas9 in its unbound apo state. The bilobed structure of the protein is well-recognizable. (c) 3D structure of Cas9 upon binding of both sgRNA and the dsDNA target. PDB: Protein Databank accession number.

The Cas9-gRNA complex "searches" for targets through random collisions with the DNA. Recognition of a PAM site is the first necessary step towards target binding. If a PAM is not encountered, the complex will rapidly dissociate from the DNA. Consequently, even perfect target sites will not be bound in the absence of a PAM (Sternberg et al., 2014). Once a PAM has been bound, the complex will probe the adjacent DNA for complementarity with the guide RNA. This probing requires the unwinding of the DNA and the formation of an R-loop structure. Cas9 does not possess a helicase domain that can unwind DNA in an ATP-dependent manner. Instead, a sharp kink turn is induced in the target strand immediately upstream of the PAM, resulting in local DNA melting and the possibility of the guide RNA invading the DNA (Figure 5c) (Anders et al., 2014; Sternberg et al., 2014). Upon sufficient hybridizing of the guide RNA with the target DNA to form a stable R-loop, generally requiring ≥ 10 nucleotide interactions (van Aelst et al., 2019), the nuclease domains are activated and will cut through both DNA strands (Sternberg et al., 2015). Cas9 will remain bound to both ends of the cleaved DNA until removed by cellular mechanisms and therefore functions as a single-turnover enzyme (Sternberg et al., 2014).

In conclusion, the CRISPR genome editing system has become an invaluable tool for molecular biologists in all fields of study. It is widely and routinely applied in basic research to study the function of genes and other genetic elements and has great potential for therapeutic applications. In 2020, Emmanuelle Charpentier and Jennifer Doudna received the Nobel Prize in Chemistry for the discovery that the CRISPR-Cas9 system could be applied as a programmable nuclease. This final breakthrough of CRISPR-Cas mediated genome editing could not have been realized without many researchers slowly adding to the knowledge about CRISPR systems over time – all starting from the human curiosity about a particularly organized stretch of DNA (Lander, 2016).

Expanding the CRISPR toolbox

New and engineered Cas nucleases for a broader targeting range

Although CRISPR-Cas9 technology brought the possibility of facile, routine genome editing, limitations still exist. As Cas9 needs the presence of an "NGG" PAM, only genomic loci preceding this motif can be targeted. To further widen the scope of CRISPR-Cas mediated genome editing, two methodologies were applied: the Cas9 enzyme was engineered to recognize other or relaxed PAM sequences, and the existing natural variation of CRISPR systems was mined for other Class II effector nucleases that might be useful for genome editing. Overall, these approaches were very successful

and using a wide variety of engineered or other Cas9 versions, it is now possible to perform genome editing practically unconstrained by PAM requirements (Ren et al., 2021; Walton et al., 2020).

Genome editing using Cas12a

The first CRISPR-Cas nuclease employed for genome editing after Cas9 was Cas12a (previously called Cpf1), which has several characteristics that set it apart from Cas9 for genome editing purposes (Zetsche et al., 2015). Cas12a requires a "TTTV" PAM at the 5' end of the protospacer, expanding the range of genetic loci that can be targeted. Additionally, Cas12a induces a staggered DSB through the two DNA strands, resulting in the formation of a 4-5 bp 5' overhang, whereas Cas9 generally induces a blunt break, leaving no sticky ends. Also, Cas12a only needs a single crRNA with no need for a tracrRNA, resulting in shorter crRNA expression systems (Figure 6). Moreover, Cas12a contains endoribonuclease activity in addition to endonuclease activity and is able to process its own CRISPR arrays into single crRNAs (Zetsche et al., 2016). Like Cas9, Cas12a has also been engineered to recognize relaxed and different PAMs (L. Gao et al., 2017).

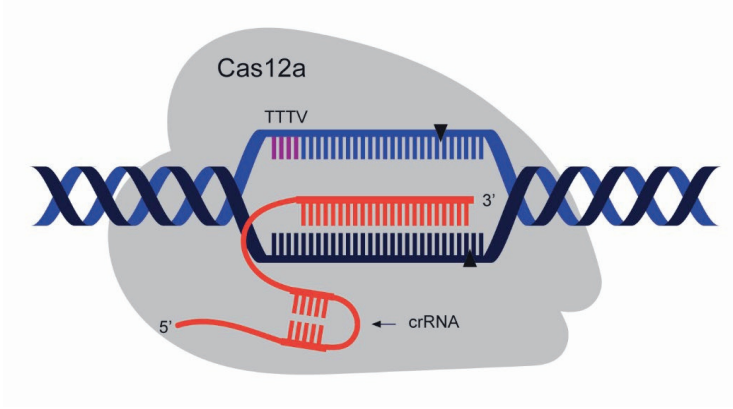


Figure 6 - The CRISPR-Cas12a system for genome editing. Unlike Cas9, Cas12a induces a staggered DSB with a 4 or 5 bp overhang. A "TTTV" PAM at the 5' end of the target site is necessary for cleavage. Spacers generally have a length of 23 bp.

Another interesting difference between Cas9 and Cas12a is that Cas12a has both *cis* and *trans*-acting DNase activity. After activation of the nuclease domain by the formation of the ternary Cas12a-RNA-DNA complex at a target site, Cas12a will first cut the bound dsDNA *in cis*. The PAM-distal end of the DNA is then released, after which ssDNA can be degraded *in trans* (J. S. Chen et al., 2018). The *trans* cleavage of non-target ssDNA is non-specific and no PAM is required. Cas12a will not cut ssDNA in *trans* without prior activation at a target site, and the *trans*-acting DNase activity can be reset by providing the protein with a fresh crRNA (Stella et al., 2018).

The Cas12a enzyme is of a similar size as Cas9 and also assumes a bilobed structure (Figure 7a,b). However, Cas9 and Cas12a have no common evolutionary ancestry, and any similarities can likely be attributed to convergent evolution (Koonin et al., 2017; Shmakov et al., 2017). Many of the Cas12a protein domains are named after their Cas9 counterparts, and the two Cas12a lobes are also called the recognition (REC) lobe and the nuclease (NUC) lobe. Unlike Cas9, which possesses both a RuvC and an HLH-like endonuclease domain, Cas12a only has a single RuvC-like endonuclease domain, which cuts both DNA strands sequentially and is present in the NUC lobe (Swarts et al., 2017). Other domains in the NUC lobe are the wedge (WED) domain, the PAM-interacting (PI) domain, the Nuclease (NUC) domain, and the bridge helix (BH) (Figure 7 a,b) (Dong et al., 2016; Swarts et al., 2017; Yamano et al., 2016). The WED domain contains the endoribonuclease activity that provides Cas12a with the ability to process its own crRNA arrays (Dong et al., 2016; Swarts et al., 2017). It recognizes the pre-crRNA structure together with the RuvC and BH domains. Once a crRNA is bound, the WED and PI domains are important for subsequent PAM recognition (Swarts et al., 2017; Yamano et al., 2016). The NUC domain is not involved in cutting the DNA, but plays an important role in the coordination of the DNA into the catalytic site. The REC lobe consists of two domains, REC1 and REC2, which are both important for coordinating the crRNA-DNA heteroduplex.

As for Cas9, binding of the crRNA induces a significant conformational change in Cas12a, pre-ordering the PI and WED domains for PAM recognition and changing the enzyme from an inactive to a scanning state (Swarts et al., 2017). Once a PAM is bound, the dsDNA will be melted, allowing strand invasion of the crRNA and R-loop formation. As for Cas9, R-loop formation progresses from PAM-proximal to PAM-distal. Upon the formation of a stable R-loop, which will generally require 17 or more crRNA-TS base pairs (Y. Jeon et al., 2018; Strohkendl et al., 2018), the final conformational change is induced. This change activates the RuvC domain, which will cut the non-target strand, followed by the target strand (Y. Jeon et al., 2018; Stella et al., 2018). Unlike Cas9, PAM-binding is not strictly necessary for the activation of the endonuclease domain. However, as PAM binding is required for dsDNA melting and subsequent R-loop formation, target sites cannot be cleaved in the absence of a PAM (Y. Jeon et al., 2018; Swarts & Jinek, 2019).

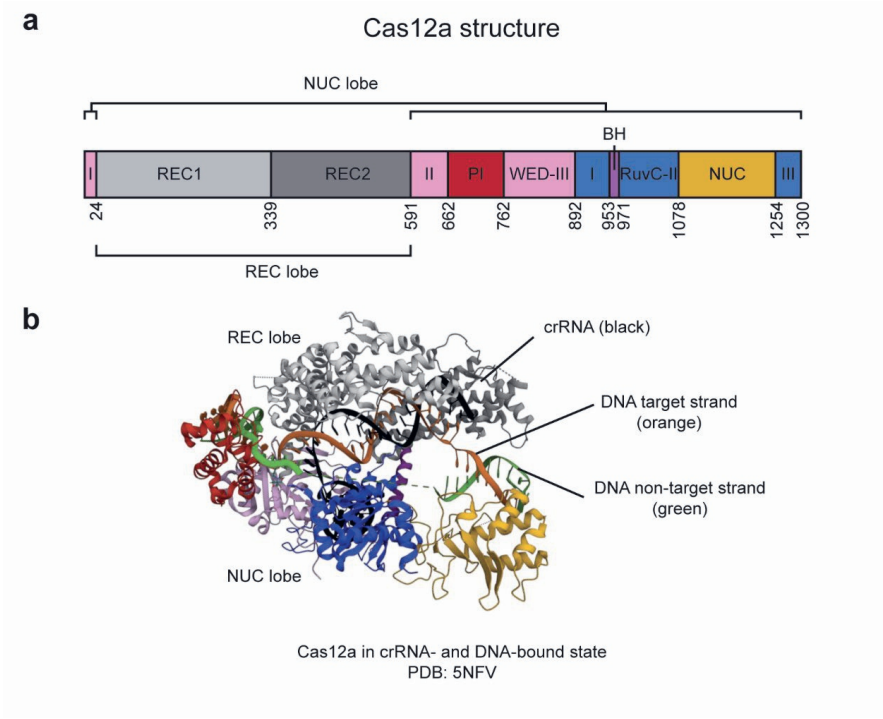


Figure 7 – Structure of Cas12a. (a) Schematic overview of the different protein domains. Domains in the 3D structure shown in (b) are coloured in agreement with (a). (b) 3D structure of Cas12a in complex with both crRNA and dsDNA target. PDB: Protein Databank accession number.

Transcriptional regulation and epigenetic modification

The ability of the Cas9-gRNA complex to recognize and bind specific sites in the genome offered the possibility of broadening the applications of Cas9. Inactivation of both catalytic domains of Cas9 results in a d(ead)Cas9. This dCas9 can be used to bring other proteins or protein domains – that either do not have a built-in targeting system or a complicated, difficult-to-engineer one – to specific positions in the genome. By fusing dCas9 to activation or repression domains, and so effectively turning the inactive nuclease into a transcription factor, gene expression can be regulated (Gilbert et al., 2014; Konermann et al., 2014). Additionally, dCas9 can be fused to DNA (de)methylases or histone modifiers to make local changes to the epigenome (X. S. Liu et al., 2016; Morita et al., 2016; Pulecio et al., 2017; X. Xu et al., 2016).

Base editing

Specific changes in genes, such as the substitution of one or a few basepairs for another, can be achieved using HDR. However, the HDR pathway is often superseded by the error-prone NHEJ

pathway. Especially in plants, HDR happens at such low frequencies that it cannot be used for routine applications. To overcome these difficulties, certain alternatives have been engineered, such as base-editing technology. The catalytically inactive dCas9 or a nickase version of Cas9 (nCas9), able to only cleave one strand of DNA due to the inactivation of one of its endonuclease domains, can be fused to cytidine deaminase domains. As the name implies, these domains can deaminase C nucleotides, effectively turning them into thymines (T). This allows for changing C's to T's (Nishida et al., 2016) and later, A's to G's using adenine deaminases (Gaudelli et al., 2017). Not yet all 12 base conversions can be achieved using these base editors, and there is the possibility of the "bystander" effect: if a targeted C is close to other C's, these other C's might also be converted (Rees & Liu, 2018).

Prime editing

A more versatile, though somewhat more complex Cas9-based method able to induce targeted basepair substitutions and small deletions or insertions is prime editing (Anzalone et al., 2019). This system uses a nCas9 which is fused to a reverse transcriptase and guided by an elongated gRNA – a so-called prime editing gRNA or pegRNA. This pegRNA not only guides nCas9 to the target position on the genome, but also functions as an RNA template with the desired changes for the reverse transcriptase fused to Cas9. Once nCas9 binds its target site, it cuts the non-target strand only. The DNA strand is released and hybridizes to a complementary stretch of RNA on the template part of the pegRNA. The reverse transcriptase uses this as a primer to further reverse transcribe the template RNA into DNA. As a result, the targeted region now has two non-complementary DNA strands: the original target strand, and the prime-edited non-target strand containing the desired edits. At this point, the DNA repair machinery will re-establish DNA complementarity by restoring the original, unedited DNA on both strands or by incorporating the prime-edited DNA on the other DNA strand. As the repair machinery prefers to use the intact, original DNA strand to restore complementarity, the first outcome is more likely. To push the equilibrium toward the second option, another normal gRNA can be included to induce an extra nick in the other DNA strand. Although this strategy is fairly efficient in mammalian cells, only low frequencies were thus far achieved in plants, and this method is therefore not yet routinely applied (Q. Lin et al., 2020; R. Xu et al., 2020; W. Xu et al., 2020).

CRISPR specificity

The occurrence of off-target mutations

An important aspect in the successful application of CRISPR-mediated genome editing – and genome editing in general - is specificity. Concerns exist about so-called "off-target" mutations induced by CRISPR-Cas, especially when used for therapeutic purposes. Off-target mutations occur at positions other than the intended target site, frequently at positions that bear a resemblance to the target site. They are the result of the ability of Cas9 to bind and cleave DNA despite one or more mismatches between gRNA and the DNA. Early studies on Cas9 specificity, focused on mammalian cells, showed that Cas9 could cleave genomic DNA at sites with as many as 6 mismatches to the target site, albeit at low frequencies (Fu et al., 2013; Pattanayak et al., 2013; Tsai et al., 2015). Investigating the occurrence and frequency of off-target mutations in plants is the main aim of this PhD study.

Methods for the detection of off-target activity

Methods for the detection of off-target activity can be broadly divided into two types: unbiased and biased. In biased methods, a selection of genomic loci is tested for the presence of off-target mutations. These genomic loci are generally selected based on their homology to the target site. The detection of mutations is frequently performed by PCR amplification of the site of interest, which is then followed by T7 endonuclease assays (Mashal et al., 1995), SURVEYOR assays (Qiu et al., 2004), Sanger sequencing and subsequent decomposition of the trace files (Brinkman et al., 2014), or next generation sequencing.

In unbiased methods, no pre-selection of genomic sites to be tested is made, and the whole genome is tested for the occurrence of off-target activity or mutations. A straightforward, *in vivo* method for unbiased off-target mutation detection is whole genome resequencing, but this can still be quite costly to perform for a lot of samples. Additionally, identified mutations may have arisen naturally rather than through CRISPR-Cas off-target activity. Therefore, several other methods have been developed.

Some of these methods, such as Digenome-seq, the related SITE-seq, and CIRCLE-seq (Cameron et al., 2017; Kim et al., 2015; Tsai et al., 2017), are based on *in vitro* digestion of genomic DNA and subsequent identification of cleavage sites by next-generation sequencing. Although informative, these *in vitro* screens may be less representative of the situation in a living cell, as the isolated

genomic DNA loses its folding and chromatin state. Using these methods, off-target sites can be identified with high sensitivity, but many false positives are found, and additional targeted deep sequencing is often necessary to validate identified off-target sites. Therefore, *in vivo* methods such as GUIDE-seq may yield more valuable information (Tsai et al., 2015). In this method, cells are transfected with the CRISPR-Cas machinery and many copies of a short, double-stranded oligodeoxynucleotide (dsODN). This dsODN can be integrated into DSBs by non-homologous end-joining. After purification of genomic DNA from the cell pool, integration sites can be identified using next-generation sequencing methods, and the location of DSBs can be retraced by the presence of the ODN sequence. It should however be kept in mind that this method identifies DSBs, which can also occur naturally – not every identified DSB is necessarily the result of off-target activity by a Cas enzyme.

CRISPR-Cas technology in plants

A role for CRISPR in plant research and plant breeding

The ability to make targeted knockouts has had an enormous influence on fundamental plant research and shows great promise for application in plant breeding. In plant breeding, breeders aim to create, by selection, new crop cultivars with desirable traits. Desirable traits include higher yield, improved flavour, longer shelf life, improved disease or pest resistance, and enhanced tolerance against drought or saline soils. Many of the grains, fruits, and vegetables on sale at (super)markets today have undergone centuries or even millennia of selective breeding and bear little resemblance to their wild ancestors. Perhaps the most striking example is the case of *Brassica oleracea*: over time, this plant, also known as wild mustard, has been bred into six different well-known vegetables: cabbage, Brussels sprouts, kohlrabi, kale, broccoli, and cauliflower. Other examples are vegetables such as tomato and eggplant, which have both significantly increased in size compared to their wild ancestors, and staple crops such as rice and maize, which have seen great improvements in grain and kernel yields and major changes in plant architecture, amongst many other traits (Q. Chen et al., 2021).

At the basis of plant breeding lies genetic variation. Breeders ideally have access to a wide range of traits, which can be selected for and combined as desired. A lot of genetic variation can readily be found in nature, but sometimes the desired trait cannot be identified, and new mutations – the basis of genetic diversity – arise only slowly.

Methods for inducing mutations and thus variation have been developed to speed up this process. In the late 20s of the previous century, it was shown that mutations could be induced by X-ray irradiation of germinating seeds in barley and maize (Stadler, 1928a, 1928b). Over time, other mutagenic treatments such as Gamma-ray irradiation and treatment with chemicals such as ethyl-methanesulfonate (EMS) have been applied to crops (Jankowicz-Cieslak & Till, 2015). The application of these methods is now known as "mutation breeding". As discussed previously, a downside of these non-specific methods is that they induce mutations all over the genome – many detrimental ones frequently accompany beneficial mutations, and often, lengthy repeated backcrossing is needed to obtain the desired elite cultivar.

In this era of widely available genomes and known gene functions, predicting which gene(s) will affect a trait is often possible. In such cases, CRISPR-mediated genome editing can be applied to modify these genes specifically. This circumvents both the need for screening mutant populations and extensive backcrossing, thus speeding up the breeding process. The power and speed of genome editing in plant breeding were illustrated by a study in which knowledge of tomato gene functions was used to guide the editing of several genes in the orphan crop groundcherry *(Physalis pruinosa)*. These efforts resulted in a more compact plant with more and bigger fruits, more suitable for larger-scale agriculture on a timescale that would be impossible using traditional mutation breeding (Lemmon et al., 2018).

CRISPR-mediated genome editing is also promising for breeding crops with more complicated genome structures. An example of this is wheat (*Triticum aestivum*), which has a hexaploid genome that combines three ancestral diploid genomes. Recombination can only occur between each pair (Feldman & Levy, 2005). To knock out a gene, all copies on all chromosomes must be perturbed, which is difficult to achieve using mutation breeding and crossing. With CRISPR-Cas9 however, such things enter the realm of possibility, and progress has even been made toward the development of gluten-free wheat (Jouanin et al., 2018).

Delivery of CRISPR components to plant cells

To obtain a CRISPR-edited plant, the components essential for CRISPR-mediated genome editing must be delivered to the plant cell. The most frequently applied method is *Agrobacterium*-mediated transformation, by which genes encoding Cas9 or another nuclease and gRNAs can be integrated into plant genomes.

Plants with this stable integration of foreign genes in their genomes are generally generated through tissue culture. In this process, small parts of plants (so-called explants) are co-cultivated with *Agrobacterium*, which will transfer its genetic payload to the plant genome. This or similar methods have successfully been applied to regenerate plants containing CRISPR mutations in Arabidopsis, tomato, rice, maize and many other species (Brooks et al., 2014; Feng et al., 2013, 2014; Liang et al., 2014). Regenerated plants can be selfed or crossed to wild-type plants to remove the integrated T-DNA encoding the CRISPR-Cas machinery from the genome while keeping the induced mutations. Unfortunately, not all plant species are amenable to *Agrobacterium*-mediated transformation or tissue culture. Additionally, plant tissue culture is a laborious and time-intensive procedure, and it can frequently take many months to go from explant to regenerated plant. This limitation complicates large-scale screens – for example, to test the efficiency or specificity of many different gRNAs - in entire plants.

Some of these issues can be overcome by making use of cell systems, such as protoplast systems. Protoplasts are plant cells that are stripped of their cell walls. This leaves them with only their cell membrane and makes them amenable to the introduction, by transfection, of plasmids encoding CRISPR systems. Alternatively, they can be transfected with pre-assembled Cas9-sgRNA complexes, also known as ribonucleoproteins or RNPs, resulting in DNA-free genome editing. Although challenging, protoplasts can be regrown into complete plants, and regeneration of CRISPR-edited protoplasts has been achieved in several plant species, such as potato, tobacco, and tomato (Andersson et al., 2018; C.-S. S. Lin et al., 2018; Y. Liu et al., 2022; Reed & Bargmann, 2021).

Tomato: an early adopter crop for CRISPR technology

Tomato (*Solanum lycopersicum)* is an economically important crop that is consumed all over the world. Additionally, it has many characteristics that make it attractive for research purposes and which allowed it to become a model species for fleshy fruit crops: it is amenable to *Agrobacterium*-mediated transformation, can readily be regenerated in tissue culture, has a relatively short generation time, can be self-fertilized, and has a relatively small diploid genome. The tomato genome has been fully sequenced and is well-annotated. Multiple projects have mapped the genetic diversity of tomato by resequencing and even constructing new genome assemblies of different tomato cultivars and wild tomato relatives (Aflitos et al., 2014; Alonge et al., 2020; Sato et al., 2012; van Rengs et al., 2022).

CRISPR-Cas9 technology was rapidly applied to tomato and was found to be an efficient method of inducing heritable gene knockouts (Brooks et al., 2014). It has been used to study the function of

genes relating to many different developmental processes and traits, such as fruit development, fruit quality, and resistance to biotic and abiotic stresses (Bate et al., 2021; Wang et al., 2019). Tomato has also been used in a showcase study demonstrating the power of CRISPR-mediated genome editing and the marked impact it can have on the speed of plant breeding – especially for crops that have well-characterized genomes. Using CRISPR-Cas9 and available knowledge about tomato gene functions, Kwon et al. set out to develop a tomato variant more suitable for urban agriculture: a tomato with a compact, "determinate" architecture and smaller fruits, without compromising on fruit yield or sweetness as compared to the wild-type (Kwon et al., 2020). This was achieved by knocking out three genes: *SELF PRUNING (SP),* its paralog *SP5G,* and *Solanum lycopersicum ERECTA (SlER)*, which they were able to do in a single step.

Due to its favourable characteristics, tomato has also been an early adopter crop for many CRISPR-Cas based technologies, such as the use of different Cas nucleases, base editors, and prime editing technology (Lu et al., 2020; Shimatani et al., 2017; Van Vu et al., 2020).

Tomato is also pioneering as the first CRISPR-edited crop to be sold on the open market. In 2021, the Tokyo-based company Sanatech Seeds started selling the "Sicilian Rouge High GABA" tomato, which produces higher amounts of gamma-aminobutyric acid (GABA) (Waltz, 2021). GABA has been claimed to be able to lower blood pressure and improve stress, mood, fatigue and sleep.

Scope of this Thesis

CRISPR-mediated genome editing is a powerful tool to induce targeted mutations in plant genomes and has the potential to speed up plant breeding. However, studies on the off-target effects of this promising technique are limited. In 2018, NWO started the research programme "Biotechnology and Safety", which aims to research safety aspects of new biotechnological techniques – among them, CRISPR-Cas. The research described in this thesis was part of this research programme. The main aim of this thesis was to acquire more insight into the specificity of CRISPR-Cas induced mutagenesis in plants. We chose to perform this research in tomato, which has all the benefits of a model species, yet is also an important food crop.

The relative lack of high-throughput research systems in plants poses a significant obstacle to obtaining the large datasets necessary to study the specificity of CRISPR-Cas in-depth. As a first step towards overcoming this, we set up a system for 96-well format transfection of tomato protoplasts. We coupled this with next-generation amplicon sequencing, which allowed us to investigate CRISPR-Cas9 and CRISPR-Cas12a induced on- and off-target mutations on a larger scale than had hitherto been performed. Although the data obtained by this approach are highly informative, a major limitation of this method is that only predicted off-target sites are investigated for the presence of mutations. Therefore, we also adapted the unbiased, *in vivo* method GUIDE-seq for off-target DSB detection in plants and succeeded at detecting on-target and off-target DSBs. Finally, as a second unbiased approach to off-target mutation identification, we apply whole genome resequencing to 38 stably transformed CRISPR-Cas9 mutagenized tomato plants. Using a combination of both biased and unbiased approaches, we have studied the specificity of especially CRISPR-Cas9 and, to a lesser extent, CRISPR-Cas12a mediated genome editing in tomato in-depth.

References

Aelst, K. van, C.J. Martínez-Santiago, S.J. Cross, and M.D. Szczelkun. 2019. "The Effect of DNA Topology on Observed Rates of R-Loop Formation and DNA Strand Cleavage by CRISPR Cas12a." *Genes* 10 (2). https://doi.org/10.3390/genes10020169.

Aflitos, S., E. Schijlen, H. de Jong, D. de Ridder, S. Smit, R. Finkers, J. Wang, … S. Peters. 2014. "Exploring Genetic Variation in the Tomato (Solanum Section Lycopersicon) Clade by Whole-Genome Sequencing." *The Plant Journal* 80 (1): 136–48. https://doi.org/10.1111/tpj.12616.

Alonge, M., X. Wang, M. Benoit, S. Soyk, L. Pereira, L. Zhang, H. Suresh, … Z.B. Lippman. 2020. "Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato." *Cell*, 1–17. https://doi.org/10.1016/j.cell.2020.05.021.

Anders, C., O. Niewoehner, A. Duerst, and M. Jinek. 2014. "Structural Basis of PAM-Dependent Target DNA Recognition by the Cas9 Endonuclease." *Nature* 513 (7519): 569–73. https://doi.org/10.1038/nature13579.

Andersson, M., H. Turesson, N. Olsson, A.S. Fält, P. Ohlsson, M.N. Gonzalez, M. Samuelsson, and P. Hofvander. 2018. "Genome Editing in Potato via CRISPR-Cas9 Ribonucleoprotein Delivery." *Physiologia Plantarum* 164 (4): 378–84. https://doi.org/10.1111/ppl.12731.

Anzalone, A. V., P.B. Randolph, J.R. Davis, A.A. Sousa, L.W. Koblan, J.M. Levy, P.J. Chen, … D.R. Liu. 2019. "Search-and-Replace Genome Editing without Double-Strand Breaks or Donor DNA." *Nature* 576 (7785): 149–57. https://doi.org/10.1038/s41586-019-1711-4.

Arnould, S., C. Perez, J.P. Cabaniols, J. Smith, A. Gouble, S. Grizot, J.C. Epinat, A. Duclert, P. Duchateau, and F. Pâques. 2007. "Engineered I-CreI Derivatives Cleaving Sequences from the Human XPC Gene Can Induce Highly Efficient Gene Correction in Mammalian Cells." *Journal of Molecular Biology* 371 (1): 49–65. https://doi.org/10.1016/j.jmb.2007.04.079.

Bae, S., J. Park, and J.-S. Kim. 2014. "Cas-OFFinder: A Fast and Versatile Algorithm That Searches for Potential off-Target Sites of Cas9 RNA-Guided Endonucleases." *Bioinformatics* 30 (10): 1473–75. https://doi.org/10.1093/BIOINFORMATICS/btu048.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. a Romero, and P. Horvath. 2007. "CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes." *Science* 315 (March): 1709–12.

Bate, N.J., C.D. Dardick, R.A. de Maagd, and R.W. Williams. 2021. "Opportunities and Challenges Applying Gene Editing to Specialty Crops." *In Vitro Cellular & Developmental Biology - Plant* 57 (4): 709–19. https://doi.org/10.1007/s11627-021-10208-x.

Belhaj, K., A. Chaparro-Garcia, S. Kamoun, N.J. Patron, and V. Nekrasov. 2015. "Editing Plant Genomes with CRISPR/Cas9." *Current Opinion in Biotechnology* 32: 76–84. https://doi.org/10.1016/j.copbio.2014.11.007.

Ben-Tov, D., F. Mafessoni, A. Cucuy, A. Honig, C. Melamed-Bessudo, and A.A. Levy. 2023. "Uncovering the Dynamics of Precise Repair at CRISPR/Cas9-Induced Double-Strand Breaks." *BioRxiv*. https://doi.org/10.1101/2023.01.10.523377.

Bétermier, M., P. Bertrand, and B.S. Lopez. 2014. "Is Non-Homologous End-Joining Really an Inherently Error-Prone Process?" *PLoS Genetics* 10 (1). https://doi.org/10.1371/journal.pgen.1004086.

Boch, J., and U. Bonas. 2010. "Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function." *Annual Review of Phytopathology* 48 (September): 419–36. https://doi.org/10.1146/annurev-phyto-080508-081936.

Bolotin, A., B. Quinquis, A. Sorokin, and S. Dusko Ehrlich. 2005. "Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) Have Spacers of Extrachromosomal Origin." *Microbiology* 151 (8): 2551–61. https://doi.org/10.1099/mic.0.28048-0.

Brinkman, E.K., T. Chen, M. Amendola, and B. Van Steensel. 2014. "Easy Quantitative Assessment of Genome Editing by Sequence Trace Decomposition." *Nucleic Acids Research* 42 (22): 1–8. https://doi.org/10.1093/nar/gku936.

Brooks, C., V. Nekrasov, Z.B. Lippman, and J. Van Eck. 2014. "Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/ CRISPR-Associated9 System." *Plant Physiology* 166 (3): 1292–1297. https://doi.org/10.1104/pp.114.247577.

Brouns, S.J.J., M.M. Jore, M. Lundgren, E.R. Westra, R.J.H. Slijkhuis, A.P.L. Snijders, M.J. Dickman, K.S. Makarova, E. V. Koonin, and J. van der Oost. 2008. "Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes." *Science* 321 (5891): 960–64. https://doi.org/10.1126/science.1159689.

Burmistrz, M., K. Krakowski, and A. Krawczyk-Balska. 2020. "RNA-Targeting CRISPR–Cas Systems and Their Applications." *International Journal of Molecular Sciences*. MDPI AG. https://doi.org/10.3390/ijms21031122.

Cameron, P., C.K. Fuller, P.D. Donohoue, B.N. Jones, M.S. Thompson, M.M. Carter, S. Gradia, … A.P. May. 2017. "Mapping the Genomic Landscape of CRISPR–Cas9 Cleavage." *Nature Methods* 14 (6). https://doi.org/10.1038/nmeth.4284.

Chatterjee, N., and G.C. Walker. 2017. "Mechanisms of DNA Damage, Repair, and Mutagenesis." *Environmental and Molecular Mutagenesis*. John Wiley and Sons Inc. https://doi.org/10.1002/em.22087.

Chen, J.S., E. Ma, L.B. Harrington, M. Da Costa, X. Tian, J.M. Palefsky, and J.A. Doudna. 2018. "CRISPR-Cas12a Target Binding Unleashes Indiscriminate Single-Stranded DNase Activity." *Science* 360 (6387): 436–39. https://doi.org/10.1126/science.aar6245.

Chen, Q., W. Li, L. Tan, and F. Tian. 2021. "Harnessing Knowledge from Maize and Rice Domestication for New Crop Breeding." *Molecular Plant*. Cell Press. https://doi.org/10.1016/j.molp.2020.12.006.

Chevalier, B.S., and B.L. Stoddard. 2001. "Homing Endonucleases: Structural and Functional Insight into the Catalysts of Intron/Intein Mobility." *SURVEY AND SUMMARY*. Vol. 29.

Concordet, J.P., and M. Haeussler. 2018. "CRISPOR: Intuitive Guide Selection for CRISPR/Cas9 Genome Editing Experiments and Screens." *Nucleic Acids Research* 46 (W1): W242–45. https://doi.org/10.1093/nar/gky354.

Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, … F. Zhang. 2013. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science* 339 (6121): 819–23. https://doi.org/10.1126/science.1231143.

Deltcheva, E., K. Chylinski, C.M. Sharma, K. Gonzales, Y. Chao, Z.A. Pirzada, M.R. Eckert, J. Vogel, and E. Charpentier. 2011. "CRISPR RNA Maturation by Trans-Encoded Small RNA and Host Factor RNase III." *Nature* 471 (7340): 602–7. https://doi.org/10.1038/nature09886.

Deveau, H., R. Barrangou, J.E. Garneau, J. Labonté, C. Fremaux, P. Boyaval, D.A. Romero, P. Horvath, and S. Moineau. 2008. "Phage Response to CRISPR-Encoded Resistance in Streptococcus Thermophilus." *Journal of Bacteriology* 190 (4): 1390–1400. https://doi.org/10.1128/JB.01412-07.

Dong, D., K. Ren, X. Qiu, J. Zheng, M. Guo, X. Guan, H. Liu, … Z. Huang. 2016. "The Crystal Structure of Cpf1 in Complex with CRISPR RNA." *Nature* 532 (7600): 522–26. https://doi.org/10.1038/nature17944.

Doudna, J.A., and E. Charpentier. 2014. "The New Frontier of Genome Engineering with CRISPR-Cas9." *Science* 346 (6213): 1258096–1258096. https://doi.org/10.1126/science.1258096.

Durai, S., M. Mani, K. Kandavelou, J. Wu, M.H. Porteus, and S. Chandrasegaran. 2005. "Zinc Finger Nucleases: Custom-Designed Molecular Scissors for Genome Engineering of Plant and Mammalian Cells." *Nucleic Acids Research* 33 (18): 5978–90. https://doi.org/10.1093/nar/gki912.

Epinat, J.C., S. Amould, P. Chames, P. Rochaix, D. Desfontaines, C. Puzin, A. Patin, A. Zanghellini, F. Pâques, and E. Lacroix. 2003. "A Novel Engineered Meganuclease Induces Homologous Recombination in Yeast and Mammalian Cells." *Nucleic Acids Research* 31 (11): 2952–62. https://doi.org/10.1093/nar/gkg375.

Feldman, M., and A.A. Levy. 2005. "Allopolyploidy - A Shaping Force in the Evolution of Wheat Genomes." *Cytogenetic and Genome Research* 109 (1–3): 250–58. https://doi.org/10.1159/000082407.

Feng, Z., Y. Mao, N. Xu, B. Zhang, P. Wei, D.-L. Yang, Z. Wang, … J.-K. Zhu. 2014. "Multigeneration Analysis Reveals the Inheritance, Specificity, and Patterns of CRISPR/Cas-Induced Gene Modifications in Arabidopsis." *Proceedings of the National Academy of Sciences of the United States of America* 111 (12): 4632–37. https://doi.org/10.1073/pnas.1400822111.

Feng, Z., B. Zhang, W. Ding, X. Liu, D.-L. Yang, P. Wei, F. Cao, … J.-K. Zhu. 2013. "Efficient Genome Editing in Plants Using a CRISPR/Cas System." *Cell Research* 23 (10): 1229–32. https://doi.org/10.1038/cr.2013.114.

Fu, Y., J.A. Foden, C. Khayter, M.L. Maeder, D. Reyon, J.K. Joung, and J.D. Sander. 2013. "High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells." *Nature Biotechnology* 31 (9): 822–26. https://doi.org/10.1038/nbt.2623.

Gao, H., J. Smith, M. Yang, S. Jones, V. Djukanovic, M.G. Nicholson, A. West, … L.A. Lyznik. 2010. "Heritable Targeted Mutagenesis in Maize Using a Designed Endonuclease." *Plant Journal* 61 (1): 176–87. https://doi.org/10.1111/j.1365-313X.2009.04041.x.

Gao, L., D.B.T. Cox, W.X. Yan, J.C. Manteiga, M.W. Schneider, T. Yamano, H. Nishimasu, O. Nureki, N. Crosetto, and F. Zhang. 2017. "Engineered Cpf1 Variants with Altered PAM Specificities." *Nature Biotechnology* 35 (8): 789–92. https://doi.org/10.1038/nbt.3900.

Garneau, J.E., M.È. Dupuis, M. Villion, D.A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A.H. Magadán, and S. Moineau. 2010. "The CRISPR/Cas Bacterial Immune System Cleaves Bacteriophage and Plasmid DNA." *Nature* 468 (7320): 67–71. https://doi.org/10.1038/nature09523.

Gaudelli, N.M., A.C. Komor, H.A. Rees, M.S. Packer, A.H. Badran, D.I. Bryson, and D.R. Liu. 2017. "Programmable Base Editing of A•T to G•C in Genomic DNA without DNA Cleavage." *Nature* 551 (7681): 1–27. https://doi.org/10.1038/nature24644.

Gilbert, L.A., M.A. Horlbeck, B. Adamson, J.E. Villalta, Y. Chen, E.H. Whitehead, C. Guimaraes, … J.S. Weissman. 2014. "Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation." *Cell* 159 (3): 647–61. https://doi.org/10.1016/j.cell.2014.09.029.

Haft, D.H., J. Selengut, E.F. Mongodin, and K.E. Nelson. 2005. "A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes." *PLoS Computational Biology* 1 (6): 0474–83. https://doi.org/10.1371/journal.pcbi.0010060.

Hale, C.R., P. Zhao, S. Olson, M.O. Duff, B.R. Graveley, L. Wells, R.M. Terns, and M.P. Terns. 2009. "RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex." *Cell* 139 (5): 945–56. https://doi.org/10.1016/j.cell.2009.07.040.

Ishino, Y., H. Shinagawa, K. Makino, M. Amemura, and A. Nakatura. 1987. "Nucleotide Sequence of the Iap Gene, Responsible for Alkaline Phosphatase Isoenzyme Conversion in Escherichia Coli, and Identification of the Gene Product." *Journal of Bacteriology* 169 (12): 5429–33. https://doi.org/10.1128/jb.169.12.5429-5433.1987.

Jankowicz-Cieslak, J., and B.J. Till. 2015. "Forward and Reverse Genetics in Crop Breeding." In *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, 1:215–40. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-22521-0_8.

Jansen, Ruud., Jan.D.A. van Embden, Wim. Gaastra, and Leo.M. Schouls. 2002. "Identification of Genes That Are Associated with DNA Repeats in Prokaryotes." *Molecular Microbiology* 43 (6): 1565–75. https://doi.org/10.1046/j.1365-2958.2002.02839.x.

Jeon, J.S., S. Lee, K.H. Jung, S.H. Jun, D.H. Jeong, J. Lee, C. Kim, … G. An. 2000. "T-DNA Insertional Mutagenesis for Functional Genomics in Rice." *Plant Journal* 22 (6): 561–70. https://doi.org/10.1046/j.1365-313X.2000.00767.x.

Jeon, Y., Y.H. Choi, Y. Jang, J. Yu, J. Goo, G. Lee, Y.K. Jeong, … S. Bae. 2018. "Direct Observation of DNA Target Searching and Cleavage by CRISPR-Cas12a." *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-018-05245-x.

Jiang, F., K. Zhou, L. Ma, S. Gressel, and J.A. Doudna. 2015. "A Cas9–Guide RNA Complex Preorganized for Target DNA Recognition." *Science* 348 (6242): 1477–81. https://doi.org/10.1126/science.aab1452.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Jinek, M., F. Jiang, D.W. Taylor, S.H. Sternberg, E. Kaya, E. Ma, C. Anders, … J.A. Doudna. 2014. "Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation." *Science* 343 (6176). https://doi.org/10.1126/science.1247997.

Jouanin, A., L.J.W.J. Gilissen, L.A. Boyd, J. Cockram, F.J. Leigh, E.J. Wallington, H.C. van den Broeck, … M.J.M. Smulders. 2018. "Food Processing and Breeding Strategies for Coeliac-Safe and Healthy Wheat Products." *Food Research International* 110 (August): 11–21. https://doi.org/10.1016/j.foodres.2017.04.025.

Joung, J.K., and J.D. Sander. 2013. "TALENs: A Widely Applicable Technology for Targeted Genome Editing." *Nature Reviews. Molecular Cell Biology* 14 (1): 49–55. https://doi.org/10.1038/nrm3486.

Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, … J. van Embden. 1997. "Simultaneous Detection and Strain Differentiation of Mycobacterium Tuberculosis for Diagnosis and Epidemiology." *Journal of Clinical Microbiology* 35 (4): 907–14. https://doi.org/10.1128/jcm.35.4.907-914.1997.

Kawakami, K., D.A. Largaespada, and Z. Ivics. 2017. "Transposons As Tools for Functional Genomics in Vertebrate Models." *Trends in Genetics*. Elsevier Ltd. https://doi.org/10.1016/j.tig.2017.07.006.

Kim, D., S. Bae, J. Park, E. Kim, S. Kim, H.R. Yu, J. Hwang, J.-I. Kim, and J.-S. Kim. 2015. "Digenome-Seq: Genome-Wide Profiling of CRISPR-Cas9 off-Target Effects in Human Cells." *Nature Methods* 12 (3): 237–44. https://doi.org/10.1038/nmeth.3284.

Koncz, C., K. N6meth, G.P. R6dei, and J. Schell. 1992. "T-DNA Insertional Mutagenesis in Arabidopsis." *Plant Molecular Biology*. Vol. 20.

Konermann, S., M.D. Brigham, a E. Trevino, J. Joung, O.O. Abudayyeh, C. Barcena, P.D. Hsu, … F. Zhang. 2014. "Genome-Scale Transcriptional Activation by an Engineered CRISPR-Cas9 Complex." *Nature* 517 (7536): 583–88. https://doi.org/10.1038/nature14136.

Koonin, E. V., and K.S. Makarova. 2019. "Origins and Evolution of CRISPR-Cas Systems." *Philosophical Transactions of the Royal Society B: Biological Sciences* 374 (1772): 20180087. https://doi.org/10.1098/rstb.2018.0087.

Koonin, E. V., K.S. Makarova, and F. Zhang. 2017. "Diversity, Classification and Evolution of CRISPR-Cas Systems." *Current Opinion in Microbiology* 37: 67–78. https://doi.org/10.1016/j.mib.2017.05.008.

Kwon, C.T., J. Heo, Z.H. Lemmon, Y. Capua, S.F. Hutton, J. Van Eck, S.J. Park, and Z.B. Lippman. 2020. "Rapid Customization of Solanaceae Fruit Crops for Urban Agriculture." *Nature Biotechnology* 38 (2): 182–88. https://doi.org/10.1038/s41587-019-0361-2.

Lander, E.S. 2016. "The Heroes of CRISPR." *Cell*. Cell Press. https://doi.org/10.1016/j.cell.2015.12.041.

Lei, Y., L. Lu, H.Y. Liu, S. Li, F. Xing, and L.L. Chen. 2014. "CRISPR-P: A Web Tool for Synthetic Single-Guide RNA Design of CRISPR-System in Plants." *Molecular Plant* 7 (9): 1494–96. https://doi.org/10.1093/mp/ssu044.

Lemmon, Z.H., N.T. Reem, J. Dalrymple, S. Soyk, K.E. Swartwood, D. Rodriguez-Leal, J. Van Eck, and Z.B. Lippman. 2018. "Rapid Improvement of Domestication Traits in an Orphan Crop by Genome Editing." *Nature Plants* 4 (10): 766–70. https://doi.org/10.1038/s41477-018-0259-x.

Liang, Z., K. Zhang, K. Chen, and C. Gao. 2014. "Targeted Mutagenesis in Zea Mays Using TALENs and the CRISPR/Cas System." *Journal of Genetics and Genomics* 41 (2): 63–68. https://doi.org/10.1016/j.jgg.2013.12.001.

Lillestøl, R., P. Redder, R.A. Garrett, and K. Brügger. 2006. "A Putative Viral Defence Mechanism in Archaeal Cells." *Archaea* 2 (1): 59–72. https://doi.org/10.1155/2006/542818.

Lin, C.-S.S., C.-T.T. Hsu, L.-H.H. Yang, L.-Y.Y. Lee, J.-Y.Y. Fu, Q.-W.W. Cheng, F.-H.H. Wu, … M.-C.C. Shih. 2018. "Application of Protoplast Technology to CRISPR/Cas9 Mutagenesis: From Single-Cell Mutation Detection to Mutant Plant Regeneration." *Plant Biotechnology Journal* 140 (6): 1295–1310. https://doi.org/10.1111/pbi.12870.

Lin, Q., Y. Zong, C. Xue, S. Wang, S. Jin, Z. Zhu, Y. Wang, … C. Gao. 2020. "Prime Genome Editing in Rice and Wheat." *Nature Biotechnology* 38 (5): 582–85. https://doi.org/10.1038/s41587-020-0455-x.

Liu, X.S., H. Wu, X. Ji, Y. Stelzer, X. Wu, S. Czauderna, J. Shu, D. Dadon, R.A. Young, and R. Jaenisch. 2016. "Editing DNA Methylation in the Mammalian Genome." *Cell* 167 (1): 233-247.e17. https://doi.org/10.1016/j.cell.2016.08.056.

Liu, Y., M. Andersson, A. Granell, T. Cardi, P. Hofvander, and A. Nicolia. 2022. "Establishment of a DNA-Free Genome Editing and Protoplast Regeneration Method in Cultivated Tomato (Solanum Lycopersicum)." *Plant Cell Reports* 41 (9): 1843–52. https://doi.org/10.1007/s00299-022-02893-8.

Loenen, W.A.M., D.T.F. Dryden, E.A. Raleigh, G.G. Wilson, and N.E. Murrayy. 2014. "Highlights of the DNA Cutters: A Short History of the Restriction Enzymes." *Nucleic Acids Research* 42 (1): 3–19. https://doi.org/10.1093/nar/gkt990.

Lu, Y., Y. Tian, R. Shen, Q. Yao, D. Zhong, X. Zhang, and J. Zhu. 2020. "Precise Genome Modification in Tomato Using an Improved Prime Editing System." *Plant Biotechnology Journal*, October, pbi.13497. https://doi.org/10.1111/pbi.13497.

Makarova, K.S., N. V. Grishin, S.A. Shabalina, Y.I. Wolf, and E. V. Koonin. 2006. "A Putative RNA-Interference-Based Immune System in Prokaryotes: Computational Analysis of the Predicted Enzymatic Machinery, Functional Analogies with Eukaryotic RNAi, and Hypothetical Mechanisms of Action." *Biology Direct* 1: 1–26. https://doi.org/10.1186/1745-6150-1-7.

Makarova, K.S., Y.I. Wolf, O.S. Alkhnbashi, F. Costa, S.A. Shah, S.J. Saunders, R. Barrangou, … E. V. Koonin. 2015. "An Updated Evolutionary Classification of CRISPR-Cas Systems." *Nature Reviews Microbiology* 13 (11): 722–36. https://doi.org/10.1038/nrmicro3569.

Makarova, K.S., Y.I. Wolf, J. Iranzo, S.A. Shmakov, O.S. Alkhnbashi, S.J.J. Brouns, E. Charpentier, … E. V. Koonin. 2020. "Evolutionary Classification of CRISPR–Cas Systems: A Burst of Class 2 and Derived Variants." *Nature Reviews Microbiology*. Nature Research. https://doi.org/10.1038/s41579-019-0299-x.

Mali, P., L. Yang, K.M. Esvelt, J. Aach, M. Guell, J.E. DiCarlo, J.E. Norville, and G.M. Church. 2013. "RNA-Guided Human Genome Engineering via Cas9." *Science* 339 (6121): 823–26. https://doi.org/10.1126/science.1232033.

Manova, V., and D. Gruszka. 2015. "DNA Damage and Repair in Plants – from Models to Crops." *Frontiers in Plant Science* 6 (October): 1–26. https://doi.org/10.3389/fpls.2015.00885.

Marraffini, L.A., and E.J. Sontheimer. 2008. "CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA." *Science* 322 (5909): 1843–45. https://doi.org/10.1126/science.1165771.

Martienssen, R.A. 1998. "Functional Genomics: Probing Plant Gene Function and Expression with Transposons." *Proceedings of the National Academy of Sciences* 95 (5): 2021–26. https://doi.org/10.1073/pnas.95.5.2021.

Mashal, R.D., J. Koontz, and J. Sklar. 1995. "Detection of Mutations by Cleavage of DNA Heteroduplexes with Bacteriophage Resolvases." *Nature Genetics* 9: 177–83. http://www.nature.com.ezproxy.library.wur.nl/ng/journal/v9/n2/pdf/ng0295-177.pdf.

Mills, S., C. Griffin, A. Coffey, W.C. Meijer, B. Hafkamp, and R.P. Ross. 2010. "CRISPR Analysis of Bacteriophage-Insensitive Mutants (BIMs) of Industrial Streptococcus Thermophilus- Implications for Starter Design." *Journal of Applied Microbiology* 108 (3): 945–55. https://doi.org/10.1111/j.1365-2672.2009.04486.x.

Mojica, F. J. M., C. Díez-Villaseñor, J. García-Martínez, and C. Almendros. 2009. "Short Motif Sequences Determine the Targets of the Prokaryotic CRISPR Defence System." *Microbiology* 155 (3): 733–40. https://doi.org/10.1099/mic.0.023960-0.

Mojica, F. J.M., C. Ferrer, G. Juez, and F. Rodríguez-Valera. 1995. "Long Stretches of Short Tandem Repeats Are Present in the Largest Replicons of the Archaea Haloferax Mediterranei and Haloferax Volcanii and Could Be Involved in Replicon Partitioning." *Molecular Microbiology* 17 (1): 85–93. https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x.

Mojica, Francisco J.M., C. Díez-Villaseñor, J. García-Martínez, and E. Soria. 2005. "Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements." *Journal of Molecular Evolution* 60 (2): 174–82. https://doi.org/10.1007/s00239-004-0046-3.

Mojica, Francisco J.M., C. Díez-Villaseñor, E. Soria, and G. Juez. 2000. "Biological Significance of a Family of Regularly Spaced Repeats in the Genomes of Archaea, Bacteria and Mitochondria." *Molecular Microbiology*. https://doi.org/10.1046/j.1365-2958.2000.01838.x.

Mokrousov, I., E. Limeschenko, A. Vyazovaya, and O. Narvskaya. 2007. "Corynebacterium Diphtheriae Spoligotyping Based on Combined Use of Two CRISPR Loci." *Biotechnology Journal* 2 (7): 901–6. https://doi.org/10.1002/biot.200700035.

Morita, S., H. Noguchi, T. Horii, K. Nakabayashi, M. Kimura, K. Okamura, A. Sakai, … I. Hatada. 2016. "Targeted DNA Demethylation in Vivo Using DCas9-Peptide Repeat and ScFv-TET1 Catalytic Domain Fusions." *Nature Biotechnology* 34 (10): 1060–65. https://doi.org/10.1038/nbt.3658.

Mullins, M.C., M. Hammerschmidt, P. Haffter, and C. Nüsslein-Volhard. 1994. "Large-Scale Mutagenesis in the Zebrafish: In Search of Genes Controlling Development in a Vertebrate." *Current Biology* 4 (3): 189–202. https://doi.org/10.1016/S0960-9822(00)00048-8.

NIH. 2023. "ClinicalTrials.Gov." 2023. https://clinicaltrials.gov/ct2/results?cond=&term=CRISPR&cntry=&state=&city=&dist=.

Nishida, K., T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, … A. Kondo. 2016. "Targeted Nucleotide Editing Using Hybrid Prokaryotic and Vertebrate Adaptive Immune Systems." *Science* 353 (6305). https://doi.org/10.1126/science.aaf8729.

Pattanayak, V., S. Lin, J.P. Guilinger, E. Ma, J.A. Doudna, and D.R. Liu. 2013. "High-Throughput Profiling of off-Target DNA Cleavage Reveals RNA-Programmed Cas9 Nuclease Specificity." *Nature Biotechnology* 31 (9): 839–43. https://doi.org/10.1038/nbt.2673.

Pourcel, C., G. Salvignol, and G. Vergnaud. 2005. "CRISPR Elements in Yersinia Pestis Acquire New Repeats by Preferential Uptake of Bacteriophage DNA, and Provide Additional Tools for Evolutionary Studies." *Microbiology* 151 (3): 653–63. https://doi.org/10.1099/mic.0.27437-0.

Pulecio, J., N. Verma, E. Mejía-Ramírez, D. Huangfu, and A. Raya. 2017. "CRISPR/Cas9-Based Engineering of the Epigenome." *Cell Stem Cell*. Cell Press. https://doi.org/10.1016/j.stem.2017.09.006.

Qiu, P., H. Shandilya, J.M. D'Alessio, K. O'Connor, J. Durocher, and G.F. Gerard. 2004. "Mutation Detection Using Surveyor Nuclease." *BioTechniques* 36 (4): 702–7. http://www.biotechniques.com/multimedia/archive/00036/BTN_A_04364PF01_O_36417a.pdf.

Ramirez, C.L., J.E. Foley, D.A. Wright, F. Müller-Lerch, S.H. Rahman, T.I. Cornu, R.J. Winfrey, … J.K. Joung. 2008. "Unexpected Failure Rates for Modular Assembly of Engineered Zinc Fingers." *Nature Methods* 5 (5): 374–75. https://doi.org/10.1038/nmeth0508-374.

Reed, K.M., and B.O.R. Bargmann. 2021. "Protoplast Regeneration and Its Use in New Plant Breeding Technologies." *Frontiers in Genome Editing* 3 (September). https://doi.org/10.3389/fgeed.2021.734951.

Rees, H.A., and D.R. Liu. 2018. "Base Editing: Precision Chemistry on the Genome and Transcriptome of Living Cells." *Nature Reviews Genetics*. Nature Publishing Group. https://doi.org/10.1038/s41576-018-0059-1.

Ren, Q., S. Sretenovic, S. Liu, X. Tang, L. Huang, Y. He, L. Liu, … Y. Zhang. 2021. "PAM-Less Plant Genome Editing Using a CRISPR–SpRY Toolbox." *Nature Plants* 7 (January). https://doi.org/10.1038/s41477-020-00827-4.

Rengs, W.M.J. van, M.H.W. Schmidt, S. Effgen, D.B. Le, Y. Wang, M.W.A.M. Zaidan, B. Huettel, H.J. Schouten, B. Usadel, and C.J. Underwood. 2022. "A Chromosome Scale Tomato Genome Built from Complementary PacBio and Nanopore Sequences Alone Reveals Extensive Linkage Drag during Breeding." *Plant Journal* 110 (2): 572–88. https://doi.org/10.1111/tpj.15690.

Sato, S., S. Tabata, H. Hirakawa, E. Asamizu, K. Shirasawa, S. Isobe, T. Kaneko, … R.M. Klein Lankhorst. 2012. "The Tomato Genome Sequence Provides Insights into Fleshy Fruit Evolution." *Nature* 485 (7400): 635–41. https://doi.org/10.1038/nature11119.

Shimatani, Z., S. Kashojiya, M. Takayama, R. Terada, T. Arazoe, H. Ishii, H. Teramura, … A. Kondo. 2017. "Targeted Base Editing in Rice and Tomato Using a CRISPR-Cas9 Cytidine Deaminase Fusion." *Nature Biotechnology* 35 (March): 441–43. https://doi.org/10.1038/nbt.3833.

Shmakov, S., A. Smargon, D. Scott, D. Cox, N. Pyzocha, W. Yan, O.O. Abudayyeh, … E. V. Koonin. 2017. "Diversity and Evolution of Class 2 CRISPR-Cas Systems." *Nature Reviews. Microbiology* 15 (3): 169–82. https://doi.org/10.1038/nrmicro.2016.184.

Stadler, L.J. 1928a. "Genetic Effects of X-Rays in Maize." *Proceedings of the National Academy of Sciences* 14 (1): 69–75. https://doi.org/10.1073/pnas.14.1.69.

Stadler, L.J.. 1928b. "Mutations in Barley Induced by X-Rays and Radium." *Science* 68 (1756): 186–87. https://doi.org/10.1126/science.68.1756.186.

Stella, S., P. Mesa, J. Thomsen, B. Paul, P. Alcón, S.B. Jensen, B. Saligram, M.E. Moses, N.S. Hatzakis, and G. Montoya. 2018. "Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity." *Cell* 175 (7): 1856-1871.e21. https://doi.org/10.1016/j.cell.2018.10.045.

Sternberg, S H, S. Redding, M. Jinek, E.C. Greene, and J.A. Doudna. 2014. "DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9." *Nature* 507 (7490): 62–67. https://doi.org/10.1038/nature13011.

Sternberg, Samuel H., B. Lafrance, M. Kaplan, and J.A. Doudna. 2015. "Conformational Control of DNA Target Cleavage by CRISPR-Cas9." *Nature* 527 (7576): 110–13. https://doi.org/10.1038/nature15544.

Strohkendl, I., F.A. Saifuddin, J.R. Rybarski, I.J. Finkelstein, and R. Russell. 2018. "Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a." *Molecular Cell* 71 (5): 816-824.e3. https://doi.org/10.1016/j.molcel.2018.06.043.

Swarts, D.C., and M. Jinek. 2019. "Mechanistic Insights into the Cis- and Trans-Acting DNase Activities of Cas12a." *Molecular Cell* 73 (3): 589-600.e4. https://doi.org/10.1016/j.molcel.2018.11.021.

Swarts, D.C., J. van der Oost, and M. Jinek. 2017. "Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a." *Molecular Cell* 66 (2): 221-233.e4. https://doi.org/10.1016/j.molcel.2017.03.016.

Tang, T.-H., J.-P. Bachellerie, T. Rozhdestvensky, M.-L. Bortolin, H. Huber, M. Drungowski, T. Elge, J. Brosius, and A. Hüttenhofer. 2002. "Identification of 86 Candidates for Small Non-Messenger RNAs from the Archaeon Archaeoglobus Fulgidus." *Proceedings of the National Academy of Sciences* 99 (11): 7536–41. https://doi.org/10.1073/pnas.112047299.

Tang, T.H., N. Polacek, M. Zywicki, H. Huber, K. Brugger, R. Garrett, J.P. Bachellerie, and A. Hüttenhofer. 2005. "Identification of Novel Non-Coding RNAs as Potential Antisense Regulators in the Archaeon Sulfolobus Solfataricus." *Molecular Microbiology* 55 (2): 469–81. https://doi.org/10.1111/j.1365-2958.2004.04428.x.

Tsai, S Q, Z. Zheng, N.T. Nguyen, M. Liebers, V. V Topkar, V. Thapar, N. Wyvekens, … J.K. Joung. 2015. "GUIDE-Seq Enables Genome-Wide Profiling of off-Target Cleavage by CRISPR-Cas Nucleases." *Nat Biotechnol* 33 (2): 187–97. https://doi.org/10.1038/nbt.3117.

Tsai, Shengdar Q, N.T. Nguyen, J. Malagon-Lopez, V. V Topkar, M.J. Aryee, and J.K. Joung. 2017. "CIRCLE-Seq: A Highly Sensitive in Vitro Screen for Genome-Wide CRISPR–Cas9 Nuclease off-Targets." *Nature Methods*, no. May. https://doi.org/10.1038/nmeth.4278.

Urnov, F.D., E.J. Rebar, M.C. Holmes, H.S. Zhang, and P.D. Gregory. 2010. "Genome Editing with Engineered Zinc Finger Nucleases." *Nature Reviews Genetics* 11 (9): 636–46. https://doi.org/10.1038/nrg2842.

Vergnaud, G., D. Zhou, M.E. Platonov, C. Pourcel, R. Yang, A.P. Anisimov, H. Neubauer, … Y. Li. 2007. "Analysis of the Three Yersinia Pestis CRISPR Loci Provides New Tools for Phylogenetic Studies and Possibly for the Investigation of Ancient DNA." In , 327–38. https://doi.org/10.1007/978-0-387-72124-8_30.

Vu, T. Van, V. Sivankalyani, E. Kim, D.T.H. Doan, M.T. Tran, J. Kim, Y.W. Sung, M. Park, Y.J. Kang, and J. Kim. 2020. "Highly Efficient Homology-directed Repair Using CRISPR/Cpf1-geminiviral Replicon in Tomato." *Plant Biotechnology Journal* 18 (10): 2133–43. https://doi.org/10.1111/pbi.13373.

Walton, R.T., K.A. Christie, M.N. Whittaker, and B.P. Kleinstiver. 2020. "Unconstrained Genome Targeting with Near-PAMless Engineered CRISPR-Cas9 Variants." *Science* 368 (6488): 290–96. https://doi.org/10.1126/science.aba8853.

Waltz, E. 2021. "GABA-Enriched Tomato Is First CRISPR-Edited Food to Enter Market." *Nature Biotechnology*, December. https://doi.org/10.1038/d41587-021-00026-2.

Wang, T., H. Zhang, and H. Zhu. 2019. "CRISPR Technology Is Revolutionizing the Improvement of Tomato and Other Fruit Crops." *Horticulture Research* 6 (1): 77. https://doi.org/10.1038/s41438-019-0159-x.

Xu, R., J. Li, X. Liu, T. Shan, R. Qin, and P. Wei. 2020. "Development of Plant Prime-Editing Systems for Precise Genome Editing." *Plant Communications* 1 (3): 100043. https://doi.org/10.1016/j.xplc.2020.100043.

Xu, W., C. Zhang, Y. Yang, S. Zhao, G. Kang, X. He, J. Song, and J. Yang. 2020. "Versatile Nucleotides Substitution in Plant Using an Improved Prime Editing System." *Molecular Plant* 13 (5): 675–78. https://doi.org/10.1016/j.molp.2020.03.012.

Xu, X., Y. Tao, X. Gao, L. Zhang, X. Li, W. Zou, K. Ruan, F. Wang, G. Xu, and R. Hu. 2016. "A CRISPR-Based Approach for Targeted DNA Demethylation." *Cell Discovery* 2 (May): 16009. https://doi.org/10.1038/celldisc.2016.9.

Yamano, T., H. Nishimasu, B. Zetsche, H. Hirano, I.M. Slaymaker, Y. Li, I. Fedorova, … O. Nureki. 2016. "Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA." *Cell* 165 (4): 949–62. https://doi.org/10.1016/j.cell.2016.04.003.

Zetsche, B., J.S. Gootenberg, O.O. Abudayyeh, I.M. Slaymaker, K.S. Makarova, P. Essletzbichler, S.E. Volz, … F. Zhang. 2015. "Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System." *Cell* 163 (3): 759–71. https://doi.org/10.1016/j.cell.2015.09.038.

Zetsche, B., M. Heidenreich, P. Mohanraju, I. Fedorova, J. Kneppers, E.M. DeGennaro, N. Winblad, … F. Zhang. 2016. "Multiplex Gene Editing by CRISPR–Cpf1 Using a Single CrRNA Array." *Nature Biotechnology* 35 (December): 31–34. https://doi.org/10.1038/nbt.3737.

Chapter 2

High-throughput sgRNA testing reveals rules for Cas9 specificity and DNA repair in tomato cells

E. Slaman[1,2], M. Lammers[2], G. C. Angenent[1,2], R. A. de Maagd[2]

[1] Wageningen University & Research, Laboratory of Molecular Biology, Wageningen, the Netherlands

[2] Wageningen University & Research, Bioscience, Wageningen, the Netherlands

This chapter was published in Frontiers in Genome Editing, 2023

https://doi.org/10.3389/fgeed.2023.1196763

Abstract

CRISPR/Cas9 technology has the potential to significantly enhance plant breeding. To determine the specificity and the mutagenic spectrum of SpCas9 in tomato, we designed 89 g(uide)RNAs targeting genes of the tomato MYB transcription factor family with varying predicted specificities. Plasmids encoding sgRNAs and Cas9 were introduced into tomato protoplasts, and target sites as well as 224 predicted off-target sites were screened for the occurrence of mutations using amplicon sequencing. Algorithms for the prediction of the efficacy of the sgRNAs had little predictive power in this system. The analysis of mutations suggested predictable identity of single base insertions. Off-target mutations were found for 13 out of 89 sgRNAs and only occurred at positions with one or two mismatches (at 14 and 3 sites, respectively). We found that PAM-proximal mismatches do not preclude low-frequency off-target mutations. Off-target mutations were not found at all 138 positions that had three or four mismatches. We compared off-target mutation frequencies obtained with plasmid encoding sgRNAs and Cas9 with those induced by ribonucleoprotein (RNP) transfections. The use of RNPs led to a significant decrease in relative off-target frequencies at 6 out of 17, no significant difference at 9, and an increase at 2 sites. Additionally, we show that off-target sequences with insertions or deletions relative to the sgRNA may be mutated, and should be considered during sgRNA design. Altogether, our data help sgRNA design by providing insight into the Cas9-induced double-strand break repair outcomes and the occurrence of off-target mutations.

Introduction

CRISPR/Cas-mediated genome editing has rapidly and decisively impacted the field of molecular biology. After the discovery of CRISPR systems functioning as an adaptive immune system in bacteria (Makarova et al. 2006; Barrangou et al. 2007; Brouns et al. 2008; Garneau et al. 2010), they have been engineered into efficient genome editing tools of unprecedented simplicity and speed (Jinek et al. 2012; Cong et al. 2013), often replacing pre-existing techniques for inducing targeted mutations, such as those using zinc-finger nucleases or transcription activator-like effector nucleases (TALEN).

CRISPR/Cas-mutagenesis has been successfully applied to many plant species (Jaganathan et al. 2018; Chen et al. 2019; Zhu et al. 2020). The technique has revolutionized plant research and has shown great potential for plant breeding (Lemmon et al. 2018; Kwon et al. 2020; Zsögön et al. 2018). In the European Union, CRISPR-edited crops are subject to the same risk assessment as transgenic crops (referred to as Genetically Modified Organisms -GMOs). This practice was partly motivated by the argument that the technology, unlike other mutagenesis techniques, does not have a long history of safe use.

In the context of CRISPR/Cas-mutagenesis, off-target mutations are induced mutations at positions other than the intended target site. These off-target sites have sequence similarity to the target site and may lead to an unintended disruption of gene function. Some still perceive off-target mutations as a hazard, even if the associated risk is low – especially when compared to spontaneously occurring mutations or genomic changes (Ossowski et al. 2010; Shirasawa et al. 2016; Bessoltane et al. 2022; Graham et al. 2020).

Early studies on CRISPR/Cas9 specificity in mammalian cells revealed that sequences with up to 4 mismatches to the target site could readily be cleaved by Cas9 (Fu et al. 2013; Pattanayak et al. 2013; Hsu et al. 2013). Since then, several studies have been performed on CRISPR/Cas9 specificity in plants, for example in *Arabidopsis*, rice, maize, cotton, tomato and grapevine (Feng et al. 2014; Q. Zhang et al. 2017; Tang et al. 2018; Lee et al. 2018; J. Li et al. 2018; Wang et al. 2021; Nekrasov et al. 2017; Sturme et al. 2022; Young et al. 2019). These studies suggest that the RNA-guided nuclease activity of Cas9 in plants is mostly specific and off-target mutations seldom occur. In tomato, a large-scale study focused on CRISPR/Cas9-induced mutagenesis of immunity genes found no evidence of off-target mutations in 144 analysed plants (N. Zhang et al. 2020). In an additional study, whole

genome resequencing of a mildew-resistant genome-edited tomato was performed. Likewise, no evidence of off-target mutations was found (Nekrasov et al. 2017).

Most studies screening for off-target mutations focus on stably transformed plants. The production of such plants often involves laborious tissue culture and regeneration protocols, which limit the number of sgRNAs that can be feasibly transformed and the number of plants that can be analysed. As a result, it does not allow for a thorough analysis of target/sgRNA-mismatch tolerance or a comprehensive analysis of double-strand break (DSB) repair outcome. Knowledge of mismatch tolerance is not only useful to assess the risk of off-target mutations but is also relevant for designing "multiplexing" mutagenesis experiments, in which highly similar alleles or paralogues are targeted simultaneously, and for allele-specific mutagenesis in heterozygous or polyploid crops. A more comprehensive analysis of DSB repair outcomes increases our understanding of repair mechanisms (in plants) and may increase the predictability of mutations. Most published studies on CRISPR/Cas-specificity and DSB repair originate from animal systems, where the use of cell cultures circumvented the limitations of studying whole organisms. Similarly, a more comprehensive overview of CRISPR/Cas9 genome editing specificity and outcomes in plants can be obtained from the study of cell cultures or of cell wall-less protoplasts that can be readily transfected with DNA encoding Cas9 and one or more sgRNAs, or with ribonucleoproteins (RNPs) composed of Cas9 and an *in vitro* produced sgRNA. Like mammalian cell lines, protoplasts allow the screening of sgRNAs in a high-throughput fashion and can give detailed information about both the types and the frequencies of induced mutations when coupled with next-generation sequencing techniques. Here we explore tomato protoplasts for this purpose, employing a flexible 96-well format for protoplast transfection and processing.

For this study, we designed 89 sgRNAs of varying predicted specificities, which together target 30 members of the *MYB* transcription factor gene family in tomato and screened for SpCas9-induced mutations at on-target sites and 224 predicted off-target sites (of which 68 in other *MYB* genes) using next-generation amplicon sequencing. Moreover, we looked at the nature of the induced mutations, the predictability of sgRNA efficacy, and the nature and predictability of insertions at the target site. The resulting data provide further insight into the range of mutations induced at on-target sites and the occurrence and frequency of mutations at off-target sites.

Materials and Methods

Selection of sgRNAs and off-target sites

Genes encoding MYB transcription factors in tomato were obtained from the Plant Transcription Factor Database Expression (Jin et al. 2017). A total of 29 *MYB* genes were selected for sgRNA design. In the selected genes, sgRNAs and potential off-target sites with a maximum of 4 mismatches were predicted using CRISPOR (Concordet and Haeussler 2018). Potential off-target sites with DNA or RNA bulges compared to the target site are not predicted by CRISPOR and were therefore predicted using Cas-OFFinder (Bae et al. 2014).

Transfection vectors and cloning of sgRNAs

The backbone of the vectors for plasmid transfections was constructed using Golden Gate cloning and parts from the MoClo Toolkit, a gift from Sylveste Marillonnet (AddGene kit #1000000044, and the MoClo Plant Parts kit, a gift from Nicola Patron (Addgene kit #1000000047) (Weber et al. 2011; Engler et al. 2008; 2014). A Level 1 vector for easy cloning of sgRNAs behind an *Arabidopsis thaliana U6-26* promoter was a gift from Marc Youles (The Sainsbury Laboratory). This vector contains an operon expressing an RFP protein that can be replaced by a spacer in a *Bsm*BI-mediated Golden Gate cut/ligate reaction, allowing pink/white screening of colonies that have successfully integrated the sgRNA ("CRISPRpink"). To allow its use in a Level 2 vector, the *Bsm*BI sites were replaced by *Bsa*I sites using restriction/ligation. Level 1 constructs *pICH4773::proNOS-NPTII-tOCS*, *pICH4772::2xproCaMV35-Cas9-tNOS, pICH47751::2xproCaMV35S-tGFP-tCaMV35S, pICH54044, pICH54055,* the adapted *pICH47781::pAtU6-26-CRISPRpink* sgRNA expression vector and end-linker *pICH41822* were then combined into Level 2 vector *pICSL4723*.

Subsequently, sgRNA transcription units were completed by replacing the CRISPRpink operon with annealed oligos encoding the spacer in a Golden Gate cut/ligate reaction using *Bsa*I. Oligonucleotide sequences used can be found in Supplementary Dataset 2.

DNA preparation

Highly pure DNA for transfection was prepared from 3 mL of overnight *E. coli* culture in LB medium using the PureYield Plasmid MiniPrep System (Promega), with the following adaptations: bacterial pellets were frozen at -20°C before processing to increase DNA yield, the column was washed twice with the endotoxin removal wash to acquire the desired purity, and plasmid DNA was eluted with 30 µL elution buffer preheated at 60°C.

Protoplast isolation

Tomato (*Solanum lycopersicum*) cv. "Moneyberg" plants were grown *in vitro* on hormone-free ½ concentration Murashige and Skoog medium (Duchefa) with 10 g/L sucrose and 8 g/L plant agar (Duchefa), pH 5.8, at 25 °C, under a 16h photoperiod. *In vitro* grown plants were used for protoplast isolation at 10-12 weeks after sowing. Reagents for protoplast isolation and transfection were prepared according to Sheen (Sheen 2002). Young, fully developed leaves were cut in a feather-like pattern and incubated overnight in the dark with the abaxial side down in 10 mL digestion solution (0.4 M mannitol, 20 mM MES, 20 mM KCl, 10 mM $CaCl_2$, 1% cellulase Onozuka R10 (Duchefa) and 0.25% macerozyme R10 (Duchefa); pH 5.7). The next day, the suspension was gently swirled to release protoplasts. The suspension was sieved through a 100 µm cell sieve into a 50 mL tube. The remaining leaf material was washed with 10 mL W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2 mM MES, pH 5.7) to release additional protoplasts and the suspensions were pooled. The pooled suspension was subsequently centrifuged for 3 min at 100 x g and the pellet was resuspended in 10 mL W5. This washing step was repeated twice, and finally pelleted protoplasts were resuspended in 10 mL magnesium-mannitol solution (0.4 M mannitol, 15 mM $MgCl_2$, 4 mM MES, pH 5.7) to prepare for transfection. Protoplast density was determined by counting in a haemocytometer and adjusted to 5.0 x 10^5 / mL.

Plasmid transfection

Protoplasts were transfected in 96 well plates, as described by Wehner et al. (Wehner et al. 2011), with the following modifications. For each replicate, 5 µL of plasmid DNA (2-3 µg per construct) was transferred to a well of a V-bottom microtiter plate and 30 µL protoplast suspension (containing 1.5 x 10^4 protoplasts) was added. Transfections were started by adding 35 µL polyethylene glycol solution (40% PEG-4000 (Fluka), 0.2 M mannitol, 0.1 M $CaCl_2$) and mixing by pipetting up and down. After a 10 min incubation, 120 µL W5 solution was added and mixed to stop transfection. Protoplasts were pelleted by centrifuging for 3 min at 200 x g. The pellet was resuspended in 100 µL WI solution (0.5 M mannitol, 20 mM KCl, 4 mM MES, pH 5.7). Protoplasts were incubated in the dark for 24h at 25 °C and GFP expression was imaged 24h after transfection by confocal microscopy (Leica DM5500) to determine transfection efficiency. Protoplasts were then pelleted by centrifuging at 200 x g for 3 min, supernatant was removed, and pellets were kept at -80°C until DNA extraction. Three biological replicates were performed on separate days.

In vitro sgRNA transcription and RNP transfection

sgRNAs were transcribed *in vitro* using the EnGen sgRNA Synthesis Kit (NEB), followed by DNAse treatment to remove the template. The sgRNAs were subsequently purified using the RNA Clean & Concentrator kit (Zymo Research). RNP complexes were formed by combining 1.5 µg (0.5 µL) EnGen SpCas9 NLS (NEB) with 1.5 µg sgRNA in Buffer 3.1 (NEB) in a total volume of 5 µL, and incubated for 20 min at room temperature. The mixture containing RNP complexes was used for the transfection of freshly isolated protoplasts as described previously. Protoplasts were incubated in the dark for 24 h at 25°C, pelleted and kept at -80°C until genomic DNA extraction.

Genomic DNA isolation and amplicon sequencing

Protoplast DNA was purified 24h after transfection in 96-well format using magnetic beads (NucleoMag Plant, Macherey-Nagel), following the manufacturer's instructions. DNA was eluted in 50 µL, of which 6 µL was subsequently used as a template in 25 µL PCR reactions using PHUSION HotStart Flex DNA polymerase (NEB) to amplify genomic DNA fragments containing target or predicted off-target sites. For the PCR, an initial denaturation for 30 s at 98 °C was followed by 38 cycles of denaturation for 10 s at 98 °C, annealing for 20 s at 58 °C, extension for 20 s at 72 °C, and a final extension step of 3 min at 72 °C. Primer sequences are listed in Supplementary Dataset 3. The resulting PCR products were visualized on a 2% agarose gel. Equal amounts of PCR products were pooled to obtain sequencing libraries. Libraries were subsequently column-purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel), following manufacturer's instructions. Illumina HiSeq sequencing (paired-end, 2x150 bp reads) was performed commercially by Eurofins Genomics Europe Sequencing GmbH, Constance, Germany.

Sequence Analysis

Paired sequencing reads were uploaded to the CLC Genomics Workbench v20, trimmed, merged and demultiplexed using standard settings. Mutation frequencies at target and predicted off-target sites were determined using Amplican (Labun et al., 2019). Amplicons amplified from the genomic DNA of mock-transfected protoplasts were used as controls to normalize on- and off-target mutation frequencies. Mutagenic spectra were determined using a custom Python script for filtering the AmpliCan output and is available from the authors upon request. Large insertions were aligned to the tomato genome and plasmid used for transfection using HiSat2 (Kim et al. 2019).

Results

sgRNA selection

We selected 29 members of the MYB (MYELOBLASTOMA) transcription factor gene family as targets. For the selected genes, sgRNAs and potential off-target sites with a maximum of 4 mismatches were predicted using CRISPOR (Concordet and Haeussler 2018). For each gene, two to six perfectly matching sgRNAs were selected, including at least one with off-target sequences having three or more mismatches ("specific") and one with one or two mismatches (and 0 mismatches in case of an alternative PAM, "non-specific"). A total of 89 sgRNAs as well as 213 of their respective predicted off-target sites (of which 68 in another *MYB* gene) were selected for the screen. Additionally, we considered potential off-target sites forming DNA or RNA bulges when hybridizing with sgRNAs. These sites are not predicted by CRISPOR and were therefore predicted using Cas-OFFinder (Bae et al. 2014). Several predicted off-target sites were selected where a bulge could not be resolved by an annealing alternative that would result in only a mismatch. This added 11 potential off-target sites to our screen, resulting in a total of 224. An overview of all selected off-target sites can be found in Supplementary Dataset 1.

Potential off-target sites with 0 (with an alternative PAM) to 4 mismatches to the sgRNA were selected. The alternative NGA or NAG Cas9 PAMs are functional at a low frequency in mammalian cells (Cho et al. 2014; Tsai et al. 2015; Hsu et al. 2013) and in rice (Meng et al. 2017). To further assess to which extent these non-canonical PAMs allow DNA cleavage in plants, we selected several sites that had either an NGA or NAG PAM (Figure 1a). Additionally, the position of the mismatch within the spacer might influence the likelihood of cleavage, resulting in an off-target mutation. Therefore, off-target site sequences with mismatches distributed over the length of the spacer were selected (Figure 1b). Notably, 20 sites were selected that had only a single mismatch in the 12 bp most proximal to the PAM (the so-called "seed sequence").

a

mismatches	NGG	NAG	NGA
0	-	1	1
1	34 (11)	3	6
2	29	4	8
3	66	12	11
4	48	0	1

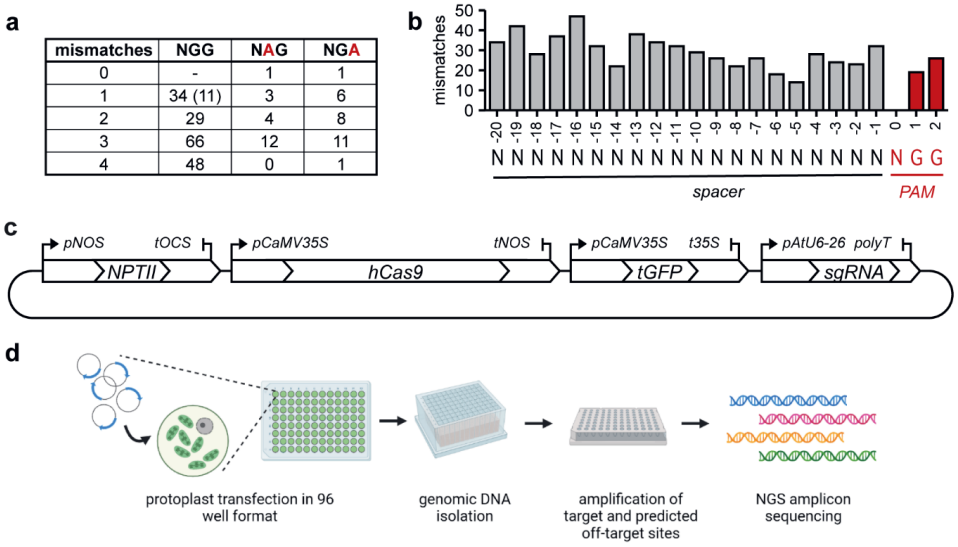


Figure 1 - Overview of selected off-target sites and experimental setup. (a) number of mismatches and PAM usage in the selected off-target sites. Number between brackets indicates the amount of predicted off-target sites that have an additional insertion or deletion compared to the gRNA in addition to a mismatch. At these sites, sgRNAs can only anneal by formation of an RNA or DNA bulge. (b) Distribution of all mismatches in selected predicted off-target sites over the spacer and PAM. (c) Graphical overview of the plasmid used for protoplast transfection. (d) Overview of the experimental steps.

CRISPR/Cas9-mediated mutagenesis of tomato protoplasts

Annealed oligonucleotides representing the spacers of the selected sgRNAs were directly ligated in a Golden Gate Level 2 vector containing *SpCas9* and *turboGFP* expression cassettes. This arrangement ensured the simultaneous delivery of all components in equimolar amounts (Figure 1c). Thus, the percentages of protoplasts to which *Cas9* was delivered could be determined by fluorescence microscopy and these were found to be similar for the three replicates (41±16%, 41±19% and 42±10%). A graphical overview of the subsequent experimental steps is shown in Figure 1d. Genomic DNA was isolated after 24 hrs, and regions containing the target sites and predicted off-target sites were amplified with barcoded primers and sequenced after pooling. Per biological replicate, approximately 7 million pooled paired-end reads were obtained, with a median of approximately 18,200 per sample (average approximately 21,500). The reads were trimmed, merged and demultiplexed and the frequency and type of CRISPR-induced mutations at target and predicted off-target sites were determined.

Correlation between predicted and actual editing efficiencies

We did not *a priori* select sgRNAs for predicted efficiency, and a wide range of mutation frequencies was obtained (Figure 2a). For only 8 targets no mutations could be detected at frequencies at least 0.1% above their wildtype control. The experimentally obtained mutation frequencies were subsequently compared to the predictions as expressed by the Doench score, currently also known as the Azimuth score (Fusi et al. 2015; Doench et al. 2016), and by the Moreno-Mateos score, also known as CrisprScan (Moreno-Mateos et al. 2015). We divided the sgRNAs into quartiles based on their prediction score. For the Azimuth score, experimentally obtained mutation frequencies of sgRNAs in the third and fourth quartiles (high score) were significantly increased compared to those in the first and second quartiles (Figure 2b). This indicates that the Azimuth score has some, albeit weak, predictive power. No correlation was found between the Moreno-Mateos score and experimentally determined mutation frequencies (Figure 2c).

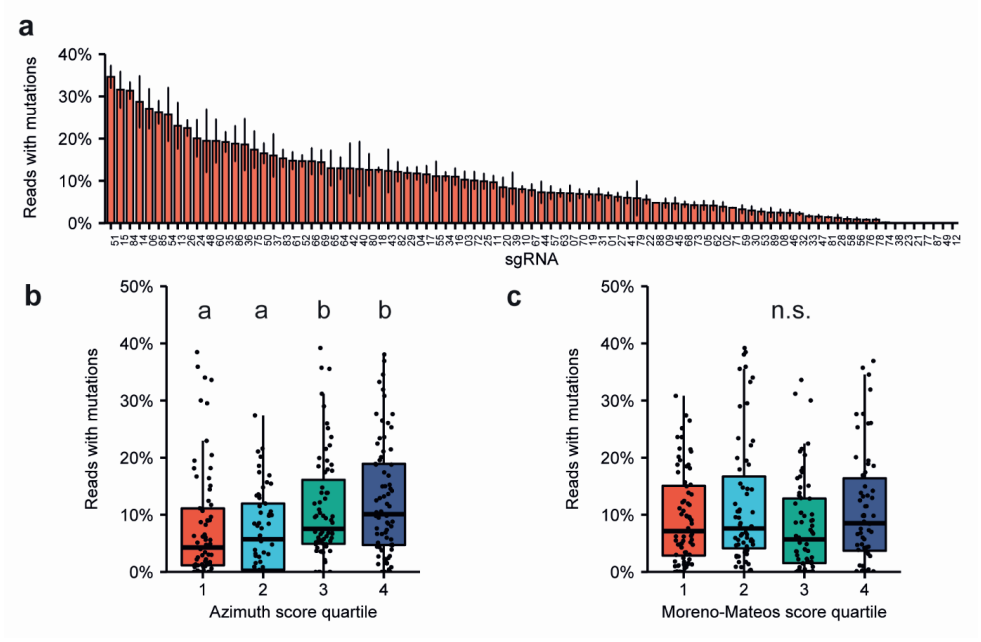


Figure 2 - On-target mutagenesis efficiency and correlation with predicted editing efficiency. (a) On-target mutation frequencies obtained for the 89 sgRNAs. Error bars indicate SE (n = 3). (b) and (c) Distribution of editing efficiencies. The sgRNAs were divided in quartiles based on their Azimuth score (b) or Moreno-Mateos score (c). The horizontal line in the box plots indicate the median, boxes represent the 2nd and 3rd quartile, and bottom and top whiskers indicate the 1st and 4th quartile, respectively. Significant differences between mutation frequencies of sgRNAs in each quartile were determined by Kruskal-Willis test, followed by Wilcoxon's Rank Sum test for pairwise comparisons.

Mutation types and the predictability of 1 bp insertions following DSB repair

We quantified the types of mutations found at the (on-)target sites (Figure 3a). As seen previously, deletions occurred at a slightly higher frequency than insertions (62%±5% and 38%±5%, respectively). Most insertions were only 1 bp in size, making that the most common single event, although larger insertions did occur at much lower frequencies. Most deletions were small-sized, and frequencies declined with increasing deletion size. However, larger deletions up to 145 bp were found.

As the 1 bp insertion was the single most favoured mutation induced by Cas9, we investigated this mutation type further. Although it is still often believed that Cas9 introduces breaks with blunt ends, multiple reports state that staggered cuts with 1-3 bp 5' overhangs frequently occur (Zuo and Liu 2016; Shou et al. 2018). These overhangs can be filled in by templated DNA synthesis, leading to two blunt ends that can be ligated, resulting in a 1 bp insertion, or more if the overhang is larger (Figure 3b). Thus, in the case of templated insertion, the inserted nucleotide can be predicted from the identity of the -4 position relative to the PAM. This has been shown for yeast (Lemos et al. 2018) and mammalian cells (Allen et al. 2019; Shen et al. 2018; Shi et al. 2019), but not yet for plants.

We investigated all on-target sequencing reads containing a 1 bp insertion. Strikingly, we found a bias for A and T insertions, as opposed to C and G insertions (Figure 3c). Additionally, for every sgRNA, we determined whether the most frequently found insertion at the target site corresponded to the predicted insertion based on the identity of the base at the -4 position. For this analysis, we disregarded sgRNAs that had no or very few (<100) sequencing reads with an identified 1 bp insertion. For only 38 out of the analysed 72 sgRNAs, the most frequently found insertion was equal to the predicted insertion, i.e. a duplication of the -4 base (Supplementary Dataset 4).

Next, all sequencing reads belonging to sgRNAs with the same nucleotide at position -4 were grouped. We then compared the frequency of A, T, C and G insertions within these groups to the frequencies of the same insertion in the other groups (e.g., the frequency of an A insertion when A is at the -4 position versus the frequency of an A insertion when T, C or G is at the -4 position) (Figure 3d). This revealed a bias towards duplication of the -4 nucleotide for each of the four nucleotides. Interestingly, only for gRNAs that were predicted to result in an A or T insertion was the most frequently observed insertion equal to the predicted insertion. Although C and G insertions occur more frequently for gRNAs with that base at the -4 position compared to the complete set of targets, the most frequently occurring inserted nucleotide was still A or T. We investigated if the bias could be caused by a higher likelihood of Cas9 inducing staggered cuts if the nucleotide at the -4 position

is an A or T. This would lead to a higher apparent percentage of 1 bp insertions out of total mutations for sgRNAs with an A or T at the -4 position. However, upon closer inspection of our data, no such significant difference could be found (Supplementary Figure 1).

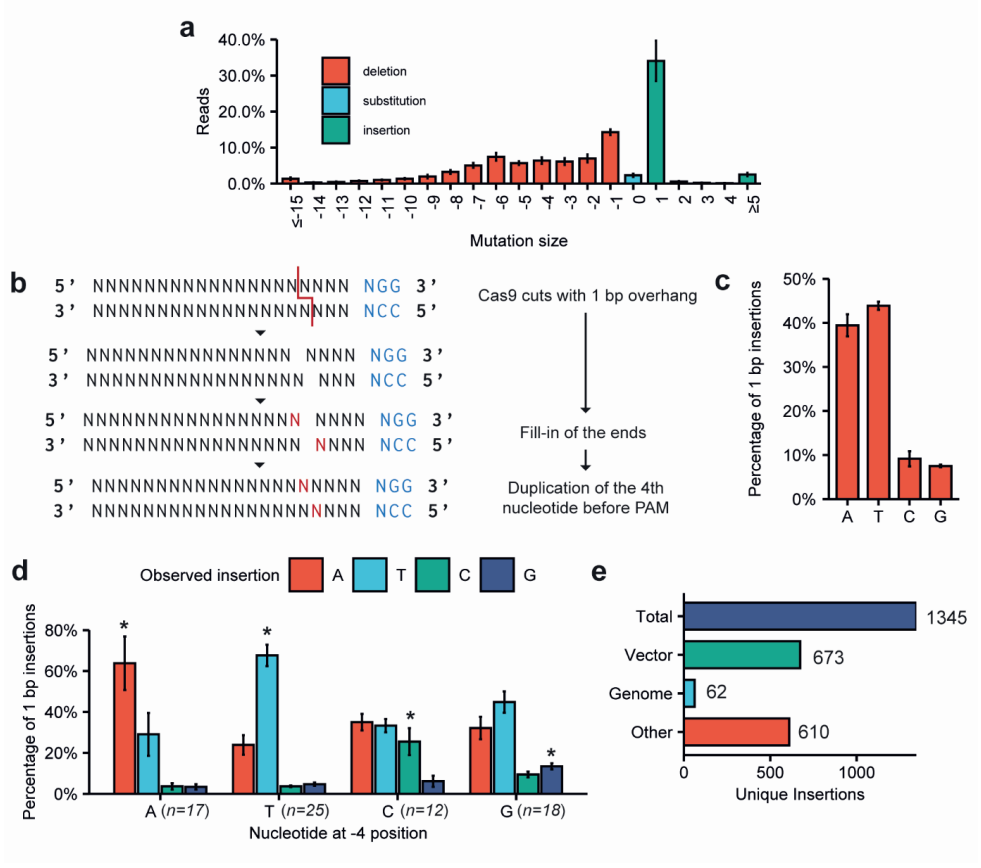


Figure 3 - Characterization of mutations found at target sites. (a) The compiled spectrum of mutations obtained for all on-target mutations induced by the 89 sgRNAs. The frequency of a mutation with the indicated size was calculated as the percentage of the total number of mutated reads. Error bars indicate SD (n = 3). (b) Cas9-induced staggered cuts with a 1 bp 5' overhang can be repaired by templated synthesis and subsequent ligation of the blunt DNA ends. This results in a duplication of the 4th nucleotide before the PAM. (c) Frequency of A, T, C and G insertions at target sites for all sgRNAs. Percentages are calculated by dividing the amount of reads with a certain 1 bp insertion by the total number of reads with a 1 bp insertion. Error bars indicate SD (n = 3). (d) Percentage of 1 bp insertions in which an A, T, C or G is inserted, observed for targets with an A, T, C or G at the -4 position. Error bars indicate SD (n = 3). For statistical analysis, the frequencies of A, T, C and G insertions were compared between the groups of sgRNAs with either an A, T, C or G at the -4 position. Statistical differences in the 4 groups were determined using one-way ANOVA. Pairwise comparisons were subsequently performed using Tukey's Honestly Significant Difference post-hoc test. Significant differences ($p < 0.05$) are indicated by asterisks (*), indicating that a specific insertion occurs more frequently with the particular -4 nucleotide than with any of the other three nucleotides at position -4. (f) Origin of large (≥11 bp) insertions. In total, 1345 unique insertion events (size ≥11bp) were identified. "Vector" indicates the plasmid used for transfection.

Origin of larger insertions at target sites

In our screen of merged paired reads we could also identify larger insertions of up to 133 bp at target sites. For the three biological replicates combined, we considered all identified insertions ≥11 bp and aligned these to both the tomato genome as well as to the sequence of the plasmid used for transfection (Figure 3e). In total, we found that 2.07% of the reads (8761 reads out of 422,323) with a mutation at an on-target site contained an insertion of ≥11 bp. These reads belonged to 1345 individual distinct events. Of these identifiable insertion events, 673 aligned to parts of the vector used for transfection and another 62 aligned to the tomato genome. For 610 insertion events, the origin of the insertion remains unknown (Figure 3e). An overview of all large insertions can be found in Supplementary Dataset 5.

Experimental identification of off-target sites

To identify off-target mutations, 224 predicted off-target sites were amplified and analyzed for the presence of InDels using AmpliCan (Labun et al. 2019). Amplicons produced from mock-transfected protoplasts were used as controls for normalizing mutation frequencies. To reliably identify true off-target mutations, excluding any sequences that might result from PCR or sequencing errors, we disregarded off-target sites for which no homogeneous control (no Cas9 added) sequences were obtained. Because it might interfere with the identification of genuine mutations, all sites for which the control showed an apparent mutation frequency over 0.1% were disregarded, which was the case for 30 sites. The remaining sites were considered to be a confirmed off-target site if the average mutation frequency was 0.1% or higher and if the mutagenic spectrum showed a pattern expected for CRISPR mutations (e.g., InDels instead of substitutions, as we considered the latter to be more likely due to polymerase or sequencing errors). Mutation patterns at the identified off-target sites can be found in Supplementary Figure 2.

At 18 of the remaining 194 analyzed sites, corresponding to 13 sgRNAs, mutations above this threshold were identified (Figure 4a, Supplementary Figure 2 for mutation patterns). Off-target activity mostly occurred at sites that had only one mismatch compared to the target (13 out of 42 tested sites), but was also found for three sites (out of 39) that had a 2 bp mismatch with the guide (Figure 4b). Interestingly, some off-target activity was identified for one site that did not contain any mismatches in the spacer, but had an alternative PAM (NGA instead of NGG, Figure 4a,b). No off-target activity was found at 89 and 49 sites that had 3 or 4 mismatches compared to the target site, respectively.

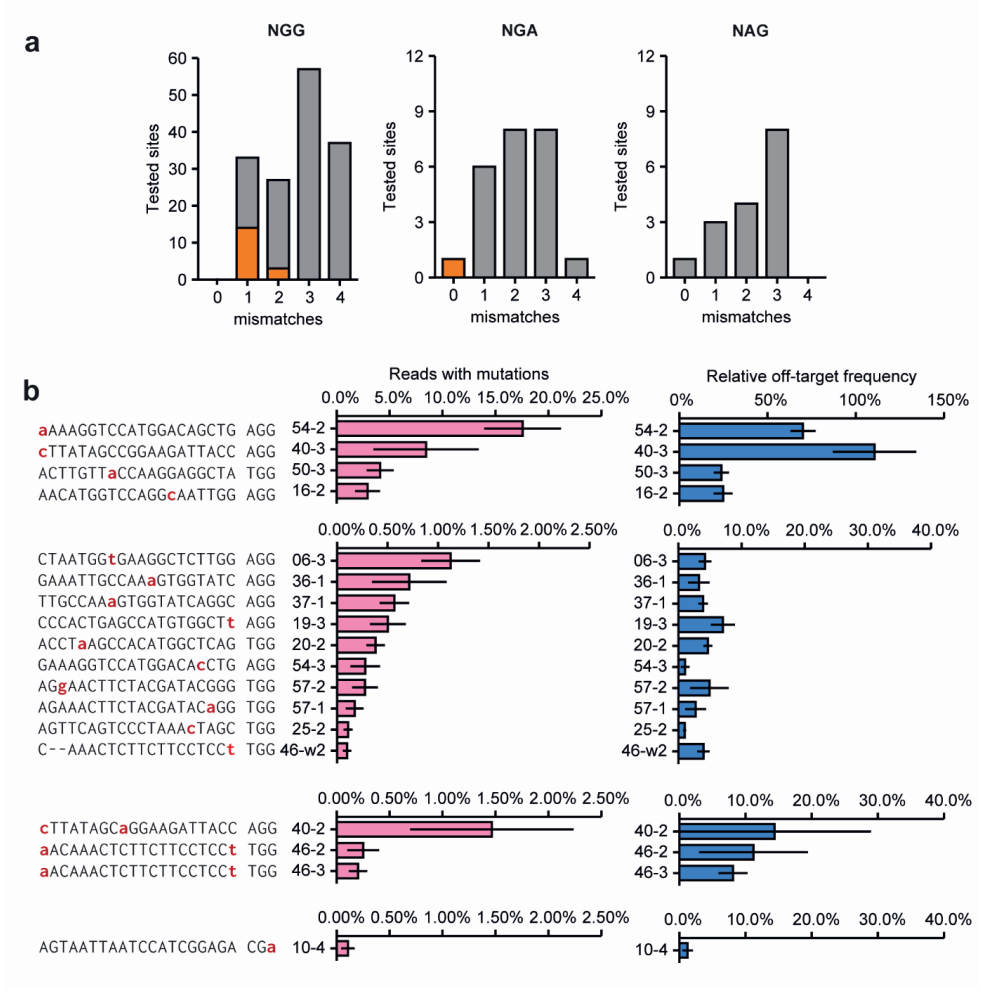


Figure 4 - Overview of experimentally confirmed off-target sites identified by screening 226 predicted off-target sites for the presence of mutations. (a) Overview showing the number and characteristics of screened predicted off-target sites with 1-4 mismatches compared to the target site sequence (grey bars). Orange indicates the number of experimentally identified off-target sites. (b) Mutation frequencies and relative off/on-target frequencies at 18 identified off-target sites that were found to be mutated. Sequencing data from mock-transfected protoplasts was used to normalize mutation frequencies. Red nucleotides indicate mismatches to the gRNA. Off-target mutations were found at sites that had 1 or 2 mismatches to the target site, or made use of an alternative PAM. Relative off-target frequencies as a measure of how often a genome will contain both a target and an off-target mutation. Percentages are calculated by dividing the off-target mutation frequency by the on-target mutation frequency. Error bars indicate SE (n = 3).

The two sites with the most frequently occurring off-target mutations (sites 54-2 and 40-3) had a single mismatch which occurred at the nucleotide most distal from the PAM (Figure 4b). Interestingly, off-target effects were only found for sites with two mismatches when one was located at the nucleotide most distal from the PAM (Figure 4b). This indicates that this nucleotide in particular adds very little to specificity and a mismatch here will most likely result in off-target

activity. Apart from the two off-target sites mentioned above, off-target mutation frequencies were generally low, never reaching more than 5% of total reads, and in most cases no more than 1% of total reads for the respective sites (Figure 4b).

Rather than the absolute mutation frequencies at off-target sites, the ratio of off-target mutation frequency relative to the on-target mutation frequency is likely to be more relevant for practical situations. Relative off-target frequencies varied widely (Figure 4b – note the different scales). Again, these were highest for the two off-target sites that had a single mismatch most distal from the PAM, even reaching a mutation frequency similar to the on-target site for sgRNA40.

Frequencies of off-target mutations produced by pre-assembled ribonucleoproteins

It is sometimes proposed that the use of ribonucleoproteins (RNPs) might decrease the frequency of off-target mutations. For RNP transfection, purified Cas9 protein preloaded with *in vitro* transcribed or synthetic sgRNAs are introduced in the protoplast.

To compare the (relative) off-target mutation frequencies for the two approaches, we synthesized the 13 sgRNAs for which we found off-target activity in our first screen (Figure 4b) by *in vitro* transcription and loaded them onto purified Cas9 protein. We again transfected protoplasts in a 96 wells format with either plasmid DNA or the corresponding RNPs. Transfection efficiencies for the three biological replicates of plasmid transfections were similar (56±7%, 64±3% and 62±1%). We amplified the on-target sites as well as the previously identified off-target sites and subjected the resulting amplicons to next-generation sequencing. After processing of the reads, the absolute and on-target and absolute and relative off-target mutation frequencies were determined.

Overall, the on-target mutation frequency of plasmid-transfected protoplast was approximately 50% higher than in the first experiment (Figures 2-4), matching the higher transfection efficiency in the latter experiment. The on-target mutation frequencies for plasmid- and RNP-transfected protoplasts were comparable for each target, with less active guides tending to give higher frequencies with plasmid transfection (Figure 5a). Overall, the on-target mutation frequencies for the two methods correlated well (Figure 5b). For most of the tested off-target sites, the use of RNPs either did not significantly affect the mutation frequency or decreased the mutation frequency (Figure 5c). There were two notable exceptions: off-target sites 36-1 and 40-2. At these sites, the RNP complex induced significantly more InDels than Cas9 and sgRNA expressed by the plasmid.

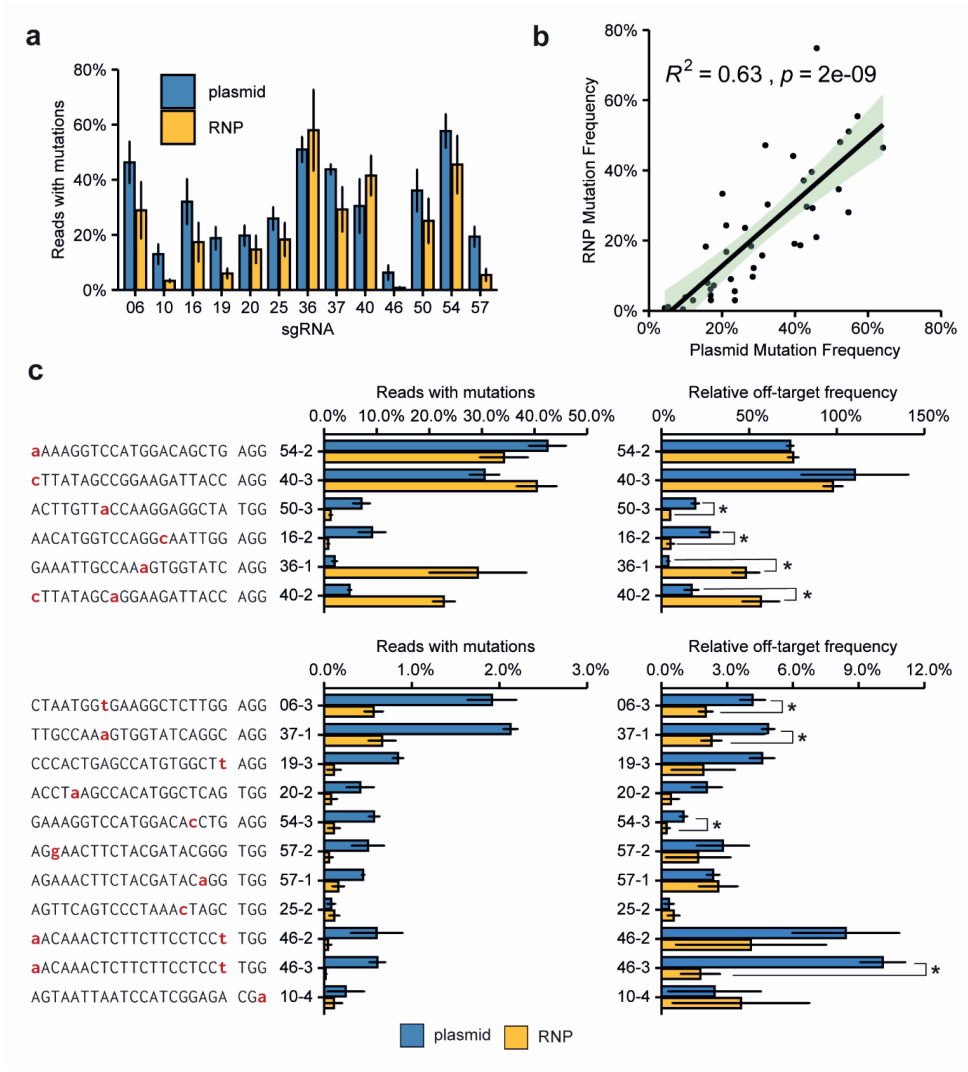


Figure 5 - Comparison of plasmid and RNP mediated on-target mutagenesis and off-target mutation frequencies. (a) Mutation frequencies at target sites. A total of 15.000 protoplasts was transfected with either 2.5 µg plasmid DNA or 1.5 µg Cas9 enzyme, loaded with 1.5 µg of in vitro synthesized sgRNA. Error bars indicate SE (n = 3). (b) Pearson correlation of mutation frequencies induced by plasmid-encoded Cas9 or RNPs. The green-shaded area indicates the 95% confidence interval. (c) Mutation frequencies and relative mutation frequencies at off-target sites. Nucleotides indicated in red indicate a mismatch to the gRNA. Relative off-target frequencies were determined by dividing the mutation frequency at the off-target site by the mutation frequency at the on-target site. Error bars indicate SE (n = 3). Asterisks (*) indicate significant differences ($p < 0.05$), as determined by two-sided Student's T Test.

Differences in on- and off-target mutation frequencies between these two types of transfection might simply be explained by varying final amounts of Cas9 protein and sgRNA present in the cells. Therefore, as discussed above for between-sgRNA comparisons, a fairer comparison between

plasmid- and RNP-based transfections is the comparison of the ratio of on/off-target mutation frequencies (Figure 5c). The use of RNPs significantly reduced the relative off-target mutation frequency for 6 out of 17 sites. For the two previously mentioned sites 36-1, and 40-2, however, RNPs led to significantly higher relative frequencies. For these two sites, the off-target mutation frequency induced by RNPs was significantly higher than for plasmids (Figure 5c, left panel). Although the on-target efficiency was also higher for RNPs than for plasmids, this difference was much smaller (Figure 5b). For 9 sites, no significant difference in the ratio of on/off-target mutation frequencies was found.

The use of *in vitro* transcribed sgRNAs in RNPs leads to low-frequency incorporation of residual dsDNA template

It was expected that the use of RNPs instead of plasmid DNA for mutagenesis would reduce the risk of integration of non-host DNA fragments in CRISPR-induced DSBs, as no plasmid DNA is introduced in the cells. To test this, we first determined the spectrum of CRISPR mutations induced by plasmid-encoded Cas9 and RNPs at on-target sites (Figure 6a). As expected, RNP-induced mutations were very similar to those from plasmid-encoded Cas9, but the percentage of reads that contained insertions larger than 5 bp, although overall low, differed substantially (Figure 6b). A significantly higher percentage of reads from plasmid-transfected protoplasts contained larger insertions, while only very few reads from RNP-transfected protoplasts contained an insertion of 32 bp or larger. In this experiment, we found that for plasmid-transfected protoplasts, 0.86% of on-target reads contained an insertion larger than or equal to 11 bp, pertaining to 181 individual events. For RNP-transfected protoplasts, this number was much lower, at 0.14% and 41 insertion events.

As these large insertion events still occurred, even for RNP-transfected protoplasts albeit at a lower frequency, we again attempted to identify their origin. For the plasmid-transfected protoplasts, 94 insertion events out of 181 could be aligned to vector DNA. Another 6 could be aligned to the tomato genome – for the remaining 81 large insertion events, the origin remains unknown (Figure 6c). The relative frequencies for these origins are very similar to those observed in the larger initial experiment (Figure 3e). Interestingly, for the RNP-transfected protoplasts, 13 out of 41 insertion events could be traced back to the dsDNA template that was used to synthesize the sgRNA *in vitro* – even though the resulting sgRNA was treated with DNase. Another 21 events could be aligned to the tomato genome. For 7 events, the origin could not be determined (Figure 6c).

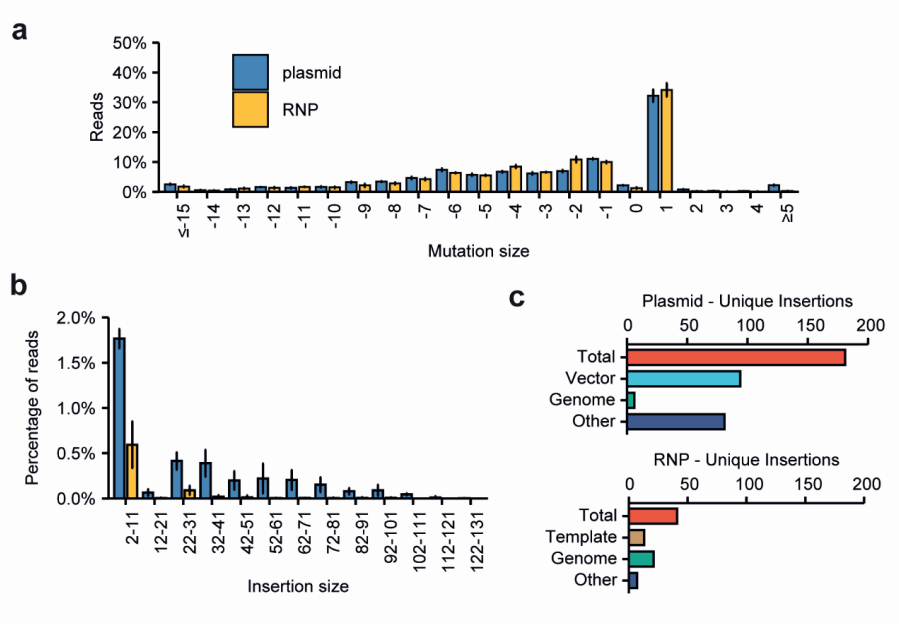


Figure 6 - Comparison of the effects of plasmid or RNP-mediated mutagenesis on the frequency and origin of large insertions found at target sites. (a) Spectrum of mutations induced by RNPs or plasmid-encoded Cas9. The frequency of an InDel of the indicated size was calculated as the percentage of the total number of mutated reads. Error bars indicate SD (n = 3). (b) Comparison of the prevalence of insertions of indicated sizes. Percentages are calculated by dividing the number of reads per bin by the total number of mutated reads. Error bars indicate SD (n = 3). (c) The origin of large (≥11 bp) deletions for plasmid and RNP-transfected protoplasts. "Vector" indicates that the origin of the inserted DNA is the plasmid used for transfection. "Template" indicates that the origin of the inserted DNA is the dsDNA template used for in vitro sgRNA synthesis.

Discussion and Conclusion

In recent years, the CRISPR/Cas9 system has proven itself to be a valuable tool for genome editing in tomato and other plant species (Zhu et al. 2020; Gao 2021). However, information about the occurrence of off-target activity in plants, in contrast to animal cells, is limited (Sturme et al. 2022). In this work, we performed a large CRISPR/Cas9 mutagenesis screen with 89 sgRNAs in tomato protoplasts to characterize on-target mutations and quantify off-target mutations. The acquired data give us more insight into the spectrum of mutations formed following Cas9 action and help provide guidelines to prevent off-target action in mutagenesis experiments.

Working with tomato protoplasts allowed us to significantly increase the speed and throughput of our experimental work, in comparison to mutagenesis in regenerating plants. We selected sgRNAs that targeted the coding sequences of 29 genes in the *SlMYB* transcription factor gene family, in order to simulate a functional screen of MYB functions in tomato. On- and selected off-target mutation frequencies were determined using next-generation sequencing of amplicons. Amplicon sequencing, yielding thousands of sequences per interrogated site, offers improved sensitivity over other methods used to determine mutation frequencies, such as restriction-enzyme polymorphism analysis, T7 endonuclease I assays and Sanger sequencing (Nadakuduti et al. 2019; Kosicki et al. 2017). However, targeted sequencing of preselected amplicons holds some limitations. Off-target mutations at sites that are not predicted by prediction algorithms will not be identified. Moreover, small amplicons of a maximum length of 270 bp were used for sequencing, meaning that large deletions that disrupt primer binding sites could not be detected.

We first investigated mutations induced by Cas9 at target sites. During the selection of the sgRNAs, predicted efficiency was not considered and thus a wide range of mutation efficiencies was acquired (Figure 2a). We attempted to correlate measured efficiency to the efficiency as predicted by the Azimuth score (formerly known as Doench score) and Moreno-Mateos score (Fusi et al. 2015; Doench et al. 2016; Moreno-Mateos et al. 2015). For the Azimuth score, the predicted efficiency had some value (Figure 2b), but no correlation could be found for the Moreno-Mateos score (Figure 2c). Generally, this indicates that a high predicted efficiency score is not a guarantee for a well-performing sgRNA, and vice versa. A possible explanation for the different performances of these algorithms might be that the Azimuth model is trained on data acquired from human cells expressing sgRNAs under control of a U6 promoter and by lentiviral transduction, whereas the Moreno-Mateos model is further trained on data acquired from zebrafish embryos. In these zebrafish, mutations were induced using RNPs. Our experimental setup, using plasmid-encoded

Cas9 and sgRNAs, more resembles the Azimuth setup. It is also not known to what extent the mutation efficiency is influenced by the genomic context, e.g. the chromatin structure around the target, which could affect the predictive value of the scores.

Further investigation of the types of mutations acquired at target sites revealed that deletions occurred at a slightly higher frequency than insertions. Nevertheless, the single most frequently occurring mutation was a 1 bp insertion (Figure 3a). As Cas9 frequently induces a DSB with a 1 bp overhang, which may be filled in by DNA polymerases before the loose DNA ends are joined again, these insertions may be predictable (Zuo and Liu 2016; Shou et al. 2018; Shi et al. 2019). We investigated this for the first time on a large scale in plant cells. Overall, A and T insertions occurred at much higher frequencies than C and G insertions (Figure 3b, N. Zhang et al. 2020). Although we were able to find a bias in repair for all -4 nucleotides, C or G insertions, when predicted, were still superseded by A or T insertions. In yeast however, more than 75% of insertions were of the predicted base (Lemos et al. 2018), and one study in mammalian cells found that in 149 out of 151 the predicted base was the most frequently occurring insertion (van Overbeek et al. 2016). Our findings indicate that in plants, repair pathways promoting A or T insertions are more frequent than in yeast and mammalian cells. Although the mechanism leading to duplication of the -4 base in mammals and yeast thus appears to be also acting in plant cells, an additional, more dominant mechanism leading to A or T insertion, exists. The existing bias might be utilized to design mutagenesis experiments resulting in specific 1 bp insertions, even if not 100% efficient, for example for the repair of an interrupted open reading frame.

It is known that DNA fragments may integrate in double stranded DNA breaks, probably through templated synthesis following the annealing of one of the free 3' DNA ends to other DNA (Salomon and Puchta 1998; Köhler et al. 1989; Gorbunova and Levy 1997). We were able to identify mutant alleles, such as large insertions, that occur at low frequencies of around 2% of total mutated reads. Although the integration of such fragments of foreign DNA can be considered undesirable in the creation of mutant plants, these events can easily be identified and avoided by routine screening of the on-target mutation by sequencing. Additionally, these findings add to evidence that NHEJ can be harnessed to specifically insert DNA fragments at target sites, as previously described (J. J. Li et al. 2016; Lu et al. 2020; Dong et al. 2020), making this a useful alternative to Homology-Directed Repair (HDR) for insertion of new sequences.

Apart from gaining a better understanding of the mutational spectrum of Cas9, we also aimed to assess the range and frequency of off-target mutations. For this purpose, we selected and

investigated 224 predicted off-target sites that had 0 to 4 mismatches to the sgRNA and used either the canonical NGG PAM, or an alternative NAG or NGA PAM. The majority of these sites occurred at positions that had only one mismatch with the sgRNA, indicating that off-target sites that only have 1 mismatch are high-risk. For two of the aforementioned off-target sites, the relative mutation frequency was very high – reaching up to 70% of the on-target mutation frequency for 54-2 and up to 110% for 40-3 (Figure 4b). Interestingly, both these off-target sites had a mismatch at the nucleotide most distal from the PAM. As previously reported, a mismatch at this position leads to a high risk of the off-target being cleaved by Cas9 (Fu et al. 2013; Hsu et al. 2013). Relative mutation frequencies for the other identified off-target sites with 1 mismatch were generally lower, rarely reaching above 10%, indicating that the majority of genomes will only contain mutations at the target site. Additionally, we found off-target activity for three sites that had two mismatches to the target. For all three sites, one mismatch was present at the most distal position, again indicating that this nucleotide adds very little to overall specificity. The last off-target site was identified at a position that had no mismatches in the spacer but made use of an alternative NGA PAM. Although cleavage activity at such sites has already been shown in mammalian cells (Cho et al. 2014; Tsai et al. 2015; Hsu et al. 2013), and for a different non-canonical NAG PAM in rice (Meng et al. 2017), activity at NGA sites had not yet been reported in plants. However, both the off-target mutation frequency and relative off-target mutation frequency at this site were very low.

We have also shown that off-target activity can occur at loci that have an insertion or deletion in comparison to the sgRNA, which can still anneal through the formation of an RNA or DNA bulge. Although our screen was limited in this respect, our results indicate that particularly the formation of an RNA or DNA bulge at the end of the spacer might allow cleavage, even in the presence of another mismatch elsewhere in the spacer (Figure 4b). This is not surprising, as it is known that truncated sgRNAs as short as 17 nt can still result in efficient cleavage of the target site (Fu et al. 2014). Popular sgRNA prediction algorithms such as CRISPR-P 2.0 (Lei et al. 2014; Liu et al. 2017) and CRISPOR (Haeussler et al. 2016) do not take off-target sites with RNA or DNA bulges into account. Therefore, it is advisable to additionally check for this type of off-target sites using one that does, such as Cas-OFFinder (Bae et al. 2014).

We found off-target activity for mismatches distributed over the complete spacer (Figure 4b). This shows that a mismatch in the proposed 7-12 bp “seed” region (Semenova et al. 2011; Jinek et al. 2012; Cong et al. 2013) directly adjacent to the PAM does not abolish mutagenic activity completely and that such sites should still be assessed for off-target activity, or avoided. This has previously

already been concluded for mammalian cells (Pattanayak et al. 2013; Hsu et al. 2013) and similar results have been published for *in planta* mutations (Raitskin et al. 2019).

We could not identify any off-target activity at positions that had three or four mismatches compared to the target site. One study in Arabidopsis did show cleavage activity at a site that had three mismatches compared to the spacer (Q. Zhang et al. 2017), as did a study in rice (Tang et al. 2018). One possible explanation for this difference is that in these studies, stable transformation was used, resulting in a prolonged exposure of the genomic DNA to Cas9 cleavage. Thus, chances of off-target activity occurring in stably transformed plants may be further decreased by selecting sgRNAs with only predicted off-target sites with four or more mismatches.

While T-DNA containing Cas9 and sgRNAs may be readily segregated out from self-compatible true-breeding cultivars or parent lines of tomato, T-DNA insertion and subsequent removal by segregation is not an option for plants that require vegetative propagation. For these plants, transient expression or transfecting protoplasts with RNP complexes provide a DNA-free mutagenesis system. Additional advantages of the use of RNPs are that the risk of vector-derived insertions in DSBs is negated, and, as RNP complexes might be degraded faster than plasmid DNA, the time in which the genome is exposed to endonuclease activity is reduced. We have shown that RNPs are a viable alternative to plasmid transfection and can induce high mutation frequencies in tomato protoplasts (Figure 5a, 5b). We also determined relative off-target frequencies at off-target sites for both methods but conclude that overall, there is no bias for more or less frequent off-target activity with either method.

Summarizing, we can propose several rules to achieve specific genome editing in tomato. To remove the risk of off-target mutations occurring altogether, sgRNAs can be chosen that only have predicted off-target sites with four or more mismatches to the target site using off-target prediction software such as Cas-OFFinder (Bae et al. 2014). However, the selection of such sgRNAs might not be possible for every experiment. Relative off-target mutation frequencies indicate that off-target mutations can generally be avoided in stably transformed plants, even if only a one bp mismatch compared to the target is present. Possible exceptions are cases where this mismatch occurs very distal from PAM. If no highly specific sgRNA can be selected, testing off-target sites with up to three mismatches is advisable to mitigate the risk. Unintended on-target events, such as plasmid integration or large deletions, can be avoided by diligently sequencing the target and high-risk off-target sites and ensuring that all alleles have been identified.

In conclusion, we have collected useful data on the specificity and mutational spectrum of 89 sgRNAs for Cas9 in tomato protoplasts. In addition to providing information about the pattern of mutations frequently caused by Cas9-mediated mutagenesis, we have presented evidence for the predictability of 1 bp insertions *in planta*. We screened predicted endogenous off-target sites and found evidence for off-target activity, of which the majority had only one mismatch compared to the spacer. Off-target activity at sites with three or four mismatches compared to the target site was not found. Thus sgRNAs with only such predicted off-target sites can generally be considered safe. Finally, we have shown that off-target sites with insertions or deletions compared to the target site do pose a risk for off-target mutations, especially if the resulting DNA or RNA bulges are formed at the end of the spacer. This risk may be mitigated by utilizing dedicated sgRNA prediction algorithms that take these types of off-target sites into account. Overall, these results can help sgRNA design and aid in fine-tuning mutagenesis experiments to specific desired outcomes.

Author Contributions

ES and RdM designed the experiments, with input from GA. ES and ML performed the experiments. ES analysed data. ES wrote the manuscript, RdM and GA reviewed and edited. RdM and GA supervised the study.

Acknowledgements

This research is supported by the Dutch Research Council (NWO), and is funded by the Ministry of Infrastructure and Water Management. Marc Youles from The Sainsbury Laboratory is kindly acknowledged for providing the CRISPR-Pink plasmid for easy sgRNA cloning. We are grateful to Jana Murovec for providing us with a practical protocol for next-generation amplicon sequencing.

Conflict of interest

The authors declare no conflict of interest.

References

Allen, F., L. Crepaldi, C. Alsinet, A.J. Strong, V. Kleshchevnikov, P. De Angeli, P. Páleníková, … L. Parts. 2019. "Predicting the Mutations Generated by Repair of Cas9-Induced Double-Strand Breaks." *Nature Biotechnology* 37 (1): 64–82. https://doi.org/10.1038/nbt.4317.

Bae, S., J. Park, and J.-S. Kim. 2014. "Cas-OFFinder: A Fast and Versatile Algorithm That Searches for Potential off-Target Sites of Cas9 RNA-Guided Endonucleases." *Bioinformatics* 30 (10): 1473–75. https://doi.org/10.1093/bioinformatics/btu048.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D.A. Romero, and P. Horvath. 2007. "CRISPR Provides Acquired Resistance against Viruses in Prokaryotes." *Science* 315 (March): 1709–12. https://doi.org/10.1126/science.1138140.

Bessoltane, N., F. Charlot, A. Guyon-Debast, D. Charif, K. Mara, C. Collonnier, P.-F. Perroud, M. Tepfer, and F. Nogué. 2022. "Genome-Wide Specificity of Plant Genome Editing by Both CRISPR–Cas9 and TALEN." *Scientific Reports* 12 (1): 1–10. https://doi.org/10.1038/s41598-022-13034-2.

Brouns, S.J.J., M.M. Jore, M. Lundgren, E.R. Westra, R.J.H. Slijkhuis, A.P.L. Snijders, M.J. Dickman, K.S. Makarova, E. V. Koonin, and J. Van Der Oost. 2008. "Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes." *Science* 321 (5891): 960–64. https://doi.org/10.1126/science.1159689.

Chen, K., Y. Wang, R. Zhang, H. Zhang, and C. Gao. 2019. "CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture." *Annual Review of Plant Biology* 70: 667–97. https://doi.org/10.1146/annurev-arplant-050718-100049.

Cho, S.W., S. Kim, Y. Kim, J. Kweon, H.S. Kim, S. Bae, and J.S. Kim. 2014. "Analysis of Off-Target Effects of CRISPR/Cas-Derived RNA-Guided Endonucleases and Nickases." *Genome Research* 24 (1): 132–41. https://doi.org/10.1101/gr.162339.113.

Concordet, J.P., and M. Haeussler. 2018. "CRISPOR: Intuitive Guide Selection for CRISPR/Cas9 Genome Editing Experiments and Screens." *Nucleic Acids Research* 46 (W1): W242–45. https://doi.org/10.1093/nar/gky354.

Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, … F. Zhang. 2013. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science* 339 (6121): 819–23. https://doi.org/10.1126/science.1231143.

Doench, J.G., N. Fusi, M. Sullender, M. Hegde, E.W. Vaimberg, K.F. Donovan, I. Smith, … D.E. Root. 2016. "Optimized SgRNA Design to Maximize Activity and Minimize Off-Target Effects of CRISPR-Cas9." *Nature Biotechnology* 34 (2): 184–91. https://doi.org/10.1038/nbt.3437.

Dong, O.X., S. Yu, R. Jain, N. Zhang, P.Q. Duong, C. Butler, Y. Li, … P.C. Ronald. 2020. "Marker-Free Carotenoid-Enriched Rice Generated through Targeted Gene Insertion Using CRISPR-Cas9." *Nature Communications* 11 (1): 1–10. https://doi.org/10.1038/s41467-020-14981-y.

Engler, C., R. Kandzia, and S. Marillonnet. 2008. "A One Pot, One Step, Precision Cloning Method with High Throughput Capability." *PLoS ONE* 3 (11). https://doi.org/10.1371/journal.pone.0003647.

Engler, C., M. Youles, R. Gruetzner, T.M. Ehnert, S. Werner, J.D.G. Jones, N.J. Patron, and S. Marillonnet. 2014. "A Golden Gate Modular Cloning Toolbox for Plants." *ACS Synthetic Biology* 3 (11): 839–43. https://doi.org/10.1021/sb4001504.

Feng, Z., Y. Mao, N. Xu, B. Zhang, P. Wei, D.-L. Yang, Z. Wang, … J.-K. Zhu. 2014. "Multigeneration Analysis Reveals the Inheritance, Specificity, and Patterns of CRISPR/Cas-Induced Gene Modifications in Arabidopsis." *Proceedings of the National Academy of Sciences of the United States of America* 111 (12): 4632–37. https://doi.org/10.1073/pnas.1400822111.

Fu, Y., J.A. Foden, C. Khayter, M.L. Maeder, D. Reyon, J.K. Joung, and J.D. Sander. 2013. "High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells." *Nature Biotechnology* 31 (9): 822–26. https://doi.org/10.1038/nbt.2623.

Fu, Y., J.D. Sander, D. Reyon, V.M. Cascio, and J.K. Joung. 2014. "Improving CRISPR-Cas Nuclease Specificity Using Truncated Guide RNAs." *Nat Biotechnol* 32 (3): 279–84. https://doi.org/10.1038/nbt.2808.

Fusi, N., I. Smith, J. Doench, and J. Listgarten. 2015. "In Silico Predictive Modeling of CRISPR/Cas9 Guide Efficiency." *BioRxiv*, 021568. https://doi.org/10.1101/021568.

Gao, C. 2021. "Genome Engineering for Crop Improvement and Future Agriculture." *Cell* 184 (6): 1621–35. https://doi.org/10.1016/j.cell.2021.01.005.

Garneau, J.E., M.È. Dupuis, M. Villion, D.A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A.H. Magadán, and S. Moineau. 2010. "The CRISPR/Cas Bacterial Immune System Cleaves Bacteriophage and Plasmid DNA." *Nature* 468 (7320): 67–71. https://doi.org/10.1038/nature09523.

Gorbunova, V., and A. Levy. 1997. "Non-Homologous DNA End Joining in Plant Cells Is Associated with Deletions and Filler DNA Insertions." *Nucleic Acids Research* 25 (22): 4650–57.

Graham, N., G.B. Patil, D.M. Bubeck, R.C. Dobert, K.C. Glenn, A.T. Gutsche, S. Kumar, … P.L. Morrell. 2020. "Plant Genome Editing and the Relevance of Off-Target Changes1[OPEN]." *Plant Physiology* 183 (4): 1453–71. https://doi.org/10.1104/pp.19.01194.

Haeussler, M., K. Schönig, H. Eckert, A. Eschstruth, J. Mianné, J.-B. Renaud, S. Schneider-Maunoury, … K. Yusa. 2016. "Evaluation of Off-Target and on-Target Scoring Algorithms and Integration into the Guide RNA Selection Tool CRISPOR." *Genome Biology* 17 (1): 148. https://doi.org/10.1186/s13059-016-1012-2.

Hsu, P.D., D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, Y. Li, … F. Zhang. 2013. "DNA Targeting Specificity of RNA-Guided Cas9 Nucleases." *Nature Biotechnology* 31 (9): 827–32. https://doi.org/10.1038/nbt.2647.

Jaganathan, D., K. Ramasamy, G. Sellamuthu, S. Jayabalan, and G. Venkataraman. 2018. "CRISPR for Crop Improvement: An Update Review." *Frontiers in Plant Science* 9 (July): 1–17. https://doi.org/10.3389/fpls.2018.00985.

Jin, J., F. Tian, D.C. Yang, Y.Q. Meng, L. Kong, J. Luo, and G. Gao. 2017. "PlantTFDB 4.0: Toward a Central Hub for Transcription Factors and Regulatory Interactions in Plants." *Nucleic Acids Research* 45 (D1): D1040–45. https://doi.org/10.1093/nar/gkw982.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Kim, D., J.M. Paggi, C. Park, C. Bennett, and S.L. Salzberg. 2019. "Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype." *Nature Biotechnology* 37 (8): 907–15. https://doi.org/10.1038/s41587-019-0201-4.

Köhler, F., G. Cardon, M. Pöhlman, R. Gill, and O. Schieder. 1989. "Enhancement of Transformation Rates in Higher Plants by Low-Dose Irradiation: Are DNA Repair Systems Involved in the Incorporation of Exogenous DNA into the Plant Genome?" *Plant Molecular Biology* 12 (2): 189–99. https://doi.org/10.1007/BF00020504.

Kosicki, M., S.S. Rajan, F.C. Lorenzetti, H.H. Wandall, Y. Narimatsu, E. Metzakopian, and E.P. Bennett. 2017. *Dynamics of Indel Profiles Induced by Various CRISPR/Cas9 Delivery Methods*. *Progress in Molecular Biology and Translational Science*. Vol. 152. Elsevier Inc. https://doi.org/10.1016/bs.pmbts.2017.09.003.

Kwon, C.T., J. Heo, Z.H. Lemmon, Y. Capua, S.F. Hutton, J. Van Eck, S.J. Park, and Z.B. Lippman. 2020. "Rapid Customization of Solanaceae Fruit Crops for Urban Agriculture." *Nature Biotechnology* 38 (2): 182–88. https://doi.org/10.1038/s41587-019-0361-2.

Labun, K., X. Guo, A. Chavez, G. Church, J.A. Gagnon, and E. Valen. 2019. "Accurate Analysis of Genuine CRISPR Editing Events with AmpliCan." *Genome Research* 29 (5): 843–47. https://doi.org/10.1101/gr.244293.118.

Lee, K., Y. Zhang, B.P. Kleinstiver, J.A. Guo, M.J. Aryee, J. Miller, A. Malzahn, … K. Wang. 2018. "Activities and Specificities of CRISPR/Cas9 and Cas12a Nucleases for Targeted Mutagenesis in Maize." *Plant Biotechnology Journal*, 1–11. https://doi.org/10.1111/pbi.12982.

Lei, Y., L. Lu, H.Y. Liu, S. Li, F. Xing, and L.L. Chen. 2014. "CRISPR-P: A Web Tool for Synthetic Single-Guide RNA Design of CRISPR-System in Plants." *Molecular Plant* 7 (9): 1494–96. https://doi.org/10.1093/mp/ssu044.

Lemmon, Z.H., N.T. Reem, J. Dalrymple, S. Soyk, K.E. Swartwood, D. Rodriguez-Leal, J. Van Eck, and Z.B. Lippman. 2018. "Rapid Improvement of Domestication Traits in an Orphan Crop by Genome Editing." *Nature Plants* 4 (10): 766–70. https://doi.org/10.1038/s41477-018-0259-x.

Lemos, B.R., A.C. Kaplan, J.E. Bae, A.E. Ferrazzoli, J. Kuo, R.P. Anand, D.P. Waterman, and J.E. Haber. 2018. "CRISPR/Cas9 Cleavages in Budding Yeast Reveal Templated Insertions and Strand-Specific Insertion/Deletion Profiles." *Proceedings of the National Academy of Sciences of the United States of America* 115 (9): E2010–47. https://doi.org/10.1073/pnas.1716855115.

Li, J., H. Manghwar, L. Sun, P. Wang, G. Wang, H. Sheng, J. Zhang, … X. Zhang. 2018. "Whole Genome Sequencing Reveals Rare Off-Target Mutations and Considerable Inherent Genetic or/and Somaclonal Variations in CRISPR/Cas9-Edited Cotton Plants." *Plant Biotechnology Journal*, 858–68. https://doi.org/10.1111/pbi.13020.

Li, J.J., X. Meng, Y. Zong, K. Chen, H. Zhang, J. Liu, J.J. Li, and C. Gao. 2016. "Gene Replacements and Insertions in Rice by Intron Targeting Using CRISPR-Cas9." *Nature Plants* 2 (10): 16139. https://doi.org/10.1038/nplants.2016.139.

Liu, H., Y. Ding, Y. Zhou, W. Jin, K. Xie, and L.L. Chen. 2017. "CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants." *Molecular Plant* 10 (3): 530–32. https://doi.org/10.1016/j.molp.2017.01.003.

Lu, Y., Y. Tian, R. Shen, Q. Yao, M. Wang, M. Chen, J. Dong, … J.K. Zhu. 2020. "Targeted, Efficient Sequence Insertion and Replacement in Rice." *Nature Biotechnology* 38 (12): 1402–7. https://doi.org/10.1038/s41587-020-0581-5.

Makarova, K.S., N. V. Grishin, S.A. Shabalina, Y.I. Wolf, and E. V. Koonin. 2006. "A Putative RNA-Interference-Based Immune System in Prokaryotes: Computational Analysis of the Predicted Enzymatic Machinery, Functional Analogies with Eukaryotic RNAi, and Hypothetical Mechanisms of Action." *Biology Direct* 1: 1–26. https://doi.org/10.1186/1745-6150-1-7.

Meng, X., X. Hu, Q. Liu, X. Song, C. Gao, J. Li, and K. Wang. 2017. "Robust Genome Editing of CRISPR-Cas9 at NAG PAMs in Rice." *Sci China Life Sci*, no. 6: 122–25. https://doi.org/10.1073/pnas.1317360111.

Moreno-Mateos, M.A., C.E. Vejnar, J.D. Beaudoin, J.P. Fernandez, E.K. Mis, M.K. Khokha, and A.J. Giraldez. 2015. "CRISPRscan: Designing Highly Efficient SgRNAs for CRISPR-Cas9 Targeting in Vivo." *Nature Methods* 12 (10): 982–88. https://doi.org/10.1038/nmeth.3543.

Nadakuduti, S.S., C.G. Starker, D.K. Ko, T.B. Jayakody, J.H. Schiemann, and J. Kühn-institut. 2019. "Evaluation of Methods to Assess in Vivo Activity of Engineered Genome-Editing Nucleases in Protoplasts." *Frontiers in Plant Science* 10 (February): 1–13. https://doi.org/10.3389/fpls.2019.00110.

Nekrasov, V., C. Wang, J. Win, C. Lanz, D. Weigel, and S. Kamoun. 2017. "Rapid Generation of a Transgene-Free Powdery Mildew Resistant Tomato by Genome Deletion." *Scientific Reports* 7 (1): 482. https://doi.org/10.1038/s41598-017-00578-x.

Ossowski, S., K. Schneeberger, J.I. Lucas-Lledó, N. Warthmann, R.M. Clark, R.G. Shaw, D. Weigel, and M. Lynch. 2010. "The Rate and Molecular Spectrum of Spontaneous Mutations in Arabidopsis Thaliana." *Science* 327 (5961): 92–94. https://doi.org/10.1126/science.1180677.

Overbeek, M. van, D. Capurso, M.M. Carter, M.S. Thompson, E. Frias, C. Russ, J.S. Reece-Hoyes, … A.P. May. 2016. "DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks." *Molecular Cell* 63 (4): 633–46. https://doi.org/10.1016/j.molcel.2016.06.037.

Pattanayak, V., S. Lin, J.P. Guilinger, E. Ma, J.A. Doudna, and D.R. Liu. 2013. "High-Throughput Profiling of off-Target DNA Cleavage Reveals RNA-Programmed Cas9 Nuclease Specificity." *Nature Biotechnology* 31 (9): 839–43. https://doi.org/10.1038/nbt.2673.

Raitskin, O., C. Schudoma, A. West, and N.J. Patron. 2019. "Comparison of Efficiency and Specificity of CRISPR-Associated (Cas) Nucleases in Plants: An Expanded Toolkit for Precision Genome Engineering." *PLoS Genetics* 14 (2). https://doi.org/https://doi.org/10.1371/journal.pone.0211598.

Salomon, S., and H. Puchta. 1998. "Capture of Genomic and T-DNA Sequences during Double-Strand Break Repair in Somatic Plant Cells." *EMBO Journal* 17 (20): 6086–95. https://doi.org/10.1093/emboj/17.20.6086.

Semenova, E., M.M. Jore, K.A. Datsenko, A. Semenova, E.R. Westra, B. Wanner, J. Van Der Oost, S.J.J. Brouns, and K. Severinov. 2011. "Interference by Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) RNA Is Governed by a Seed Sequence." *Proceedings of the National Academy of Sciences of the United States of America* 108 (25): 10098–103. https://doi.org/10.1073/pnas.1104144108.

Sheen, J. 2002. "A Transient Expression Assay Using Arabidopsis Mesophyll Protoplasts," 1–6.

Shen, M.W., M. Arbab, J.Y. Hsu, D. Worstell, S.J. Culbertson, O. Krabbe, C.A. Cassa, D.R. Liu, D.K. Gifford, and R.I. Sherwood. 2018. "Predictable and Precise Template-Free CRISPR Editing of Pathogenic Variants." *Nature* 563 (7733): 646–51. https://doi.org/10.1038/s41586-018-0686-x.

Shi, X., J. Shou, M.M. Mehryar, J. Li, L. Wang, M. Zhang, H. Huang, X. Sun, and Q. Wu. 2019. "Cas9 Has No Exonuclease Activity Resulting in Staggered Cleavage with Overhangs and Predictable Di- and Tri-Nucleotide CRISPR Insertions without Template Donor." *Cell Discovery*. Nature Publishing Groups. https://doi.org/10.1038/s41421-019-0120-z.

Shirasawa, K., H. Hirakawa, T. Nunome, S. Tabata, and S. Isobe. 2016. "Genome-Wide Survey of Artificial Mutations Induced by Ethyl Methanesulfonate and Gamma Rays in Tomato." *Plant Biotechnology Journal* 14 (1): 51–60. https://doi.org/10.1111/pbi.12348.

Shou, J., J. Li, Y. Liu, and Q. Wu. 2018. "Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion." *Molecular Cell* 71 (4): 498-509.e4. https://doi.org/10.1016/j.molcel.2018.06.021.

Sturme, M.H.J., J.P. Van Der Berg, L.M.S. Bouwman, A. De Schrijver, R.A. De Maagd, G.A. Kleter, and E. Battaglia-De Wilde. 2022. "Occurrence and Nature of Off-Target Modifications by CRISPR-Cas Genome Editing in Plants." *ACS Agricultural Science and Technology* 2 (2): 192–201. https://doi.org/10.1021/acsagscitech.1c00270.

Tang, X., G. Liu, J. Zhou, Q. Ren, Q. You, L. Tian, X. Xin, … Y. Zhang. 2018. "A Large-Scale Whole-Genome Sequencing Analysis Reveals Highly Specific Genome Editing by Both Cas9 and Cpf1 (Cas12a) Nucleases in Rice." *Genome Biology* 19 (1): 84. https://doi.org/10.1186/s13059-018-1458-5.

Tsai, S.Q., Z. Zheng, N.T. Nguyen, M. Liebers, V. V Topkar, V. Thapar, N. Wyvekens, … J.K. Joung. 2015. "GUIDE-Seq Enables Genome-Wide Profiling of off-Target Cleavage by CRISPR-Cas Nucleases." *Nature Biotechnology* 33 (2): 187–97. https://doi.org/10.1038/nbt.3117.

Wang, X., M. Tu, Y. Wang, W. Yin, Y. Zhang, H. Wu, Y. Gu, Z. Li, Z. Xi, and X. Wang. 2021. "Whole-Genome Sequencing Reveals Rare off-Target Mutations in CRISPR/Cas9-Edited Grapevine." *Horticulture Research* 8 (1). https://doi.org/10.1038/s41438-021-00549-4.

Weber, E., C. Engler, R. Gruetzner, S. Werner, and S. Marillonnet. 2011. "A Modular Cloning System for Standardized Assembly of Multigene Constructs." Edited by J. Peccoud. *PLoS ONE* 6 (2): e16765. https://doi.org/10.1371/journal.pone.0016765.

Wehner, N., L. Hartmann, A. Ehlert, S. Böttner, L. Oñate-Sánchez, and W. Dröge-Laser. 2011. "High-Throughput Protoplast Transactivation (PTA) System for the Analysis of Arabidopsis Transcription Factor Function." *The Plant Journal* 68 (3): 560–69. https://doi.org/10.1111/j.1365-313X.2011.04704.x.

Young, J., G. Zastrow-Hayes, S. Deschamps, S. Svitashev, M. Zaremba, A. Acharya, S. Paulraj, … S. Kumar. 2019. "CRISPR-Cas9 Editing in Maize: Systematic Evaluation of Off-Target Activity and Its Relevance in Crop Improvement." *Scientific Reports* 9 (1). https://doi.org/10.1038/s41598-019-43141-6.

Zhang, N., H.M. Roberts, J. Van Eck, and G.B. Martin. 2020. "Generation and Molecular Characterization of CRISPR/Cas9-Induced Mutations in 63 Immunity-Associated Genes in Tomato Reveals Specificity and a Range of Gene Modifications." *Frontiers in Plant Science* 11 (February): 1–13. https://doi.org/10.3389/fpls.2020.00010.

Zhang, Q., H.-L. Xing, Zhi-Ping Wang, Hai-Yan Zhang, Fang Yang, Yun Zhou, Xue-Chen Wang, and, and Q.-J. Chen. 2017. "High-Frequency off-Target Mutagenesis Induced by CRISPR/Cas9 in Arabidopsis and Its Prevention by Improving Specificity of the Tools." *Plant Molecular Biology* 96 (4): 445–56. https://doi.org/10.1101/203489.

Zhu, H., C. Li, and C. Gao. 2020. "Applications of CRISPR–Cas in Agriculture and Plant Biotechnology." *Nature Reviews Molecular Cell Biology* 21 (11): 661–77. https://doi.org/10.1038/s41580-020-00288-9.

Zsögön, A., T. Čermák, E.R. Naves, M.M. Notini, K.H. Edel, S. Weinl, L. Freschi, D.F. Voytas, J. Kudla, and L.E.P. Peres. 2018. "De Novo Domestication of Wild Tomato Using Genome Editing." *Nature Biotechnology*, no. October. https://doi.org/10.1038/nbt.4272.

Zuo, Z., and J. Liu. 2016. "Cas9-Catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations." *Scientific Reports* 5 (October): 1–9. https://doi.org/10.1038/srep37584.

Supplementary Information

Supplementary Datasets

Five Supplementary Datasets belong to this chapter:

Supplementary Dataset 1 – Overview targets, off-targets and genes

Supplementary Dataset 2 – Oligonucleotide sequences for sgRNA cloning

Supplementary Dataset 3 – Barcoded primer sequences for target and off-target amplification

Supplementary Dataset 4 – Predicted and actual most frequent 1 bp insertions

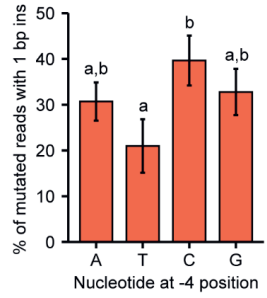
Supplementary Dataset 5 – Large insertions identified in this screen

They can be downloaded from the published version of this chapter at:

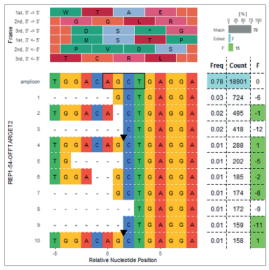
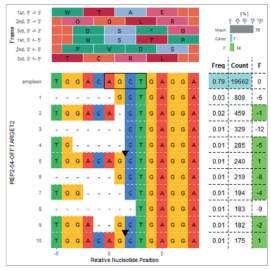
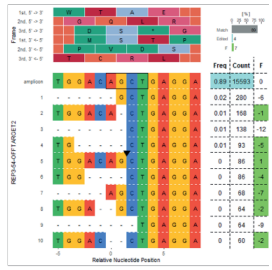
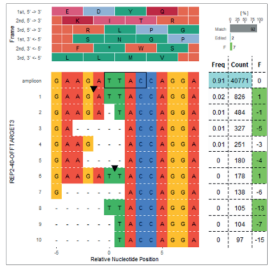
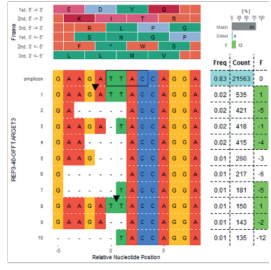
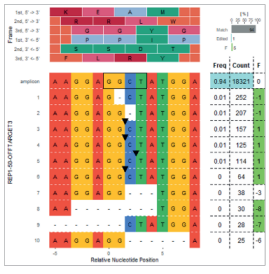
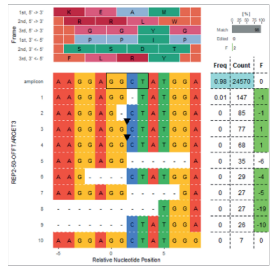
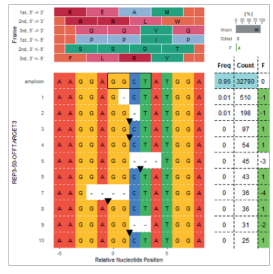
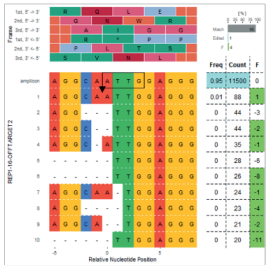
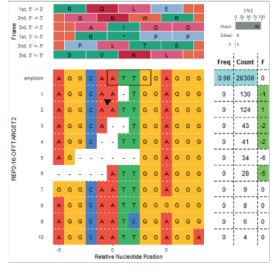
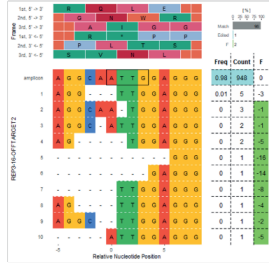
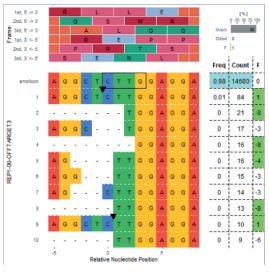
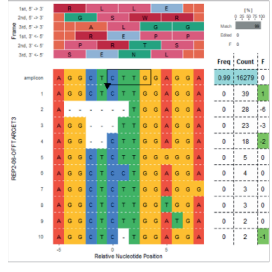
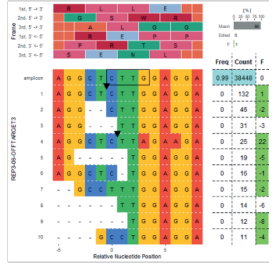
https://www.frontiersin.org/articles/10.3389/fgeed.2023.1196763/full#supplementary-material

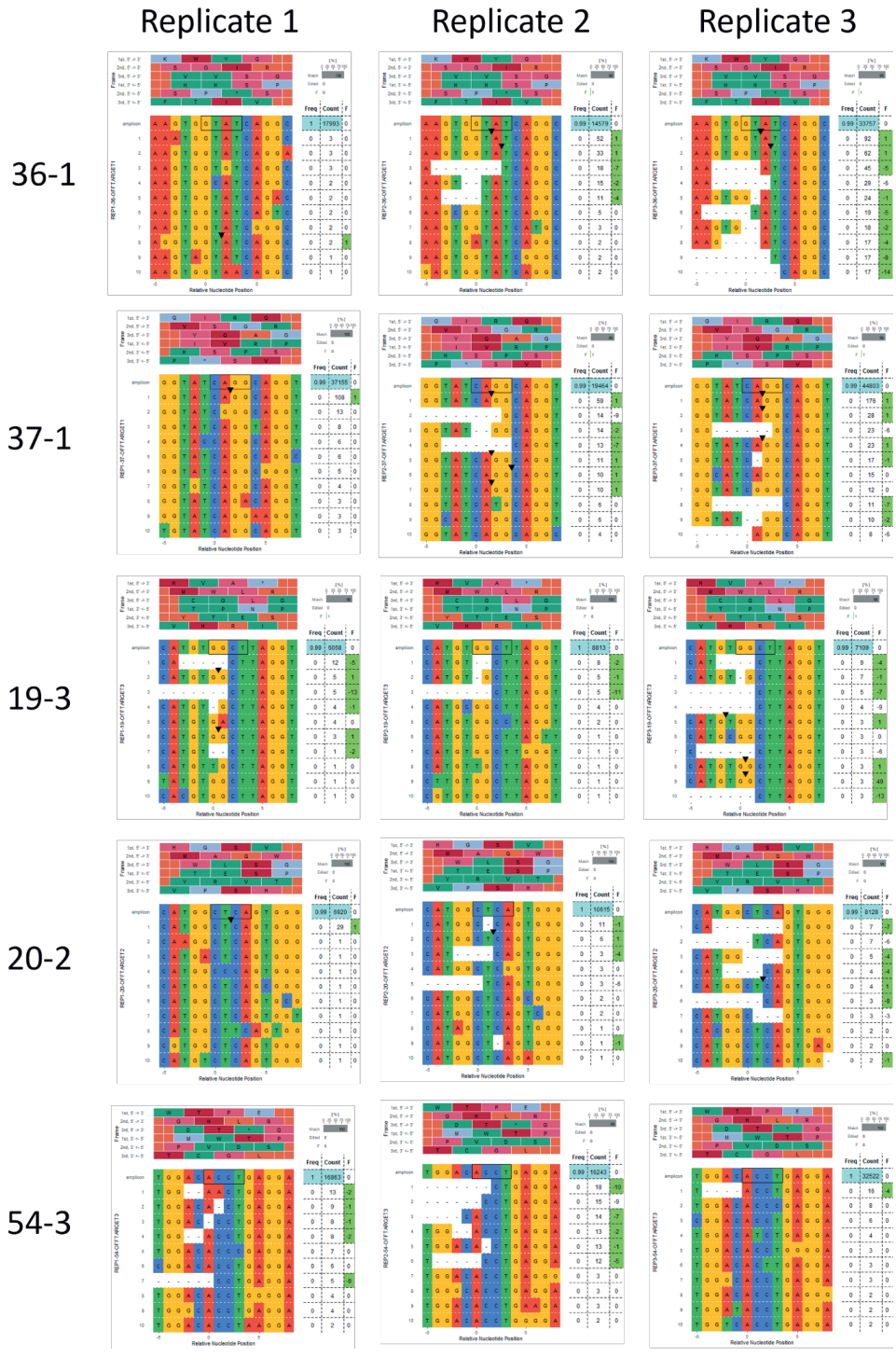
The raw sequencing data are available in the NCBI Sequencing Read Archive, BioProject accession number PRJNA975109.

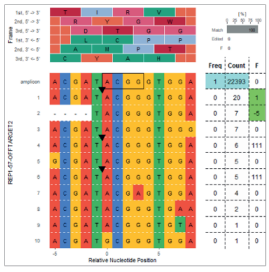
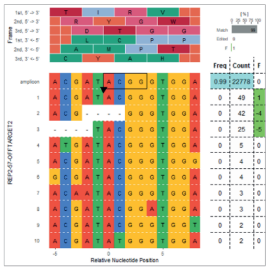
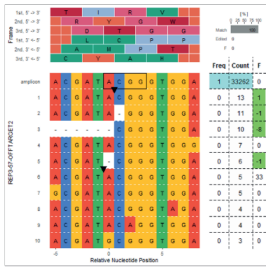
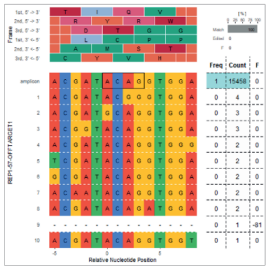
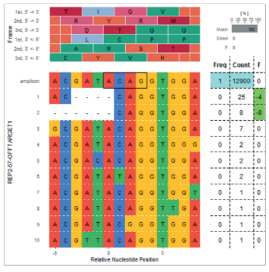
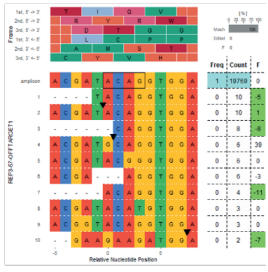
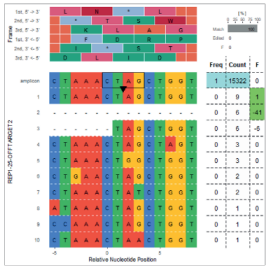
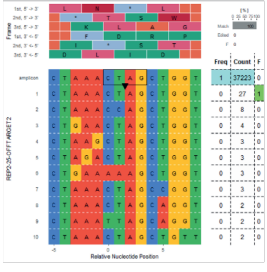
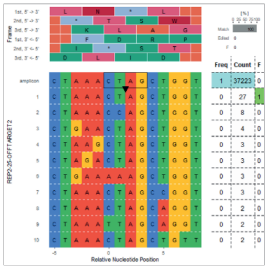
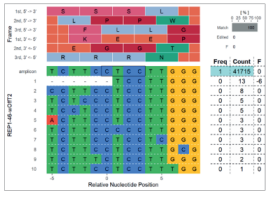
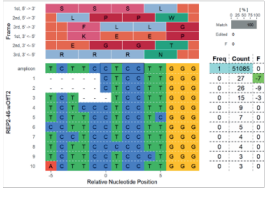
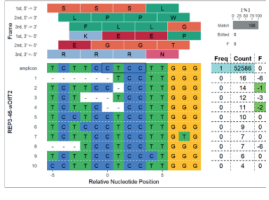
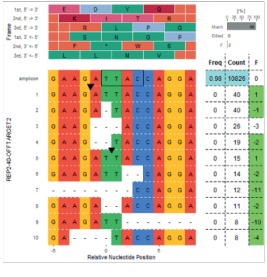
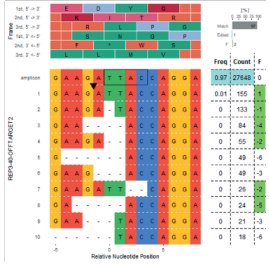
Supplementary Figures

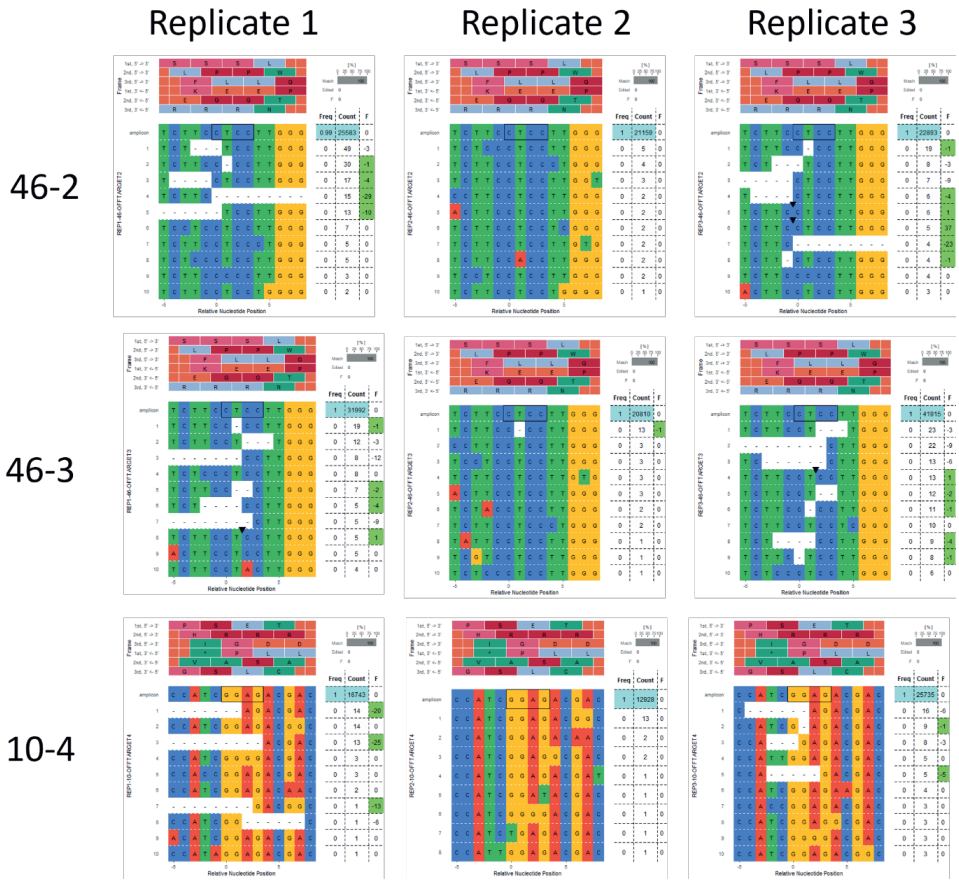


Supplementary Figure 1 - Influence of the nucleotide at the -4 position on the percentage of mutant reads with a 1 bp insertion. Different letters indicate statistically significant differences between groups, as determined by one-way ANOVA followed by a post-hoc Tukey test. Although A and T insertions occur more frequently overall, the presence of an A or T at the -4 positions does not result in a much higher frequency of 1 bp insertions. This indicates that staggered cuts do not occur more frequently if the -4 base is an A or T.

	Replicate 1	Replicate 2	Replicate 3
54-2			
40-3			
50-3			
16-2			
06-3			


Replicate 1
Replicate 2
Replicate 3
36-1
37-1
19-3
20-2
54-3

	Replicate 1	Replicate 2	Replicate 3
57-2			
57-1			
25-2			
46-w2			
40-2			



Supplementary Figure 2 - Variation plots of identified off-target sites. Variation plots for off-target sites were checked to ensure variation was not solely due to sequencing errors, which frequently result in substitutions.

Chapter 3

Comparison of Cas12a- and Cas9-mediated mutagenesis in tomato cells

E. Slaman[1,2], L. Kottenhagen[1], W. de Martines[1],

G. C. Angenent[1,2], R. A. de Maagd[2]

[1] Wageningen University & Research, Laboratory of Molecular Biology, Wageningen, the Netherlands

[2] Wageningen University & Research, Bioscience, Wageningen, the Netherlands

Abstract

The Cas12a nuclease is a promising addition to the CRISPR toolbox, offering versatility due to its TTTV-protospacer adjacent motif (PAM) and the fact that it induces double-stranded breaks (DSBs) with single-stranded overhangs. Using data from high-throughput amplicon sequencing experiments performed on tomato protoplasts, we optimized and characterized Cas12a-mediated genome editing in tomato. We showed that, of three tested variants, the *Lachnospiraceae* (Lb) Cas12a is the most efficient for genome editing in tomato and developed an easy-to-use and effective Golden-Gate-based system for crRNA cloning. Additionally, we compared LbCas12a to SpCas9 by investigating the on-target efficacy and specificity at 35 overlapping target sites and 57 (LbCas12a) or 100 (SpCas9) predicted off-target sites. We found LbCas12a an efficient, robust addition to SpCas9, with similar overall – though target-dependent – efficiencies, expanding the range of genomic targets that CRISPR methods may mutate. LbCas12a induced more and larger deletions than SpCas9, which may be advantageous for specific genome editing applications. Off-target activity for LbCas12a was found at 10 out of 57 investigated sites. In all cases, one or two mismatches were present in the last nine nucleotides of the spacer, distal from the PAM. We conclude that Cas12a-mediated genome editing is generally precise as long as such predictable off-target sites can be avoided. In conclusion, we have determined the mutation pattern and efficacy of Cas12a-mediated CRISPR mutagenesis in tomato and developed a cloning system for the routine application of Cas12a for tomato genome editing.

Introduction

The application of the CRISPR-Cas9 system in plants has enabled targeted mutagenesis with unprecedented speed and simplicity (Jaganathan et al. 2018; K. Chen et al. 2019; Zhu et al. 2020; Lemmon et al. 2018; Kwon et al. 2020; Zsögön et al. 2018). However, the number of mutable genomic targets is limited by the requirement for an "NGG" protospacer adjacent motif (PAM) for cleavage. A promising addition to the CRISPR toolbox is the Cas12a nuclease, which differs from Cas9 in several aspects (Zetsche et al. 2015). First, it requires a 5' "TTTV" PAM instead of the 3' "NGG" of Cas9. The alternative PAM may make it easier to find Cas12a target sites than Cas9 sites in A/T-rich genomic regions, such as promoters. Additionally, Cas12a induces double-stranded breaks (DSBs) with a 4-5 bp overhang, in contrast to the blunt breaks induced by Cas9. These overhangs ("sticky ends") may prove helpful in targeted integration approaches. Recently, they were shown to be advantageous for achieving precise integrations in the genomes of mammalian cells through a combined mechanism of microhomology-mediated end joining (MMEJ) and homology-directed repair (HDR) (Sage and Geijsen 2022). Furthermore, Cas12a CRISPR-RNAs (crRNAs) only need a short, 21-36 bp direct repeat 5' of the spacer to provide the correct structure to the crRNA for proper loading in the nuclease. Cas9 needs two RNAs: a crRNA and a transactivating-crRNA, often combined in a single guide RNA (sgRNA) with a combined length of ~100 bp (Jinek et al. 2012). Finally, Cas12a is also a ribonuclease able to process its CRISPR arrays, allowing the application of such arrays for multiplexing (Fonfara et al. 2016; Zetsche et al. 2016)

Cas12a variants from *Acidaminococcus, Francisella novicida,* and *Lachnospiraceae* (AsCas12a, FnCas12a, and LbCas12a, respectively) were shown to reliably induce mutations in mammalian cell lines, and Cas12a was quickly adopted as an efficient genome editing tool (Hur et al. 2016; Y. Kim et al. 2016; H. K. Kim et al. 2017). As an added benefit, Cas12a seemed to induce fewer off-target mutations than Cas9 (Kleinstiver et al. 2016; D. Kim et al. 2016). In plants, however, the nuclease was less readily applied. Early reports of the application of Cas12a for rice genome editing – the first plant species reported to be edited using Cas12a - revealed low editing efficiencies (Hu et al. 2017; Xu et al. 2017). Editing efficiencies were subsequently improved by using specific methods for crRNA expression and increasing Cas12a nuclease activity (Tang et al. 2017; Schindele and Puchta 2020; M. Wang et al. 2018).

Tomato is an economically important crop, as well as a model species for research on fleshy fruits. Cas9-mediated mutagenesis has been readily and frequently applied to tomato and was used to study, among other traits, plant architecture, fruit development, and (a)biotic stress tolerance(Xia

et al. 2021; Chandrasekaran et al. 2021). However, only a few reports using Cas12a have been published (Vu et al. 2020; 2021; Bernabé-Orts et al. 2019). The slow and limited adoption of Cas12a might be due to limited data on the performance of Cas12a in the tomato genome, and to the absence of an efficient, easy-to-use cloning system for crRNA expression.

The components needed for CRISPR-Cas mutagenesis in tomato are often delivered to the plant through *Agrobacterium tumefaciens*-mediated transformation (van Roekel et al. 1993). Although effective, regenerating stably transformed plants through tissue culture is laborious and, therefore, not particularly suitable for the optimization of CRISPR-mediated genome editing techniques. Consequently, we focused our efforts on protoplasts. In Chapter 2, we developed a method for 96 well-format protoplast transfections and coupled this to next-generation amplicon sequencing to study the characteristics and specificity of CRISPR-Cas9-mediated genome editing in tomato. In this work, we used a similar approach to compare multiplex crRNA expression strategies and developed an efficient, easy-to-use Golden Gate-based system for crRNA expression. Additionally, we compared Cas12a and Cas9 for efficacy, mutational pattern, and specificity on a set of overlapping targets. To achieve this, we selected 35 overlapping target sites for Cas9 and Cas12a in the *bHLH* transcription factor gene family and determined on- and off-target mutations for the corresponding crRNAs and sgRNAs. We found Cas12a a reliable and robust addition to Cas9 genome editing. Additionally, our study revealed that Cas12a preferentially induces more and larger deletions than Cas9 – a trait that may be useful when specific mutational outcomes are desired. These data pave the way for the routine application of Cas12a in mutagenesis experiments in tomato.

Materials and Methods

Selecting target sites and off-target sites

For the initial Cas12a optimization experiments, CRISPOR (Concordet and Haeussler 2018) was used to identify Cas12a target sites in the first exons of the tomato *PHYTOENE DESATURASE (PDS)* gene (*Solyc03g123760*).

To identify overlapping target sites in transcription factor gene families, coordinates from all exons that are part of coding sequences were extracted from the ITAG4.0_gene_models.gff file, obtained from solgenomics.net (grep -w "CDS" ITAG4.0_gene_models.gff). The resulting file was converted to a BED-file, and corresponding DNA sequences were extracted from the ITAG4.0 tomato genome build using BEDtools. Using a list of transcription factors obtained from the Plant Transcription Factor Database (Guo et al. 2008) and a regular expression describing the target sites, overlapping target sites were identified in coding sequences for transcription factors using a Python script. All off-target sites for the identified target sites with a maximum of 3 mismatches and a maximum of one-nucleotide DNA/RNA bulge were predicted using CasOFF-Finder (S. Bae et al. 2014), for both enzymes. Thirty-five target sites in the *bHLH* gene family with predicted off-target sites for both Cas12a and Cas9 were selected for further testing. Primers for amplifying on-target sequences and a selection of predicted off-target sequences were designed using BatchPrimer3 (Ali and Al-Koofee 2019) in DNA sequences surrounding the target and predicted off-target sites extracted from the ITAG4.0 tomato genome build using BEDtools.

Vector construction

All vector assembly was done with Golden Gate cloning, using parts from the MoClo toolkit (Weber et al. 2011) (Addgene #1000000044) and from the MoClo Plant Parts kit (Engler et al. 2014) (Addgene #1000000047), unless otherwise described. Used primer sequences can be found in Supplementary Dataset 3. Schematic overviews of the cloned plasmids can be found in Supplementary Figure 1.

Plasmids encoding human codon-optimized AsCas12a, FnCas12a, and LbCas12a containing a nuclear localization signal and 3xHA tag at the 3'end were gifts from the Feng Zhang lab (Zetsche et al. 2015). They were obtained through Addgene (accession numbers 69982, 69988, and 69976, respectively). The nucleases were amplified with primers adding flanking *Bpi*I sites (Supplementary Dataset 3) and subsequently inserted in the level 0 vector for coding sequences, pICH41308, using

restriction-ligation. The coding sequences were then combined with the CaMV35S promoter and NOS terminator (pICH51288 and pICH41421, respectively) in pICH47742.

To create vector backbones for crRNA expression, a Cas9-based CRISPR-Pink cassette was used as a basis (a gift from Marc Youles, The Sainsbury Laboratory). The AtU6-26 promoter was amplified using a reverse primer adding the direct repeat sequence for either AsCas12a, FnCas12a, or LbCas12a, with either the mature or the pre-crRNA sequence, and a flanking *Bsa*I site introducing an overhang to allow seamless cloning to the CRISPR-Pink RFP operon. This RFP was then amplified with primers adding *Bsa*I sites with compatible overhangs to fuse this part to the AtU6-26 promoter with direct repeat sequence. The two amplicons were then combined into level 1, position 1 to 7 backbone vectors (pICH47732, pICH47742, pICH47751, pICH47761, pICH47772, pICH47781, and pICH47791) using restriction-ligation to create the final crRNA expression cassettes. Primer sequences can be found in Supplementary Dataset 3.

For our initial crRNA expression optimization experiments, we selected three target sites in tomato *SlPDS (Solyc03g123760)*, designed and annealed oligonucleotides (Supplementary Dataset 2), and ligated these into the previously constructed crRNA expression cassettes, following the protocol as described in Supplementary Information 2. These crRNA expression vectors were then combined with pICSL7004 (*NPTII*), the constructed *AsCas12a*, *FnCas12a,* or *LbCas12a* expression vector, a *tGFP* marker (a combination of pICH41414, pICH51288, and pICH41414 in pICH47751) and end-linker pICH41822 in pICSL4723 to form binary multiplexing level 2 vectors. Additionally, arrays encoding the three selected crRNAs each transcribed from their own AtU6-26 promoter in both pre-crRNA form and mature form were synthesized for all nucleases (GenScript, sequences can be found in Supplementary Information 1). These arrays were subsequently cloned to a level 1, position 4 backbone (pICH47761) and again combined with pICSL7004, the nuclease, a *tGFP* marker, and end linker pICH41780 into pICSL4723 to form level 2 binary vectors.

To clone the vector expressing the mature LbCas12a crRNA array using a PolII promoter, the array was amplified with primers adding overhangs to allow cloning into a level 0 vector for coding sequences (pICH41308). The array was subsequently combined with a CsVMV promoter (pICSL12006) and MAS terminator (pICH77901) into a level 1, position 4 backbone (pICH47761). The crRNA expression cassette was combined with NPTII, LbCas12a and tGFP into a binary level 2 vector as described above.

For the expression system using ribozymes, the LbCas12a crRNA array was amplified and subsequently cloned in pGEM-T Easy (Promega), according to the manufacturer's instructions. This

allowed it to function as a level -1 part. The Hepatitis Delta Virus (HDV) and Hammerhead (HH) ribozymes were amplified from Addgene plasmid #86197, which was a gift from Tang et al Nature Plants 2017, and similarly cloned to pGEM-T Easy (Promega). The three parts were subsequently combined in a level 0 vector for coding sequences (pICH41308), subsequently combined with the CsVMV promoter and MAS terminator, and next combined with *NPTII*, *LbCas12a,* and *tGFP* as described above.

For later experiments with the 35 overlapping target sites, we constructed level 2 backbones in which a single crRNA or sgRNA could easily be inserted. A schematic overview can be found in Supplementary Figure 1c. For the LbCas12a variant, a thermotolerant, Arabidopsis codon-optimized LbCas12a was used, which was a gift from the Puchta lab (Schindele and Puchta 2020). Modifications were made to this ttLbCas12a to include two 5' SV40 nuclear localization signals. A potato IV2 intron was added after the second NLS to prevent bacterial expression of the Cas12a protein. Additionally, a 3' nucleoplasmin NLS and a third SV40 NLS were added. This modified ttLbCas12a was combined with a CaMV35S promoter (pICH51288) and NOS terminator (pICH41414) into pICH47742. For SpCas9, the same two 5' SV40 NLS with the potato IV20 intron were added, and the nuclease was then combined with a CaMV35S promoter and NOS terminator into pICH47742. To be able to clone crRNAs directly in binary level 2 vectors, the *Bsm*BI sites in both the level 1, position 6 CRISPR-Pink backbones for Cas9 sgRNA and LbCas12a mature crRNA expression were replaced by *Bsa*I sites. For the final vectors, *NPTII* (pICSL7004) was combined with either the modified *ttLbCas12a* or *SpCas9*, *tGFP*, pICH54055, pICH54066, the *Bsa*I-adapted CRISPR-Pink vector for crRNA or sgRNA expression, and end-linker pICH41822 into pICSL4723. The 35 sgRNAs and crRNAs were subsequently cloned into their respective backbones by introducing the spacer, as annealed oligonucleotides, in the CRISPR-Pink module by restriction/ligation using *Bsa*I. Sequences of the oligonucleotides can be found in Supplementary Dataset 2.

DNA preparation

Highly pure DNA for transfection was prepared from 3 mL of overnight *E. coli* culture in LB medium using the PureYield Plasmid MiniPrep System (Promega), with the following adaptations: bacterial pellets were frozen at -20°C before processing to increase DNA yield, the column was washed twice with the endotoxin removal wash to acquire the desired purity, and plasmid DNA was eluted with 30 uL elution buffer preheated at 60°C.

Protoplast isolation and transfection

Protoplast isolation and transfection in 96-well format were performed as described in Chapter 2.

Genomic DNA isolation and amplicon sequencing

Protoplast DNA was purified 24h after transfection using magnetic beads (NucleoMag Plant, Macherey-Nagel), following the manufacturer's instructions. DNA was eluted in 50 µL, of which 6 µL was subsequently used as a template in 25 µL PCR reactions using PHUSION HotStart Flex DNA polymerase (NEB) to amplify genomic DNA fragments containing target or predicted off-target sites using barcoded primers. For the PCR, an initial denaturation for 30 s at 98 °C was followed by 38 cycles of denaturation for 10 s at 98 °C, annealing for 20 s at 58 °C, extension for 20 s at 72 °C, and a final extension step of 3 min at 72 °C. Primer sequences are listed in Supplementary Dataset 3. The resulting PCR products were visualized by electrophoresis on a 2% agarose gel. Equal amounts of PCR products were pooled to obtain sequencing libraries. Libraries were subsequently column-purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel), following the manufacturer's instructions. Illumina HiSeq sequencing (paired-end, 2x150 bp reads) was performed by Eurofins Genomics Europe Sequencing GmbH, Constance, Germany.

Sequence Analysis

Paired sequencing reads were uploaded to the CLC Genomics Workbench v22, trimmed, merged, and demultiplexed using default settings. Mutation frequencies at target and predicted off-target sites were determined using Amplican (Labun et al. 2019).

Results

A Golden-Gate crRNA cloning system

For mutagenesis with Cas12a in plants, we determined the most efficient Cas12a orthologue and the best method for crRNA expression. In mammalian cells, three different Cas12a orthologues, AsCas12a, FnCas12a and LbCas12a, were initially found capable of inducing mutations. Therefore, we compared these three orthologues' efficiencies in causing mutations in tomato cells.

As Cas12a is capable of processing its own crRNA arrays, the individual crRNAs can be expressed in two different forms: as a longer, unprocessed pre-crRNA, which still needs additional processing by Cas12a before complex formation, or the shorter, mature version, skipping the first step and facilitating direct loading of the crRNA into the Cas12a-crRNA complex (Figure 1a). Additionally, the processing abilities of Cas12a raise the opportunity to express crRNAs as an array. In this case, multiple crRNAs are expressed in tandem as a single transcript, without additional provisions for processing as required for Cas9 sgRNAs (Figure 1b and 1c).

To facilitate easy cloning of single crRNAs, each expressed by its own U6-26 promoter, we adapted Golden-Gate compatible CRISPR-Pink sgRNA cassettes (a gift from Marc Youles, The Sainsbury Laboratory). These plasmids contain an AtU6-26 promoter for sgRNA expression, followed by an operon expressing an RFP protein. This operon can easily be replaced by a spacer in a Golden Gate cut/ligate reaction, allowing pink/white screening of colonies that have successfully integrated the sgRNA. We constructed two of these plasmids for use with each of the three Cas12a orthologues: one to express the unprocessed pre-crRNA, and one for the mature crRNA. To insert new spacers into these plasmids, two oligonucleotides encoding the spacer and compatible overhangs are annealed and subsequently cloned into the plasmid (Figure 1d, protocol in Supplementary Information 2).

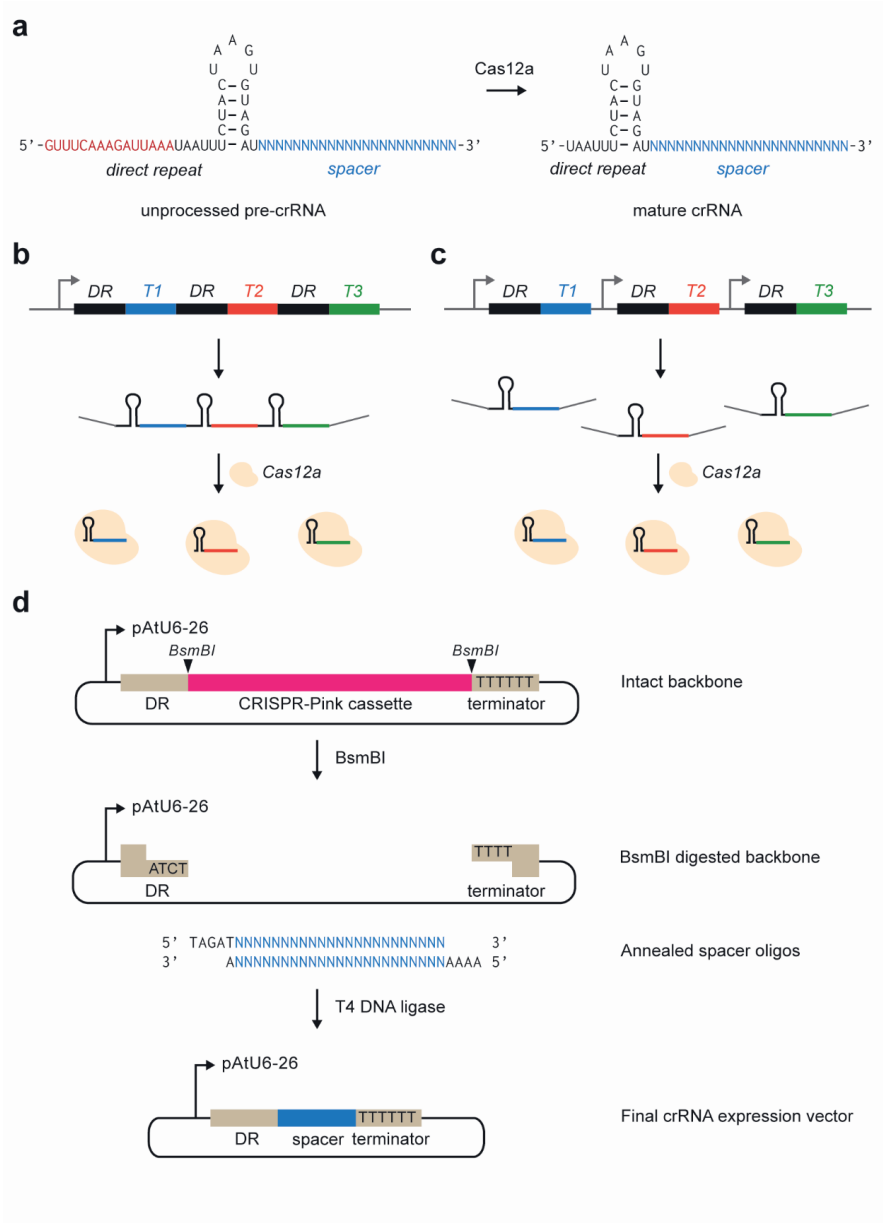


Figure 1 - Expression of crRNAs for Cas12a. (a) Cas12a is capable of processing its own crRNA arrays. The sequences of the unprocessed, pre-crRNA and processed, mature crRNA for LbCas12a are shown. (b and c) Cas12a crRNAs can be expressed in an array (b), driven by a single promoter, or individually, each under control of their own promoter (c) DR: direct repeat, T1-3: spacer sequences. (d) Cloning strategy for new crRNA expression vectors. Forward and reverse oligos encoding a new spacer are annealed and can then be cloned into the backbone vector using a BsmBI-based restriction-ligation process.

LbCas12a is the most effective orthologue for mutation induction in tomato

For our initial optimization, we selected three Cas12a target sites in tomato *PHYTOENE DESATURASE (PDS) (Solyc03g123760)* (Figure 2a). Spacers targeting these sites were cloned in the vectors for pre-crRNA and for mature crRNA expression for all three Cas12a orthologues. Each set of three level 1 vectors was then combined with the nuclease gene and a tGFP marker into a binary level 2 vector. Additionally, expression cassettes in which the three crRNAs were expressed as a single transcript from an AtU6-26 promoter – both in the pre-crRNA and mature form - were synthesized as level 1 vectors. These arrays were likewise combined with the nuclease and tGFP marker into level 2 vectors. In total, we thus made twelve level 2 vectors – each one expressing either AsCas12a, LbCas12a, or FnCas12a from a *2XCaMV35S* promoter and the three crRNAs using one of the four expression methods (see Supplementary Figure 1a for a graphical overview). These constructs were then transfected into tomato protoplasts. The presence of the tGFP marker allowed for determining the transfection efficiency, which was similar across the three replicates and was approximately 50% (see also Figure 4b). After the purification of DNA from the protoplast pools, the three target sites were amplified by PCR, and the resulting amplicons were subjected to next-generation amplicon sequencing. The percentage of reads with a mutation was determined using AmpliCan (Figure 2b) (Labun et al. 2019).

The observed mutation frequencies varied strongly per target site and per orthologue (Figure 2b). For target 1, none of the orthologues performed well, and mutation frequencies never reached over 2.5%. For target 2, both FnCas12a and LbCas12a performed well, whereas AsCas12a resulted in significantly lower mutation frequencies. For target 3, LbCas12a performed best, significantly outperforming AsCas12a and FnCas12a. From these results, we concluded that LbCas12a was the best-performing choice for Cas12a mutagenesis in tomato and therefore, we aimed to further optimize crRNA expression for this orthologue.

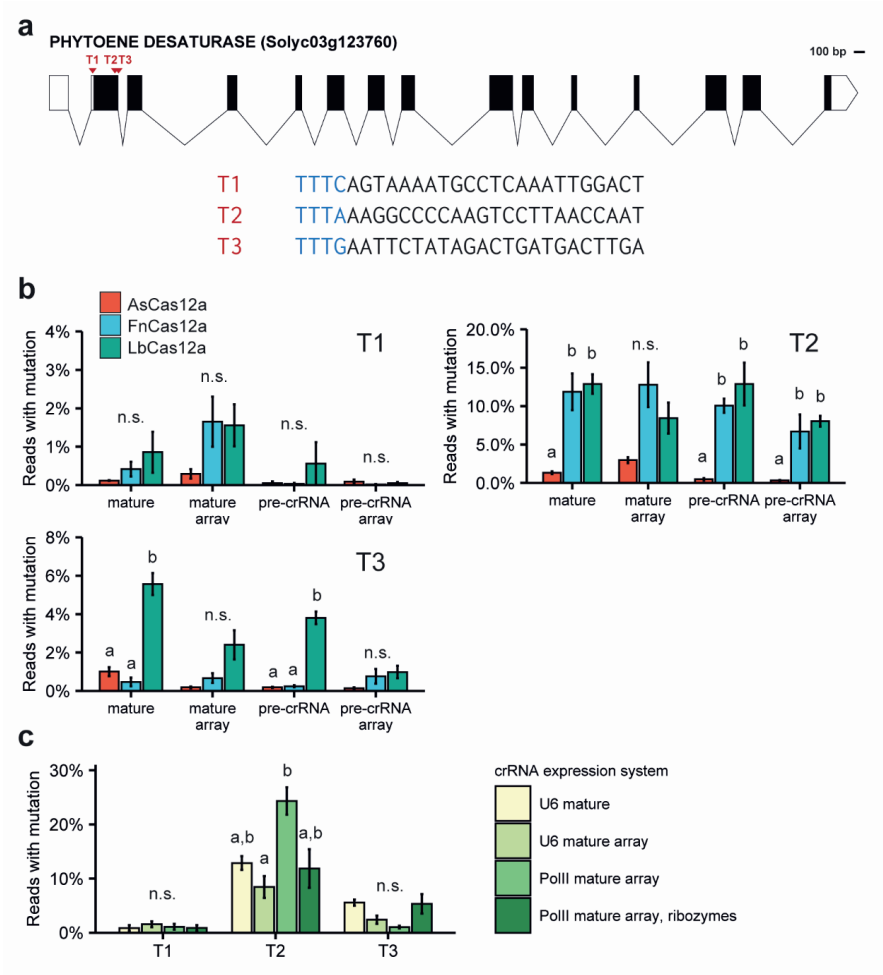


Figure 2 - Identification of the most efficient Cas12a orthologue and method for crRNA expression. (a) Three target sites in tomato PHYTOENE DESATURASE (SlPDS) were selected. (b) Mutation frequencies for three Cas12a orthologues and four methods of crRNA expression at three targets (T1-T3). Note the different y-axis scales for T1, T2 and T3. Error bars indicate standard error (n=3). Different letters indicate significant differences ($p < 0.05$) between mutation frequencies induced by the different orthologues, as determined using Two-Way ANOVA followed by Tukey's HSD Post-Hoc test. (c) For LbCas12a, two additional expression methods using a PolII promoter, and a PolII promoter combined with ribozymes were tested. Error bars indicate standard error (n=3). Different letters indicate significant differences ($p < 0.05$) between mutation frequencies obtained using different crRNA expression methods, as determined using Two-Way ANOVA followed by Tukey's HSD Post Hoc test.

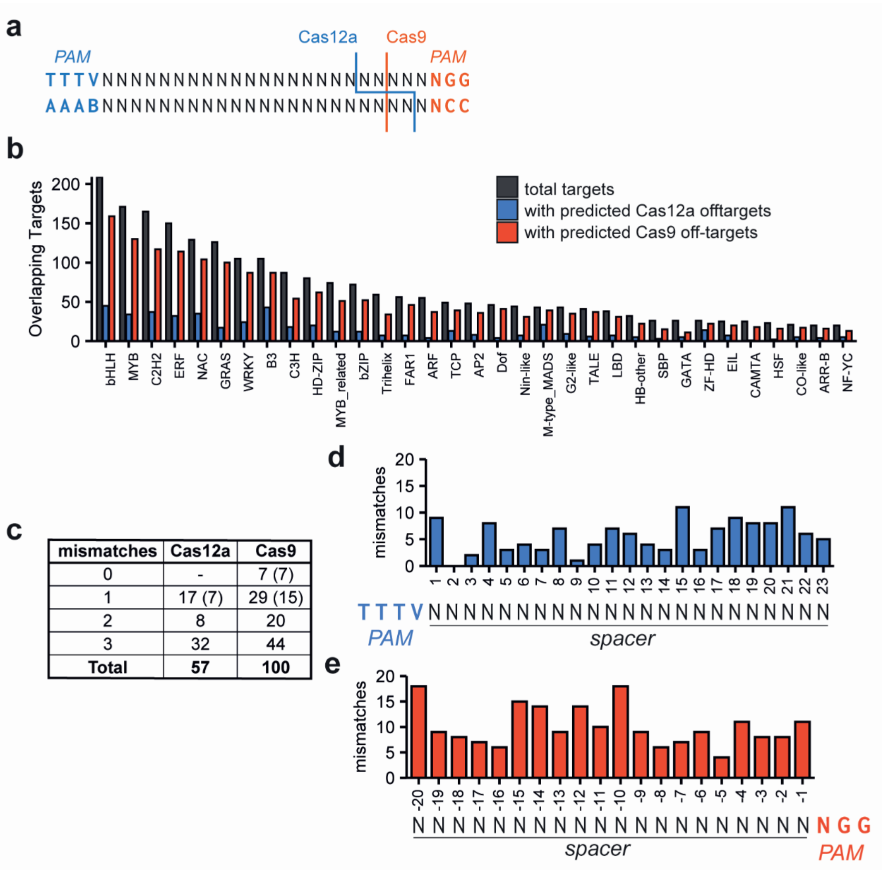
Several different methods of crRNA expression resulted in efficient mutagenesis

It was previously reported that the use of a PolII promoter instead of a PolIII promoter for crRNA expression improved Cas12a editing efficiency, as did the use of self-cleaving ribozymes flanking the crRNA array (Tang et al. 2017; M. Wang et al. 2018). As mature crRNAs performed comparably to pre-crRNAs in our previous experiment, we tested these additional expression systems only for the

combination of LbCas12a and mature crRNAs (Supplementary Figure 1b). The Cassava Vein Mosaic Virus (*CsVMV*) promoter was selected as the PolII promoter for crRNA expression. Significant differences between mutation efficiencies of the different expression systems were only found for target 2, for which the array-based crRNA expression system with the PolII promoter resulted in significantly higher mutation frequencies than the same crRNA array expression driven by the AtU6-26 promoter (Figure 2c). As the use of arrays and ribozymes made the system more complex, but did not significantly improve mutation frequencies, we used a combination of LbCas12a with mature, individually expressed crRNAs for Cas12a-mediated mutagenesis in subsequent experiments.

Comparing Cas12a and Cas9 performance at overlapping target sites

We next compared Cas9 and Cas12a efficiency, specificity, and the mutations produced. For this comparison, we identified sites where targets for Cas9 and Cas12a overlap, thus removing variation caused by differences in genomic context for the two enzymes (Figure 3a). We identified these overlapping sites in several gene families encoding transcription factors. The use of gene families allows for the selection of target sites that have predicted off-target sites with a range of mismatching nucleotides. This provides insight into the number and position of mismatches that will enable Cas-mediated double-strand breaks and mutagenesis at off-target sites. We predicted these off-target sites with up to 3 mismatches for both Cas9 and Cas12a and for all identified overlapping target sites in all transcription factor families using Cas-OFFinder (Figure 3b) (Sangsu Bae et al. 2014). In general, Cas12a target sites had fewer predicted off-target sites than Cas9 target sites, probably due to the longer spacer (23 nt for Cas12a and 20 nt for Cas9). We selected 35 overlapping target sites in the *bHLH* gene family as this family had the highest number of available overlapping targets and off-target sites (Figure 3b). For the Cas9 sgRNAs, we selected 100 potential off-target sites with varying amounts of mismatches and, in 22 cases, an insertion or deletion compared to the target site (Lin et al. 2014). For the Cas12a crRNAs, we selected 55 potential off-target sites, of which 7 had an additional insertion or deletion compared to the target site (Figure 3c). Up to four potential off-target sites per target site were selected for the study. We aimed to select potential off-target sites that resulted in an as equal as possible distribution of mismatches over the length of the spacer (Figure 3d,e). Selected target and predicted off-target sites are listed in Supplementary Dataset 1.



mismatches	Cas12a	Cas9
0	-	7 (7)
1	17 (7)	29 (15)
2	8	20
3	32	44
Total	**57**	**100**

Figure 3 - Overlapping Cas12a crRNA and Cas9 sgRNA sites for unbiased comparison of the two nucleases. (a) Sequence of overlapping target sites. (b) The number of overlapping targets found in different transcription factor families, and the number of overlapping target sites that have predicted off-target sites with a maximum of three mismatches for Cas12a and Cas9. (c) The number of predicted off-target sites with 0, 1, 2 or 3 mismatches for the 35 selected overlapping target sites for Cas12a and Cas9. Numbers between brackets indicate the amount of predicted off-target sites that have an insertion or deletion, leading to the formation of an RNA or DNA bulge, in addition to the number of mismatches indicated in the first column. (d) and (e) Distribution of mismatches over the length of the spacer for Cas12a (d) and Cas9 (e).

To facilitate easy cloning of these single crRNAs or sgRNAs, we constructed level 2 vectors containing either the Cas12a or Cas9 nuclease expression cassette, a turboGFP expression cassette for monitoring transfection efficiency, and a CRISPR-Pink cassette in which the spacer can be inserted using *BsaI*-mediated restriction-ligation. For this expression cassette, we used an improved, *Arabidopsis*-codon optimized and thermotolerant version of LbCas12a (Schindele and Puchta 2020). As we had noticed that *E. coli* liquid cultures with Cas12a-containing plasmids sometimes grew poorly, we inserted an intron in *LbCas12a* to prevent expression of the Cas12a protein in *E. coli*, and did the same for Cas9. The 35 spacers were subsequently ligated in both

vectors, resulting in 70 vectors. Tomato protoplasts were simultaneously transfected in a 96-well format. Transfection efficiencies were determined using confocal microscopy for Cas9- and Cas12a-transfected protoplasts and were found to be similar (Figure 4a,b). Target site and predicted off-target site fragments were PCR-amplified, and the pooled, barcoded amplicons were sequenced.

Cas12a and Cas9 have similar overall efficiencies, but strongly different efficiencies at individual targets

We first determined and compared the on-target mutation efficiencies at every target site. In this experiment, Cas12a performed slightly – though not significantly - better overall (Figure 4c). However, the best-performing nuclease varied per target site. Cas12a performed significantly better than Cas9 at 13 sites, and Cas9 performed better than Cas12a at 9 sites (Figure 4d). The correlation between Cas9 and Cas12a activity at target sites was low (Supplementary Figure 2). Previously, we correlated predicted to measured Cas9 activity and found that the so-called Azimuth score (Fusi et al. 2015; Doench et al. 2016) had some, albeit limited predictive value (Chapter 2). Here, we calculated the DeepCpf1 score (Luo et al. 2019) for each tested Cas12a target site and correlated this score to the obtained mutation frequency (Supplementary Figure 3). Although DeepCpf1 was able to predict the top and bottom performers to some extent, the correlation was generally low.

Cas12a induces more and larger mutations than Cas9

We compared all obtained on-target mutations for both nucleases, first by type. For Cas12a, insertions occurred at a frequency of 1.8±0.4%, and deletions at 94.1±0.7%. For Cas9, insertions occurred at a frequency of 17.4±1.5%, of which most (89±4%) were one bp, and deletions at 80.5±1.6%. The distributions of mutation sizes for both enzymes are shown in Figure 4e. Cas12a-induced deletions tend to be larger than Cas9-induced mutations. Interestingly, the characteristic peak for one bp insertions induced by Cas9 is absent for Cas12a. Likely as a result of this single difference, Cas9 caused more frameshift mutations than Cas12a (Figure 4f).

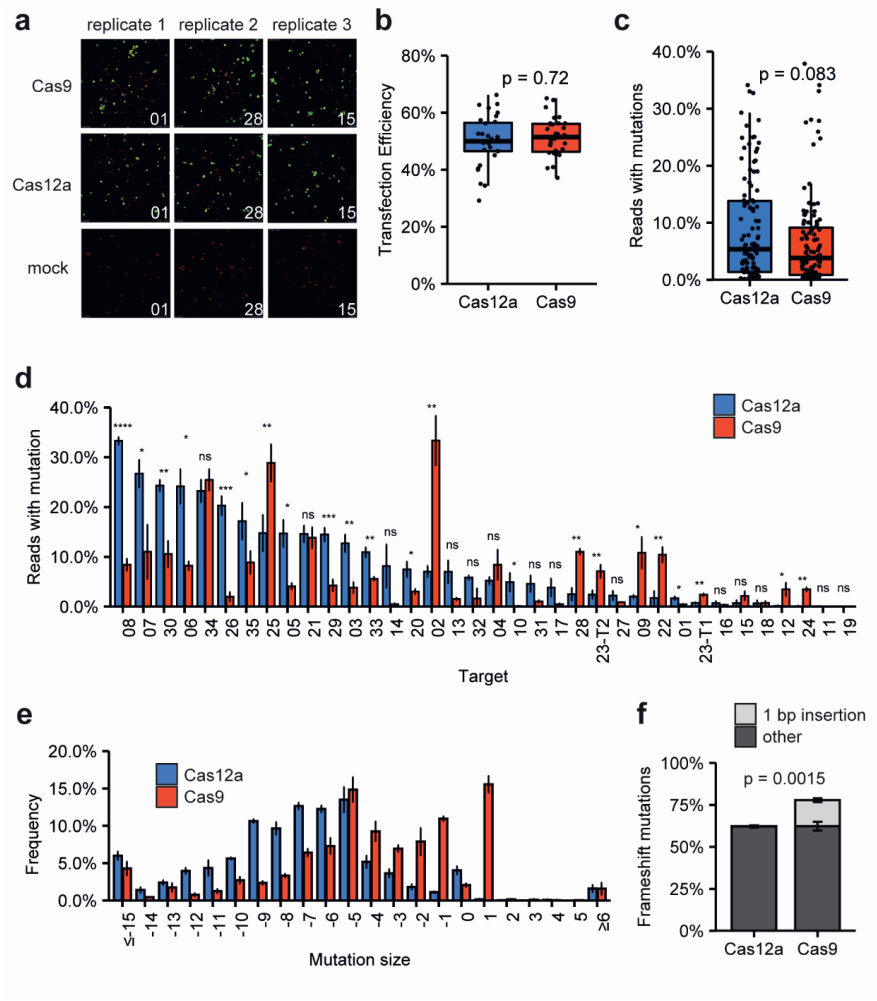


Figure 4 - Comparison of on-target mutations for Cas12a and Cas9. (a) and (b) Comparison of transfection efficiencies of protoplasts transfected with Cas9- or Cas12a- encoding plasmids. Representative images of Cas9-, Cas12a- and mock-transfected protoplasts are shown in (a). Successfully transfected protoplasts show green GFP fluorescence, while untransfected protoplasts only show red autofluorescence of chloroplasts. Pictures shown belong to targets 1, 28 and 15. (b) Transfection efficiencies were determined by counting the number of successfully transfected protoplasts and dividing it by the total number of protoplasts. For every nuclease in every replicate, the transfection efficiency of 11 individual transfections in one 96w plate was determined. Significance of difference was determined using the Wilcoxon Test. (c) Comparison of mutation frequencies for Cas12a and Cas9. Every point represents the mutation frequency at a target site. Significance of differences were determined using the Wilcoxon test. (d) Comparison of mutation frequencies for Cas12a and Cas9 per individual target site. Error bars indicate standard deviation (n = 3). Significances were determined using Student's T Test. Asterisks indicate significant differences (: p < 0.05, **: p < 0.01, ***: p < 0.001, ****: p < 0.0001, ns: not significant). (e) Mutation pattern of Cas9 and Cas12a at target sites. Error bars indicate standard deviation (n = 3). (f) Percentage of total mutations that are frameshift mutations for both nucleases. Frameshift mutations resulting from 1 bp insertions are shown in light grey, all others in dark grey. Error bars indicate standard deviation (n = 3). Significant difference between the total percentage of frameshift mutations was determined by Student's T Test.*

The frequency of off-target mutations is low for both Cas12a and Cas9

To gain more insight into the specificity of Cas12a and Cas9, we determined mutation frequencies at amplified predicted off-target sites using AmpliCan. To identify genuine Cas-induced off-target mutations as opposed to sequencing or PCR errors, we considered all off-target sites at which mutations occurred at a frequency of 0.1% of total reads or more. We disregarded any tested off-target sites for which we did not obtain a reliable wild-type control consensus sequence. Finally, we inspected the obtained mutation patterns to ensure they showed characteristics of CRISPR-induced mutations, such as insertions and deletions instead of substitutions, likely caused by PCR or sequencing errors (Supplementary Figures 4, 5).

For Cas12a, we identified 10 sites with genuine off-target mutations out of 55 tested sites; for Cas9, we identified 7 sites out of the 97 sites for which an amplicon could successfully be obtained (Figure 5a). To estimate how often an off-target mutation and an on-target mutation would occur in the same genome, we calculated the relative off-target frequencies by dividing the off-target frequency by the on-target frequency. The absolute and relative off-target mutation frequencies are shown in Figure 5b (Cas12a) and Figure 5c (Cas9).

For Cas12a, off-target mutations frequently occurred when 1 or 2 mismatches to the target occurred. However, none of these mismatches occurred in the first 14 nucleotides of the spacer. Additionally, relative off-target frequencies seem to decrease as mismatches are present closer to the PAM (Fig. 5b). No mutations were found in predicted off-target sequences with 3 mismatches.

Cas9 showed activity at sites with 1 or 3 mismatches to the target site. The off-target site with the highest mutation frequency contained only one mismatch, at the position most distal from the PAM. Interestingly, mutations were also found at sites that had mismatches close to the PAM.

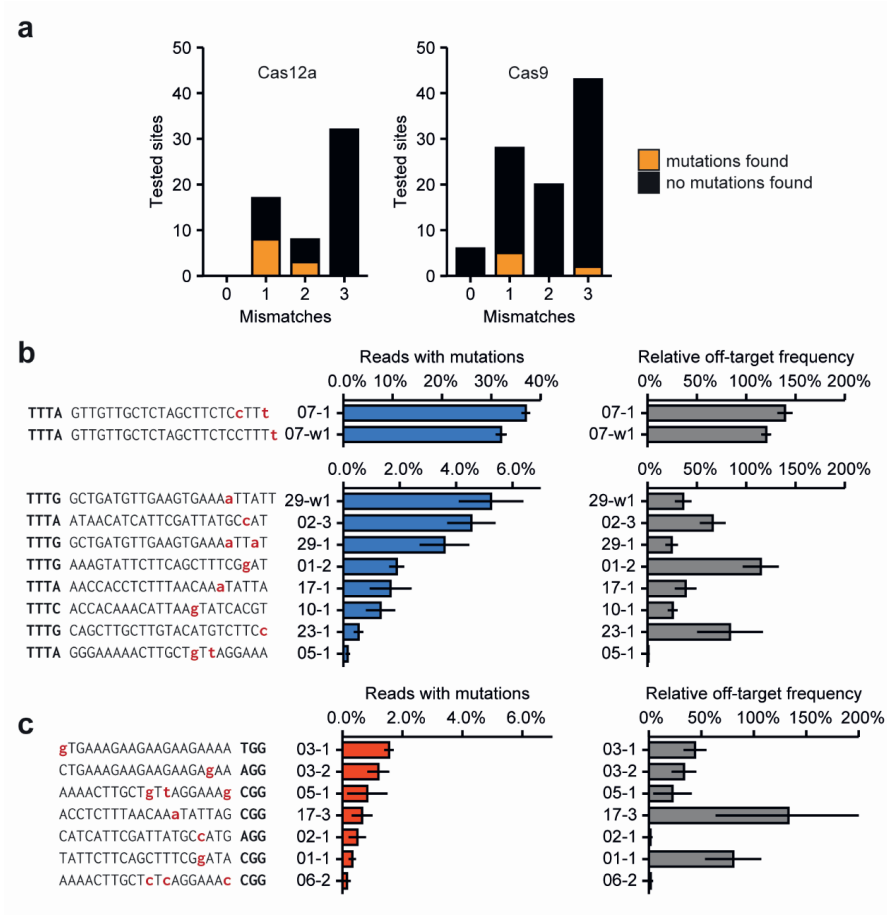


Figure 5 - Mutated off-target sites for Cas12a and Cas9. (a) Overview of the number of tested off-target sites. For Cas9, the off-target sites with 0 mismatches contain an insertion or deletion leading to the formation of an RNA or DNA bulge. The number of sites at which genuine off-target mutations were identified are indicated by yellow. (b) and (c) Identified off-target sites for Cas12a (b) and Cas9 (c). Mismatches to the target site are indicated in lowercase red. Protospacer adjacent motif is indicated in bold. Mutation frequencies as well as relative off-target frequencies are shown. Relative off-target frequencies were calculated by dividing the mutation frequency at the off-target site by the mutation frequency at the target site and give a measure of the likelihood of an on- and off-target mutation occurring in the same genome. Error bars indicate standard error (n = 3).

Discussion and Conclusion

In this study, we aimed at optimizing CRISPR-Cas12a mediated genome editing in tomato cells and comparing its performance to the frequently used CRISPR-Cas9. To achieve this, we first compared different Cas12a orthologues and methods for crRNA expression. We found LbCas12a to be the most efficient and robust orthologue for inducing mutations, in agreement with previous reports (Hu et al. 2017; M. Wang et al. 2017; H. Kim et al. 2017; Tang et al. 2017). Although FnCas12a was also capable of inducing mutations at high frequency, one of the target sites used for testing (T3, Figure 2b) that was successfully mutated by LbCas12a gave only low mutation frequencies for FnCas12a. AsCas12a performed poorly at all three tested target sites. It was shown previously that the efficiency of Cas12a-mediated genome editing, like that of Cas9-mediated genome editing, increases with temperature (Moreno-Mateos et al. 2017; LeBlanc et al. 2017; Malzahn et al. 2019). AsCas12a seemed to be more sensitive to temperature than LbCas12a (Moreno-Mateos et al. 2017). Tomato protoplast experiments as well as tissue culture were routinely performed at 25°C, which may be too low for AsCas12a.

As it was found that LbCas12a was the most efficient at mutating the tomato genome, we further investigated the best method for crRNA expression for this nuclease. Mutations could reliably be obtained with all tested crRNA expression systems, in contrast to earlier studies in rice and soybean where the use of mature crRNAs in combination with PolIII promoters resulted in no or very low mutation frequencies (Tang et al. 2017; Xu et al. 2017; H. Kim et al. 2017). As the individual crRNA expression cassettes we created in this study offer greater flexibility of cloning than arrays and ribozyme-based systems, this is the method that is routinely applied in our laboratory for the construction of binary vectors for stable transformation. Although the study presented here focuses on protoplasts, we have successfully generated a large number of stably transformed tomato plants with Cas12a-induced mutations using a combination of thermotolerant LbCas12a (Schindele and Puchta 2020) and crRNAs expressed in the mature form, cloned in the vectors as described in Figure 1d.

To compare the performance of Cas9 and Cas12a as fairly as possible, we selected 35 overlapping target sites in the coding sequence of genes from the bHLH gene family. Overall, Cas12a showed editing at these sites at a similar level as Cas9. However, mutation rates varied strongly depending on the target site, as has been previously reported (Bernabé-Orts et al. 2019). As overlapping target sites were used, characteristics such as G/C content, chromatin conformation, and epigenetic marks such as DNA methylation or histone modifications are mostly similar for the Cas12a and Cas9 target

sites. However, Cas9 and Cas12a might have different preferences or tolerances for such features, affecting their efficacy. Additionally, the exact nucleosome localization might affect the availability of the target sites (Horlbeck et al. 2016; Yarrington et al. 2018).

Interestingly, some target sites showed hardly any editing for Cas9, whereas Cas12a could reliably induce mutations, such as targets 10 and 26 (Figure 4d). For these two specific targets, Cas9 inactivity might be explained by the presence of a "TT" motif in the 3' end of the spacer, resulting in low expression of the sgRNA (Graf et al. 2019). This might be overcome by using pre-assembled Cas9-sgRNA complexes (RNPs) or mutated scaffold RNAs (Graf et al. 2019).

Reliably predicting which nuclease would perform best at a specific target site before proceeding to stable transformation is desirable. For Cas9, several algorithms for efficiency prediction exist, which are implemented in frequently used tools for sgRNA prediction, such as CRISPR-P (Lei et al. 2014; Doench et al. 2016; Moreno-Mateos et al. 2015). Information for Cas12a, especially about activity in plants, is more limited. We tested the correlation between the DeepCpf1 prediction score (Luo et al. 2019) and the obtained mutation frequencies from our dataset (Supplementary Figure 3). crRNAs were divided into quartiles based on their DeepfCpf1 score and plotted against mutation frequencies per quartile. Although the first and fourth quartiles gave significantly lower and higher actual activities respectively, the variation of mutation frequencies within quartiles was large, and the correlation between the DeepCpf1 score and obtained mutation frequencies low (Supplementary Figure 3). In the future, more high-throughput data could be obtained to specifically train algorithms to predict efficient crRNAs in plants.

Apart from the obtained mutation frequencies, we also compared the mutation patterns induced by Cas9 and Cas12a at target sites. For Cas9, a significant portion of induced mutations is a one bp insertion. Previously, we and others have shown that these characteristic one bp insertions are likely often the result of the fill-in of Cas9-induced staggered DSBs with a one bp 5' overhang, followed by subsequent ligation of the, now blunt, DNA strands (Zuo and Liu 2016; Shou et al. 2018; Allen et al. 2019; W. Chen et al. 2019; Shi et al. 2019; Lemos et al. 2018). LbCas12a is also known to induce staggered cuts, with a larger 4-5 bp 5' overhang (Zetsche et al. 2015). Interestingly, no peak is found in the mutagenic spectrum for Cas12a at the +4 and +5 positions (Figure 4e), indicating that the fill-in of these staggered overhangs and subsequent ligation of ends is not frequently employed for the repair of Cas12a-induced DSBs.

Cas12a-induced deletions are frequently larger than Cas9-induced deletions: for Cas9, the majority of deletions range from 1-5 bp, whereas for Cas12a, most deletions range between 5-10 bp. It has

been suggested that this difference may be caused by the fact that LbCas12a cuts distal from the PAM, outside of its "seed" sequence. As a consequence, small mutations may be tolerated by the recognition of the target site, and the target site may be cleaved again until a large enough deletion finally precludes recognition and cleavage (Wolter and Puchta 2019). An alternative explanation for the difference in mutation patterns between Cas12a and Cas9 is that a larger fraction of the Cas12a-induced mutations are caused by microhomology-mediated end-joining (MMEJ, also called alternative end joining or alt-EJ), which always induces deletions. This could either be explained by the fact that the staggered DSB caused by Cas12a preferentially triggers end-resectioning and thus MMEJ, or by the fact that NHEJ-mediated repair is more often perfect as a result of the overhangs produced by Cas12a. In that case, DSBs will keep being induced until either NHEJ is unsuccessful (which only happens at low frequencies) or, more likely, the break is repaired following end-resectioning, such as in MMEJ. Either way, this bias towards end resectioning can explain why Cas12a is generally found to be more successful than Cas9 in inducing homology-directed repair (HDR) (Wolter and Puchta 2019; Van Vu et al. 2020), as end-resectioning is the first step required for this repair outcome (Truong et al. 2013; Ceccaldi et al. 2016; Puchta 2005; Manova and Gruszka 2015).

We found Cas9 to induce frameshift mutations at higher frequencies than Cas12a, which is predominantly caused by the 1 bp insertion mutations. This difference in mutation pattern may make Cas9 the more suitable option for producing knock-out mutants in protein-coding genes. Desirable phenotypes can also be obtained by tweaking the expression of genes through modification of cis-regulatory elements (Rodríguez-Leal et al. 2017; X. Wang et al. 2021). Cas12a could be the enzyme of choice for this type of genome editing because it has an A/T rich PAM and propensity to induce slightly larger deletions, which disrupt or delete the often short regulatory motifs present in promoters.

In studies of mammalian cells, Cas12a is generally reported to be more specific than Cas9 (D. Kim et al. 2016; Kleinstiver et al. 2016). Studies on Cas12a-induced off-target mutations in plants have so far been conducted on a small scale (Tang et al. 2018; Lee et al. 2018; Raitskin et al. 2019) but do indicate that Cas12a does not frequently induce off-target mutations. To acquire more data on Cas12a specificity, we selected 57 predicted off-target sites with 1-3 mismatches to the spacer and investigated them for the presence of off-target mutations. At 10 out of 57 sites, off-target mutations were identified. For Cas9, we investigated 100 predicted off-target sites and found evidence of off-target mutation at seven sites. Cas12a off-target activity was strongly linked to mismatches at the 3' end of the spacer, distal to the so-called "seed sequence." Conversely, Cas9 off-target sites with

mismatches proximal to the PAM were still found to be mutagenized, albeit at low frequencies (Figure 5c, Chapter 2) (Hsu et al. 2013). Based on these results, the existence of a seed sequence may be more applicable to Cas12a than Cas9. This would make potential high-risk Cas12a off-target sites easier to predict and, therefore, to avoid.

In this study, we selected spacers with a length of 23 nucleotides. However, previous research has shown that spacers as short as 19 nucleotides retain almost complete activity (H. K. Kim et al. 2017; Zetsche et al. 2015). It is therefore not surprising that the four nucleotides most distal to the PAM add little to nothing to editing specificity, resulting in high relative off-target frequencies at these sites (Figure 5b).

Concluding, we have shown that LbCas12a can reliably and specifically induce mutations in the tomato genome. Our high-throughput testing methods allowed us to test Cas12a orthologues, crRNA expression systems, efficiency at target sites, and specificity on a larger scale than previously reported in plants. Together with constructing a convenient, Golden Gate-compatible cloning system for crRNAs, this work helps lay the foundation for routine application of Cas12a to induce mutations in the tomato genome.

Acknowledgements

This research is supported by the Dutch Research Council (NWO) and is funded by the Ministry of Infrastructure and Water Management (Biotechnology and Safety project #15792). Kai Thoris is kindly acknowledged for constructing the LbCas12a-Pink position 1, 2, 3, and 7 vectors.

References

Ali, S.K., and D.A.F. Al-Koofee. 2019. "BatchPrimer3: A Free Web Application for Allele Specific (SBE and Allele Flanking) Primer Design for SNPs Genotyping in Molecular Diagnostics: A Bioinformatics Study." *Gene Reports* 17 (September). https://doi.org/10.1016/j.genrep.2019.100524.

Allen, F., L. Crepaldi, C. Alsinet, A.J. Strong, V. Kleshchevnikov, P. De Angeli, P. Páleníková, ... L. Parts. 2019. "Predicting the Mutations Generated by Repair of Cas9-Induced Double-Strand Breaks." *Nature Biotechnology* 37 (1): 64–82. https://doi.org/10.1038/nbt.4317.

Bae, S., J. Park, and J.-S. Kim. 2014. "Cas-OFFinder: A Fast and Versatile Algorithm That Searches for Potential off-Target Sites of Cas9 RNA-Guided Endonucleases." *Bioinformatics* 30 (10): 1473–75. https://doi.org/10.1093/bioinformatics/btu048.

Bae, Sangsu, J. Park, and J.-S. Kim. 2014. "Cas-OFFinder: A Fast and Versatile Algorithm That Searches for Potential off-Target Sites of Cas9 RNA-Guided Endonucleases." *Bioinformatics* 30 (10): 1473–75. https://doi.org/10.1093/BIOINFORMATICS/btu048.

Bernabé-Orts, J.M., I. Casas-Rodrigo, E.G. Minguet, V. Landolfi, V. Garcia-Carpintero, S. Gianoglio, M. Vázquez-Vilar, ... D. Orzaez. 2019. "Assessment of Cas12a-Mediated Gene Editing Efficiency in Plants." *Plant Biotechnology Journal* 17 (10): 1971–84. https://doi.org/10.1111/pbi.13113.

Ceccaldi, R., B. Rondinelli, and A.D. D'Andrea. 2016. "Repair Pathway Choices and Consequences at the Double-Strand Break." *Trends in Cell Biology* 26 (1): 52–64. https://doi.org/10.1016/j.tcb.2015.07.009.

Chandrasekaran, M., T. Boopathi, and M. Paramasivan. 2021. "A Status-Quo Review on CRISPR-Cas9 Gene Editing Applications in Tomato." *International Journal of Biological Macromolecules*. Elsevier B.V. https://doi.org/10.1016/j.ijbiomac.2021.08.169.

Chen, K., Y. Wang, R. Zhang, H. Zhang, and C. Gao. 2019. "CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture." *Annual Review of Plant Biology* 70: 667–97. https://doi.org/10.1146/annurev-arplant-050718-100049.

Chen, W., A. McKenna, J. Schreiber, M. Haeussler, Y. Yin, V. Agarwal, W.S. Noble, and J. Shendure. 2019. "Massively Parallel Profiling and Predictive Modeling of the Outcomes of CRISPR/Cas9-Mediated Double-Strand Break Repair." *Nucleic Acids Research* 47 (15): 7989–8003. https://doi.org/10.1093/nar/gkz487.

Concordet, J.P., and M. Haeussler. 2018. "CRISPOR: Intuitive Guide Selection for CRISPR/Cas9 Genome Editing Experiments and Screens." *Nucleic Acids Research* 46 (W1): W242–45. https://doi.org/10.1093/nar/gky354.

Doench, J.G., N. Fusi, M. Sullender, M. Hegde, E.W. Vaimberg, K.F. Donovan, I. Smith, ... D.E. Root. 2016. "Optimized SgRNA Design to Maximize Activity and Minimize Off-Target Effects of CRISPR-Cas9." *Nature Biotechnology* 34 (2): 184–91. https://doi.org/10.1038/nbt.3437.

Engler, C., M. Youles, R. Gruetzner, T.M. Ehnert, S. Werner, J.D.G. Jones, N.J. Patron, and S. Marillonnet. 2014. "A Golden Gate Modular Cloning Toolbox for Plants." *ACS Synthetic Biology* 3 (11): 839–43. https://doi.org/10.1021/sb4001504.

Fonfara, I., H. Richter, M. BratoviÄ, A. Le Rhun, and E. Charpentier. 2016. "The CRISPR-Associated DNA-Cleaving Enzyme Cpf1 Also Processes Precursor CRISPR RNA." *Nature* 532 (7600): 517–21. https://doi.org/10.1038/nature17945.

Fusi, N., I. Smith, J. Doench, and J. Listgarten. 2015. "In Silico Predictive Modeling of CRISPR/Cas9 Guide Efficiency." *BioRxiv*, 021568. https://doi.org/10.1101/021568.

Graf, R., X. Li, V.T. Chu, and K. Rajewsky. 2019. "SgRNA Sequence Motifs Blocking Efficient CRISPR/Cas9-Mediated Gene Editing." *Cell Reports* 26 (5): 1098-1103.e3. https://doi.org/10.1016/j.celrep.2019.01.024.

Guo, A.Y., X. Chen, G. Gao, H. Zhang, Q.H. Zhu, X.C. Liu, Y.F. Zhong, X. Gu, K. He, and J. Luo. 2008. "PlantTFDB: A Comprehensive Plant Transcription Factor Database." *Nucleic Acids Research* 36 (SUPPL. 1): 966–69. https://doi.org/10.1093/nar/gkm841.

Horlbeck, M.A., L.B. Witkowsky, B. Guglielmi, J.M. Replogle, L.A. Gilbert, J.E. Villalta, S.E. Torigoe, R. Tjian, and J.S. Weissman. 2016. "Nucleosomes Impede Cas9 Access to DNA in Vivo and in Vitro." *ELife* 5 (MARCH2016): 1–21. https://doi.org/10.7554/eLife.12677.

Hsu, P.D., D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, Y. Li, … F. Zhang. 2013. "DNA Targeting Specificity of RNA-Guided Cas9 Nucleases." *Nature Biotechnology* 31 (9): 827–32. https://doi.org/10.1038/nbt.2647.

Hu, X., C. Wang, Q. Liu, Y. Fu, and K. Wang. 2017. "Targeted Mutagenesis in Rice Using CRISPR-Cpf1 System." *Journal of Genetics and Genomics* 44 (1): 71–73. https://doi.org/10.1016/j.jgg.2016.12.001.

Hur, J.K., K. Kim, K.W. Been, G. Baek, S. Ye, J.W. Hur, S.M. Ryu, Y.S. Lee, and J.S. Kim. 2016. "Targeted Mutagenesis in Mice by Electroporation of Cpf1 Ribonucleoproteins." *Nature Biotechnology* 34 (8): 807–8. https://doi.org/10.1038/nbt.3596.

Jaganathan, D., K. Ramasamy, G. Sellamuthu, S. Jayabalan, and G. Venkataraman. 2018. "CRISPR for Crop Improvement: An Update Review." *Frontiers in Plant Science* 9 (July): 1–17. https://doi.org/10.3389/fpls.2018.00985.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Kim, D., J. Kim, J.K. Hur, K.W. Been, S.H. Yoon, and J.S. Kim. 2016. "Genome-Wide Analysis Reveals Specificities of Cpf1 Endonucleases in Human Cells." *Nature Biotechnology* 34 (8): 863–68. https://doi.org/10.1038/nbt.3609.

Kim, H., S.T. Kim, J. Ryu, B.C. Kang, J.S. Kim, and S.G. Kim. 2017. "CRISPR/Cpf1-Mediated DNA-Free Plant Genome Editing." *Nature Communications* 8: 1–7. https://doi.org/10.1038/ncomms14406.

Kim, H.K., M. Song, J. Lee, A.V. Menon, S. Jung, Y.M. Kang, J.W. Choi, … H. Kim. 2017. "In Vivo High-Throughput Profiling of CRISPR-Cpf1 Activity." *Nature Methods* 14 (2): 153–59. https://doi.org/10.1038/nmeth.4104.

Kim, Y., S.A. Cheong, J.G. Lee, S.W. Lee, M.S. Lee, I.J. Baek, and Y.H. Sung. 2016. "Generation of Knockout Mice by Cpf1-Mediated Gene Targeting." *Nature Biotechnology* 34 (8): 808–10. https://doi.org/10.1038/nbt.3614.

Kleinstiver, B.P., S.Q. Tsai, M.S. Prew, N.T. Nguyen, M.M. Welch, J.M. Lopez, Z.R. McCaw, M.J. Aryee, and J.K. Joung. 2016. "Genome-Wide Specificities of CRISPR-Cas Cpf1 Nucleases in Human Cells." *Nature Biotechnology* 34 (8): 869–74. https://doi.org/10.1038/nbt.3620.

Kwon, C.T., J. Heo, Z.H. Lemmon, Y. Capua, S.F. Hutton, J. Van Eck, S.J. Park, and Z.B. Lippman. 2020. "Rapid Customization of Solanaceae Fruit Crops for Urban Agriculture." *Nature Biotechnology* 38 (2): 182–88. https://doi.org/10.1038/s41587-019-0361-2.

Labun, K., X. Guo, A. Chavez, G. Church, J.A. Gagnon, and E. Valen. 2019. "Accurate Analysis of Genuine CRISPR Editing Events with AmpliCan." *Genome Research* 29 (5): 843–47. https://doi.org/10.1101/gr.244293.118.

LeBlanc, C., F. Zhang, J. Mendez, Y. Lozano, K. Chatpar, V. Irish, and Y. Jacob. 2017. "Increased Efficiency of Targeted Mutagenesis by CRISPR/Cas9 in Plants Using Heat Stress." *The Plant Journal* 12 (10): 3218–21. https://doi.org/10.1111/tpj.13782.

Lee, K., Y. Zhang, B.P. Kleinstiver, J.A. Guo, M.J. Aryee, J. Miller, A. Malzahn, … K. Wang. 2018. "Activities and Specificities of CRISPR/Cas9 and Cas12a Nucleases for Targeted Mutagenesis in Maize." *Plant Biotechnology Journal*, 1–11. https://doi.org/10.1111/pbi.12982.

Lei, Y., L. Lu, H.Y. Liu, S. Li, F. Xing, and L.L. Chen. 2014. "CRISPR-P: A Web Tool for Synthetic Single-Guide RNA Design of CRISPR-System in Plants." *Molecular Plant* 7 (9): 1494–96. https://doi.org/10.1093/mp/ssu044.

Lemmon, Z.H., N.T. Reem, J. Dalrymple, S. Soyk, K.E. Swartwood, D. Rodriguez-Leal, J. Van Eck, and Z.B. Lippman. 2018. "Rapid Improvement of Domestication Traits in an Orphan Crop by Genome Editing." *Nature Plants* 4 (10): 766–70. https://doi.org/10.1038/s41477-018-0259-x.

Lemos, B.R., A.C. Kaplan, J.E. Bae, A.E. Ferrazzoli, J. Kuo, R.P. Anand, D.P. Waterman, and J.E. Haber. 2018. "CRISPR/Cas9 Cleavages in Budding Yeast Reveal Templated Insertions and Strand-Specific Insertion/Deletion Profiles." *Proceedings of the National Academy of Sciences of the United States of America* 115 (9): E2010–47. https://doi.org/10.1073/pnas.1716855115.

Lin, Y., T.J. Cradick, M.T. Brown, H. Deshmukh, P. Ranjan, N. Sarode, B.M. Wile, P.M. Vertino, F.J. Stewart, and G. Bao. 2014. "CRISPR/Cas9 Systems Have off-Target Activity with Insertions or Deletions between Target DNA and Guide RNA Sequences." *Nucleic Acids Research* 42 (11): 7473–85. https://doi.org/10.1093/nar/gku402.

Luo, J., W. Chen, L. Xue, and B. Tang. 2019. "Prediction of Activity and Specificity of CRISPR-Cpf1 Using Convolutional Deep Learning Neural Networks." *BMC Bioinformatics* 20 (1): 332. https://doi.org/10.1186/s12859-019-2939-6.

Malzahn, A.A., X. Tang, K. Lee, Q. Ren, S. Sretenovic, Y.Y. Zhang, H. Chen, … Y. Qi. 2019. "Application of CRISPR-Cas12a Temperature Sensitivity for Improved Genome Editing in Rice, Maize, and Arabidopsis." *BMC Biology* 17 (1): 9. https://doi.org/10.1186/s12915-019-0629-5.

Manova, V., and D. Gruszka. 2015. "DNA Damage and Repair in Plants – from Models to Crops." *Frontiers in Plant Science* 6 (October): 1–26. https://doi.org/10.3389/fpls.2015.00885.

Moreno-Mateos, M.A., J.P. Fernandez, R. Rouet, C.E. Vejnar, M.A. Lane, E. Mis, M.K. Khokha, J.A. Doudna, and A.J. Giraldez. 2017. "CRISPR-Cpf1 Mediates Efficient Homology-Directed Repair and Temperature-Controlled Genome Editing." *Nature Communications* 8 (1): 2024. https://doi.org/10.1038/s41467-017-01836-2.

Moreno-Mateos, M.A., C.E. Vejnar, J.D. Beaudoin, J.P. Fernandez, E.K. Mis, M.K. Khokha, and A.J. Giraldez. 2015. "CRISPRscan: Designing Highly Efficient SgRNAs for CRISPR-Cas9 Targeting in Vivo." *Nature Methods* 12 (10): 982–88. https://doi.org/10.1038/nmeth.3543.

Puchta, H. 2005. "The Repair of Double-Strand Breaks in Plants: Mechanisms and Consequences for Genome Evolution." *Journal of Experimental Botany* 56 (409): 1–14. https://doi.org/10.1093/jxb/eri025.

Raitskin, O., C. Schudoma, A. West, and N.J. Patron. 2019. "Comparison of Efficiency and Specificity of CRISPR-Associated (Cas) Nucleases in Plants: An Expanded Toolkit for Precision Genome Engineering." *PloS One* 14 (2): e0211598. https://doi.org/10.1371/journal.pone.0211598.

Rodríguez-Leal, D., Z.H. Lemmon, J. Man, M.E. Bartlett, and Z.B. Lippman. 2017. "Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing." *Cell* 171 (2): 470-480.e8. https://doi.org/10.1016/j.cell.2017.08.030.

Roekel, J.S.C. van, B. Damm, L.S. Melchers, and A. Hoekema. 1993. "Factors Influencing Transformation Frequency of Tomato (Lycopersicon Esculentum)." *Plant Cell Reports* 12 (11): 644–47. https://doi.org/10.1007/BF00232816.

Sage, F., and N. Geijsen. 2022. "Ligation-Assisted Homologous Recombination Enables Precise Genome Editing by Deploying Both MMEJ and HDR," 1–14. https://zenodo.org/deposit/4754116.

Schindele, P., and H. Puchta. 2020. "Engineering CRISPR/LbCas12a for Highly Efficient, Temperature-Tolerant Plant Gene Editing." *Plant Biotechnology Journal* 18 (5): 1118–20. https://doi.org/10.1111/pbi.13275.

Shi, X., J. Shou, M.M. Mehryar, J. Li, L. Wang, M. Zhang, H. Huang, X. Sun, and Q. Wu. 2019. "Cas9 Has No Exonuclease Activity Resulting in Staggered Cleavage with Overhangs and Predictable Di- and Tri-Nucleotide CRISPR Insertions without Template Donor." *Cell Discovery* 5 (1): 4–7. https://doi.org/10.1038/s41421-019-0120-z.

Shou, J., J. Li, Y. Liu, and Q. Wu. 2018. "Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion." *Molecular Cell* 71 (4): 498-509.e4. https://doi.org/10.1016/j.molcel.2018.06.021.

Tang, X., G. Liu, J. Zhou, Q. Ren, Q. You, L. Tian, X. Xin, … Y. Zhang. 2018. "A Large-Scale Whole-Genome Sequencing Analysis Reveals Highly Specific Genome Editing by Both Cas9 and Cpf1 (Cas12a) Nucleases in Rice." *Genome Biology* 19 (1): 84. https://doi.org/10.1186/s13059-018-1458-5.

Tang, X., L.G. Lowder, T. Zhang, A.A. Malzahn, X. Zheng, D.F. Voytas, Z. Zhong, … Y. Qi. 2017. "A CRISPR-Cpf1 System for Efficient Genome Editing and Transcriptional Repression in Plants." *Nature Plants* 3 (February): 1–5. https://doi.org/10.1038/nplants.2017.18.

Truong, L.N., Y. Li, L.Z. Shi, P.Y.H. Hwang, J. He, H. Wang, N. Razavian, M.W. Berns, and X. Wu. 2013. "Microhomology-Mediated End Joining and Homologous Recombination Share the Initial End Resection Step to Repair DNA Double-Strand Breaks in Mammalian Cells." *Proceedings of the National Academy of Sciences of the United States of America* 110 (19): 7720–25. https://doi.org/10.1073/pnas.1213431110.

Vu, T. Van, D.T.H. Doan, M.T. Tran, Y.W. Sung, Y.J. Song, and J.Y. Kim. 2021. "Improvement of the LbCas12a-CrRNA System for Efficient Gene Targeting in Tomato." *Frontiers in Plant Science* 12 (August). https://doi.org/10.3389/fpls.2021.722552.

Vu, T. Van, V. Sivankalyani, E. Kim, D.T.H. Doan, M.T. Tran, J. Kim, Y.W. Sung, M. Park, Y.J. Kang, and J. Kim. 2020. "Highly Efficient Homology-directed Repair Using CRISPR/Cpf1-geminiviral Replicon in Tomato." *Plant Biotechnology Journal* 18 (10): 2133–43. https://doi.org/10.1111/pbi.13373.

Vu, T. Van, V. Sivankalyani, E.J. Kim, D.T.H. Doan, M.T. Tran, J. Kim, Y.W. Sung, M. Park, Y.J. Kang, and J.Y. Kim. 2020. "Highly Efficient Homology-Directed Repair Using CRISPR/Cpf1-Geminiviral Replicon in Tomato." *Plant Biotechnology Journal* 18 (10): 2133–43. https://doi.org/10.1111/pbi.13373.

Wang, M., Y. Mao, Y. Lu, X. Tao, and J. kang Zhu. 2017. "Multiplex Gene Editing in Rice Using the CRISPR-Cpf1 System." *Molecular Plant* 10 (7): 1011–13. https://doi.org/10.1016/j.molp.2017.03.001.

Wang, M., Y. Mao, Y. Lu, Z. Wang, X. Tao, and J.-K. Zhu. 2018. "Multiplex Gene Editing in Rice with Simplified CRISPR-Cpf1 and CRISPR-Cas9 Systems." *Journal of Integrative Plant Biology* 60 (8): 1–11. https://doi.org/10.1111/jipb.12667.

Wang, X., L. Aguirre, D. Rodríguez-Leal, A. Hendelman, M. Benoit, and Z.B. Lippman. 2021. "Dissecting Cis-Regulatory Control of Quantitative Trait Variation in a Plant Stem Cell Circuit." *Nature Plants* 7 (4): 419–27. https://doi.org/10.1038/s41477-021-00898-x.

Weber, E., C. Engler, R. Gruetzner, S. Werner, and S. Marillonnet. 2011. "A Modular Cloning System for Standardized Assembly of Multigene Constructs." Edited by J. Peccoud. *PLoS ONE* 6 (2): 1–11. https://doi.org/10.1371/journal.pone.0016765.

Wolter, F., and H. Puchta. 2019. "*In Planta* Gene Targeting Can Be Enhanced by the Use of <scp>CRISPR</Scp> /Cas12a." *The Plant Journal*, tpj.14488. https://doi.org/10.1111/tpj.14488.

Xia, X., X. Cheng, R. Li, J. Yao, Z. Li, and Y. Cheng. 2021. "Advances in Application of Genome Editing in Tomato and Recent Development of Genome Editing Technology." *Theoretical and Applied Genetics* 134 (9): 2727–47. https://doi.org/10.1007/s00122-021-03874-3.

Xu, R., R. Qin, H. Li, D. Li, L. Li, P. Wei, and J. Yang. 2017. "Generation of Targeted Mutant Rice Using a CRISPR-Cpf1 System." *Plant Biotechnology Journal* 15 (6): 713–17. https://doi.org/10.1111/pbi.12669.

Yarrington, R.M., S. Verma, S. Schwartz, J.K. Trautman, and D. Carroll. 2018. "Nucleosomes Inhibit Target Cleavage by CRISPR-Cas9 in Vivo." *Proceedings of the National Academy of Sciences* 115 (38): 201810062. https://doi.org/10.1073/PNAS.1810062115.

Zetsche, B., J.S. Gootenberg, O.O. Abudayyeh, I.M. Slaymaker, K.S. Makarova, P. Essletzbichler, S.E. Volz, ... F. Zhang. 2015. "Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System." *Cell* 163 (3): 759–71. https://doi.org/10.1016/j.cell.2015.09.038.

Zetsche, B., M. Heidenreich, P. Mohanraju, I. Fedorova, J. Kneppers, E.M. DeGennaro, N. Winblad, ... F. Zhang. 2016. "Multiplex Gene Editing by CRISPR–Cpf1 Using a Single CrRNA Array." *Nature Biotechnology* 35 (December): 31–34. https://doi.org/10.1038/nbt.3737.

Zhu, H., C. Li, and C. Gao. 2020. "Applications of CRISPR–Cas in Agriculture and Plant Biotechnology." *Nature Reviews Molecular Cell Biology* 21 (11): 661–77. https://doi.org/10.1038/s41580-020-00288-9.

Zsögön, A., T. Čermák, E.R. Naves, M.M. Notini, K.H. Edel, S. Weinl, L. Freschi, D.F. Voytas, J. Kudla, and L.E.P. Peres. 2018. "De Novo Domestication of Wild Tomato Using Genome Editing." *Nature Biotechnology*, no. October. https://doi.org/10.1038/nbt.4272.

Zuo, Z., and J. Liu. 2016. "Cas9-Catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations." *Scientific Reports* 5 (October): 1–9. https://doi.org/10.1038/srep37584.

Supplementary Information

Supplementary Information

Supplementary Information 1 – Sequences of arrays synthesized by GenScript for crRNA expression

In all sequences:

yellow : BpiI recognition and cut sites
turquoise: AtU6-26 promoter
pink : direct repeat
gray : spacer

(1/6) Array for AsCas12a, mature crRNAs

gaagacaattacttcaaaagtcccacatcgcttagataagaaaacgaagctgagtttatatacagcta
gagtcgaagtagtgattGTAATTTCTACTCTTGTAGATTGCCACTCGTTTAACTTCTGAGGTAATTTC
TACTCTTGTAGATAGTAAAATGCCTCAAATTGGACTTAATTTCTACTCTTGTAGATAAGGCCCCAAGT
CCTTAACCAATTAATTTCTACTCTTGTAGATAATTCTATAGACTGATGACTTGAtttttttcagatt
gtcttc

(2/6) Array for AsCas12a, precrRNAs

gaagacaattacttcaaaagtcccacatcgcttagataagaaaacgaagctgagtttatatacagcta
gagtcgaagtagtgattGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGATTGCCACTCGTTTAACT
TCTGAGGGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGATAGTAAAATGCCTCAAATTGGACTGTC
AAAAGACCTTTTTAATTTCTACTCTTGTAGATAAGGCCCCAAGTCCTTAACCAATGTCAAAAGACCTT
TTTAATTTCTACTCTTGTAGATAATTCTATAGACTGATGACTTGAtttttttcagattgtcttc

(3/6) Array for FnCas12a, mature crRNAs

gaagacaattacttcaaaagtcccacatcgcttagataagaaaacgaagctgagtttatatacagcta
gagtcgaagtagtgattGTAATTTCTACTGTTGTAGATTGCCACTCGTTTAACTTCTGAGGTAATTTC
TACTGTTGTAGATAGTAAAATGCCTCAAATTGGACTTAATTTCTACTGTTGTAGATAAGGCCCCAAGT
CCTTAACCAATTAATTTCTACTGTTGTAGATAATTCTATAGACTGATGACTTGAtttttttcagatt
gtcttc

(4/6) Array for FnCas12a, precrRNAs

gaagacaattacttcaaaagtcccacatcgcttagataagaaaacgaagctgagtttatatacagcta
gagtcgaagtagtgattGTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATTGCCACTCGTTTAAC
TTCTGAGGGTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATAGTAAAATGCCTCAAATTGGACTG
TCTAAGAACTTTAAATAATTTCTACTGTTGTAGATAAGGCCCCAAGTCCTTAACCAATGTCTAAGAAC
TTTAAATAATTTCTACTGTTGTAGATAATTCTATAGACTGATGACTTGAtttttttcagattgtctt
c

(5/6) Array for LbCas12a, mature crRNAs

gaagacaattacttcaaaagtcccacatcgcttagataagaaaacgaagctgagtttatatacagcta
gagtcgaagtagtgattGTAATTTCTACTAAGTGTAGATTGCCACTCGTTTAACTTCTGAGGTAATTT
CTACTAAGTGTAGATAGTAAAATGCCTCAAATTGGACTTAATTTCTACTAAGTGTAGATAAGGCCCCA
AGTCCTTAACCAATTAATTTCTACTAAGTGTAGATAATTCTATAGACTGATGACTTGAtttttttca
gattgtcttc

(6/6) Array for LbCas12a, precrRNAs

gaagacaattacttcaaaagtcccacatcgcttagataagaaaacgaagctgagtttatatacagcta
gagtcgaagtagtgattGTTTCAAAGATTAAATAATTTCTACTAAGTGTAGATTGCCACTCGTTTAAC
TTCTGAGGGTTTCAAAGATTAAATAATTTCTACTAAGTGTAGATAGTAAAATGCCTCAAATTGGACTG
TTTCAAAGATTAAATAATTTCTACTAAGTGTAGATAAGGCCCCAAGTCCTTAACCAATGTTTCAAAGA
TTAAATAATTTCTACTAAGTGTAGATAATTCTATAGACTGATGACTTGAtttttttcagattgtctt
c

Supplementary Information 2 – Protocol for cloning new spacers in crRNA backbones

Step 1/3. Ordering oligos

Once you've selected your 23 nt target (websites such as http://crispor.tefor.net/ can help you with this), you can order oligos. You need two oligos: one forward, one reverse. To create the correct overhangs, attach 5'-TAGAT-3' at the 5' end of your forward oligo. Attach 5'-AAAA-3' to the 5' end of your reverse oligo and 5'-A-3' to the 3' end of your reverse oligo. The T/A pair (indicated in red below) is needed because the nucleotide was accidentally omitted from the LbCas12a direct repeat sequence while cloning the backbone vectors. It is necessary for proper function, so please make sure to include it.

Example:

You selected the following target (including PAM, underlined here): 5'TTTAAAGGCCCCAAGTCCTTAACCAAT-3'

Your forward oligo will be: 5'-TAGATAAGGCCCCAAGTCCTTAACCAAT-3'.

Your reverse oligo will be: 5'-AAAAATTGGTTAAGGACTTGGGGCCTTA-3'

Step 2/3. Annealing oligos

Annealing oligos can be performed in multiple ways. This is one method:

1. Dilute your oligo's to 10 mM

2. To a PCR tube, add:

Ingredient	**Amount**
FW oligo (10mM)	2 uL
REV oligo (10mM)	2 uL
Annealing buffer*/PCR buffer (10x)	2 uL
MilliQ to 20 uL	14 uL

* 10x annealing buffer: 100mM Tris pH 7.5, 10 mM EDTA, 500 mM NaCl. You can also use a PCR buffer.

3. Mix well, and in a thermocycler, cook the mix and slowly let it cool down to room temperature:

95 degrees - 5 minutes

90 degrees - 1 minute

go back to previous step 40x, each time decrease the temperature 1.5 degrees (=cool to room temperature)

hold forever at 12 degrees

A lot of variations to the abovementioned thermocycler protocol will work.You can even make your mix in a 1.5 mL Eppendorf tube, cook for ~5 min at 95 degrees in a heat block, take it out and let slowly cool to room temperature on a lab bench (~20 min).

Step 3/3. Restriction/Ligation and transformation to E. coli

Pipette & mix together in a PCR tube:

Ingredient	Amount
Cas12aPINK backbone vector (100-200ng)	1 uL
Ligated oligo	2 uL
Buffer Tango*	2 uL
ATP (10 mM)	2 uL
BsmBI*	1 uL
T4 DNA ligase	1 uL
MilliQ to 20 uL	11 uL

* BsmBI is available from different manufacturers. Check which one you have and use the corresponding buffer. The BsmBI (Esp3I) available from ThermoFisher, which cuts optimally in Tango buffer, is very efficient as it cuts at 37 degrees. The BsmBI-v2 available from NEB, which cuts optimally in Buffer 3.1, is less efficient as it has a temperature optimum at 55 degrees (but do NOT use 55 degrees in your restriction/ligation protocol, as it will inactivate the T4 ligase!). This one can still be used, but it might be advisable to plate a bit more of your E. coli transformation.

Perform 10 cycles of restriction/ligation in a thermocycler, followed by a final restriction and heat inactivation steps:

37 degrees - 5 minutes (10x)
16 degrees - 5 minutes (10x)
50 degrees - 10 minutes
80 degrees - 10 minutes
hold at 12 degrees forever

Transform to E. coli (we usually transform 2 uL to 50 uL of electrocompetent E. coli). Plate on LB plates with added carbenicillin (or ampicillin; carbenicillin prevents the formation of satellite colonies). White colonies are correct; pink colonies have not successfully taken up the spacer.

Supplementary Datasets

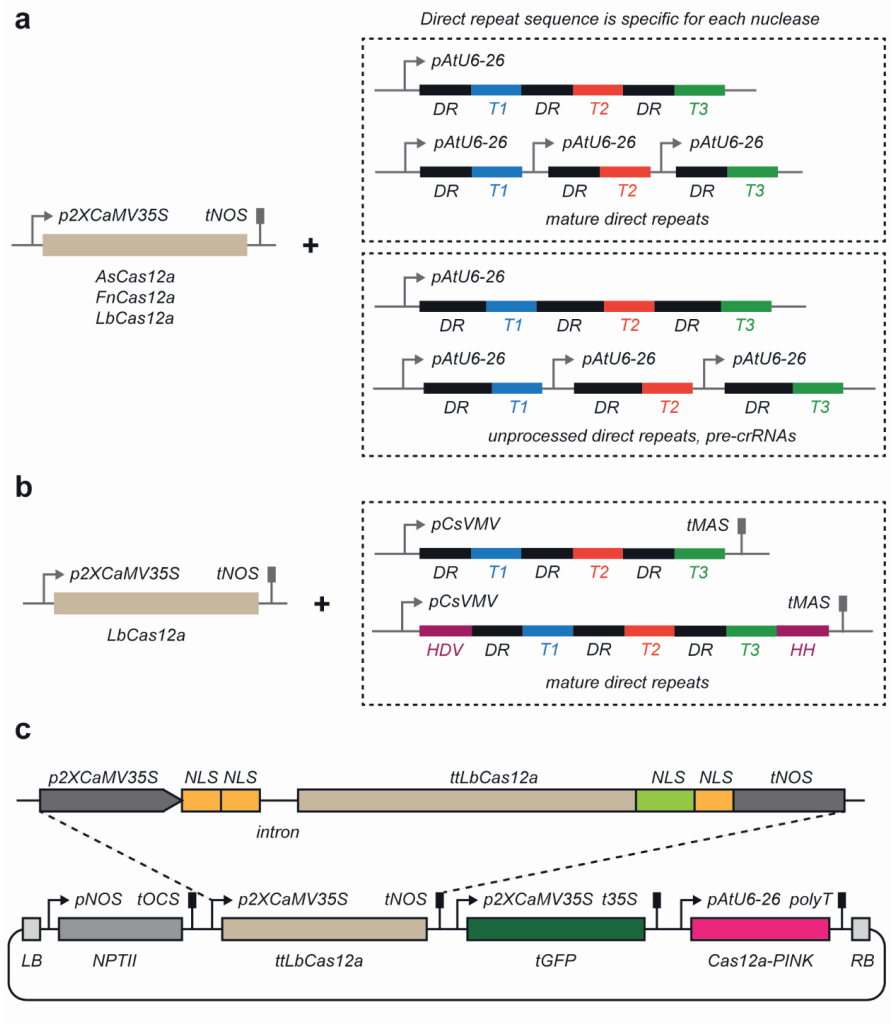
Three Supplementary Datasets belong to this chapter:

Supplementary Dataset 1 – Overview of all selected overlapping target sites and selected off-target sites
Supplementary Dataset 2 – Oligonucleotide sequences for cloning crRNAs and sgRNAs in expression vectors
Supplementary Dataset 3 – Primers used in this study

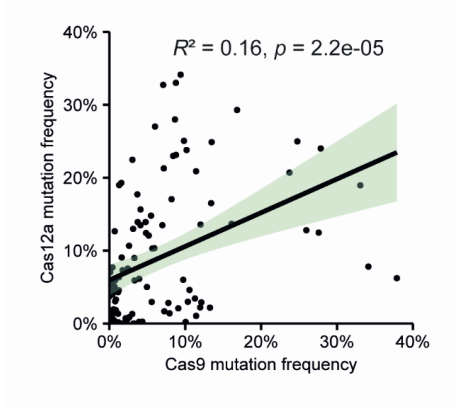
These are available from the author upon request.

Raw sequencing data has been uploaded to the NCBI Sequencing Read Archive, BioProject accession number PRJNA980545.

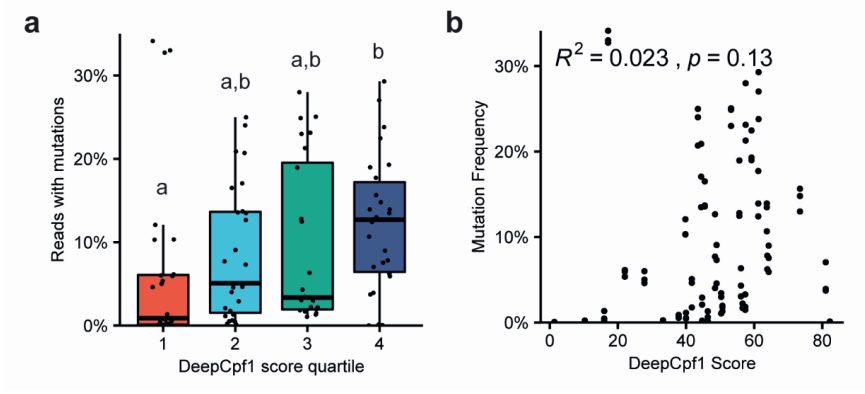
Supplementary Figures



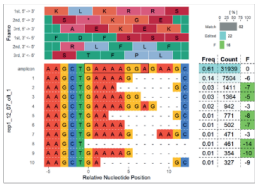
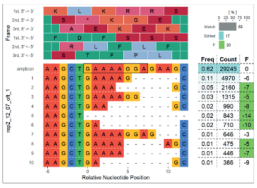
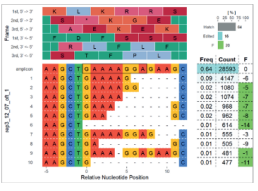
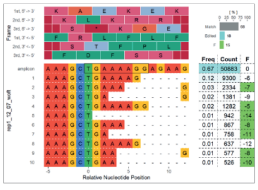
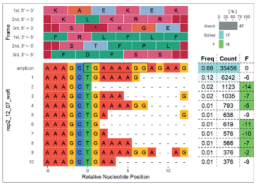
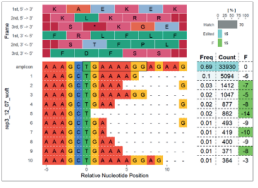
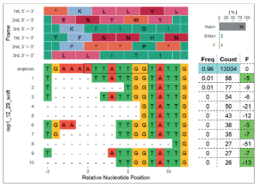
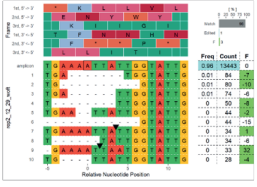
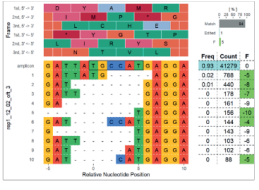
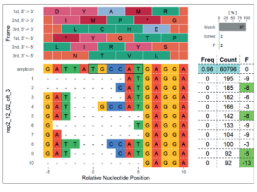
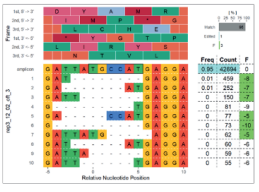
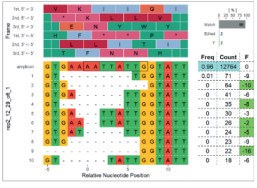
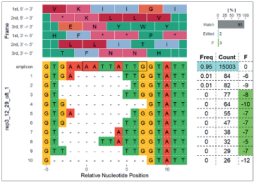
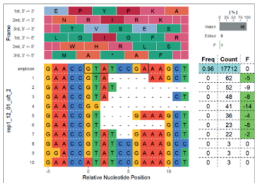
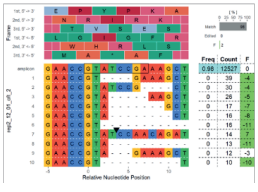
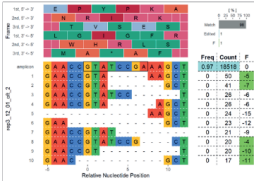
Supplementary Figure 1 - Schematic overview of plasmids used in this study. DR: Direct Repeat, T1-T3: spacers for T1-T3. (a) Plasmids used to identify the most efficient Cas12a orthologue and method for crRNA expression. Each Cas12a orthologue (AsCas12a, FnCas12a and LbCas12a) was tested with four methods of crRNA expression. Either the mature, or the unprocessed pre-crRNA form of the direct repeat was used. The crRNAs were either expressed together as an array, or each separately by an AtU6-26 promoter. Twelve plasmids were constructed in total. (b) For LbCas12a, two additional methods of crRNA expression were tested: an array expressed by the PolII CsVMV promoter, and an array expressed by the PolII CsVMV promoter and flanked by ribozymes. HDV: Hepatitis Delta Virus ribozyme, HH: hammerhead ribozyme. (c) Binary level 2 vector used for direct cloning of a single crRNA. In this vector, the thermotolerant version of LbCas12a was used (Schindele and Puchta, 2020). To this version, two extra SV40 nuclear localization sites (NLS, yellow) were added at the 5' end, together with a potato IV2 intron. At the 3' end, there is one nucleoplasmin NLS (green) and another SV40 (yellow) NLS. Single crRNA spacers can be cloned in the Cas12a-PINK cassette using BsaI, as NPTII contains a BsmBI site. This is different from the level 1 Cas12a-PINK cassettes, where BsmBI is used.



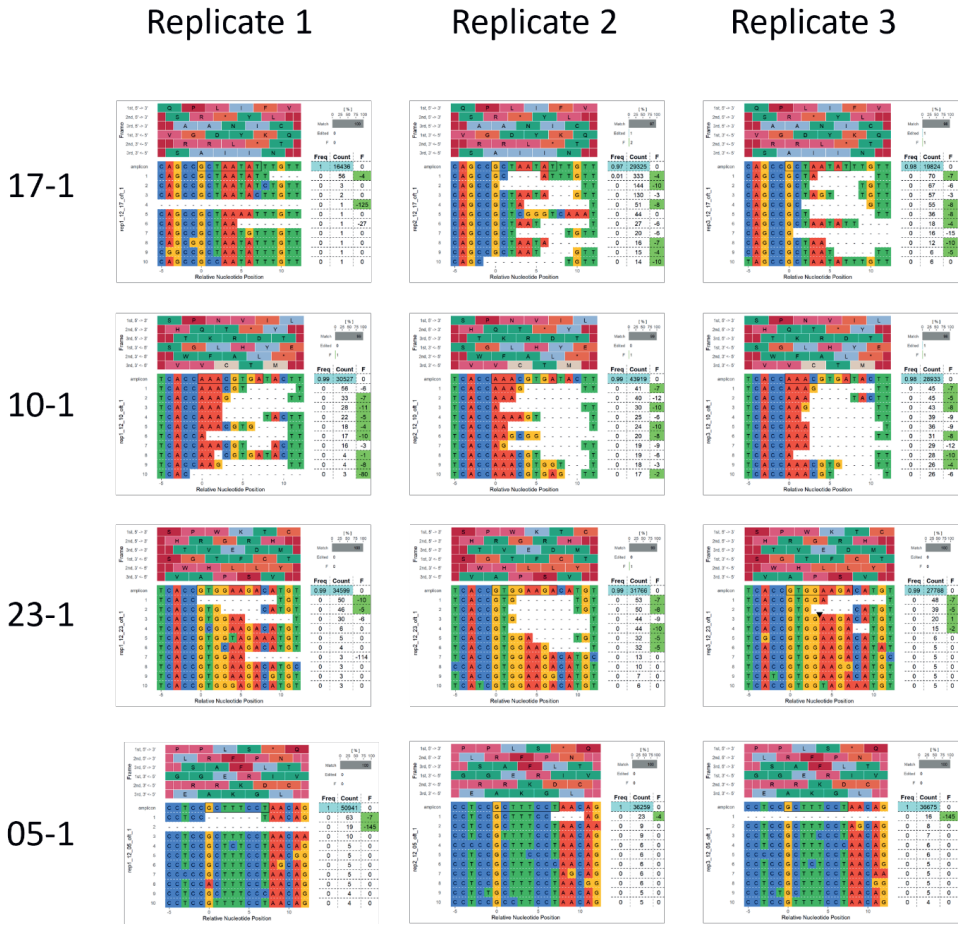
Supplementary Figure 2 - Pearson correlation of Cas12a and Cas9 mutation frequencies. The green shaded area indicates the 95% confidence interval.



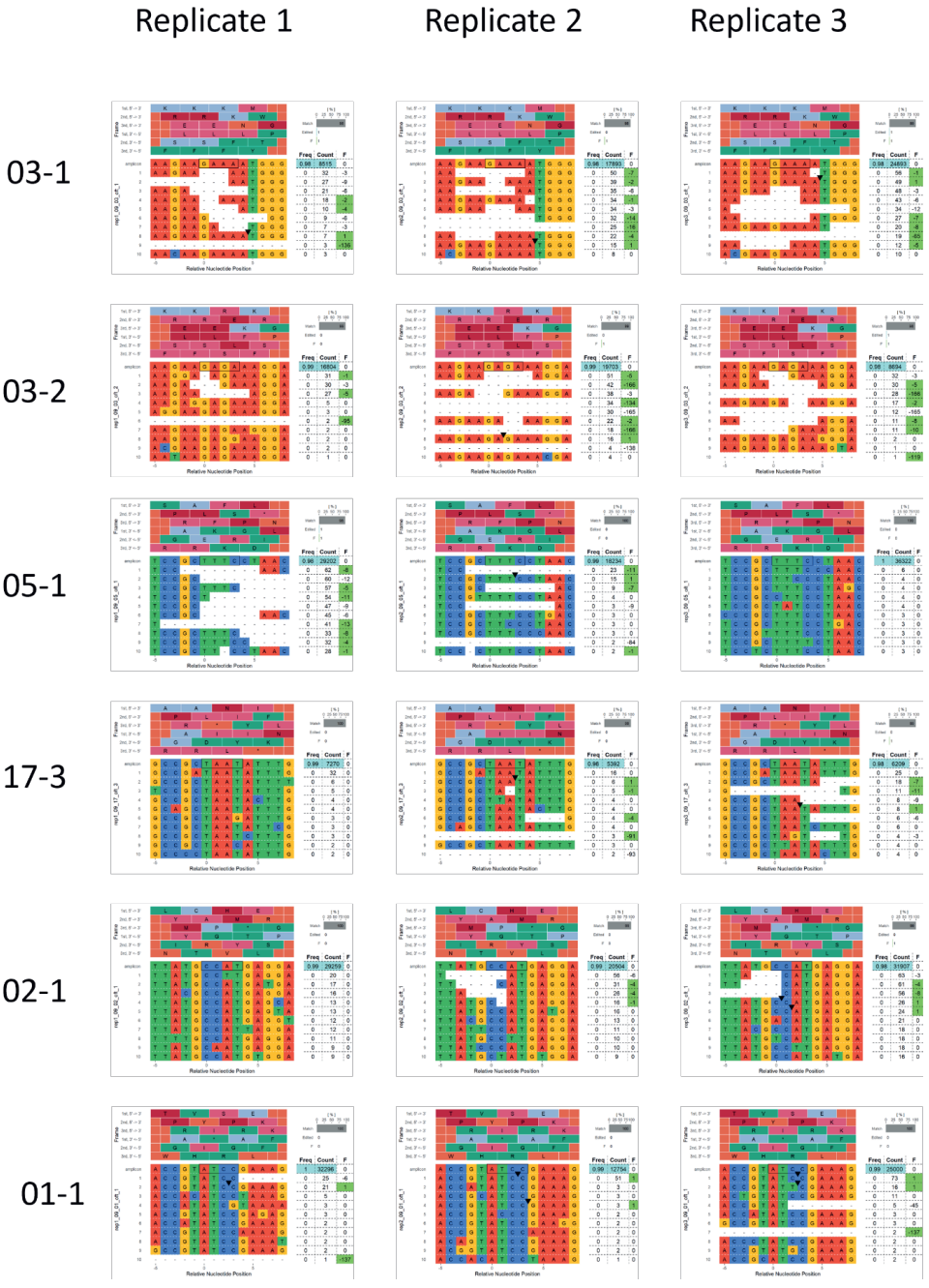
Supplementary Figure 3 - Correlation of predicted and measured activity of Cas12a crRNAs. (a) The crRNAs were divided in quartiles based on their DeepCpf1 score. The horizontal line in the box plots indicates the median, boxes represent the 2nd and 3rd quartile, and bottom and top whiskers indicate the 1st and 4th quartile, respectively. Significant differences between mutation frequencies of sgRNAs in each quartile were determined by Kruskal-Willis test, followed by Wilcoxon's Rank Sum test for pairwise comparison. Different letters indicate significant differences between groups (p < 0.05). (b) Scatterplot showing the distribution and Pearson correlation of DeepCpf1 score and obtained mutation frequencies.

	Replicate 1	Replicate 2	Replicate 3
07-1			
07-w1			
29-w1			
02-3			
29-1			
01-2			

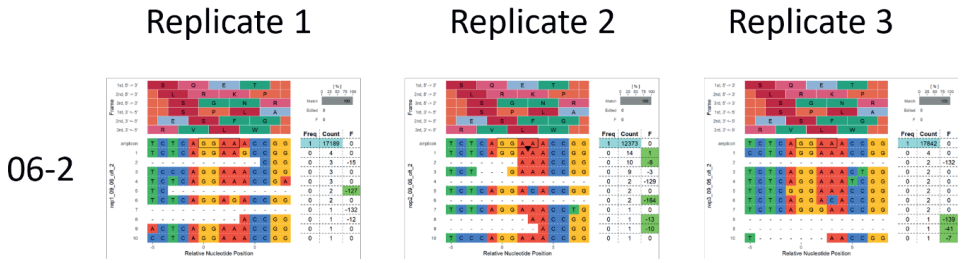
(Figure continues on next page)



Supplementary Figure 4 - Variation plots of identified Cas12a off-target sites.



(Figure continues on next page)



Supplementary Figure 5 – Variation plots of identified Cas9 off-target sites.

Chapter 4

Unbiased CRISPR-Cas9 off-target activity detection in tomato protoplasts using GUIDE-seq

E. Slaman[1,2], J. Muraoka[2], G. C. Angenent[1,2], R. A. de Maagd[2]

[1] Wageningen University & Research, Laboratory of Molecular Biology, Wageningen, the Netherlands

[2] Wageningen University & Research, Bioscience, Wageningen, the Netherlands

Abstract

In most applications, CRISPR-Cas-mediated genome editing relies on the induction of DNA double-strand breaks (DSBs). DSBs are mutagenic, and therefore, it is desirable to prevent the occurrence of off-target DSBs. Most off-target mutation detection is biased: only the positions that are predicted to be likely off-target sites are studied. In plants, unbiased off-target activity detection methods have only been applied on a limited scale. GUIDE-seq, developed for cultured mammalian cells, is an unbiased off-target activity detection method that relies on capturing a double-stranded oligodeoxynucleotide (dsODN) in DSBs. It has the advantage of being a high-throughput, *in vivo* method capable of detecting all DSBs occurring in an experiment, whether by CRISPR or otherwise. In this study, we optimized and applied GUIDE-seq to detect off-target activity in tomato protoplasts. Only previously predicted off-target sites were identified for all six tested Cas9 sgRNAs, demonstrating the specificity of CRISPR-Cas9 mediated genome editing and the value of specificity prediction algorithms. We detected off-target mutations with a minimum mutation frequency of 0.7% or more in protoplast pools. This detection limit is higher than previously reported in mammalian cells (~0.1%). Nevertheless, we show for the first time that GUIDE-seq can be used for unbiased off-target activity detection in plants.

Introduction

CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated protein) mediated genome editing has been widely applied to many species in the tree of life (Doudna and Charpentier 2014; Schiml and Puchta 2016; Jaganathan et al. 2018). Integral to these systems is the induction of a double-stranded break (DSB) at a specific position in a genome. This break is produced by a Cas endonuclease and a complexed guide RNA that can recognize and bind target DNA through Watson-and-Crick base pairing. The most frequently-used Cas endonucleases are SpCas9 and various Cas12a orthologues (the latter previously called Cpf1). SpCas9 requires a 3' "NGG" protospacer adjacent motif (PAM) immediately following a target site to be able to bind and cleave DNA and produces mostly blunt DSBs (Jinek et al. 2012). Cas12a uses a T-rich 5' "TTTV" PAM and induces a staggered DSB (Zetsche et al. 2015).

DSBs are detrimental to a cell's survival, so DSB repair mechanisms have evolved. The two most well-known repair pathways are canonical or classic non-homologous end-joining (cNHEJ) and homology-directed repair (HDR). In cNHEJ, DNA strands are joined together in a Ku70-Ku80 and ligase IV-mediated way, in which the binding of Ku70 and Ku80 to the free DNA ends prevents end-resectioning of the DNA (Tamura et al. 2002; H. Puchta 2004). Although mostly precise, small insertions and deletions (indels) may arise in the process, which can result in the loss of the gene's function. In homology-directed repair, which is regarded as precise, extensive resectioning of the DNA ends occurs. Strand invasion of the sister chromatid and subsequent DNA synthesis will result in perfect, templated repair. Instead of a sister chromatid, donor DNA can be introduced as a repair template, resulting in targeted insertion or specific replacement of genetic material. HDR is often considered the "holy grail" of genome editing, but especially in plants remains elusive and is usually superseded by the NHEJ pathway (H. Puchta 2002; H. Puchta 2004; Steinert et al. 2016).

Two other DSB repair mechanisms that result in imperfect repair are the single-strand annealing (SSA) and alternative end-joining (a-EJ) pathways, of which the latter is also known as microhomology-mediated end joining (MMEJ). In both pathways, the mechanism results in the loss of genetic information. For SSA, extensive resectioning of the DNA ends takes place. Instead of annealing with template DNA, the single-stranded DNA will anneal with a homologous repeat sequence that flanks the DSB. Nucleotides between these repeat sequences are lost in subsequent DNA repair (H. Puchta 2004). The a-EJ pathway functions very similarly, but end resectioning is less extensive, and used homologies are usually short (1-20 bp, hence the name "microhomology-mediated end joining") and closer to the DSB (Vu et al. 2021). Both pathways, especially a-EJ, are

active in plants in addition to NHEJ and HDR (Vu et al. 2021; Tan et al. 2020; Bernabé-Orts et al. 2019; Ohtsuki et al. 2021; Shen et al. 2016).

In conclusion, DSBs are intrinsically mutagenic and may have genome-wide effects. In genome editing, this is generally undesirable, particularly at positions other than the intended target site (so-called off-target sites). Every tool for genome editing should therefore be as specific as possible in its DSB induction.

In mammalian research, many large-scale studies on the specificity of CRISPR-Cas mediated genome editing have been performed, especially given possible future therapeutic applications (Hsu et al. 2013; Pattanayak et al. 2013; Lin et al. 2014; Fu et al. 2013; Cho et al. 2014; Kim et al. 2016). In plants, available studies are more limited (Sturme et al. 2022).

Reliable methods for detecting DSBs generated at off-target sites are essential for determining the specificity of genome editing tools. These methods can be broadly divided into two categories: biased and unbiased. In biased off-target detection, off-target sites of a chosen sgRNA are predicted using readily available prediction algorithms such as CRISPR-P, CRISPOR, or CasOFFinder (Liu et al. 2017; Concordet and Haeussler 2018; Bae et al. 2014). After the induction of CRISPR mutations, these predicted off-target sites can be screened for the presence of actual mutations. Frequently used methods are T7 endonuclease assays (Mashal et al. 1995), Sanger sequencing and subsequent decomposition of the trace files (Brinkman et al. 2014; Dehairs et al. 2016; Bloh et al. 2021), or high-throughput amplicon sequencing (see Chapters 2 and 3 of this thesis). These methods are easy and cheap to perform, yield valuable information, and are applicable to plants. However, the detection of off-target mutations is limited to predicted off-target sites, based on the assumption that off-target mutations mostly or only occur at sites that show close sequence similarity to the target site. Additionally, larger deletions or rearrangements may be overlooked, as they may cause the loss of primer binding sites and thus prevent amplification of the region of interest.

Unbiased off-target activity detection methods screen the entire genome for off-target mutations and are not limited by prediction algorithms. An obvious straightforward approach is whole genome resequencing (WGS), which has been applied to various plant species such as *Arabidopsis*, tomato, grapevine, and cotton (Tang et al. 2018; Li et al. 2019; Z. Feng et al. 2014; Wang et al. 2021; Peterson et al. 2016). This technique is excellent for getting an overview of on- and off-target mutations that have occurred in a single individual but less suitable for getting a complete overview of all off-target DSBs that can be induced by a single guide RNA in a larger group of individuals (see Chapter 5 of this thesis). Particularly for guide RNAs that only induce off-target mutations at low frequencies, many

individuals need to be screened to establish those frequencies. Although sequencing costs are decreasing rapidly, they are still prohibitive for screening large populations.

Higher throughputs can be obtained by other methods such as Digenome-Seq (Kim et al. 2015), CIRCLE-seq (Shengdar Q Tsai et al. 2017), SITE-seq (Cameron et al. 2017), LAM-HTGTS (Hu et al. 2016), BLISS (Yan et al. 2017) and GUIDE-seq (S Q Tsai et al. 2015). GUIDE-seq is based on the integration, by NHEJ, of a short, double-stranded oligodeoxynucleotide (dsODN) at DSBs in the genome. These dsODNs can be delivered to cells simultaneously with ribonucleoproteins (RNPs) or plasmid encoding the CRISPR-Cas components. After transfection and incubation of the cells, genomic DNA is purified, fragmented, and Y-adapters are ligated to the free ends. Enrichment for DNA fragments with integrated dsODNs is then achieved by performing two nested rounds of PCR using one primer that hybridizes with the single-stranded part of the Y-adapter and one that fits on the integrated dsODN (Figure 1).

In this study, we optimized the GUIDE-seq protocol for use in tomato protoplasts. We optimized dsODN integration at DSBs induced at Cas9 and Cas12a target sites. We reliably detected dsODN integration at target sites and known off-target sites using a targeted approach for all three tested sgRNAs. In unbiased off-target detection of six sgRNAs, activities at target sites and known off-target sites were confirmed, but no new, unpredicted off-target sites were identified. We conclude that GUIDE-seq is applicable to plant protoplast systems and that unpredicted off-target mutations do not occur in tomato protoplasts at frequencies above 0.7%. Additionally, we propose several methods to increase the sensitivity of this method.

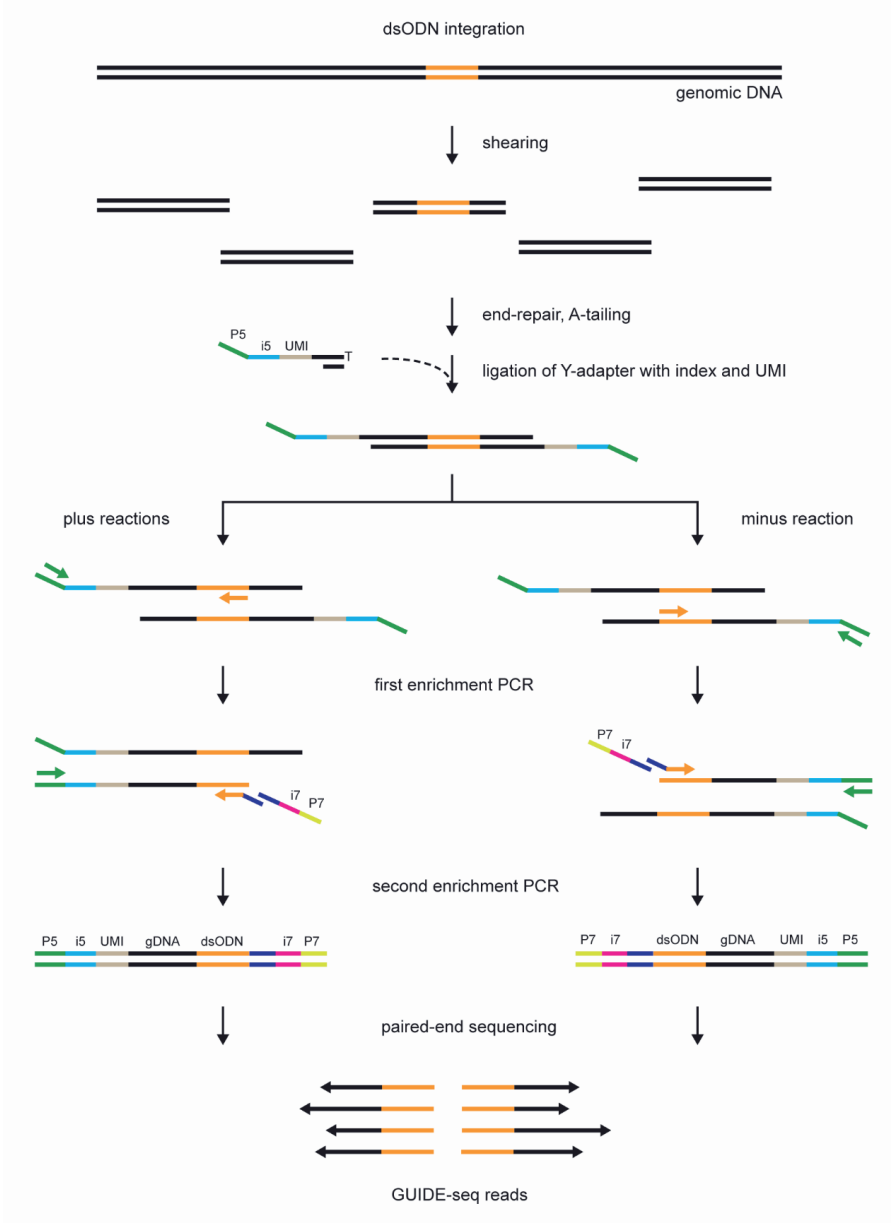


Figure 1 - Graphical overview of GUIDE-seq library preparation. Double-stranded oligodeoxynucleotides (dsODNs) are incorporated at DSBs in the genomic DNA. Next, genomic DNA is purified, fragmented, end-repaired, A-tailed, and Y-adapters containing unique molecular identifiers (UMIs) and an i5 index are ligated. Two rounds of nested PCR enrich for DNA fragments with dsODN integration and add an i7 index. As dsODNs can integrate in forward and reverse orientations, the sample is split in "plus" and "minus" reactions. This graphical overview is based on Malinin et al., 2021. "gDNA": genomic DNA flanking the DSB

Materials and Methods

Plasmid construction and DNA preparation

The construction of plasmids targeting *SlPDS* or tomato MYB genes by SpCas9 was described in Chapter 3 and Chapter 2, respectively. Highly pure DNA for transfection was prepared from 25 mL of overnight *E. coli* culture in LB medium using the Qiagen MidiPrep kit.

Protoplast isolation and transfection

Protoplast isolation was performed as described in Chapter 2. Transfection was based on the protocol from Sheen and colleagues (Sheen 2002). Briefly, for plasmid transfection, 10 ug of plasmid DNA and the desired amount of dsODN were added to a 2 mL tube and brought to a total volume of 20 uL. To the transfection mix, 200 µL of protoplast suspension in magnesium-mannitol solution (0.4 M mannitol, 15 mM $MgCl_2$, 4 mM MES, pH 5.7) at a density of 6.0 x 10^5 / mL was added. Subsequently, 200 µL of PEG solution (4.0 g PEG-4000 (Fluka), 3.0 mL MilliQ, 2.5 mL 0.8M mannitol, 1.0 mL 1M $CaCl_2$) was added and mixed with the protoplast suspension by slowly inverting the tube to start transfection. After incubating for 10 minutes, the transfection was stopped by slowly adding 1 mL of W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2 mM MES, pH 5.7). Protoplasts were pelleted by centrifugation for 3 min at 200xg. The supernatant was removed, and protoplasts were resuspended in 1 mL W5 solution to wash away the remaining PEG. Protoplasts were again pelleted by centrifugation, resuspended in 150 µL WI solution (0.5 M mannitol, 20 mM KCl, 4 mM MES, pH 5.7), and incubated for 24 hours, in the dark, at 25°C.

Direct PCR to detect dsODN integration

After a 24h incubation, transfected protoplasts were pelleted. Subsequently, 1 uL of the pellet was used as a template in a Phire Plant Direct PCR (Thermo Scientific) using primer combinations P1 and dsODN1, and P2 and dsODN1 (see Supplementary Table 1). The thermocycler protocol was as follows: initial denaturation for 5 min at 98°C, followed by 40 cycles of denaturation at 98°C for 5 s, annealing for 5 s at 58°C, extension for 25 s at 72°C, and a final extension step of 1 min at 72°C.

Protoplast DNA isolation and amplicon sequencing

Protoplast DNA was purified 24h after transfection using magnetic beads (NucleoMag Plant, Macherey-Nagel), following the manufacturer's instructions. DNA was eluted in 50 µL, of which 6 µL was subsequently used as a template in 25 µL PCR reactions using PHUSION HotStart Flex DNA

polymerase (NEB) to amplify genomic DNA fragments containing target or predicted off-target sites. For the PCR, an initial denaturation for 30 s at 98 °C was followed by 38 cycles of denaturation for 10 s at 98 °C, annealing for 20 s at 58 °C, extension for 20 s at 72 °C, and a final extension step of 3 min at 72 °C. Primer sequences are listed in Supplementary Table 1. The resulting PCR products were visualized by electrophoresis on a 2% agarose gel. Equal amounts of PCR products were pooled for the production of sequencing libraries. Libraries were subsequently column-purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel), following the manufacturer's instructions. Illumina NovaSeq sequencing (paired-end, 2x150 bp reads) was performed by Eurofins Genomics Europe Sequencing GmbH, Constance, Germany.

Amplicon Sequencing Analysis

Paired sequencing reads were uploaded to the CLC Genomics Workbench v20, trimmed, merged, and demultiplexed using standard settings. Mutation frequencies at target and predicted off-target sites were determined using Amplican (Labun et al. 2019). The frequency of dsODN integrations was determined using a custom Python script. For Cas9 target sites, a version of the script was used that uses the Amplican-output file "events_filtered_shifted_normalized.csv" as input. For Cas12a target sites, this script version did not correctly identify integration lengths, as Amplican failed to correctly align the deletion-integration combinations that often arose in Cas12a-mediated integration of dsODNs. Therefore, a different version of the script was used that directly used fastq files as input data.

GUIDE-seq library preparation, sequencing and data analysis

For unbiased off-target detection, protoplast transfections were performed as described previously. For every sgRNA, four transfections were performed and in each transfection, 5 pmol of end-protected dsODN was added, followed by 24h incubation at 25°C. Subsequently, the four transfections per sgRNA were pooled, and genomic DNA was isolated using the NucleoMag Plant kit (Macherey-Nagel). Library preparation and MiSeq sequencing were subsequently performed as described by Malinin et al. 2021, from step 20 onwards. For data analysis, the GUIDE-seq pipeline available at github.com/tsailabSJ/guideseq was used. PCR duplicates were removed using UMIs, reads aligned to the reference genome, DSB sites were determined and visualized, and (off-)target sites were visualized. Minor changes were made to the script for read consolidation as MiSeq output data was obtained that contained UMI tags in the read header instead of in a separate index fastq file. Additionally, we changed the use of "itertools.izip" to the built-in Python3 "zip". As a reference

genome for read alignment, build Sl4.00 of the tomato reference genome was used (available on solgenomics.net). The plasmid sequence used for transfection was manually added to the genome fasta file as an additional chromosome. MACS2 was used for additional peak calling on aligned GUIDE-seq reads (Zhang et al. 2008; J. Feng et al. 2012). As too few peaks were available for modelling, the options -g 7.5e+08 --nomodel --extsize147 were used.

Results

dsODNs are integrated at CRISPR-Cas induced DSBs

For the unbiased detection of DSBs, including those at off-target positions, GUIDE-seq relies on the high-frequency integration of an intact dsODN in those DSBs (Figure 1). Therefore, we first investigated whether intact dsODNs could integrate at Cas-targeted sites of tomato protoplasts. We selected two Cas9 and one Cas12a target site in tomato *PHYTOENE DESATURASE (SlPDS, Solyc03g123760)* (Figure 2a). We then used PCR to determine if a dsODN had been integrated by using PCR with one primer that anneals to the dsODN, and the other upstream or downstream of the targeted site (Figure 2a,b, primers *P1* and *P2*). We transfected protoplasts with a plasmid encoding the CRISPR-Cas machinery and co-transfected with a 5' phosphorylated dsODN that was either end-protected with phosphorothioate bonds or unprotected (Figure 2c). PCR was performed directly on the transfected protoplasts 24h after transfection, without a DNA purification step. Subsequent gel analysis revealed that both types of dsODN could integrate at target sites, in both orientations, for Cas9- and Cas12a-induced DSBs (Figure 2d). Additionally, using the fluorescence of the tGFP marker, we checked the transfection efficiency of protoplasts with these plasmids, both with and without dsODN (Figure 2e). Transfection efficiencies were similar, indicating that the addition of the dsODN did not negatively affect transfection or survival.

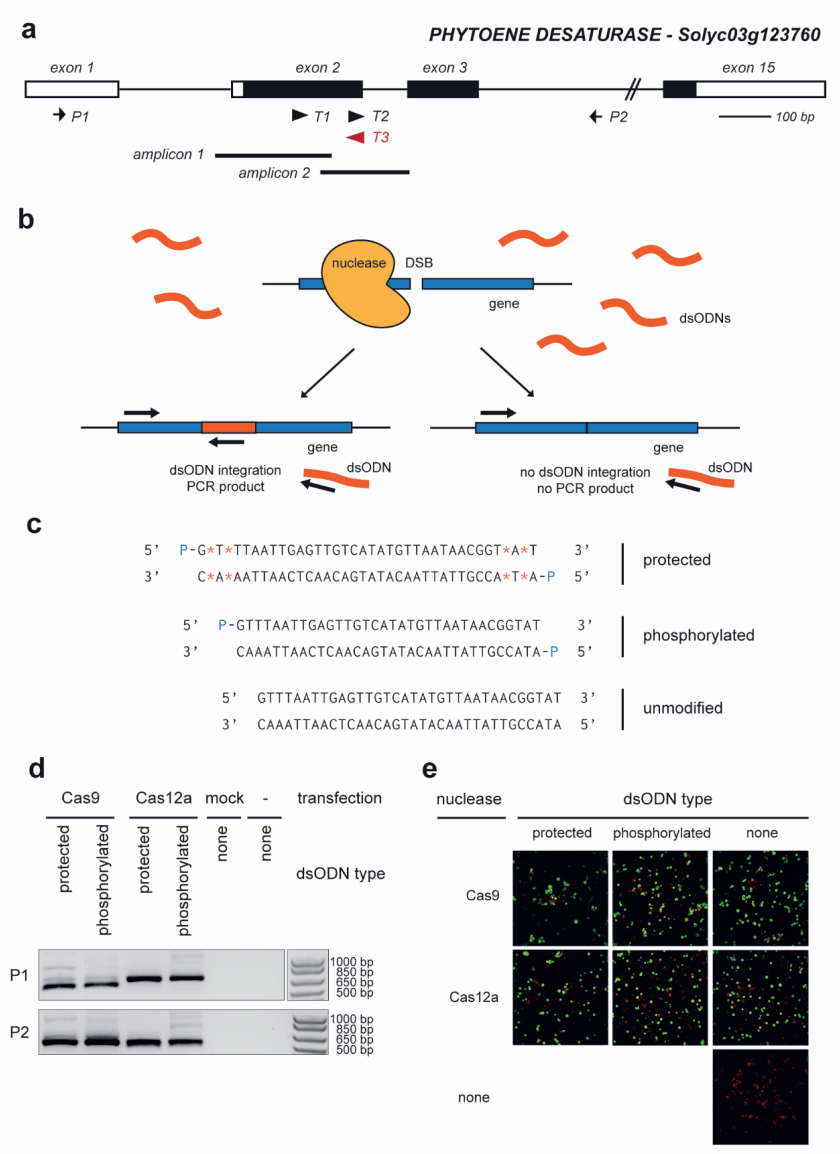


Figure 2 - dsODNs are integrated at CRISPR-Cas9 and CRISPR-Cas12a target sites. (a) Overview of tomato PHYTOENE DESATURASE. Black arrows indicate CRISPR-Cas9 target sites T1 and T2, a red arrow indicates the CRISPR-Cas12a target site T3. Primers P1 and P2 were used for the experiment, as shown in (b). Amplicons used for sequencing are indicated. (b) PCR strategy for demonstrating successful integration of a dsODN at CRISPR-Cas target sites. (c) Overview of the different types of dsODNs used in this study. Orange asterisks indicate phosphorothioate bonds between nucleotides. Blue "P" indicates phosphorylation. (d) Agarose gel image showing the integration of of dsODNs at CRISPR-Cas9 and CRISPR-Cas12a target sites, using the strategy as depicted in (b). Binding sites for primers P1 and P2 are shown in (a). In addition to either P1 or P2, primer dsODN1, binding to the integrated dsODN, was used (Supplementary Table 2). (e) Overlay confocal images of protoplast transfections using Cas9- or Cas12a-encoding plasmids, with and without protected or phosphorylated dsODN. Successfully transfected protoplasts show green turboGFP fluorescence. Non-transfected protoplasts only show red chlorophyll autofluorescence.

The addition of dsODN decreases mutation frequency

To determine more exact frequencies of dsODN integration, we prepared amplicons of the two Cas9 target sites in Sl*PDS* for high-throughput sequencing (Figure 2a). We transfected protoplasts with a plasmid targeting *SlPDS* and 100 pmol of 5'-phosphorylated and end-protected, 5'-phosphorylated-only, and unmodified dsODNs. We then isolated the genomic DNA of the obtained protoplast pools and amplified the CRISPR-Cas9 target sites using barcoded primers. Mutant alleles were determined using AmpliCan (Labun et al. 2019). A custom Python script was used to identify sequencing reads that contained a (partial) dsODN integration. Interestingly, the inclusion of dsODN in the transfection significantly decreased the mutation frequency at both target sites. This effect was most severe for the end-protected version (Supplementary Figure 1a). The dsODNs that were merely 5'-phosphorylated or unmodified had less severe and comparable effects on mutation frequency. As a result, only a very small percentage of total reads belonging to protoplast pools contained a (partial) dsODN integration (Supplementary Figure 1b). We hypothesized that if we could restore the mutation frequency, we could significantly increase the amount of dsODN integrated at target sites. Therefore, we performed another next-generation sequencing experiment, titrating the amount of dsODN added to the protoplasts. We determined the mutation frequencies at *T1*, which was found to be the more efficiently mutated target in the previous experiment. We observed that decreasing amounts of added dsODN indeed resulted in restored mutation frequencies (Figure 3a).

The use of end-protected dsODNs results in the highest frequency of complete integrations

To allow primer annealing during GUIDE-seq library preparation (Figure 1), almost the entire length of the dsODN needs to be incorporated at the DSB. Closer inspection of sequencing reads containing dsODN insertions revealed that imperfect, truncated insertions frequently occurred (Figure 3b), suggesting that the dsODN are partially degraded before or during integration at the DSB. The percentage of reads in each protoplast pool that contained perfect integrations was highest for protected dsODN (Figure 3c). Overall, 5 to 10 pmol of end-protected dsODN gave the best combination of integration and integrity of the dsODN at DSBs. For phosphorylated and unmodified dsODN the optimum integration frequency was obtained with a ten times higher concentration (50 to 100 pmol) of dsODN, but the majority of the integrations contained partial dsODN sequences.

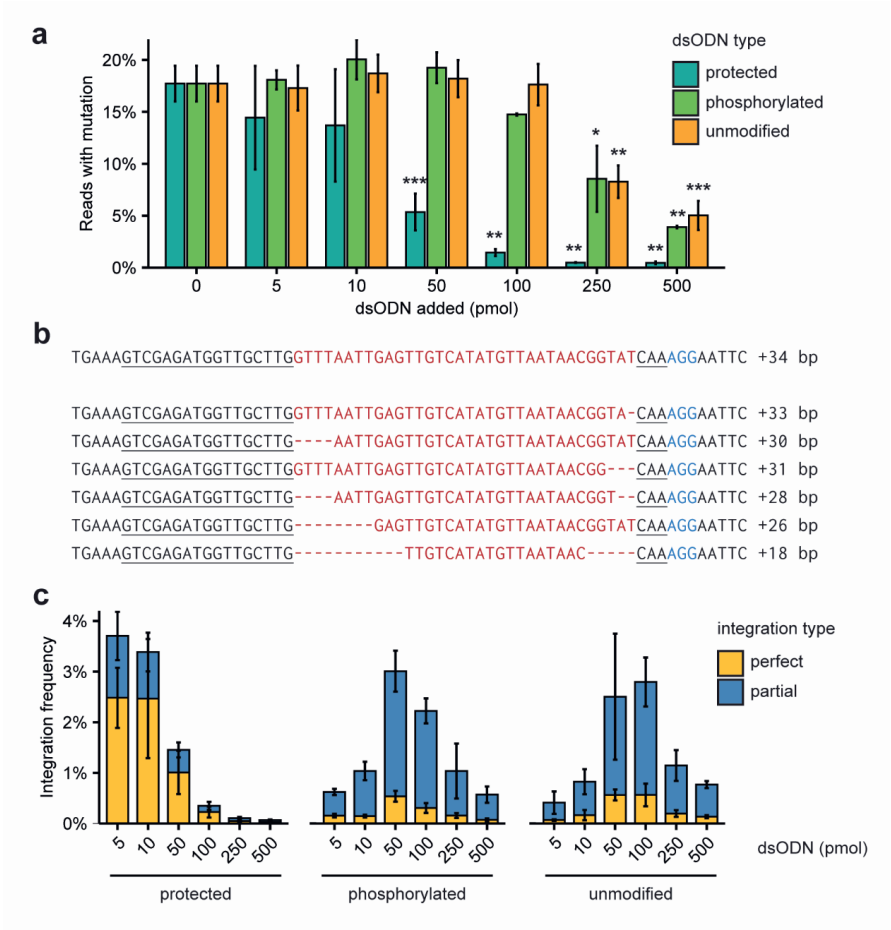


*Figure 3 - Optimization of perfect dsODN integration at CRISPR-Cas9 target site T1. (a) The effect of dsODN concentration on mutation frequency. Error bars indicate SD (n = 3). Significant differences to the control (0 pmol dsODN) are indicated by asterisks and were calculated using Student's T-Test. * p < 0.05; ** p < 0.01; *** p <0.001. (b) Examples of (partial) dsODN integrations as found in protoplast pools transfected with 5 pmol of unmodified dsODN. Red nucleotides indicate the dsODN integration, blue nucleotides indicate the PAM, and underlined nucleotides indicate the protospacer. (c) Percentage of complete and partial integrations in protoplast pools with different types and concentrations of dsODN. Error bars indicate SD (n = 3).*

We repeated this optimization for Cas12a. Based on the results from the Cas9 experiment, we added a range of 0.5 – 50 pmol of end-protected dsODN to the transfections. As for Cas9, the mutation frequency decreased significantly at 50 pmol (Figure 4a). Again, the highest percentage of perfect integrations was achieved with 5 to 10 pmol of protected dsODN (Figure 4b). Possibly as a consequence of the Cas12a-induced staggered breaks with overhangs, the precise integration site of the dsODN relative to the cut site was variable, and integrations were frequently accompanied by deletions in the DNA flanking the cut site (Figure 4c).

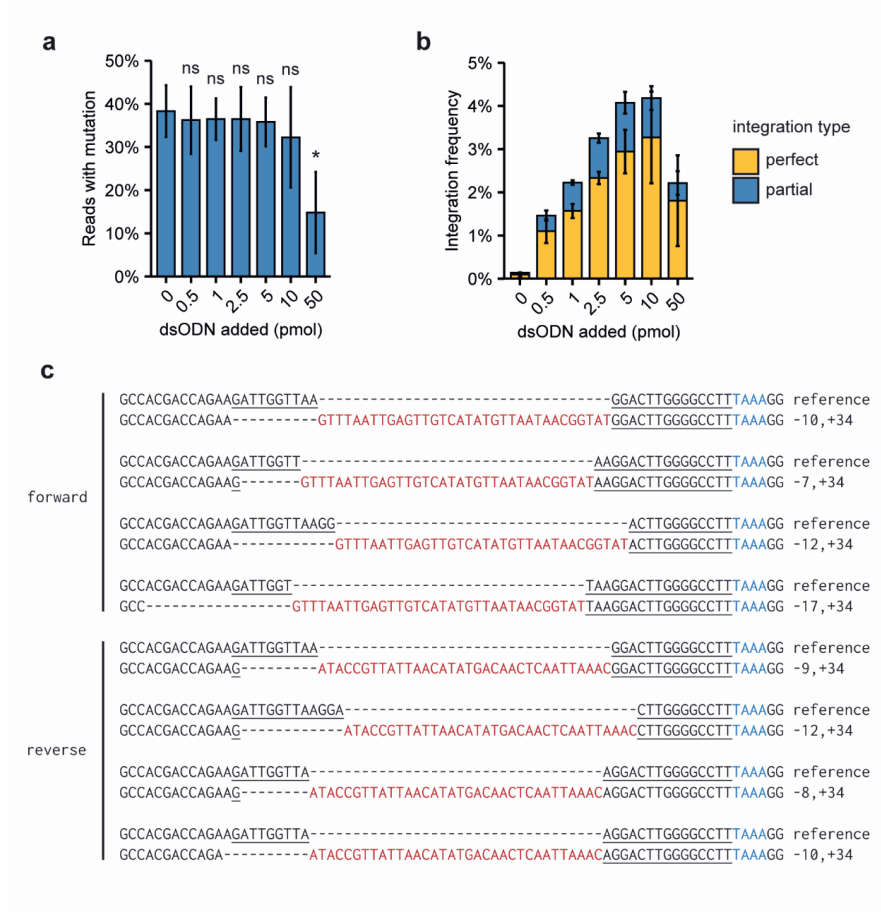


*Figure 4 - Optimization of dsODN integration at Cas12a target site T3. (a) The effect of the amount of end-protected dsODN added to transfection on the mutation frequency. Error bars indicate SD (n = 3). Significant differences to the control (0 pmol) were determined using Student's T-Test and are indicated by an asterisk. * $p < 0.05$. (b) Percentage of complete and partial dsODN integrations in protoplast pools at different dsODN concentrations. (c) Examples of perfect dsODN integration in the forward and reverse directions in samples transfected with 5 pmol end-protected dsODN. Red nucleotides indicate the dODN integration, blue nucleotides indicate the PAM, and underlined nucleotides indicate the protospacer.*

Successful off-target activity detection using dsODN integration

After determining the optimal conditions for dsODN integration, we performed a proof-of-principle experiment to test our ability to identify CRISPR-Cas9 sites with off-target activity. Previously, we identified several *MYB* gene-targeting gRNAs that showed evidence of off-target activity (Chapter 2 of this thesis). Three of these gRNAs (Figure 5a) were used for the detection of dsODN integration at the previously identified on- and off-target mutation sites using targeted amplicon sequencing. Protoplasts were transfected with a plasmid encoding Cas9, an sgRNA, and 5 pmol of end-protected dsODN in three independently replicated experiments. Genomic DNA was isolated from these

protoplast pools, and the target site and previously identified off-target sites were amplified and submitted to high-throughput amplicon sequencing. First, we confirmed the occurrence of on- and off-target mutations (Figure 5b). Next, we determined the frequency of perfect dsODN integration at these sites (Figure 5c). For most off-target sites, complete dsODN integrations could reliably be detected at a frequency of 0.1% or higher. The only exception was MYB54-OffT2, whose integration frequency varied around 0.01%.

Interestingly, although the on-target mutation frequency was higher for MYB54 than for MYB19 (51% ± 8.7% and 21% ± 4%, respectively, Figure 5b), perfect dsODN integration frequencies at those sites were much more similar (13% ± 2.0% and 12% ± 1.4%, respectively) (Figure 5c). To investigate this further, we calculated the percentage of reads that contained perfect integrations as a fraction of all mutated reads for the *MYB*-targeting sgRNAs, as well as the previously tested Cas9-target T2 and the Cas12a-target T3 in *SlPDS* (Figure 2a). We observed that these percentages could differ significantly depending on the guide (Figure 5d).

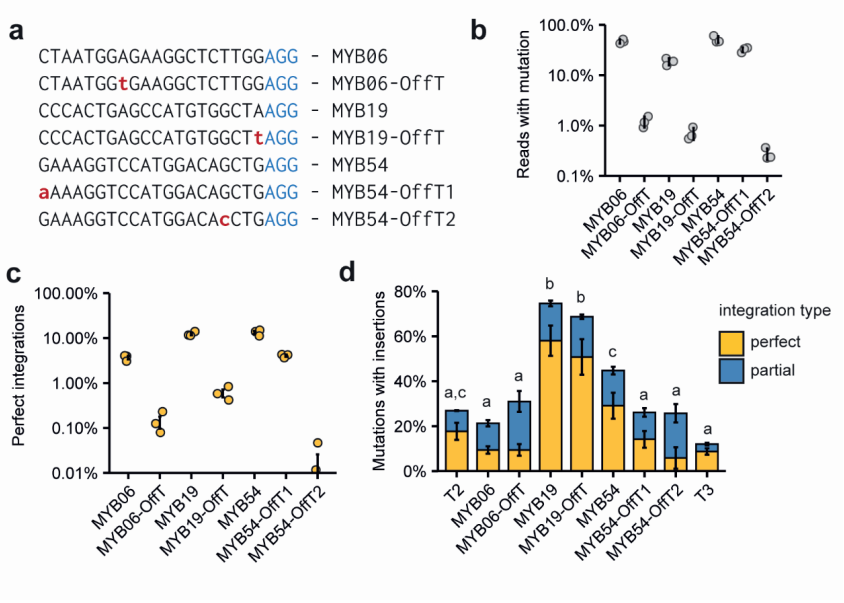


Figure 5 - Detection of dsODN integration at known off-target sites. (a) Overview of the sequence of the target sites of the three sgRNAs used, and of their previously identified off-target sites. The PAM is indicated in blue. Mismatches to the target site are indicated by bold red lowercase. (b) Mutation frequencies at target and off-target sites. Individual data points are shown, bars indicate SD. (c) Percentage of complete dsODN integration at target and off-target sites. Individual data points are shown, and bars indicate SD. (d) The percentage of mutations found for all tested target- and off-target sites containing a dsODN integration. T2 and T3 are the previously tested Cas9 and Cas12a target sites in SlPDS, respectively (see Figure 2a). Error bars indicate SD (n = 3). Different letters indicate statistically significant differences ($p < 0.05$) between groups, as determined by two-way ANOVA, followed by Tukey's HSD.

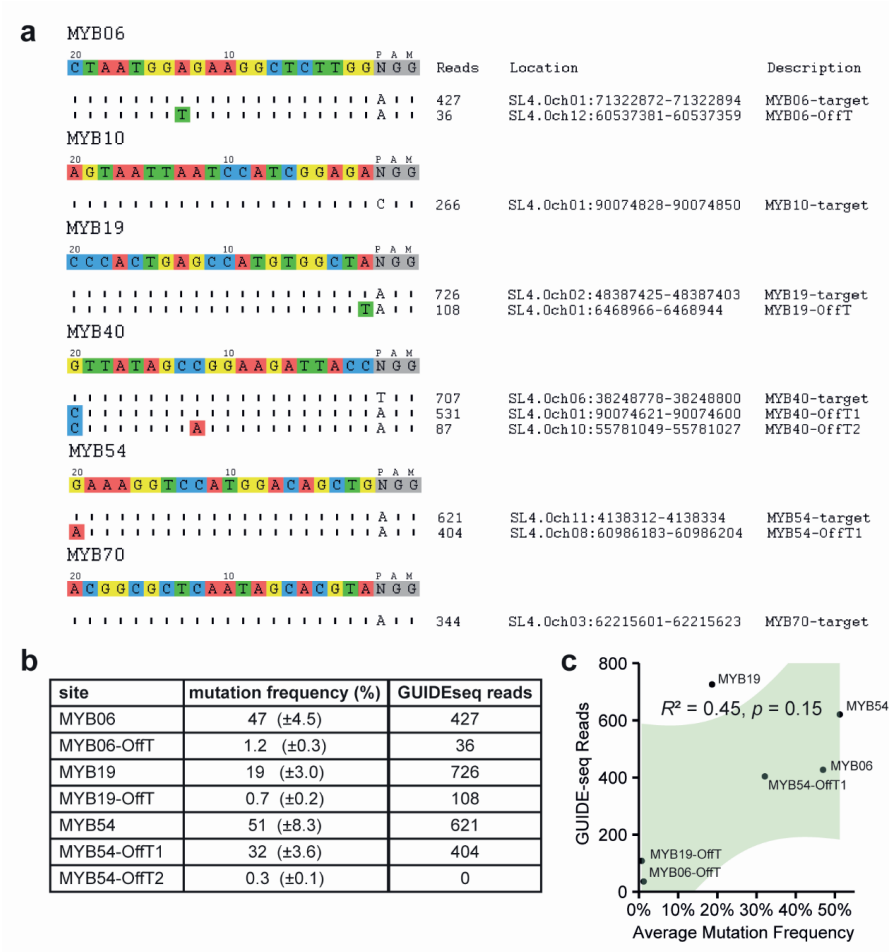
Unbiased off-target activity detection with GUIDE-seq reveals that off-target mutations are predictable

We then set out to perform unbiased detection of off-target activity for six sgRNAs in tomato protoplasts. For this, we chose five sgRNAs that were shown to produce off-target mutations in our previous work (Chapter 2). These sgRNAs include the three sgRNAs (MYB06, MYB19 and MYB54) we used in the above proof-of-principle experiment and two additional ones, MYB10 and MYB40. We also included sgRNA MYB70, which had no predicted off-target sites. The sequences of the target and previously verified off-target sites for these sgRNAs are listed in Supplementary Table 2. Protoplasts were transfected with plasmids encoding Cas9 and a single sgRNA and dsODNs, and GUIDE-seq libraries were prepared as described (Malinin et al. 2021). Additionally, we included a control sample that was transfected with dsODNs only. The resulting sequencing reads were analyzed using the GUIDE-seq pipeline (github.com/tsailab/guideseq). All identified DSB sites are shown in Figure 6a. All on-target sites were readily detected (Figure 6a). Likewise, known off-target sites were confirmed, except MYB54-OffT2 and MYB10-OffT. For MYB54-OffT2, we previously showed that the dsODN integration frequency was low (Figure 5c). Our previous work also showed that mutations at the MYB10-OffT site occur at a very low frequency and could only barely be detected using highly sensitive targeted amplicon sequencing (Chapter 2). Therefore, DSBs at these sites likely occur too infrequently to be detected using GUIDE-seq under our conditions. For all tested sgRNAs, no additional, previously unknown off-target activity sites were identified. In the control samples, we could identify low numbers of reads mapping to the on-target sites of sgRNAs *MYB54* and *MYB19*, which are most likely the result of a low amount of cross-contamination. No additional DSB sites were identified in these samples (Supplementary Table 3).

After read mapping, during the filtering and visualization of off-target sites by the GUIDE-seq pipeline, only potential off-target sites are considered that have at most 6 mismatches to the target site. We additionally used MACS2 (Zhang et al. 2008) for peak calling on aligned reads to make the procedure truly unbiased and to ensure that no DSB hotspots that might indicate off-target activity by Cas9 were missed (Supplementary Table 4). We then compared this output to the GUIDE-seq output for visualization. No additional peaks were identified, supporting the conclusion that no further off-target DSBs occurred in our samples.

Next, we assessed the detection limit for DSBs in this experiment. For this purpose, we compared the obtained average mutation frequencies at target and off-target sites in the proof-of-principle experiment, as shown in Figure 5b, and the obtained GUIDE-seq reads (Figure 6a,b). As we could not

detect MYB54-OffT2, but could identify MYB19-OffT, we conclude that (off-target) mutations should occur with a frequency of at least 0.3 - 0.7% in our experiment to be detected. Additionally, we investigated the correlation between the obtained mutation frequencies and the amount of obtained GUIDE-seq reads (Figure 6c), and found that the correlation is low (R^2=0.45).



site	mutation frequency (%)	GUIDEseq reads
MYB06	47 (±4.5)	427
MYB06-OffT	1.2 (±0.3)	36
MYB19	19 (±3.0)	726
MYB19-OffT	0.7 (±0.2)	108
MYB54	51 (±8.3)	621
MYB54-OffT1	32 (±3.6)	404
MYB54-OffT2	0.3 (±0.1)	0

Figure 6 - DSB positions identified by GUIDE-seq. (a) Overview of identified target and off-target DSB positions for the six tested sgRNAs. Indicated read counts are a summation of read counts obtained in both the plus and minus reactions. (b) Overview of average mutation frequencies shown in Figure 5b and the amount of obtained GUIDE-seq sequencing reads. Mutation frequencies are calculated by dividing the number of mutated reads by the total number of reads. The percentages shown are averages of 3 biological replicates, and SD is indicated between parentheses. (c) Pearson correlation of mutation frequencies with the amount of obtained GUIDE-seq reads. The green-shaded area indicates the 95% confidence interval.

Discussion and Conclusion

Few unbiased off-target activity detection methods have been applied to plants (Sturme et al. 2022). Nevertheless, high throughput unbiased off-target detection methods provide valuable information on nuclease and sgRNA specificity.

Many of the unbiased off-target detection methods that were developed, such as Digenome-seq (Kim et al. 2015), CIRCLE-seq (Shengdar Q Tsai et al. 2017), and SITE-seq (Cameron et al. 2017) are *in vitro* methods in which CRISPR-Cas mediated digestion of genomic DNA is performed in a test tube. Although these methods are very sensitive, they may be less representative of the CRISPR-induced DSB and subsequent DNA repair as it occurs in living cells due to the absence of DNA folding, chromatin state, and DNA repair enzymes. In plants, these *in vitro* techniques have been applied on a limited scale (Lee et al. 2018; Young et al. 2019).

In vivo methods such as LAM-HTGTS (Hu et al. 2016) and BLISS (Yan et al. 2017) detect DSBs in their native environment and have been successfully applied in mammalian cells, but are not easily transferable to plants. For most plant species, cellular assays are limited to protoplasts. Additionally, for many species, protoplasts cannot readily be cultured (Reed and Bargmann 2021). Obtaining the large amount of DNA necessary for LAM-HGTGS is, therefore, challenging. BLISS requires protocols for cell fixation and only detects DSBs that are present at the moment of labelling. Consequently, we set out to apply the *in vivo* GUIDE-seq method to tomato protoplasts.

Successful execution of GUIDE-seq is dependent on the efficient integration of a dsODN in DSBs. We observed that adding dsODNs to protoplast transfections – specifically end-protected dsODNs – inhibits mutagenesis. This observation is in accordance with a study on potato protoplasts (Nadakuduti et al. 2019). Like us, they observe that transfection efficiency is not affected by the dsODN, supporting the idea that mutagenesis is impaired instead. This effect was strongest for end-protected dsODNs, which exonucleases cannot readily degrade due to their 5' and 3' phosphorothioate bonds (Figure 2c). Like Nadakuduti et al., we hypothesize that the large number of free DNA ends introduced with the dsODN saturates the DNA repair machinery. As native exonucleases are bound to the dsODN, they will act to a lesser extent on the genomic DSBs, potentially resulting in more perfect repair. Alternatively, a saturation of the DSB repair machinery as a whole will result in genomic DSBs remaining unrepaired and thus undetectable using PCR. In this study, we managed to titrate the dsODN concentration such that the perfect integration of end-

protected dsODN was not compromised by its inhibitory effect on mutagenesis. This optimization paved the way for the application of GUIDE-seq in plant cells.

The targeted integration of dsODNs may have applications beyond DSB detection. In plant cells, homology-directed repair (HDR) remains challenging. Integrating end-protected dsODNs at CRISPR-induced target sites may provide an alternative to HDR-mediated integration of small pieces of genetic information, such as regulatory motifs or protein tags (eg, FLAG-tags or HIS-tags) by NHEJ. Of course, regenerating entire plants from such experiments remains challenging (Reed and Bargmann 2021). A more simple application of dsODN integration could be to confirm on-target activity of a selected sgRNA. Integration of a dsODN at a target site can be detected using a simple PCR (as in Figure 1b) to affirm sgRNA activity before proceeding to time-consuming *Agrobacterium*-mediated transformation (Nadakuduti et al. 2019).

We also provide data proving that dsODNs integrate at Cas12a-induced DSBs in plants and that GUIDE-seq for Cas12a off-target activity detection is feasible. This observation is in agreement with data from human cell lines (Kleinstiver et al. 2016). In this study, it was observed that dsODN integration at Cas12a target sites is less efficient than at Cas9 target sites. We tested only one Cas12a target site in this study (T3), and found that it performed similarly to the less efficient Cas9 sgRNAs (Figure 5d). Integration efficiency affects sensitivity and should be considered when performing GUIDE-seq to compare the specificities of different nucleases. Inspection of dsODN integrations at the Cas12a target and comparison to the Cas9 targets revealed that Cas12a integrations are frequently accompanied by deletions in the flanking DNA, unlike Cas9 integrations (Figure 3b, 4c). This raises the hypothesis that integration at Cas12a targets requires more complicated or slower repair processes than at Cas9 targets – likely as a result of the overhangs introduced by Cas12a needing to be processed before NHEJ can proceed. Nevertheless, the fact that dsODNs integrate into both blunt and staggered breaks indicates that this method can be applied to detect DSBs induced by a wide range of nucleases. Using dsODNs with overhangs that are compatible with those of the DSB in the target site may enhance directional insertion without the need to process the flanking DNA, thus increasing the fidelity of insertion of specific sequence elements.

When it was published in 2015, GUIDE-seq was proclaimed to be a relatively cheap method for gaining an in-depth overview of CRISPR-Cas mediated DSB induction. However, the cost-per-read of running an Illumina MiSeq is currently much higher than, for example, the more modern Illumina NovaSeq. Due to the current design of the GUIDE-seq sequencing adapters, specific NGS sequencing primers have to be added, and data needs to be exported from the sequencing machine in a specific

way to retain the UMI information. Both issues preclude the cheaper commercial sequencing of GUIDE-seq libraries. To increase the accessibility of GUIDE-seq, the adapter structure and PCR scheme might be changed to allow in-line UMI sequencing, for example, as proposed by Palani, 2018 (Palani 2018).

In this study, we found that GUIDE-seq could detect (off-target) mutations that occurred at frequencies between 0.3 and 0.7% in the protoplast pool (Figure 6b). We did not detect any off-target activity at sites that were not previously identified using amplicon sequencing (Chapter 2 of this thesis), and we can thus conclude that additional off-target mutations did not occur at frequencies higher than 0.7%. Additionally, this observation leads us to conclude that off-target mutations in tomato protoplasts can reliably be predicted based on sequence similarity.

GUIDE-seq is a method for DSB identification. These DSBs can be induced by CRISPR-Cas mediated genome editing, but may also occur naturally. In the first report of GUIDE-seq, several DSB hotspots were identified in human U2OS and HEK293 cell lines by using control samples that were transfected only with dsODNs (S Q Tsai et al. 2015). We could not identify such hotspots in our control samples (Supplementary Tables 3, 4). Likely, higher sensitivity is needed to detect these events.

Our off-target detection limit was higher than the ~ 0.1% limit previously reported for GUIDE-seq in mammalian cells (S Q Tsai et al. 2015; Kleinstiver et al. 2016). Multiple approaches may increase sensitivity. We noticed that 61%-66% of the obtained sequencing reads aligned to the plasmid used for transfection (Supplementary Table 5), indicating that dsODNs frequently ligate to degraded fragments of plasmid DNA. This ligation could happen in the protoplast and during the A-tailing of fragmented genomic DNA and adapter ligation steps of library preparation, as dsODNs and plasmid DNA are likely copurified with the genomic DNA. This problem could be overcome by using RNPs to deliver the CRISPR-Cas machinery to the cell instead of a plasmid. Likely, this would necessitate additional optimization of dsODN integration at different concentrations of RNPs and dsODNs.

Sensitivity could also be increased by improving transfection and mutagenesis efficiencies in protoplast pools. We routinely incubate protoplasts for 24h after transfection, but improved results may be obtained by an extended incubation of 48-72h. Additionally, successfully transfected protoplasts could be selected using fluorescence-assisted cell sorting (FACS) (Petersen et al. 2019). Furthermore, the amount of input genomic DNA might be increased, and deeper sequencing could be applied.

It is also important to note that the detection limit of the GUIDE-seq method can differ depending on the sgRNA used. We found that the percentage of mutations containing a dsODN can vary significantly between sgRNAs, and even for one sgRNA, this percentage may differ between target and off-target sites (Figure 5d). These differences result in a low correlation between mutation frequencies and the amounts of corresponding GUIDE-seq reads (Figure 6b,c). The chance of dsODN integration is variable and may depend on the characteristics of the sequence flanking the DSB. Therefore, the relative specificity of a sgRNA should not be determined by using GUIDE-seq results alone. Instead, GUIDE-seq should be used to identify potential off-target sites, followed by targeted amplicon sequencing of these sites to more precisely determine off-target mutation frequencies relative to on-target mutation frequencies.

In conclusion, we implemented unbiased off-target activity detection using GUIDE-seq for the first time in plant cells. We could not detect any unpredicted off-target DSBs, indicating that these either do not occur or at very low frequencies. This observation emphasizes the value of algorithms that predict off-target sites and supports the notion that CRISPR-Cas9 mediated genome editing in tomato is generally specific when well-designed. We propose several approaches to improve this method's sensitivity and ease of use. Further research might lead to comparing the specificities of alternative endonucleases, the application of GUIDE-seq to additional plant species, and the use of dsODN integration to generate targeted knock-in mutants or to study DSB repair processes.

Acknowledgements

This research was supported by the Dutch Research Council (NWO) and is funded by the Ministry of Infrastructure and Water Management. We are grateful to Elio Schijlen, Femke Bults, and Sven Warris for their sequencing and bioinformatics expertise and for performing the sequencing of the GUIDE-seq libraries.

References

Bae, S., J. Park, and J.-S. Kim. 2014. "Cas-OFFinder: A Fast and Versatile Algorithm That Searches for Potential off-Target Sites of Cas9 RNA-Guided Endonucleases." *Bioinformatics* 30 (10): 1473–75. https://doi.org/10.1093/BIOINFORMATICS/btu048.

Bernabé-Orts, J.M., I. Casas-Rodrigo, E.G. Minguet, V. Landolfi, V. Garcia-Carpintero, S. Gianoglio, M. Vázquez-Vilar, … D. Orzaez. 2019. "Assessment of Cas12a-Mediated Gene Editing Efficiency in Plants." *Plant Biotechnology Journal* 17 (10): 1971–84. https://doi.org/10.1111/pbi.13113.

Bloh, K., R. Kanchana, P. Bialk, K. Banas, Z. Zhang, B.C. Yoo, and E.B. Kmiec. 2021. "Deconvolution of Complex DNA Repair (DECODR): Establishing a Novel Deconvolution Algorithm for Comprehensive Analysis of CRISPR-Edited Sanger Sequencing Data." *CRISPR Journal* 4 (1): 120–31. https://doi.org/10.1089/crispr.2020.0022.

Brinkman, E.K., T. Chen, M. Amendola, and B. Van Steensel. 2014. "Easy Quantitative Assessment of Genome Editing by Sequence Trace Decomposition." *Nucleic Acids Research* 42 (22): 1–8. https://doi.org/10.1093/nar/gku936.

Cameron, P., C.K. Fuller, P.D. Donohoue, B.N. Jones, M.S. Thompson, M.M. Carter, S. Gradia, … A.P. May. 2017. "Mapping the Genomic Landscape of CRISPR–Cas9 Cleavage." *Nature Methods* 14 (6). https://doi.org/10.1038/nmeth.4284.

Cho, S.W., S. Kim, Y. Kim, J. Kweon, H.S. Kim, S. Bae, and J.S. Kim. 2014. "Analysis of Off-Target Effects of CRISPR/Cas-Derived RNA-Guided Endonucleases and Nickases." *Genome Research* 24 (1): 132–41. https://doi.org/10.1101/gr.162339.113.

Concordet, J.P., and M. Haeussler. 2018. "CRISPOR: Intuitive Guide Selection for CRISPR/Cas9 Genome Editing Experiments and Screens." *Nucleic Acids Research* 46 (W1): W242–45. https://doi.org/10.1093/nar/gky354.

Dehairs, J., A. Talebi, Y. Cherifi, and J. V. Swinnen. 2016. "CRISP-ID: Decoding CRISPR Mediated Indels by Sanger Sequencing." *Scientific Reports* 6 (August 2015): 28973. https://doi.org/10.1038/srep28973.

Doudna, J.A., and E. Charpentier. 2014. "The New Frontier of Genome Engineering with CRISPR-Cas9." *Science* 346 (6213): 1258096–1258096. https://doi.org/10.1126/science.1258096.

Feng, J., T. Liu, B. Qin, Y. Zhang, and X.S. Liu. 2012. "Identifying ChIP-Seq Enrichment Using MACS." *Nature Protocols* 7 (9): 1728–40. https://doi.org/10.1038/nprot.2012.101.

Feng, Z., Y. Mao, N. Xu, B. Zhang, P. Wei, D.-L. Yang, Z. Wang, … J.-K. Zhu. 2014. "Multigeneration Analysis Reveals the Inheritance, Specificity, and Patterns of CRISPR/Cas-Induced Gene Modifications in Arabidopsis." *Proceedings of the National Academy of Sciences of the United States of America* 111 (12): 4632–37. https://doi.org/10.1073/pnas.1400822111.

Fu, Y., J.A. Foden, C. Khayter, M.L. Maeder, D. Reyon, J.K. Joung, and J.D. Sander. 2013. "High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells." *Nature Biotechnology* 31 (9): 822–26. https://doi.org/10.1038/nbt.2623.

Hsu, P.D., D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, Y. Li, … F. Zhang. 2013. "DNA Targeting Specificity of RNA-Guided Cas9 Nucleases." *Nature Biotechnology* 31 (9): 827–32. https://doi.org/10.1038/nbt.2647.

Hu, J., R.M. Meyers, J. Dong, R.A. Panchakshari, F.W. Alt, and R.L. Frock. 2016. "Detecting DNA Double-Stranded Breaks in Mammalian Genomes by Linear Amplification-Mediated High-Throughput Genome-Wide Translocation Sequencing." *Nature Protocols* 11 (5): 853–71. https://doi.org/10.1038/nprot.2016.043.

Jaganathan, D., K. Ramasamy, G. Sellamuthu, S. Jayabalan, and G. Venkataraman. 2018. "CRISPR for Crop Improvement: An Update Review." *Frontiers in Plant Science* 9 (July): 1–17. https://doi.org/10.3389/fpls.2018.00985.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier. 2012. "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337: 816–21. http://www.ask-force.org/web/Genomics/Jinek-Programmable-Dual-RNA-Guided-DNA-Endonuclease-Adaptive-Bacterial-Immunity-2012.pdf.

Kim, D., S. Bae, J. Park, E. Kim, S. Kim, H.R. Yu, J. Hwang, J.-I. Kim, and J.-S. Kim. 2015. "Digenome-Seq: Genome-Wide Profiling of CRISPR-Cas9 off-Target Effects in Human Cells." *Nature Methods* 12 (3): 237–44. https://doi.org/10.1038/nmeth.3284.

Kim, D., J. Kim, J.K. Hur, K.W. Been, S.H. Yoon, and J.S. Kim. 2016. "Genome-Wide Analysis Reveals Specificities of Cpf1 Endonucleases in Human Cells." *Nature Biotechnology* 34 (8): 863–68. https://doi.org/10.1038/nbt.3609.

Kleinstiver, B.P., S.Q. Tsai, M.S. Prew, N.T. Nguyen, M.M. Welch, J.M. Lopez, Z.R. McCaw, M.J. Aryee, and J.K. Joung. 2016. "Genome-Wide Specificities of CRISPR-Cas Cpf1 Nucleases in Human Cells." *Nature Biotechnology* 34 (8): 869–74. https://doi.org/10.1038/nbt.3620.

Labun, K., X. Guo, A. Chavez, G. Church, J.A. Gagnon, and E. Valen. 2019. "Accurate Analysis of Genuine CRISPR Editing Events with AmpliCan." *Genome Research* 29 (5): 843–47. https://doi.org/10.1101/gr.244293.118.

Lee, K., Y. Zhang, B.P. Kleinstiver, J.A. Guo, M.J. Aryee, J. Miller, A. Malzahn, … K. Wang. 2018. "Activities and Specificities of CRISPR/Cas9 and Cas12a Nucleases for Targeted Mutagenesis in Maize." *Plant Biotechnology Journal*, 1–11. https://doi.org/10.1111/pbi.12982.

Li, J., H. Manghwar, L. Sun, P. Wang, G. Wang, H. Sheng, J. Zhang, … X. Zhang. 2019. "Whole Genome Sequencing Reveals Rare Off-target Mutations and Considerable Inherent Genetic or/and Somaclonal Variations in <scp>CRISPR</Scp> /Cas9-edited Cotton Plants." *Plant Biotechnology Journal* 17 (5): 858–68. https://doi.org/10.1111/pbi.13020.

Lin, Y., T.J. Cradick, M.T. Brown, H. Deshmukh, P. Ranjan, N. Sarode, B.M. Wile, P.M. Vertino, F.J. Stewart, and G. Bao. 2014. "CRISPR/Cas9 Systems Have off-Target Activity with Insertions or Deletions between Target DNA and Guide RNA Sequences." *Nucleic Acids Research* 42 (11): 7473–85. https://doi.org/10.1093/nar/gku402.

Liu, H., Y. Ding, Y. Zhou, W. Jin, K. Xie, and L.L. Chen. 2017. "CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants." *Molecular Plant* 10 (3): 530–32. https://doi.org/10.1016/j.molp.2017.01.003.

Malinin, N.L., G. Lee, C.R. Lazzarotto, Y. Li, Z. Zheng, N.T. Nguyen, M. Liebers, … S.Q. Tsai. 2021. "Defining Genome-Wide CRISPR–Cas Genome-Editing Nuclease Activity with GUIDE-Seq." *Nature Protocols* 16 (December). https://doi.org/10.1038/s41596-021-00626-x.

Mashal, R.D., J. Koontz, and J. Sklar. 1995. "Detection of Mutations by Cleavage of DNA Heteroduplexes with Bacteriophage Resolvases." *Nature Genetics* 9: 177–83. http://www.nature.com.ezproxy.library.wur.nl/ng/journal/v9/n2/pdf/ng0295-177.pdf.

Nadakuduti, S.S., C.G. Starker, D.K. Ko, T.B. Jayakody, C.R. Buell, D.F. Voytas, and D.S. Douches. 2019. "Evaluation of Methods to Assess in Vivo Activity of Engineered Genome-Editing Nucleases in Protoplasts." *Frontiers in Plant Science* 10 (February): 1–13. https://doi.org/10.3389/fpls.2019.00110.

Ohtsuki, N., K. Kizawa, A. Mori, A. Nishizawa-Yokoi, T. Komatsuda, H. Yoshida, K. Hayakawa, S. Toki, and H. Saika. 2021. "Precise Genome Editing in MiRNA Target Site via Gene Targeting and Subsequent Single-Strand-Annealing-Mediated Excision of the Marker Gene in Plants." *Frontiers in Genome Editing* 2 (January). https://doi.org/10.3389/fgeed.2020.617713.

Palani, N. 2018. "GUIDE-Seq Simplified Library Preparation Protocol (CRISPR/Cas9 off-Target Cleavage Detection)." https://doi.org/10.17504/protocol.

Pattanayak, V., S. Lin, J.P. Guilinger, E. Ma, J.A. Doudna, and D.R. Liu. 2013. "High-Throughput Profiling of off-Target DNA Cleavage Reveals RNA-Programmed Cas9 Nuclease Specificity." *Nature Biotechnology* 31 (9): 839–43. https://doi.org/10.1038/nbt.2673.

Petersen, B.L., S.R. Möller, J. Mravec, B. Jørgensen, M. Christensen, Y. Liu, H.H. Wandall, E.P. Bennett, and Z. Yang. 2019. "Improved CRISPR/Cas9 Gene Editing by Fluorescence Activated Cell Sorting of Green Fluorescence Protein Tagged Protoplasts." *BMC Biotechnology* 19 (1). https://doi.org/10.1186/s12896-019-0530-x.

Peterson, B.A., D.C. Haak, M.T. Nishimura, P.J.P.L. Teixeira, S.R. James, J.L. Dangl, and Z.L. Nimchuk. 2016. "Genome-Wide Assessment of Efficiency and Specificity in CRISPR/Cas9 Mediated Multiple Site Targeting in Arabidopsis." Edited by B. Zhang. *PLOS ONE* 11 (9): e0162169. https://doi.org/10.1371/journal.pone.0162169.

Puchta, H. 2004. "The Repair of Double-Strand Breaks in Plants: Mechanisms and Consequences for Genome Evolution." *Journal of Experimental Botany* 56 (409): 1–14. https://doi.org/10.1093/jxb/eri025.

Puchta, H. 2002. "Gene Replacement by Homologous Recombination in Plants." *Plant Molecular Biology* 48: 173–82. https://search-proquest-com.ezproxy.library.wur.nl/docview/768110637?OpenUrlRefId=info:xri/sid:wcdiscovery&accountid=27871.

Reed, K.M., and B.O.R. Bargmann. 2021. "Protoplast Regeneration and Its Use in New Plant Breeding Technologies." *Frontiers in Genome Editing* 3 (September). https://doi.org/10.3389/fgeed.2021.734951.

Schiml, S., and H. Puchta. 2016. "Revolutionizing Plant Biology: Multiple Ways of Genome Engineering by CRISPR/Cas." *Plant Methods* 12 (1): 1–9. https://doi.org/10.1186/s13007-016-0103-0.

Sheen, J. 2002. "A Transient Expression Assay Using Arabidopsis Mesophyll Protoplasts," 1–6. http://genetics.mgh.harvard.edu/sheenweb/.

Shen, H., G.D. Strunks, B.J.P.M. Klemann, P.J.J. Hooykaas, and S. de Pater. 2016. "CRISPR/Cas9-Induced Double Strand Break Repair in Arabidopsis Non-Homologous End-Joining Mutants." *G3 (Bethesda, Md.)* 7 (January): 1–22. https://doi.org/10.1534/g3.116.035204.

Steinert, J., S. Schiml, and H. Puchta. 2016. "Homology-Based Double-Strand Break-Induced Genome Engineering in Plants." *Plant Cell Reports* 35 (7): 1429–38. https://doi.org/10.1007/s00299-016-1981-3.

Sturme, M.H.J., J.P. Van Der Berg, L.M.S. Bouwman, A. De Schrijver, R.A. De Maagd, G.A. Kleter, and E. Battaglia-De Wilde. 2022. "Occurrence and Nature of Off-Target Modifications by CRISPR-Cas Genome Editing in Plants." *ACS Agricultural Science and Technology* 2 (2): 192–201. https://doi.org/10.1021/acsagscitech.1c00270.

Tamura, K., Y. Adachi, K. Chiba, K. Oguchi, and H. Takahashi. 2002. "Identification of Ku70 and Ku80 Homologues in Arabidopsis Thaliana: Evidence for a Role in the Repair of DNA Double-Strand Breaks." *Plant Journal* 29 (6): 771–81. https://doi.org/10.1046/j.1365-313X.2002.01258.x.

Tan, J., Y. Zhao, B. Wang, Y. Hao, Y. Wang, Y. Li, W. Luo, … Q. Zhu. 2020. "Efficient CRISPR/Cas9-based Plant Genomic Fragment Deletions by Microhomology-mediated End Joining." *Plant Biotechnology Journal*, 1–3. https://doi.org/10.1111/pbi.13390.

Tang, X., G. Liu, J. Zhou, Q. Ren, Q. You, L. Tian, X. Xin, … Y. Zhang. 2018. "A Large-Scale Whole-Genome Sequencing Analysis Reveals Highly Specific Genome Editing by Both Cas9 and Cpf1 (Cas12a) Nucleases in Rice." *Genome Biology* 19 (1): 84. https://doi.org/10.1186/s13059-018-1458-5.

Tsai, S Q, Z. Zheng, N.T. Nguyen, M. Liebers, V. V Topkar, V. Thapar, N. Wyvekens, … J.K. Joung. 2015. "GUIDE-Seq Enables Genome-Wide Profiling of off-Target Cleavage by CRISPR-Cas Nucleases." *Nat Biotechnol* 33 (2): 187–97. https://doi.org/10.1038/nbt.3117.

Tsai, Shengdar Q, N.T. Nguyen, J. Malagon-Lopez, V. V Topkar, M.J. Aryee, and J.K. Joung. 2017. "CIRCLE-Seq: A Highly Sensitive in Vitro Screen for Genome-Wide CRISPR–Cas9 Nuclease off-Targets." *Nature Methods*, no. May. https://doi.org/10.1038/nmeth.4278.

Vu, T., D. Doan, J. Kim, Y.W. Sung, M. Tran, Y.J. Song, S. Das, and J. Kim. 2021. "CRISPR/Cas-based Precision Genome Editing via Microhomology-mediated End Joining." *Plant Biotechnology Journal* 19 (2): 230–39. https://doi.org/10.1111/pbi.13490.

Wang, X., M. Tu, Y. Wang, W. Yin, Y. Zhang, H. Wu, Y. Gu, Z. Li, Z. Xi, and X. Wang. 2021. "Whole-Genome Sequencing Reveals Rare off-Target Mutations in CRISPR/Cas9-Edited Grapevine." *Horticulture Research* 8 (1). https://doi.org/10.1038/s41438-021-00549-4.

Yan, W.X., R. Mirzazadeh, S. Garnerone, D. Scott, M.W. Schneider, T. Kallas, J. Custodio, … N. Crosetto. 2017. "BLISS Is a Versatile and Quantitative Method for Genome-Wide Profiling of DNA Double-Strand Breaks." *Nature Communications* 8 (May): 15058. https://doi.org/10.1038/ncomms15058.

Young, J., G. Zastrow-Hayes, S. Deschamps, S. Svitashev, M. Zaremba, A. Acharya, S. Paulraj, … S. Kumar. 2019. "CRISPR-Cas9 Editing in Maize: Systematic Evaluation of Off-Target Activity and Its Relevance in Crop Improvement." *Scientific Reports* 9 (1). https://doi.org/10.1038/s41598-019-43141-6.

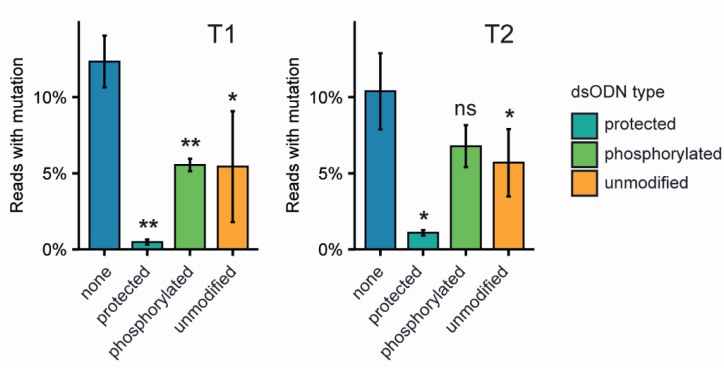
Zetsche, B., J.S. Gootenberg, O.O. Abudayyeh, I.M. Slaymaker, K.S. Makarova, P. Essletzbichler, S.E. Volz, … F. Zhang. 2015. "Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System." *Cell* 163 (3): 759–71. https://doi.org/10.1016/j.cell.2015.09.038.

Zhang, Y., T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nussbaum, … X.S. Shirley. 2008. "Model-Based Analysis of ChIP-Seq (MACS)." *Genome Biology* 9 (9). https://doi.org/10.1186/gb-2008-9-9-r137.

Supplementary Information

Supplementary Figure

a



b

T1

dsODN type	replicate	reads	integrations	percentage
protected	1	54499	76	0.14%
protected	2	78591	196	0.25%
protected	3	73498	284	0.39%
phosphorylated	1	47522	668	1.4%
phosphorylated	2	66050	1247	1.9%
phosphorylated	3	59432	1301	2.2%
unmodified	1	31113	252	0.81%
unmodified	2	51452	189	0.37%
unmodified	3	41855	463	1.1%

T2

dsODN type	replicate	reads	integrations	percentage
protected	1	66571	0	0.00%
protected	2	48298	0	0.00%
protected	3	50315	1	0.00%
phosphorylated	1	53262	0	0.00%
phosphorylated	2	35581	1	0.00%
phosphorylated	3	31852	41	0.13%
unmodified	1	79355	52	0.07%
unmodified	2	63408	1	0.00%
unmodified	3	60187	43	0.07%

Supplementary Figure 1 - The addition of dsODNs to the transfection mixture decreases mutation frequency. (a) Mutation frequencies at CRISPR-Cas9 target sites T1 and T2 upon addition of 100 pmol of either protected, phosphorylated or unmodified dsODN. (b) Total amount of reads and the amount of reads containing a dsODN integration longer than 20 bp in each sample.

Supplementary Tables

Supplementary Table 1 - Overview of primers used in this study. To distinguish samples and replicates in next-generation amplicon sequencing, 8 bp barcodes were attached to the 5' end of the primer.

primer	application	sequence (5'-3')
P1	dsODN integration at target site	AACTGCCAAACCACCACAAA
P2	dsODN integration at target site	AGAGAGTGATGGCTGAATAGTC
dsODN1	dsODN integration at target site	GTTTAATTGAGTTGTCATATGTTAATAAC
amp1-F	amplicon NGS amplicon 1 *PDS*	TTCTTACTGTGAAATATCCTTATGG
amp1-R	amplicon NGS amplicon 1 *PDS*	AATCTTTAACTTATGACCCATTGAT
amp2-F	amplicon NGS amplicon 2 *PDS*	GGGTCATAAGTTAAAGATTCGTACTCCC
amp2-R	amplicon NGS amplicon 2 *PDS*	CATACGACCTGAATGACAAGATAGTTCC
MYB06-F	amplicon NGS MYB06 target	ATAATGGGAAGAACACCTTGTTG
MYB06-R	amplicon NGS MYB06 target	AGTAGGTAATCGTACCGGCATTT
MYB06-OffT-F	amplicon NGS MYB06-OffT	TTTTGAATGAGATGATTATGCGTTA
MYB06-OffT-R	amplicon NGS MYB06-OffT	TGGTAAGAACACCTTGTTGTGAG
MYB19-F	amplicon NGS MYB19 target	AGAACGATGCCTTGTTGTCC
MYB19-R	amplicon NGS MYB19 target	CGAGTGAATTTTGGAAACTATTGG
MYB19-OffT-F	amplicon NGS MYB19-OffT	AGAATTATTGGAACACGCACTTG
MYB19-OffT-R	amplicon NGS MYB19-OffT	TTCGAAATAATTTGTTGGTGTTG
MYB54-F	amplicon NGS MYB54 target	AAAATATTATCATCATGGGGAGACA
MYB54-R	amplicon NGS MYB54 target	AGACCTGCTAATTTAGGGACAGC
MYB54-OffT1-F	amplicon NGS MYB54-OffT1	TAATAGGGATGGGAAGACAACCT
MYB54-OffT1-R	amplicon NGS MYB54-OffT1	ACTCTTACCGCAACGCCTAAG
MYB54-OffT2-F	amplicon NGS MYB54-OffT2	AGAGAAGAAACCTGCAAGATGTG
MYB54-OffT2-R	amplicon NGS MYB54-OffT2	GTGTATGTGTGTGTAAGGGATGG

Supplementary Table 2 - Overview of sgRNA target and known off-target sites used for unbiased off-target detection. Mismatches to the target site are indicated in bold red. PAM is underlined.

sgRNA	on/off-target	sequence
MYB06	target	CTAATGGAGAAGGCTCTTGGAGG
	off-target	CTAATGG**T**GAAGGCTCTTGGAGG
MYB10	target	AGTAATTAATCCATCGGAGACGG
	off-target	AGTAATTAATCCATCGGAGACG**A**
MYB19	target	CCCACTGAGCCATGTGGCTAAGG
	off-target	CCCACTGAGCCATGTGGCT**T**AGG
MYB40	target	GTTATAGCCGGAAGATTACCTGG
	off-target 1	**C**TTATAGCCGGAAGATTACCAGG
	off-target 2	**C**TTATAGC**A**GGAAGATTACCAGG
MYB54	target	GAAAGGTCCATGGACAGCTGAGG
	off-target 1	**A**AAAGGTCCATGGACAGCTGAGG
	off-target 2	GAAAGGTCCATGGACA**C**CTGAGG
MYB70	target	ACGGCGCTCAATAGCACGTAAGG

Supplementary Table 3 - Overview of all DSB sites identified by the GUIDE-seq pipeline in control samples.

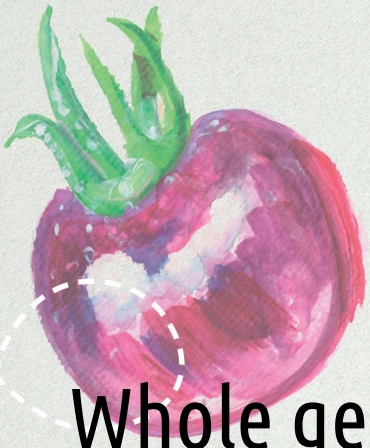
sample	chr	start	end	description
ctrl_plus	SL4.0ch02	48387409	48387413	target site MYB19
ctrl_plus	SL4.0ch11	4138311	4138351	target site MYB54
ctrl_minus	SL4.0ch02	48387409	48387411	target site MYB19
ctrl_minus	SL4.0ch11	4138304	4138356	target site MYB54

Supplementary Table 4 - Peaks called on aligned GUIDE-seq reads using MACS2. No peaks were identified in both the plus and minus reactions of the control sample.

sample	chr	peak start	peak end	fold enrichment	description
MYB06_plus	SL4.0ch01	71322606	71323291	19.3	MYB06-target
MYB06_minus	SL4.0ch01	71322590	71323308	19.9	MYB06-target
MYB06_minus	SL4.0ch12	60537262	60537449	7.6	MYB06-OffT
MYB10_plus	SL4.0ch01	90074529	90075046	15.5	MYB10-target
MYB10_minus	SL4.0ch01	90074505	90075210	14.0	MYB10-target
MYB19_plus	SL4.0ch01	6468723	6469053	9.3	MYB19-OffT
MYB19_plus	SL4.0ch02	48387086	48387804	20.4	MYB19-target
MYB19_minus	SL4.0ch01	6468739	6469235	9.9	MYB19-OffT
MYB19_minus	SL4.0ch02	48387010	48387765	19.2	MYB19-target
MYB40_plus	SL4.0ch01	90074259	90074916	20.7	MYB40-OffT1
MYB40_plus	SL4.0ch06	38248499	38249135	19.4	MYB40-target
MYB40_plus	SL4.0ch10	55780856	55781262	9.0	MYB40-OffT2
MYB40_minus	SL4.0ch01	90074193	90074911	18.9	MYB40-OffT1
MYB40_minus	SL4.0ch06	38248516	38249137	20.9	MYB40-target
MYB40_minus	SL4.0ch10	55780803	55781181	12.0	MYB40-OffT2
MYB54_plus	SL4.0ch08	60985976	60986455	19.6	MYB54-OffT1
MYB54_plus	SL4.0ch11	4138099	4138706	18.9	MYB54-target
MYB54_minus	SL4.0ch08	60985953	60986576	16.5	MYB54-OffT1
MYB54_minus	SL4.0ch11	4138028	4138689	22.6	MYB54-target
MYB70_plus	SL4.0ch03	62215293	62215923	16.0	MYB70-target
MYB70_minus	SL4.0ch03	62215278	62215758	18.6	MYB70-target

Supplementary Datasets

The raw sequencing data has been uploaded to the NCBI Sequencing Read Archive, BioProject accession number PRJNA980592.



Chapter 5

Whole genome resequencing for the detection of Cas9-induced off-target mutations in tomato plants

E. Slaman[1,2], S. Warris[2], J. Busscher-Lange[2], R. Dekker[2],

G. C. Angenent[1,2], R. A. de Maagd[2]

[1] Wageningen University & Research, Laboratory of Molecular Biology, Wageningen, the Netherlands

[2] Wageningen University & Research, Plant Developmental Systems, Bioscience, Wageningen, the Netherlands

Abstract

CRISPR-Cas9-mediated genome editing is a powerful tool with the potential to speed up plant breeding and research. However, knowledge about the occurrence or absence of off-target mutations is limited. In this study, we performed whole genome resequencing at an average of ~34x coverage on 38 progeny of CRISPR-Cas9 genome-edited plants, five transformation and tissue culture control plants, and three wild type plants, each from two consecutive generations. In all plants, including the wild types, we found, on average, ~71.400 single nucleotide variants (SNVs), ~28.900 indels, and ~450 structural variants (SVs) compared to the reference genome. We concluded that CRISPR-Cas9-mediated genome editing did not significantly increase the number of indels and SVs found in plants. Additionally, we tested 29.803 predicted off-target sites with up to six mismatches to the three used target sites for the presence of mutations. Genuine off-target mutations were found at two previously identified off-target sites, both with only one mismatch to the target. A third site with five mismatches contained a 1 bp deletion, but could not be unequivocally identified as an off-target site due to the presence of a homopolymer, which may result in erroneous indel calling. Overall, we conclude that CRISPR-mediated genome editing is precise. This study confirms, in plants, that Cas9 action is very sensitive to mismatches between the guide RNA and its genomic target. Additionally, we illustrate that applying WGS for unbiased off-target detection is challenging due to pre-existing genomic variation – even within breeding lines - and the inherent uncertainty of variant-calling algorithms.

Introduction

CRISPR-Cas mediated genome editing has allowed researchers to induce mutations at almost every desired position in the genome. In this system, an endonuclease such as Cas9 or Cas12a is guided to a user-defined position in the genome by an easily programmable guide RNA (gRNA) (Jinek et al. 2012; Cong et al. 2013; Mali et al. 2013). After binding to the target site by the nuclease-gRNA complex, the nuclease will induce a double-strand break (DSB) in the genomic DNA. DSBs threaten the cell's survival; therefore, its DNA repair machinery will activate and repair the damage. The most frequently employed DNA repair pathway – especially in plants – is non-homologous end joining (NHEJ) (Puchta 2004). During NHEJ, small mistakes may be introduced, resulting in insertions or deletions (indels) of one or a few base pairs. This can disrupt open reading frames or other sequences that are essential for gene function. Alternatively, the homology-directed repair pathway can be activated, resulting in targeted genomic changes or gene insertions if a DNA repair template is present. In plants, genome editing can potentially speed up research and breeding significantly (Jaganathan et al. 2018; Zhu et al. 2020; Kwon et al. 2020; Lemmon et al. 2018).

Putative undesired outcomes of CRISPR-mediated genome editing are so-called "off-target" mutations. In those events, the nuclease-gRNA complex induces DSBs at positions other than the intended target site. This off-target activity is particularly frequent at sites with high sequence similarity to the target site (Fu et al. 2013; Pattanayak et al. 2013; Hsu et al. 2013). In plants, relatively little research has been performed on the specificity of CRISPR-mediated genome editing (reviewed by Sturme et al. 2022). More data would be valuable for regular users of CRISPR-Cas9 and regulatory officials.

Detection of off-target mutations can be performed using biased or unbiased approaches. In biased approaches, custom algorithms predict potential off-target sites (Bae, Park, and Kim 2014; Lei et al. 2014; Concordet and Haeussler 2018). These sites can be subsequently screened for the presence of mutations. Unbiased methods are not limited to off-target site predictions and screen the entire genome. In plants, most off-target research has been performed in a biased manner. Predicted off-target sites are screened for mutations using T7 endonuclease assays, Sanger sequencing, or amplicon sequencing (Chapter 2, Chapter 3, Shan et al. 2013; Svitashev et al. 2016; Veillet et al. 2019). These methods are relatively fast and easy to perform. However, they can only feasibly scan a limited number of off-target sites. Additionally, all of these approaches contain a PCR step. Therefore, larger or more complex mutations that disrupt primer binding sites may be missed.

A multitude of unbiased off-target detection methods were developed for mammalian cells. Examples are Digenome-seq (Kim et al. 2015), CIRCLE-seq (Shengdar Q Tsai et al. 2017), SITE-seq (Cameron et al. 2017), BLISS (Yan et al. 2017), and GUIDE-seq (S Q Tsai et al. 2015). Few of these detection methods have reached the plant research field. One maize study has used the *in vitro* method CIRCLE-seq to identify off-target DSBs (Lee et al. 2018), and another made use of a SITE-seq based method (Young et al. 2019). A study in *Arabidopsis* used the *in vitro* method Digenome-seq (Xu et al. 2019). We have previously applied the *in vivo* method GUIDE-seq for DSB identification in tomato (Chapter 4). Although these methods are suitable for gaining a general overview of all sites at which guided nucleases may induce a DSB, they are not suited to identify all changes that have occurred in an individual, genome-edited plant.

Whole genome sequencing (WGS) is an unbiased approach that can identify off-target sites and the exact mutation that occurred, even if this mutation is large or more complex. In plants, the WGS approach to investigate CRISPR-Cas genome editing specificity has been applied to *Arabidopsis*, rice, maize, cotton, grapevine, and the moss *Physcomitrium patens* (Z. Feng et al. 2014; H. Zhang et al. 2014; Peterson et al. 2016; C. Feng et al. 2018; J. Li et al. 2019; Wang et al. 2021; Bessoltane et al. 2022). In these studies, CRISPR-Cas mediated genome editing was generally found to be specific. Although tomato is a well-studied model for fleshy fruits and an economically important crop, only one published study reported WGS of a CRISPR-Cas9 edited tomato line (Nekrasov et al. 2017).

We previously identified several sgRNAs for SpCas9 that showed off-target activity in protoplast pools (Chapter 2 of this thesis). To further investigate their activities in entire plants, we generated mutant plant lines using three of these sgRNAs and performed WGS at ~34x coverage of 38 genome-edited T_1 progeny plants. Additionally, we sequenced five plants that underwent transformation and tissue culture without introduced nucleases like Cas9, and three seed-grown wild type plants each from two consecutive generations as controls. Our results allow us to draw conclusions about the specificity of CRISPR-Cas9 mediated genome editing in tomato. Additionally, we discuss the suitability of WGS as a method for unbiased off-target detection.

Materials and Methods

Plant material and growth conditions

Solanum lycopersicum cultivar "Moneyberg GT" was used for transformations (obtained from Nunhems BV, Netherlands). Cv. Moneyberg GT is derived from cv. Moneyberg by introgression of a large fragment of chromosome 9 from *S. peruvianum,* conferring TMV resistance, as described earlier (van Rengs et al. 2022). Tomato seedlings for transformation, explants, callus, and shoots were grown in a growth chamber at 25°C under a 16h light/8h dark photoperiod. After shoots had rooted, they were transferred to stone wool blocks and grown in a 21°C growth chamber with a 16h light/8h dark photoperiod. After approximately two months, plants were transferred to a greenhouse (Unifarm, Wageningen, The Netherlands), where they were grown under ambient temperature (> 20 °C) and under natural light supplemented by artificial sodium lights to obtain a 16h light/8h dark photoperiod.

Binary vector construction

All plasmids were constructed using Golden Gate Cloning and the MoClo toolkit (Engler et al. 2008; Weber et al. 2011). The construction of binary vectors targeting *SlMYB12 (Solyc01g079620), SlMYB7 (Solyc01g111500),* and *SlMYB43* (*THM16*) *(Solyc11g011050)* was described in Chapter 2 of this thesis. These were guides MYB06, MYB10, and MYB54, respectively. Briefly, binary vectors used for stable tomato transformation encode the kanamycin-resistance gene *NPTII*, a human codon-optimized *SpCas9*, a fluorescent *turboGFP* gene, and the sgRNA under the control of the *AtU6_26* promoter.

Tomato transformation

Tomato transformation was done using *Agrobacterium tumefaciens* strain C58C1, as described previously (van Roekel et al. 1993) with modifications: in Medium B, 0.05 mg/L 2.4D was used instead of IAA, explants were first cultured on Medium C without IAA for two weeks, after which they were transferred to Medium C with IAA as described, and no feeder layers were used.

Tomato genotyping

Genotyping of primary (T_0) transformants was performed using Phire Direct PCR (Thermo Fisher). Primers for genotyping can be found in Supplementary Table 6. The thermocycler protocol used was as follows: initial denaturation at 98°C for 5 min, followed by 35 cycles of denaturation at 98°C for 5 sec, annealing at 60°C for 10 sec and extension at 72°C for 20 sec. Then, a final extension step

of 72°C for 1 minute was performed. PCR products were purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey Nagel) and sequenced using Sanger sequencing. For some regions, it was difficult to successfully obtain a PCR product long enough for successful direct sequencing. The short PCR product was amplified for these regions and cloned into pJET (CloneJET PCR Cloning Kit, Thermo Fisher). At least five clones were grown for every plant. Their plasmids were isolated (NucleoSpin Plasmid Kit, Macherey-Nagel) and the insert was subsequently sequenced using M13 primers.

T_1 individuals from CRISPR lines were first selected for the absence of the T-DNA by screening for GFP fluorescence using a UV torch as seedlings. GFP-negative plants were genotyped for on-target mutations, as described previously.

T_1 individuals from tissue culture and transformation controls were selected for the absence of the T-DNA using T-DNA-specific primers (Supplementary Table 6). PCR was performed directly on young leaf material using Phire Direct PCR (Thermo Fisher). The thermocycler protocol used was as follows: initial denaturation at 98°C for 5 min, followed by 35 cycles of denaturation at 98°C for 5 sec, annealing at 60°C for 5 sec and extension at 72°C for 25 sec. Then, a final extension step of 72°C for 1 minute was performed.

Genomic DNA isolation

Genomic DNA for next-generation amplicon sequencing or whole genome resequencing was extracted using the cetyltrimethylammonium bromide (CTAB) method. Frozen leaf material was finely ground using bead beating. To every sample, 500 µL of CTAB buffer (2% CTAB, 1% polyvinylpyrrolidone, 100 mM Tris, 1.4 M NaCl, 20 mM EDTA), and 2 µg/mL RNAse A was added, and samples were incubated at 60°C for 1h. After cooling, 250 µL of a 24:1 mixture of chloroform and isoamyl alcohol was added. Mixing was performed by inversion, and phases were separated by centrifuging at 15.000 g for 7 minutes. To 400 uL of the aqueous phase, 200 uL of cold isopropanol was added and mixed by inversion. DNA was pelleted by centrifuging at 15.000 g for 10 minutes. The supernatant was removed, and the DNA pellet was washed by adding 300 uL of 70% ethanol. DNA was again pelleted by centrifuging at 15.000 g for 10 minutes. The supernatant was subsequently removed, and the DNA pellet was air-dried and dissolved in 50 uL 5mM Tris, pH 8.5.

Amplicon sequencing and data analysis

Genomic DNA was isolated from T0 lines using the CTAB method outlined above for amplicon sequencing. Target and predicted off-target sites were amplified using barcoded primers

(Supplementary Table 6), and amplicons were visualized on a 2% agarose gel. Equal amounts of PCR products were pooled and column-purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). Libraries were subsequently sequenced by Eurofins Genomics GmbH (Germany) on the Illumina NovaSeq platform, using 2x150 bp reads. Quality control, read trimming, and merging of overlapping reads was performed using CLC Genomics Workbench v22. The R package Amplican was used to identify mutant reads and alleles (Labun et al. 2019).

Whole genome sequencing and data analysis

The genomic DNA of selected plants was extracted from the young leaves of five- to seven-week-old plants using the CTAB method described previously. Library preparation and whole genome resequencing were performed by NEOGEN (Scotland, UK) on the Illumina NovaSeq platform, yielding 2x150 bp paired-end reads.

The published reference genome that was used for comparison in this study is called "MbTMV" and is, to our knowledge, similar if not identical to our "Moneyberg GT" used for all sequencing. However, we could not establish the number of generations that separate them. Sequencing reads were trimmed using trimmomatic (Bolger et al. 2014) and aligned to the published MbTMV reference genome, available on plabipd.de/portal/moneyberg, using bwa-mem (H. Li and Durbin 2009). SNVs and indels were called using GATK Best Practices and filtered using default settings (Van der Auwera et al. 2013). Structural variation was called using smoove (github.com/brentp/smoove). For SNVs and indels, the overlap between samples was determined using vcftools (Danecek et al. 2011). Bcftools was used to filter out SNVs and indels in CRISPR lines that overlap with wild type controls or transformation and tissue culture controls (H. Li 2011). Structural variation also present in the wild types and controls was excluded from further analysis by first converting the VCF files to a BEDfile using sv2bed (github.com/dantaki/sv2bed). Breakends (BNDs), which only have one coordinate, were recorded as a 20 bp region spanning 10 bp on each side of the identified BND. SVs from all eleven control lines were concatenated and subsequently merged (bedtools merge -d 200) using bedtools (Quinlan 2014). Unique structural variation in the CRISPR lines was then identified as the difference between the sample SV file and this reference SV file using bedtools intersect -v.

Potential off-target sites in the reference Heinz 1706 SL4.0 (available on solgenomics.net) or MbTMV were predicted using CasOFFinder (Bae et al. 2014), with up to 6 mismatches and with either NGG, NAG, or NGA as PAM. Overlap between identified SNVs, indels, and SV break ends (20 bp regions surrounding every called BND and every start and end coordinate of all other called SVs) with a 40 bp window centered on the predicted off-target sites was determined using bedtools. A 40 bp

window should cover most CRISPR-Cas9 induced indels and indels called in the genome sequencing data, as they are mostly shorter than 20 bp. Visual inspection of identified SNVs and indels was performed using the Integrative Genomics Viewer (IGV).

Results

Tomato mutagenesis with Cas9 and three non-specific sgRNAs in regenerated plants

We previously identified several sgRNAs for SpCas9, targeting *MYB* genes that induced off-target mutations (Chapter 2). This study was performed in protoplasts and in a biased manner as we only checked for mutations at a limited number of predicted off-target sites. We aimed to gain more insight into how non-specific sgRNAs behave in regenerated mutant plants. For this purpose, we selected three sgRNAs with known off-target activity: sgRNA06, targeting *MYB12*, sgRNA10, targeting *MYB07*, and sgRNA54, targeting *MYB43* (Figure 1a). The off-target mutations associated with these sgRNAs occur at varying frequencies in protoplast pools: at the first off-target site of the sgRNA targeting *MYB43,* mutations frequently occur, whereas mutations at the off-target site with a non-canonical NGA PAM associated with the sgRNA targeting *MYB7* rarely occur. We then transformed and regenerated plants with constructs containing Cas9 and one of these sgRNAs. The obtained transformants were genotyped using Sanger sequencing of the PCR-amplified target sites and subsequent deconvolution of the trace files. Lines with mutations at the target site were selected and grown in the greenhouse (Figure 1b, Figure 2). Additionally, we performed amplicon sequencing to check for mutations at the previously identified off-target sites for the respective sgRNAs. Off-target mutations were identified in four out of six *MYB12*-targeted lines and at off-target site 1, but not at off-target site 2, in the *MYB43*-targeted line. No off-target mutations were found in the two other MYB12 lines or in the three MYB07 lines (Figure 2, Supplementary Figure 1). These results are consistent with the off-target mutation frequency in protoplast pools: we only found mutations at the two most frequently edited off-target sites (Figure 1a).

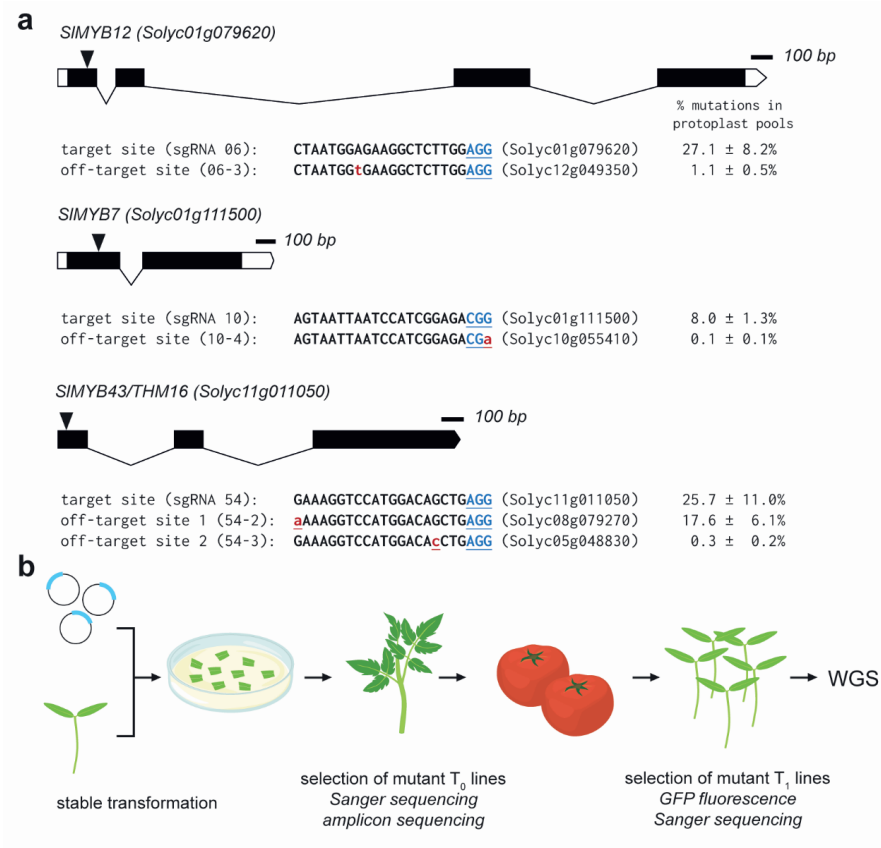


Figure 1 - Experimental overview. (a) The sgRNAs used in creating stable transformants for mutagenesis of SlMYB12, SlMYB7, and SlMYB43, with their previously verified off-target sites. Numbers of sgRNAs and off-target sites as used in Chapter 2, Figure 4 of this thesis are indicated, as are the average mutation frequencies and their standard deviation (n = 3) in protoplast pools as shown in Chapter 2, Figure 4. (b) Target sites of T_0 transformants were genotyped using Sanger sequencing. Mutant lines were subjected to amplicon sequencing to detect off-target mutations at sites shown in (a) and moved to the greenhouse. T_1 lines were selected to be GFP-negative as an indication for the absence of T-DNA and genotyped at the target sites using Sanger sequencing; mutant lines were selected for WGS.

Selection of plants for whole genome resequencing

Phenotyping of CRISPR-Cas genome-edited tomato plants is generally done for the T_1 generation. It is preferable to work with T-DNA- and Cas9-free plants with a fixed genotype and avoid the effects of tissue culture that sometimes occur in the primary T_0 transformants. For this reason, we chose to perform whole genome sequencing on T_1 plants. We included a turboGFP marker in the T-DNA to easily select for transformation and identification of T-DNA-free progeny in later generations. For every T_0 line, we selected three or four GFP-negative T_1 offspring with mutations at the target site for whole genome resequencing, resulting in 38 plants. No new on-target or off-target alleles were identified in the T_1 generation, even if one of the T_0 alleles was wild type, indicating that initial

genotypes were stably transmitted to the next generation. After whole genome resequencing, we found that, although seemingly GFP-negative, eight plants still contained the T-DNA (indicated by red exclamation marks, Figure 2).

Apart from CRISPR-Cas mediated genome editing, mutations can also be induced by transformation and tissue culture alone (Phillips et al. 1994; Evans and Sharp 1983; Han et al. 2018; X. Zhang et al. 2022). Therefore, for comparison, we selected five T-DNA-less T_1 plants originating from three T_0 plants that had been transformed with an unrelated, no nuclease-containing overexpression construct as controls. While these plants have been subjected to *Agrobacterium*–mediated transformation and tissue culture, their genomes have not been exposed to endonucleases such as Cas9.

Finally, we selected six wild type plants from two generations (three plants from an original cv. Moneyberg GT seed batch, used for all transformations in this study, and three offspring from selfings from this batch) as wild type controls. An overview of all 49 selected plants and their genotypes can be found in Figure 2.

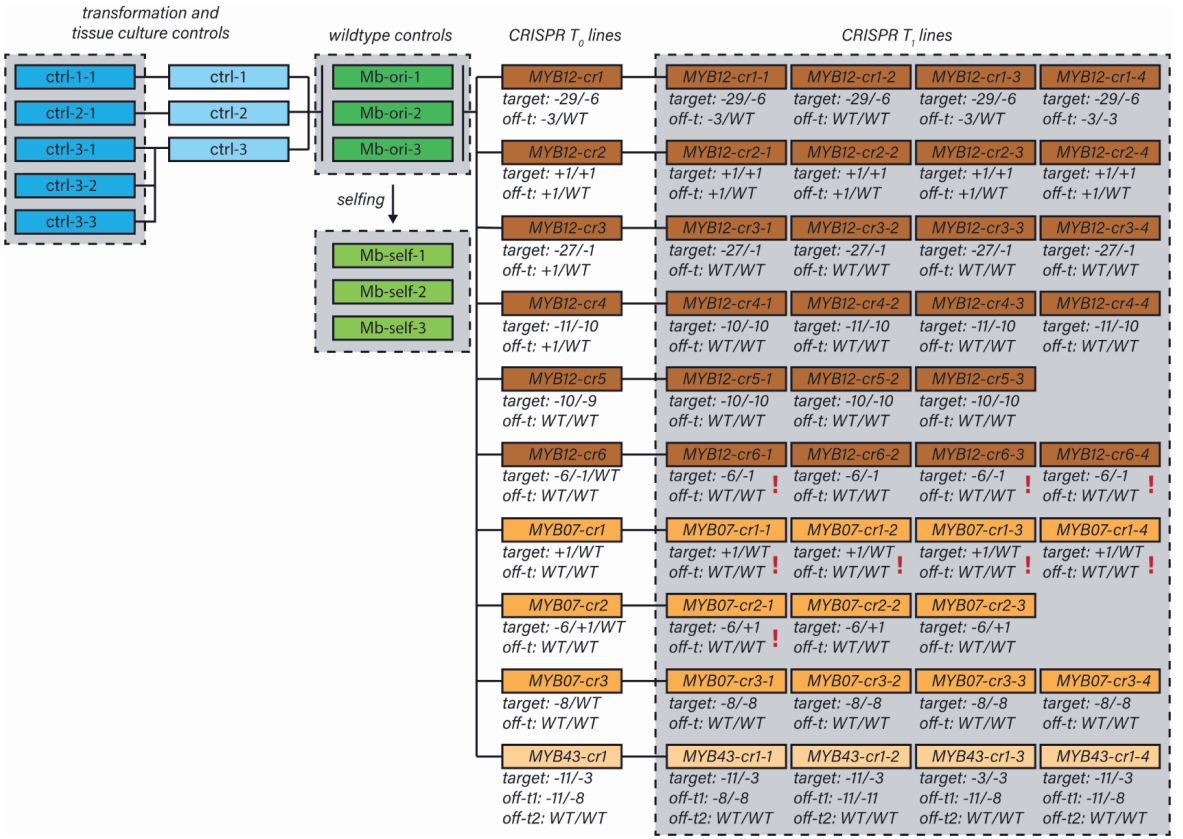


Figure 2 - Overview of plants used for WGS and their genotypes. Three wild type controls (dark green) from an original cv. Moneyberg GT seed batch were selected for WGS, as were three offspring (light green) resulting from selfings of this original seed batch. Five T-DNA free T_1 plants derived (blue) from 3 T_0 lines (light blue) that were originally transformed with an unrelated, nuclease-gene free construct were selected as tissue culture and transformation controls. 38 mutant CRISPR T_1 lines, resulting from 10 T_0 lines, were selected. All plants selected for WGS are in grey-shaded boxes. Genotypes of target sites and known off-target sites are shown for both T_0 and T_1 plants. Red exclamation marks indicate that T-DNA was still found to be present in the plant after WGS.

CRISPR-Cas9 genome editing has no apparent effects on indel and SV counts

Genomic DNA from all 49 plants was sequenced at an average read depth of 34x (Supplementary Table 1). Reads were mapped to the recently published long-read genome assembly of the Tomato Mosaic Virus resistant Moneyberg cultivar (MbTMV), which is the cultivar Moneyberg GT used in our research (van Rengs et al. 2022). On average, ~71.400 SNVs, ~28.900 indels, and ~450 SVs were identified in every genome, including the wild type controls (Figure 3a,c,e).

We compared the number of SNVs, indels, and SVs identified in genome-edited CRISPR lines to the tissue culture/transformation and wild type controls (Figure 3b,d,f). The number of SNVs in CRISPR lines was significantly, though moderately, higher than in wild type lines, but not different from the numbers in tissue culture control lines (Figure 3b). For indels and SVs, no such difference was found (Figure 3d,f).

Most CRISPR-Cas-induced mutations are indels (Chapter 2, Chapter 3, N. Zhang et al. 2020; Bernabé-Orts et al. 2019). For Cas9, these indels occur with recognizable patterns, with the single most frequent event being a one-basepair insertion (Chapter 2, Chapter 3). Therefore, we investigated the pattern of indels obtained for CRISPR lines, tissue culture, transformation controls, and wild types (Figure 3g). The patterns were very similar, and no significant differences were found.

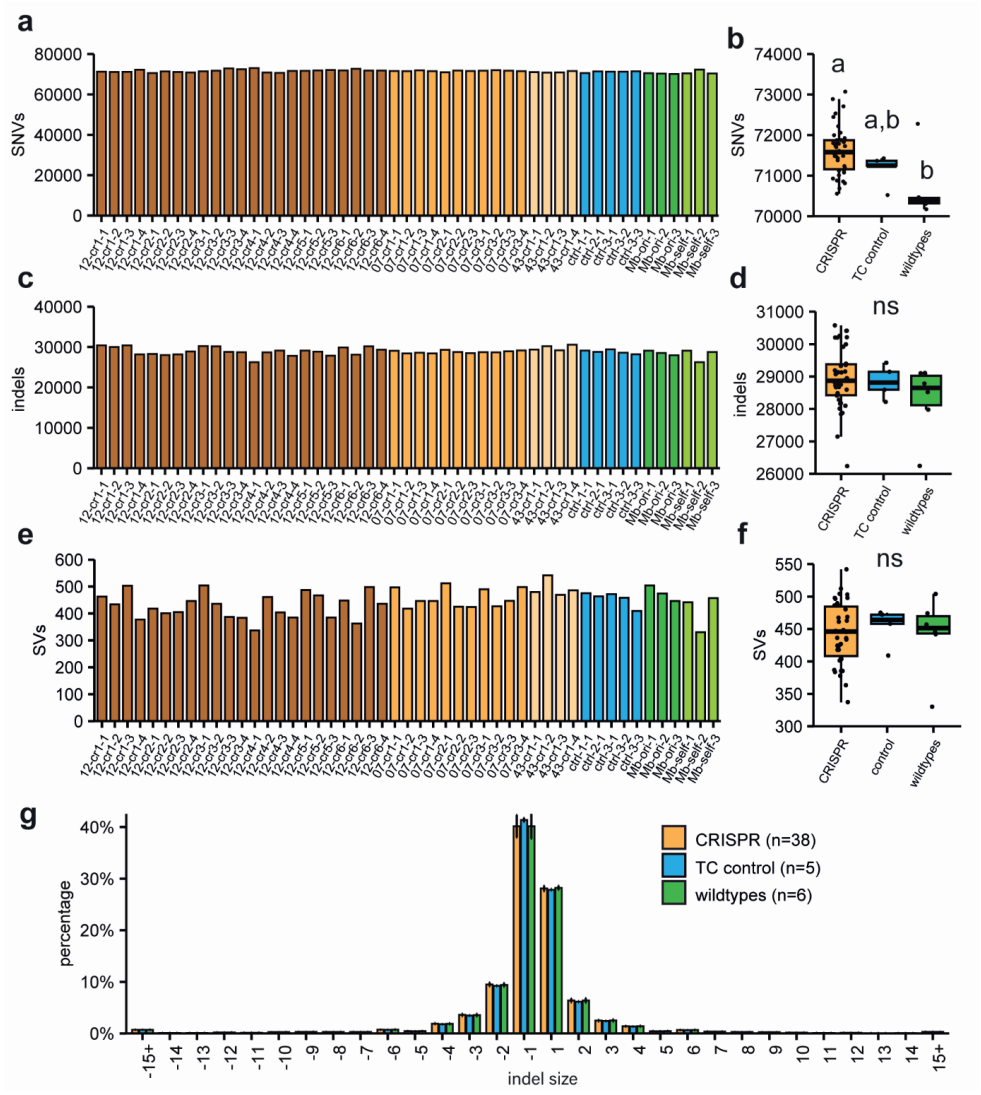


Figure 3 - Overview of the number of SNVs, indels and SV identified in all 49 sequenced plants. (a, b) SNVs, (c,d) indels, (e,f) SVs. For the boxplots in (b,d,f): the box indicates the 2nd and 3rd quartile. The median is indicated by the horizontal black line. 1st and 4th quartile are indicated by bottom and top whiskers, respectively. CRISPR: genome edited T_1 lines (n = 38). TC control: transformation and tissue culture controls (n = 5). Wildtypes: wild type plants from two consecutive generations (n = 6). Significant differences between the number of SNVs, indels or SVs for plants in the different categories were determined by a Kruskal-Willis test, followed by Wilcoxon's Rank Sum Test for pairwise comparison. (g) indel patterns obtained for genome-edited plants, transformation and tissue culture control plants, and wild type plants. Error bars indicate standard deviation.

Prolonged *Cas9* presence did not increase SNV, indel, and SV counts

In the binary vector for tomato transformation, a *tGFP* marker was included to help select successfully transformed T_0 plants from tissue culture. Additionally, the absence or presence of GFP was used in the T_1 generation to select seedlings that either segregated or retained the T-DNA. Although we selected seemingly GFP-negative seedlings for WGS, we found that for eight plants, reads aligned to the complete T-DNA used for transformation (Figure 4a). No evidence of (partial) T-DNA presence was found in the other thirty plants. As we used a constitutive Cauliflower Mosaic Virus 35S (CaMV35S) promoter for Cas9 expression, the T-DNA positive plants may have been exposed to the endonuclease activity for a longer time than the other, T-DNA free plants. For four of these plants (*MYB07-cr1-1* to *-4*), initial genotyping of the target site had revealed heterozygous on-target mutations, with one allele being wild type and thus still amenable to Cas9-mediated mutagenesis (Figure 2). We inspected the mutations found at these target sites in the genome sequencing data. In *MYB07-cr1-1*, we found a single read that contained a new 8 bp deletion (indicated by a dashed box, Figure 4b) that might result from a somatic mutation. No additional alleles were identified in the other three lines (Figure 4b). Additionally, we compared the numbers of SNVs, indels, and SVs between T-DNA positive and T-DNA negative genome-edited plants (Figure 4c). No significant differences were found.

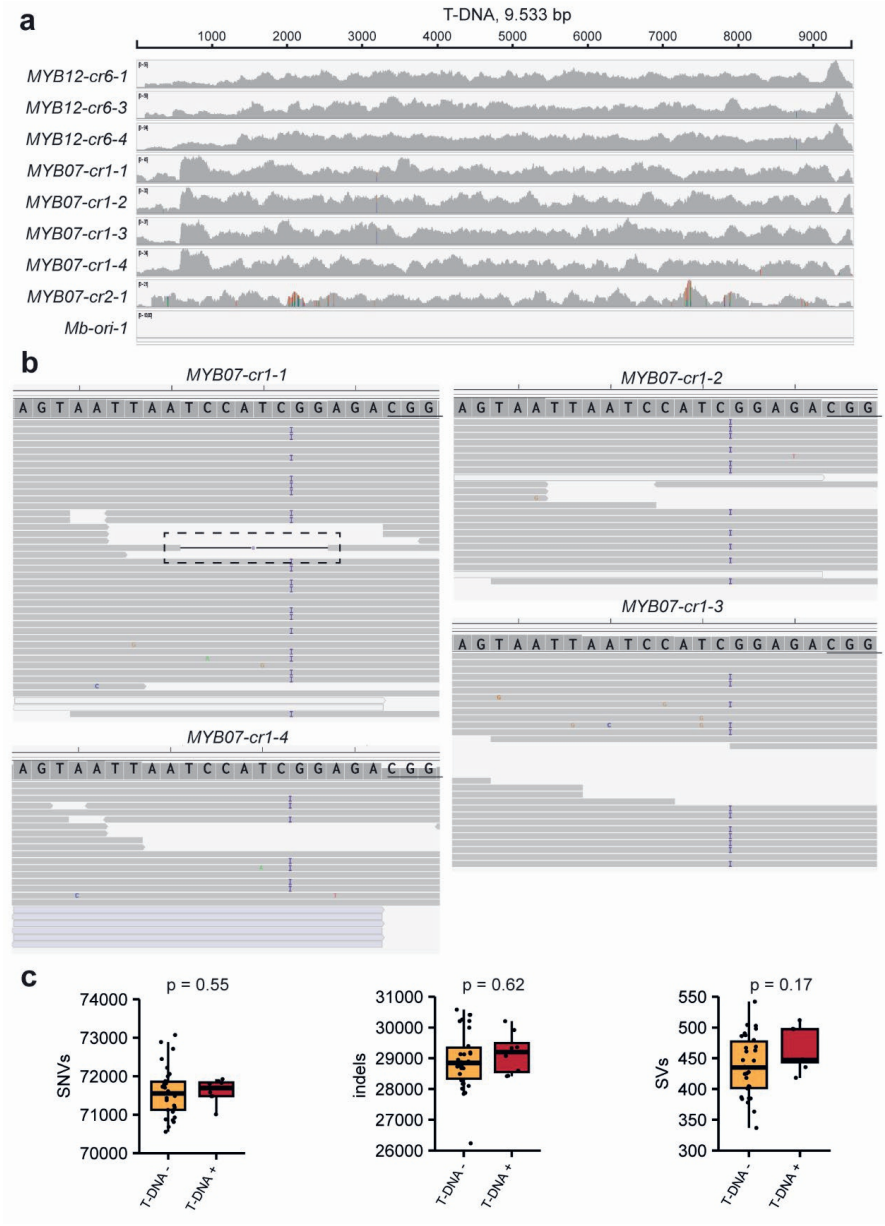


Figure 4 - Effects of prolonged presence of T-DNA in CRISPR-Cas9 genome edited T_1 lines. (a) Coverage plot showing the presence of T-DNA in eight lines. Mb-ori-1 is shown as wild type control. (b) Aligned WGS reads for the MYB07 target site. The complete target site is shown and the PAM is underlined. A +1bp "G" insertion, when present, is indicated by a vertical purple line. For MYB07-cr1-1, one read is found that contains a -8 bp deletion. This read is indicated by a dashed box. (c) Amount of SNVs, indels and SVs found in T-DNA negative (-) and T-DNA positive (+) lines. The median is represented by the black horizontal line. Boxes indicate 2nd and 3rd quartile, top and bottom whiskers the 1st and 4th quartile, respectively. The p-value shown is determined by Wilcoxon's test.

Variation within cultivar Moneyberg GT

In the three plants from our original Moneyberg GT seed batch, we found more indels and SNVs compared to the reference MbTMV genome than expected. When comparing the three genomes, we found most of these differences in all three plants (Figure 5a). This observation could indicate minor errors in the MbTMV genome assembly or differences between the exact plant used to generate this assembly and those used in our study. On average, these wild type plants contained 4348 unique indels and 3135 unique SNVs, suggesting that variation within the cultivar exists. We additionally inspected SNVs and indels found in the offspring of selfed Moneyberg GT plants from the original seed batch. Most of the identified variation was also present in the original parent plants (Figure 5b). After excluding the common SNVs and indels, we still found SNVs and indels that overlapped for all three offspring. These variants are unlikely to have independently arisen from one generation to the next. Therefore, they strongly indicate that not all standing variation within the cultivar has been captured by the sequenced plants (Figure 5c). Remarkably, thousands of unique indels and SNVs can still be identified in these offspring plants.

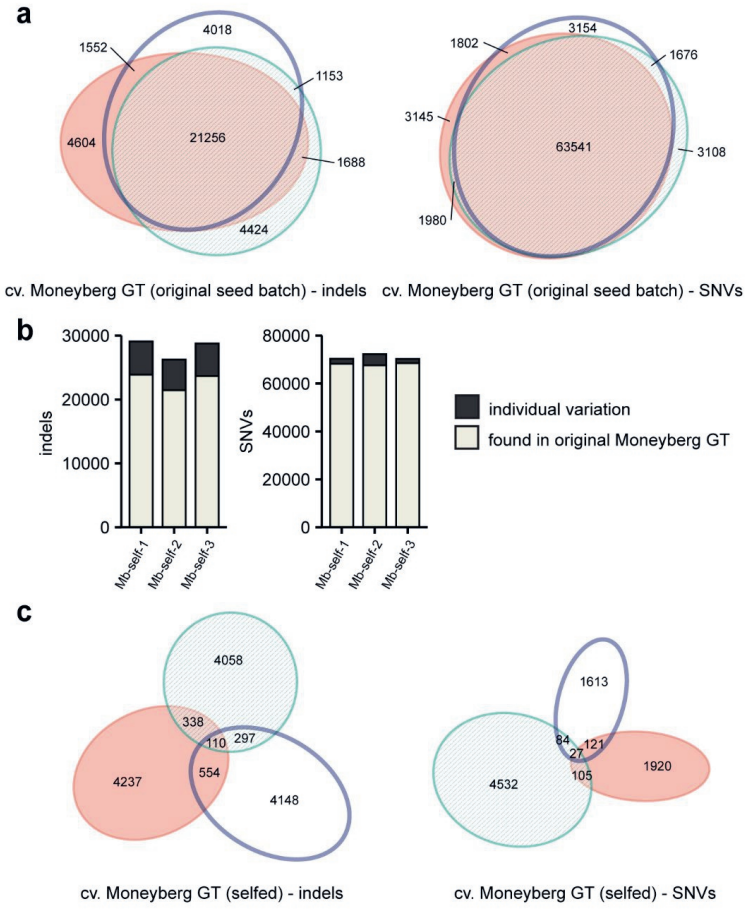


Figure 5 - The majority of identified variation in wild type lines overlaps. (a) Venn diagrams showing the overlap between identified indels and SNVs of the three sequenced plants from the original seed batch. (b) Variation found for the three sequenced offspring of the original MbTMV seed batch. Variation that is also identified in the original MbTMV plants is indicated. (c) Venn diagrams showing the overlap between individual indels and SNVs of the three sequenced offspring. Individual indels and SNVs are the indels and SNVs remaining after filtering out all variation that also occurs in the original sequenced MbTMV plants.

Identification of potential off-target mutations

To identify potential CRISPR-induced off-target mutations, we predicted all off-target sites with up to 6 mismatches and NGG, NAG, and NGA PAMs for all three sgRNAs. We used CasOFFinder (Bae et al. 2014), with the MbTMV assembly as the reference genome. The complete pipeline is shown in Figure 6a. In total, we screened 29.803 predicted off-target sites for the presence of mutations.

To avoid false positive calling of off-target mutations in CRISPR lines, we aimed to exclude as much naturally occurring variation as possible. Therefore, we disregarded all SNVs, indels, and SVs from

these lines that were found in any of the six wild types or the five transformation and tissue culture control plants. This action resulted in an average number of 3691 remaining indels, 2418 remaining SNVs, and 11 remaining SVs per genome-edited plant (Figure 6b, c, d, Supplementary Figures 2, 3).

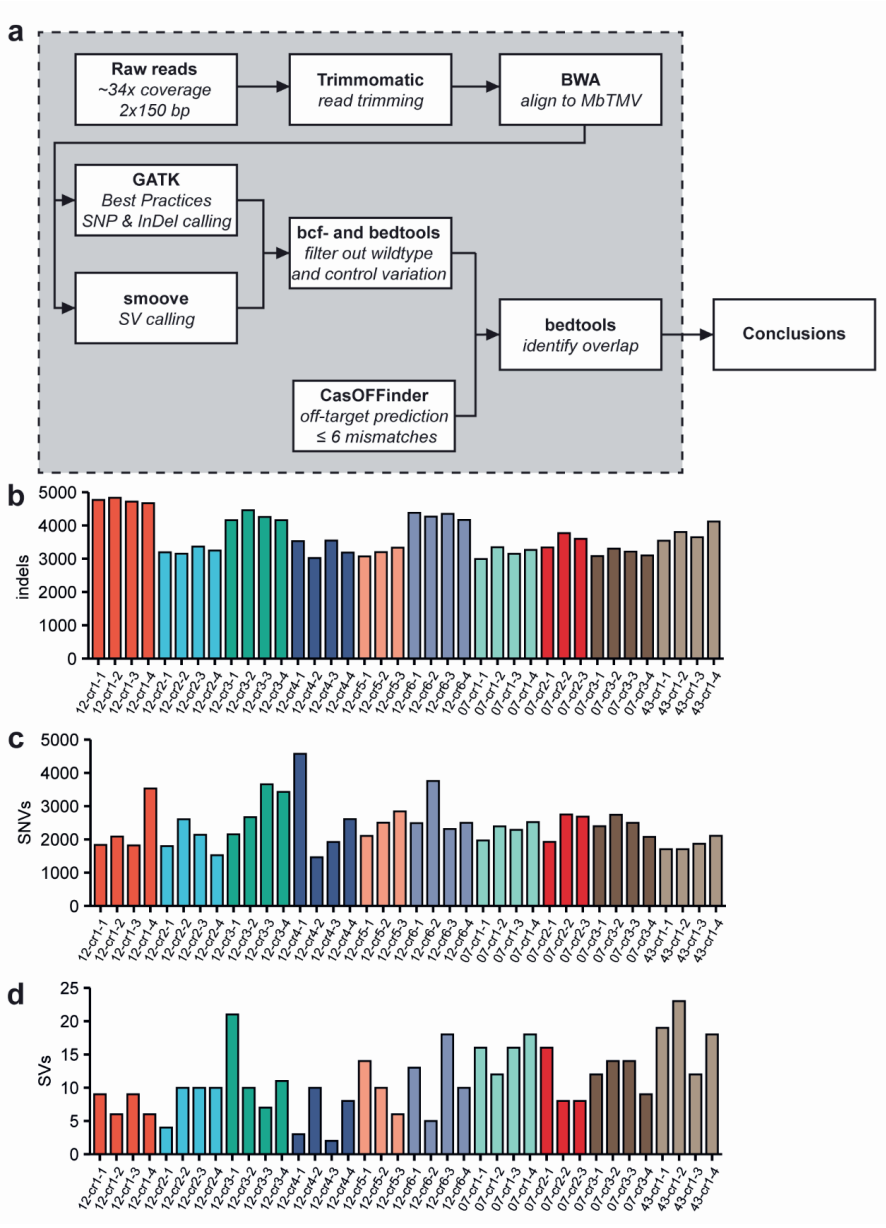


Figure 6 - Filtering of variation for off-target mutation identification. (a) Pipeline used to identify and filter variation. (b,c,d) The number of indels (b), SNVs (c) and SVs (d) remaining in the CRISPR-Cas9 genome edited lines after filtering out all variation that occurs in wild type plants or transformation and tissue culture control plants.

We then determined which of these SNVs, indels, or SV-breakends was located in a 40 bp window surrounding the predicted off-target sites. We could readily identify all on-target mutations and the previously predicted and discovered off-target mutations. The latter all occur at off-target positions that have a single mismatch to the target site (Table 1). In addition, 36 sequence variants in *MYB12*-targeted plants, 15 in *MYB07*-targeted plants, and 4 in *MYB43*-targeted plants, respectively, were located in such a 40 bp window. Most of these (50 out of 55) occurred at positions with six mismatches to the target site (Supplementary Tables 2,3,4,5). We then visually inspected the read alignments to determine if the identified SNV or indel (no SV-breakends were found to overlap) was consistent with a CRISPR-induced mutation. For most mutations, this was not the case, as the SNV or indel did not overlap the -3 position relative to the PAM where Cas9 is known to cut (Jinek et al. 2012; Zuo and Liu 2016; Shou et al. 2018) (see IGV snapshots in Supplementary Dataset 1). Only one sequence variant could not be excluded as being an off-target mutation based on its position and could have been introduced by Cas9 (Table 1, Figure 7). However, the exact same variant was also found at low frequencies in several wild type and transformation and tissue culture control lines (Figure 7, Supplementary Figure 4).

Table 1 - The number of predicted off-target sites that contain mutations that can be ascribed to CRISPR-Cas9 mediated genome editing.

	MYB12			*MYB07*			*MYB43*		
PAM / mism	NGG	NAG	NGA	NGG	NAG	NGA	NGG	NAG	NGA
0	**1/1**	-	-	**1/1**	-	0/1	**1/1**	-	-
1	**1/1**	-	-	-	-	-	**1/2**	-	-
2	0/1	-	-	-	-	0/3	-	0/2	-
3	0/1	0/2	-	0/7	0/3	0/13	0/2	0/10	0/5
4	0/16	0/23	0/29	0/34	0/40	0/85	0/7	0/33	0/9
5	**1/149**	0/560	0/370	0/536	0/528	0/758	0/99	0/207	0/165
6	0/1261	0/5225	0/2808	0/2772	0/4949	0/4818	0/818	0/1710	0/1741

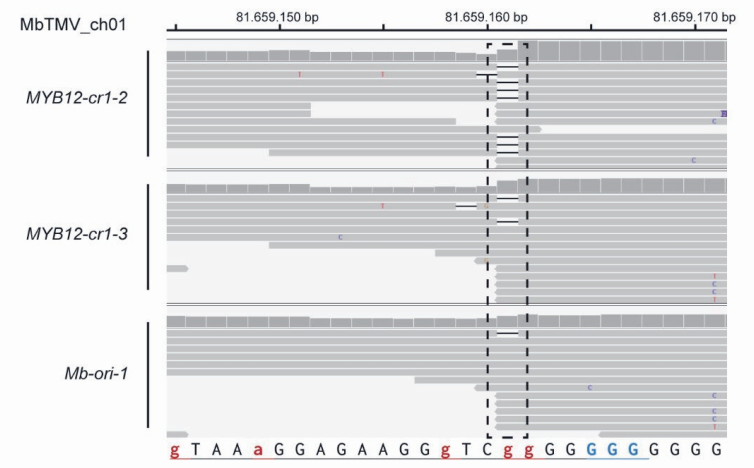


Figure 7 - Potential off-target site identified in MYB12-cr1-2 and MYB12-cr1-3. Aligned reads are shown for both lines, as well as for Mb-ori-1 as a wild type control. The potential off-target mutation, a 1 bp deletion, is indicated by a dashed box. The potential off-target site is underlined. The five mismatches to the target site are indicated by red lowercase. The PAM is indicated in blue.

Comparison of off-target predictions between two reference genomes

Our previous research selected potential off-target sites for sgRNAs using the Heinz SL4.0 genome as a reference, available on solgenomics.net. As more recently, the MbTMV reference became available and was used in the current study, we determined whether any difference exists in predicted off-target sites for the three sgRNAs targeting *MYB12*, *MYB07* and *MYB43* used in this study and compared the CasOFFinder output for MbTMV with SL4.0 (Table 2). Although results for high-homology putative off-target sites (up to 3 mismatches) were generally identical, considerable differences were identified for sites with more mismatches. This underscores the importance of high-quality reference genome assemblies for off-target site prediction.

Table 2 - Comparison of the number of predicted off-target sites for the three sgRNAs used in this study in the MbTMV and (in brackets) SL4.0 genome assemblies. Bold numbers indicate different outcomes.

	MYB12			*MYB07*			*MYB43*		
PAM / mism	NGG	NAG	NGA	NGG	NAG	NGA	NGG	NAG	NGA
0	1 (1)	0 (0)	0 (0)	1 (1)	0 (0)	1 (1)	1 (1)	0 (0)	0 (0)
1	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (2)	0 (0)	0 (0)
2	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	3 (3)	0 (0)	2 (2)	0 (0)
3	1 (1)	**2 (3)**	0 (0)	7 (7)	3 (3)	**13 (12)**	2 (2)	10 (10)	**5 (4)**
4	**16 (15)**	**23 (22)**	**29 (30)**	**34 (37)**	**40 (37)**	**85 (81)**	7 (7)	33 (33)	**9 (8)**
5	**149 (142)**	**560 (535)**	**370 (379)**	**536 (532)**	**528 (520)**	**758 (740)**	**99 (88)**	**207 (196)**	**165 (164)**
6	**1261 (1220)**	**5225 (5052)**	**2808 (2733)**	**2772 (2702)**	**4949 (4894)**	**4818 (4690)**	**818 (769)**	**1710 (1623)**	**1741 (1666)**

Discussion and Conclusion

In this study, we set out to determine the specificity of CRISPR-Cas9-mediated genome editing using whole genome resequencing while simultaneously assessing the suitability and limits of that approach. We resequenced the genomes of 38 CRISPR-Cas9 edited T_1 plants, five non-edited transformation and tissue culture control plants, and six wild type plants at an average coverage of ~34x.

Previous research has shown that mutations formed by CRISPR-Cas9 mediated genome editing are predominantly indels (Chapter 2, Chapter 3, N. Zhang et al. 2020; Bernabé-Orts et al. 2019). Importantly, similar numbers of indels were identified in all genome-edited lines compared to the reference genome (Figure 3), indicating that CRISPR-Cas9 does not cause dramatic numbers of *de novo* variation. Further proof of this was found by inspecting the size pattern of identified indels (Figure 3g). The indel size pattern was similar for CRISPR lines, transformation and tissue culture control plants, and wild type plants and did not show the characteristic pattern expected for CRISPR-Cas9 induced mutations: a peak at +1 bp with the highest frequency, followed by deletions of increasing size. In contrast, we find a peak at -1 bp, and relatively high frequencies of 2 bp and 3 bp insertions, which are rarely found for Cas9-mediated genome editing (Chapter 2, Chapter 3). Additionally, similar numbers of SVs identified in all lines indicate that CRISPR-Cas9 does not induce more large deletions, insertions or inversions than occur in the non-CRISPR-Cas9 plants (Figure 3c,f).

Interestingly, the number of SNVs was slightly higher in CRISPR lines than in wild type control plants (Figure 3d). A similar trend was observed in rice (Tang et al. 2018). As CRISPR-Cas9 mediated genome editing rarely induces SNVs, it is unlikely that this difference is caused by off-target mutations. Since a slightly elevated number of SNVs was also identified in the tissue culture and transformation control plants compared to wild type plants, it is possible that the process of transformation and tissue culture induces these SNVs, rather than CRISPR-Cas9 activity.

In the binary vector used for transfection, a tGFP marker was included to aid in the selection of successful transformants during the tissue culture process and transgene-free seedlings in the T_1 generation. Although we selected seemingly GFP-negative T_1 plants for whole genome sequencing, eight plants out of 38 were shown to contain the T-DNA (Figure 4a). Previously, it has been shown that small fragments of T-DNA can be integrated into the genome (Schouten et al. 2017), but we could not find evidence of this in the other 30 genome-edited T_1 lines.

As we used the constitutive CaMV35S promoter for Cas9 expression, the T-DNA-positive plants have potentially been exposed to Cas9 for longer than their T-DNA-negative counterparts. However, this did not increase SNV, indel, or SV numbers (Figure 4c). Although we found limited evidence for the appearance of an additional somatic CRISPR-induced allele in one line, three other lines showed no evidence of further Cas9 editing at the target site (Figure 4b). This observation might indicate that mutations rarely arise outside of the initial transformation and tissue culture process, which was previously also observed in rice (Tang et al. 2018). However, as we have used the same promoter for both tGFP and Cas9 expression (Brooks et al. 2014), the absence of GFP expression and the lack of additional mutations can also be explained by the silencing of the CaMV35S promoter. Nevertheless, a cautionary note to users of fluorescent markers is that it would be prudent to confirm the absence of T-DNA in GFP-negative seedlings, for example, using PCR.

A long-read genome assembly was recently published for the Moneyberg (MbTMV) tomato cultivar used in this study (van Rengs et al. 2022). Yet, we identified substantial variation (~71.400 SNVs, ~28.900 indels and ~450 SVs) in our wild type plants compared to this reference. Possible explanations for these identified differences are minor errors in the MbTMV assembly, false positives in our SNV and indel calling, and standing variation within the cultivar (Figure 5) (Durand et al. 2010; Barchenger et al. 2022). The MbTMV assembly was produced using long-read sequencing data (PacBio HiFi and Oxford Nanopore). However, repetitive regions remain challenging to assemble correctly (van Rengs et al. 2022). Additionally, reliably calling SNVs and indels in repetitive regions, and indels in homopolymer stretches, is difficult (Lefouili et al. 2022; Nakamura et al. 2011; Minoche et al. 2011). Such false positives can partially be prevented by applying repeat masking. However, as this might have hidden actual CRISPR-Cas9-induced off-target mutations, we chose not to do this in this research. Taken together, these factors complicate the detection of low-frequency events and make it challenging to apply WGS for truly unbiased off-target mutation detection.

Encountering the extent of variation found in this study is not exceptional, and it can be argued that our reported SNV and indel numbers are even relatively low due to the availability of a highly related, high-quality reference genome. Before filtering out variation occurring in wild type resequenced plants, a WGS study in cotton reported on average ~1.249.000 SNVs and ~149.100 indels in the genome of CRISPR-Cas9-edited plants compared to a reference genome (J. Li et al. 2019), and a study in grapevine ~7.514.000 SNVs and ~663.700 indels (Wang et al. 2021). Much lower numbers were reported in rice (maximum ~150 SNVs and ~75 indels) (Tang et al. 2018). However, the authors of this study applied much more stringent filtering to their SNVs and indels, only

selecting variants that were collectively called by three different software tools for further inspection. Individually, the software tools identified hundreds to tens of thousands of variants.

Apart from minimizing the number of background variations, the need for an appropriate reference genome is also exemplified by the different numbers of potential off-target sites predicted for the three used sgRNAs in the MbTMV and SL4.0 assembly (Table 2). This observation was previously also illustrated by a study in grapevine (Wang et al. 2021).

After excluding as many common variants from the genome-edited lines as possible, we identified three sites out of 29.803 with (potential) off-target mutations. Two of these sites, which have only a single mismatch to the target site, were previously identified by both amplicon sequencing and GUIDE-seq (Figure 1, Chapters 2 and 4 of this thesis) and were verified off-target sites. The other mutation is a one bp deletion called in a homopolymer stretch (Figure 7) and belonged to a predicted off-target site with five mismatches to the target site. Both these characteristics make it unlikely – though not impossible - that this indel results from actual off-target activity. Off-target activity at sites with as many as five mismatches has only been reported once in plants (Wang et al. 2021). Additionally, homopolymer stretches are notoriously challenging to sequence and may result in false positive indel calling (Minoche et al. 2011; Nakamura et al. 2011). An argument for the latter explanation is that low numbers of reads with the same deletion were also found in several wild type and tissue culture and transformation control plants (Figure 7, Supplementary Figure 4). This example illustrates one of the limitations of WGS for (off-target) mutation discovery. Sanger sequencing of this site could resolve this issue.

The results on off-target mutations reported here are in agreement with other whole-genome resequencing studies performed in plants. Studies in *Arabidopsis*, rice, tomato and *Physcomitrium patens* did not identify off-target mutations at all (Z. Feng et al. 2014; Peterson et al. 2016; Nekrasov et al. 2017; C. Feng et al. 2018; Bessoltane et al. 2022). Likewise, we could identify T_1 plants without any off-target mutations for the sgRNAs targeting *MYB12* and *MYB07*. Limited numbers of off-target mutations were previously found in rice, cotton and grapevine (H. Zhang et al. 2014; Tang et al. 2018; J. Li et al. 2019; Wang et al. 2021). The corresponding off-target sites had high homology to the target site and, with the exception of one off-target site reported in grapevine that had 5 mismatches, contained only 1-3 mismatches to the intended target. Furthermore, these mismatches were frequently distal from the PAM. We and others have shown that such off-target sites are high-risk (Chapters 2 and 3 of this thesis, Hsu et al. 2013; Fu et al. 2013). Accordingly, we found similar high-

risk off-target sites to be mutagenized in T_1 plants in which *MYB12* and *MYB43* were targeted by CRISPR-Cas9.

Although we performed whole genome resequencing and obtained a complete overview of all variation in 38 genome-edited T_1 lines, we could not identify off-target mutations in a truly unbiased manner. While WGS allows the easy scanning of tens of thousands of predicted off-target sites for mutations and will thus give a good measure of specificity, it is, in our case, impossible to determine the origin of other sequence variants with certainty. In our study, every genome-edited plant contained, on average, 3691 individual indels – the variant type that is associated with CRISPR-Cas9 activity (Figure 6b). Only a few of these were present in proximity of predicted off-target sites (Supplementary Table 2). We assume that much of the remaining variants are either the result of inaccurate variant calling, natural variation that has not been captured by the eleven plants used as controls, or new variation induced by transformation or tissue culture. However, we cannot determine their origin with certainty. To allow off-target detection as unbiased as possible, a genome assembly or at least WGS data for the wild type parent plants that were used to create the T_0 lines would ideally be available, as well as WGS data for the T_0 and the resulting T_1 generations. Although this might be feasible for studies explicitly focusing on CRISPR specificity, it is not practical for most general users of CRISPR-Cas9 due to the involved cost and the difficulty of working with large sequencing datasets. Furthermore, most studies investigating the specificity of CRISPR-Cas9 mediated genome editing thus far have reported low numbers of off-target mutations or no off-target mutations at all (Tang et al. 2018; J. Li et al. 2019; Wang et al. 2021; Bessoltane et al. 2022; reviewed by Sturme et al. 2022). Therefore, whole genome resequencing is unlikely to yield significantly more informative data on potential off-target mutations than easier, biased methods for off-target mutation detection such as amplification of predicted off-target sites followed by Sanger sequencing.

Concluding, we find that CRISPR-Cas9-mediated genome editing does not result in elevated numbers of indels or SVs in the genome. Furthermore, we tested 29,803 predicted off-target sites from three sgRNAs for the presence of mutations. We only found mutations at two previously verified off-target sites with only one mismatch and possibly at one additional site with five mismatches. At the latter site, an indel might have been erroneously called due to the presence of a homopolymer stretch. We could not detect off-target mutations in a truly unbiased manner due to the large number of identified variations, which are also abundantly present in control plants. Nevertheless, we can conclude that if unpredicted off-target mutations occur, they occur at frequencies that are insignificant when compared to pre-existing variation and variation resulting

from transformation and tissue culture. Therefore, we conclude that CRISPR-Cas9-mediated genome editing in tomato is specific.

Acknowledgements

This research was supported by the Dutch Research Council (NWO), and is funded by the Ministry of Infrastructure and Water Management. We thank Eveline Da Rocha Tavano for the tomato plants that were used as transformation and regenerations controls in this study.

References

Auwera, G.A. Van der, M.O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, … M.A. DePristo. 2013. "From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline." *Current Protocols in Bioinformatics*, no. SUPL.43. https://doi.org/10.1002/0471250953.bi1110s43.

Bae, S., J. Park, and J.-S. Kim. 2014. "Cas-OFFinder: A Fast and Versatile Algorithm That Searches for Potential off-Target Sites of Cas9 RNA-Guided Endonucleases." *Bioinformatics* 30 (10): 1473–75. https://doi.org/10.1093/BIOINFORMATICS/btu048.

Barchenger, D.W., Y. ming Hsu, J. yang Ou, Y. ping Lin, Y. cheng Lin, M.A.O. Balendres, Y. che Hsu, R. Schafleitner, and P. Hanson. 2022. "Whole Genome Resequencing and Complementation Tests Reveal Candidate Loci Contributing to Bacterial Wilt (Ralstonia Sp.) Resistance in Tomato." *Scientific Reports* 12 (1). https://doi.org/10.1038/s41598-022-12326-x.

Bernabé-Orts, J.M., I. Casas-Rodrigo, E.G. Minguet, V. Landolfi, V. Garcia-Carpintero, S. Gianoglio, M. Vázquez-Vilar, … D. Orzaez. 2019. "Assessment of Cas12a-Mediated Gene Editing Efficiency in Plants." *Plant Biotechnology Journal* 17 (10): 1971–84. https://doi.org/10.1111/pbi.13113.

Bessoltane, N., F. Charlot, A. Guyon-Debast, D. Charif, K. Mara, C. Collonnier, P.-F. Perroud, M. Tepfer, and F. Nogué. 2022. "Genome-Wide Specificity of Plant Genome Editing by Both CRISPR–Cas9 and TALEN." *Scientific Reports* 12 (1): 1–10. https://doi.org/10.1038/s41598-022-13034-2.

Bolger, A.M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics (Oxford, England)* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Brooks, C., V. Nekrasov, Z.B. Lippman, and J. Van Eck. 2014. "Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/ CRISPR-Associated9 System." *Plant Physiology* 166 (3): 1292–1297. https://doi.org/10.1104/pp.114.247577.

Cameron, P., C.K. Fuller, P.D. Donohoue, B.N. Jones, M.S. Thompson, M.M. Carter, S. Gradia, … A.P. May. 2017. "Mapping the Genomic Landscape of CRISPR–Cas9 Cleavage." *Nature Methods* 14 (6). https://doi.org/10.1038/nmeth.4284.

Concordet, J.P., and M. Haeussler. 2018. "CRISPOR: Intuitive Guide Selection for CRISPR/Cas9 Genome Editing Experiments and Screens." *Nucleic Acids Research* 46 (W1): W242–45. https://doi.org/10.1093/nar/gky354.

Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, … F. Zhang. 2013. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science* 339 (6121): 819–23. https://doi.org/10.1126/science.1231143.

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, … R. Durbin. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27 (15): 2156–58. https://doi.org/10.1093/bioinformatics/btr330.

Durand, E., M.I. Tenaillon, C. Ridel, D. Coubriche, P. Jamin, S. Jouanne, A. Ressayre, A. Charcosset, and C. Dillmann. 2010. "Standing Variation and New Mutations Both Contribute to a Fast Response to Selection for Flowering Time in Maize Inbreds." *BMC Evolutionary Biology* 10 (1): 2. https://doi.org/10.1186/1471-2148-10-2.

Engler, C., R. Kandzia, and S. Marillonnet. 2008. "A One Pot, One Step, Precision Cloning Method with High Throughput Capability." *PLoS ONE* 3 (11). https://doi.org/10.1371/journal.pone.0003647.

Evans, D.A., and W.R. Sharp. 1983. "Single Gene Mutations in Tomato Plants Regenerated from Tissue Culture." *New Series*. Vol. 221.

Feng, C., H. Su, H. Bai, R. Wang, Y. Liu, X. Guo, C. Liu, … F. Han. 2018. "High-Efficiency Genome Editing Using a Dmc1 Promoter-Controlled CRISPR/Cas9 System in Maize." *Plant Biotechnology Journal* 16 (11): 1848–57. https://doi.org/10.1111/pbi.12920.

Feng, Z., Y. Mao, N. Xu, B. Zhang, P. Wei, D.-L. Yang, Z. Wang, … J.-K. Zhu. 2014. "Multigeneration Analysis Reveals the Inheritance, Specificity, and Patterns of CRISPR/Cas-Induced Gene Modifications in Arabidopsis." *Proceedings of the National Academy of Sciences of the United States of America* 111 (12): 4632–37. https://doi.org/10.1073/pnas.1400822111.

Fu, Y., J.A. Foden, C. Khayter, M.L. Maeder, D. Reyon, J.K. Joung, and J.D. Sander. 2013. "High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells." *Nature Biotechnology* 31 (9): 822–26. https://doi.org/10.1038/nbt.2623.

Han, Z., P.A. Crisp, S. Stelpflug, S.M. Kaeppler, Q. Li, and N.M. Springer. 2018. "Heritable Epigenomic Changes to the Maize Methylome Resulting from Tissue Culture." *Genetics* 209 (4): 983–95. https://doi.org/10.1534/genetics.118.300987.

Hsu, P.D., D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, Y. Li, … F. Zhang. 2013. "DNA Targeting Specificity of RNA-Guided Cas9 Nucleases." *Nature Biotechnology* 31 (9): 827–32. https://doi.org/10.1038/nbt.2647.

Jaganathan, D., K. Ramasamy, G. Sellamuthu, S. Jayabalan, and G. Venkataraman. 2018. "CRISPR for Crop Improvement: An Update Review." *Frontiers in Plant Science* 9 (July): 1–17. https://doi.org/10.3389/fpls.2018.00985.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Kim, D., S. Bae, J. Park, E. Kim, S. Kim, H.R. Yu, J. Hwang, J.-I. Kim, and J.-S. Kim. 2015. "Digenome-Seq: Genome-Wide Profiling of CRISPR-Cas9 off-Target Effects in Human Cells." *Nature Methods* 12 (3): 237–44. https://doi.org/10.1038/nmeth.3284.

Kwon, C.T., J. Heo, Z.H. Lemmon, Y. Capua, S.F. Hutton, J. Van Eck, S.J. Park, and Z.B. Lippman. 2020. "Rapid Customization of Solanaceae Fruit Crops for Urban Agriculture." *Nature Biotechnology* 38 (2): 182–88. https://doi.org/10.1038/s41587-019-0361-2.

Labun, K., X. Guo, A. Chavez, G. Church, J.A. Gagnon, and E. Valen. 2019. "Accurate Analysis of Genuine CRISPR Editing Events with AmpliCan." *Genome Research* 29 (5): 843–47. https://doi.org/10.1101/gr.244293.118.

Lee, K., Y. Zhang, B.P. Kleinstiver, J.A. Guo, M.J. Aryee, J. Miller, A. Malzahn, … K. Wang. 2018. "Activities and Specificities of CRISPR/Cas9 and Cas12a Nucleases for Targeted Mutagenesis in Maize." *Plant Biotechnology Journal*, 1–11. https://doi.org/10.1111/pbi.12982.

Lefouili, M., K. Nam, and N. Kiwong. 2022. "The Evaluation of Bcftools Mpileup and GATK HaplotypeCaller for Variant Calling in Non-Human Species." *Scientific Reports*, no. 1: 10. https://doi.org/10.1038/s41598-022-15563-2ï.

Lei, Y., L. Lu, H.Y. Liu, S. Li, F. Xing, and L.L. Chen. 2014. "CRISPR-P: A Web Tool for Synthetic Single-Guide RNA Design of CRISPR-System in Plants." *Molecular Plant* 7 (9): 1494–96. https://doi.org/10.1093/mp/ssu044.

Lemmon, Z.H., N.T. Reem, J. Dalrymple, S. Soyk, K.E. Swartwood, D. Rodriguez-Leal, J. Van Eck, and Z.B. Lippman. 2018. "Rapid Improvement of Domestication Traits in an Orphan Crop by Genome Editing." *Nature Plants* 4 (10): 766–70. https://doi.org/10.1038/s41477-018-0259-x.

Li, H. 2011. "A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation from Sequencing Data." *Bioinformatics* 27 (21): 2987–93. https://doi.org/10.1093/bioinformatics/btr509.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Li, J., H. Manghwar, L. Sun, P. Wang, G. Wang, H. Sheng, J. Zhang, … X. Zhang. 2019. "Whole Genome Sequencing Reveals Rare Off-target Mutations and Considerable Inherent Genetic or/and Somaclonal Variations in <scp>CRISPR</Scp> /Cas9-edited Cotton Plants." *Plant Biotechnology Journal* 17 (5): 858–68. https://doi.org/10.1111/pbi.13020.

Mali, P., L. Yang, K.M. Esvelt, J. Aach, M. Guell, J.E. DiCarlo, J.E. Norville, and G.M. Church. 2013. "RNA-Guided Human Genome Engineering via Cas9." *Science* 339 (6121): 823–26. https://doi.org/10.1126/science.1232033.

Minoche, A.E., J.C. Dohm, and H. Himmelbauer. 2011. "Evaluation of Genomic High-Throughput Sequencing Data Generated on Illumina HiSeq and Genome Analyzer Systems." *Genome Biology* 12 (11). https://doi.org/10.1186/gb-2011-12-11-r112.

Nakamura, K., T. Oshima, T. Morimoto, S. Ikeda, H. Yoshikawa, Y. Shiwa, S. Ishikawa, … S. Kanaya. 2011. "Sequence-Specific Error Profile of Illumina Sequencers." *Nucleic Acids Research* 39 (13). https://doi.org/10.1093/nar/gkr344.

Nekrasov, V., C. Wang, J. Win, C. Lanz, D. Weigel, and S. Kamoun. 2017. "Rapid Generation of a Transgene-Free Powdery Mildew Resistant Tomato by Genome Deletion." *Scientific Reports* 7 (1): 482. https://doi.org/10.1038/s41598-017-00578-x.

Pattanayak, V., S. Lin, J.P. Guilinger, E. Ma, J.A. Doudna, and D.R. Liu. 2013. "High-Throughput Profiling of off-Target DNA Cleavage Reveals RNA-Programmed Cas9 Nuclease Specificity." *Nature Biotechnology* 31 (9): 839–43. https://doi.org/10.1038/nbt.2673.

Peterson, B.A., D.C. Haak, M.T. Nishimura, P.J.P.L. Teixeira, S.R. James, J.L. Dangl, and Z.L. Nimchuk. 2016. "Genome-Wide Assessment of Efficiency and Specificity in CRISPR/Cas9 Mediated Multiple Site Targeting in Arabidopsis." Edited by B. Zhang. *PLOS ONE* 11 (9): e0162169. https://doi.org/10.1371/journal.pone.0162169.

Phillips, R.L., S.M. Kaeppler, and P. Olhoft. 1994. "Genetic Instability of Plant Tissue Cultures: Breakdown of Normal Controls." *Proceedings of the National Academy of Sciences of the United States of America* 91 (12): 5222–26. https://doi.org/10.1073/pnas.91.12.5222.

Puchta, H. 2004. "The Repair of Double-Strand Breaks in Plants: Mechanisms and Consequences for Genome Evolution." *Journal of Experimental Botany* 56 (409): 1–14. https://doi.org/10.1093/jxb/eri025.

Quinlan, A.R. 2014. "BEDTools: The Swiss-Army Tool for Genome Feature Analysis." *Current Protocols in Bioinformatics* 2014: 11.12.1-11.12.34. https://doi.org/10.1002/0471250953.bi1112s47.

Rengs, W.M.J. van, M.H.W. Schmidt, S. Effgen, D.B. Le, Y. Wang, M.W.A.M. Zaidan, B. Huettel, H.J. Schouten, B. Usadel, and C.J. Underwood. 2022. "A Chromosome Scale Tomato Genome Built from Complementary PacBio and Nanopore Sequences Alone Reveals Extensive Linkage Drag during Breeding." *Plant Journal* 110 (2): 572–88. https://doi.org/10.1111/tpj.15690.

Rensing, S.A., B. Goffinet, R. Meyberg, S.-Z. Wu, and M. Bezanilla. 2020. "The Moss Physcomitrium (Physcomitrella) Patens: A Model Organism for Non-Seed Plants." *The Plant Cell* 32 (5): 1361–76. https://doi.org/10.1105/tpc.19.00828.

Roekel, J.S.C. van, B. Damm, L.S. Melchers, and A. Hoekema. 1993. "Factors Influencing Transformation Frequency of Tomato (Lycopersicon Esculentum)." *Plant Cell Reports* 12 (11): 644–47. https://doi.org/10.1007/BF00232816.

Schouten, H.J., H. vande Geest, S. Papadimitriou, M. Bemer, J.G. Schaart, M.J.M. Smulders, G.S. Perez, and E. Schijlen. 2017. "Re-Sequencing Transgenic Plants Revealed Rearrangements at T-DNA Inserts, and Integration of

a Short T-DNA Fragment, but No Increase of Small Mutations Elsewhere." *Plant Cell Reports* 0 (0): 1–12. https://doi.org/10.1007/s00299-017-2098-z.

Shan, Q., Y. Wang, J. Li, Y. Zhang, K. Chen, Z. Liang, K. Zhang, … C. Gao. 2013. "Targeted Genome Modification of Crop Plants Using a CRISPR-Cas System." *Nature Biotechnology*. Nature Publishing Group. https://doi.org/10.1038/nbt.2650.

Shou, J., J. Li, Y. Liu, and Q. Wu. 2018. "Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion." *Molecular Cell* 71 (4): 498-509.e4. https://doi.org/10.1016/j.molcel.2018.06.021.

Sturme, M.H.J., J.P. Van Der Berg, L.M.S. Bouwman, A. De Schrijver, R.A. De Maagd, G.A. Kleter, and E. Battaglia-De Wilde. 2022. "Occurrence and Nature of Off-Target Modifications by CRISPR-Cas Genome Editing in Plants." *ACS Agricultural Science and Technology* 2 (2): 192–201. https://doi.org/10.1021/acsagscitech.1c00270.

Svitashev, S., C. Schwartz, B. Lenderts, J.K. Young, and A. Mark Cigan. 2016. "Genome Editing in Maize Directed by CRISPR–Cas9 Ribonucleoprotein Complexes." *Nature Communications* 7: 13274. https://doi.org/10.1038/ncomms13274.

Tang, X., G. Liu, J. Zhou, Q. Ren, Q. You, L. Tian, X. Xin, … Y. Zhang. 2018. "A Large-Scale Whole-Genome Sequencing Analysis Reveals Highly Specific Genome Editing by Both Cas9 and Cpf1 (Cas12a) Nucleases in Rice." *Genome Biology* 19 (1): 84. https://doi.org/10.1186/s13059-018-1458-5.

Tsai, S Q, Z. Zheng, N.T. Nguyen, M. Liebers, V. V Topkar, V. Thapar, N. Wyvekens, … J.K. Joung. 2015. "GUIDE-Seq Enables Genome-Wide Profiling of off-Target Cleavage by CRISPR-Cas Nucleases." *Nat Biotechnol* 33 (2): 187–97. https://doi.org/10.1038/nbt.3117.

Tsai, Shengdar Q, N.T. Nguyen, J. Malagon-Lopez, V. V Topkar, M.J. Aryee, and J.K. Joung. 2017. "CIRCLE-Seq: A Highly Sensitive in Vitro Screen for Genome-Wide CRISPR–Cas9 Nuclease off-Targets." *Nature Methods*, no. May. https://doi.org/10.1038/nmeth.4278.

Veillet, F., L. Perrot, L. Chauvin, M.P. Kermarrec, A. Guyon-Debast, J.E. Chauvin, F. Nogué, and M. Mazier. 2019. "Transgene-Free Genome Editing in Tomato and Potato Plants Using Agrobacterium-Mediated Delivery of a CRISPR/Cas9 Cytidine Base Editor." *International Journal of Molecular Sciences* 20 (2). https://doi.org/10.3390/ijms20020402.

Wang, X., M. Tu, Y. Wang, W. Yin, Y. Zhang, H. Wu, Y. Gu, Z. Li, Z. Xi, and X. Wang. 2021. "Whole-Genome Sequencing Reveals Rare off-Target Mutations in CRISPR/Cas9-Edited Grapevine." *Horticulture Research* 8 (1). https://doi.org/10.1038/s41438-021-00549-4.

Weber, E., C. Engler, R. Gruetzner, S. Werner, and S. Marillonnet. 2011. "A Modular Cloning System for Standardized Assembly of Multigene Constructs." Edited by J. Peccoud. *PLoS ONE* 6 (2): e16765. https://doi.org/10.1371/journal.pone.0016765.

Xu, W., W. Fu, P. Zhu, Z. Li, C. Wang, C. Wang, Y. Zhang, and S. Zhu. 2019. "Comprehensive Analysis of CRISPR/Cas9-Mediated Mutagenesis in Arabidopsis Thaliana by Genome-Wide Sequencing." *International Journal of Molecular Sciences* 20 (17). https://doi.org/10.3390/ijms20174125.

Yan, W.X., R. Mirzazadeh, S. Garnerone, D. Scott, M.W. Schneider, T. Kallas, J. Custodio, … N. Crosetto. 2017. "BLISS Is a Versatile and Quantitative Method for Genome-Wide Profiling of DNA Double-Strand Breaks." *Nature Communications* 8 (May): 15058. https://doi.org/10.1038/ncomms15058.

Young, J., G. Zastrow-Hayes, S. Deschamps, S. Svitashev, M. Zaremba, A. Acharya, S. Paulraj, … S. Kumar. 2019. "CRISPR-Cas9 Editing in Maize: Systematic Evaluation of Off-Target Activity and Its Relevance in Crop Improvement." *Scientific Reports* 9 (1). https://doi.org/10.1038/s41598-019-43141-6.

Zhang, H., J. Zhang, P. Wei, B. Zhang, F. Gou, Z. Feng, Y. Mao, … J.K. Zhu. 2014. "The CRISPR/Cas9 System Produces Specific and Homozygous Targeted Gene Editing in Rice in One Generation." *Plant Biotechnology Journal* 12 (6): 797–807. https://doi.org/10.1111/pbi.12200.

Zhang, N., H.M. Roberts, J. Van Eck, and G.B. Martin. 2020. "Generation and Molecular Characterization of CRISPR/Cas9-Induced Mutations in 63 Immunity-Associated Genes in Tomato Reveals Specificity and a Range of Gene Modifications." *Frontiers in Plant Science* 11 (February): 1–13. https://doi.org/10.3389/fpls.2020.00010.

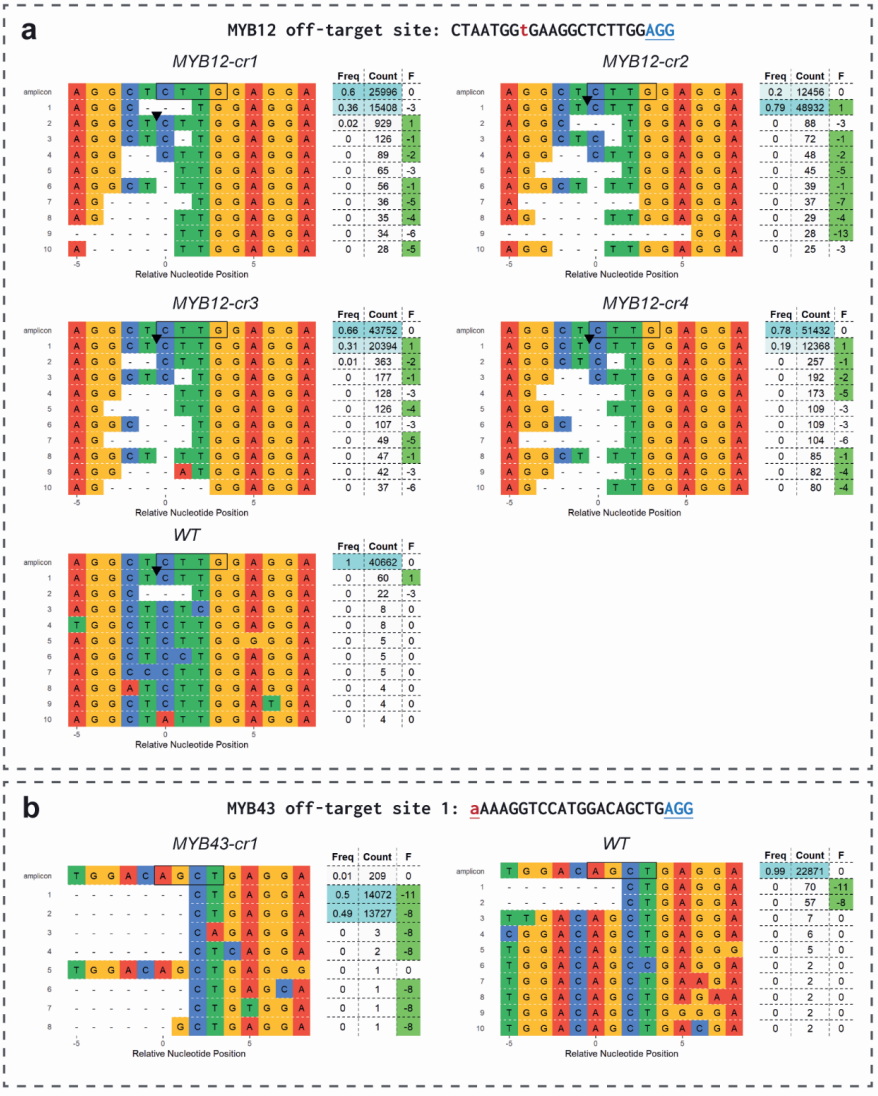
Zhang, X., M.J.G. Hooykaas, G. Paul Van Heusden, and P.J.J. Hooykaas. 2022. "The Translocated Virulence Protein VirD5 Causes DNA Damage and Mutation during Agrobacterium-Mediated Transformation of Yeast." *Sci. Adv*. Vol. 8. https://www.science.org.

Zhu, H., C. Li, and C. Gao. 2020. "Applications of CRISPR–Cas in Agriculture and Plant Biotechnology." *Nature Reviews Molecular Cell Biology* 21 (11): 661–77. https://doi.org/10.1038/s41580-020-00288-9.

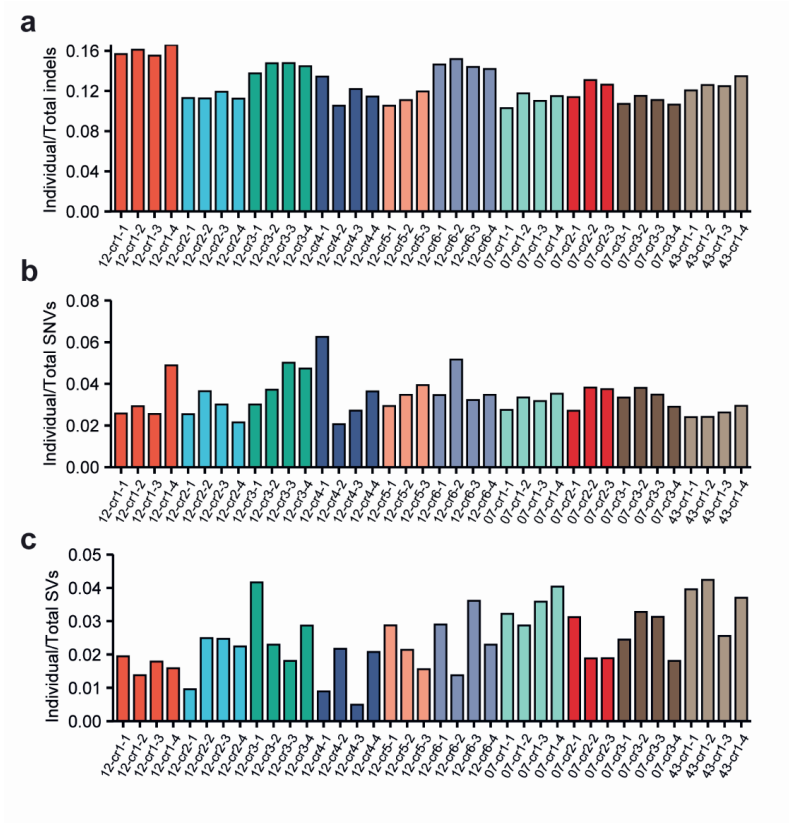
Zuo, Z., and J. Liu. 2016. "Cas9-Catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations." *Scientific Reports* 5 (October): 1–9. https://doi.org/10.1038/srep37584.

Supplementary Information

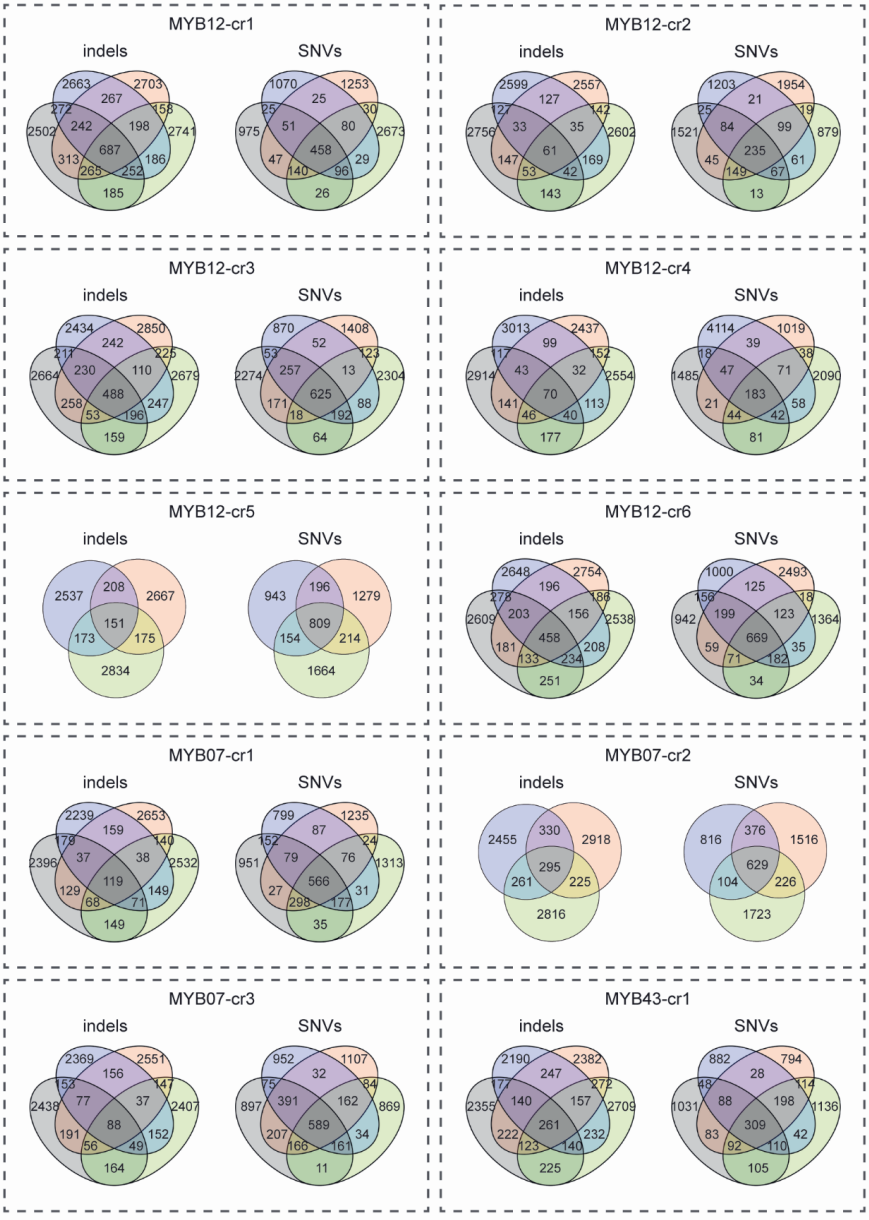
Supplementary Figures



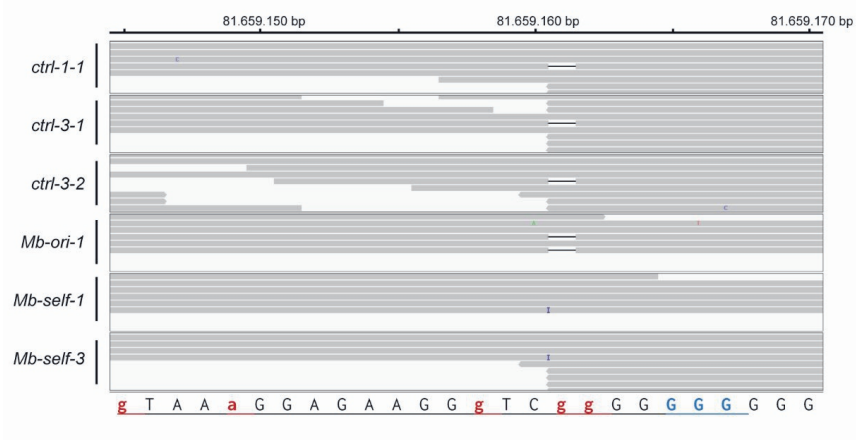
Supplementary Figure 1 - Off-target mutations identified by amplicon sequencing of T0 lines. (a) Off-targets identified for lines in which MYB12 was targeted. (b) Off-targets identified for lines in which MYB43 was targeted. (a) and (b): Freq = frequency of the allele, count = read count of the allele, F = change of the open reading frame, WT = wild type control.



Supplementary Figure 2 - Individual indels (a), SNVs (b) and SVs (c) as a fraction of the total amount of indels, SNVs and SV, respectively, identified for every CRISPR line. Individual variation is variation that is not present in one of the six wild type controls or one of the five transformation and tissue culture controls.



Supplementary Figure 3 - Overlap of indels and SNVs between CRISPR T_1 lines originating from the same T_0 parent.



Supplementary Figure 4 - Indels found in wildtype and transformation and tissue culture control plants at a potential off-target site. The potential off-target site is underlined. The five mismatches to the target site are indicated by red lowercase. The PAM is indicated in blue.

Supplementary Tables

Supplementary Table 1 - Read coverage of different samples.

Sample	Coverage	Sample	Coverage
MYB12-cr1-1	36.0146	MYB07-cr1-3	34.262
MYB12-cr1-2	32.5802	MYB07-cr1-4	30.9283
MYB12-cr1-3	37.8828	MYB07-cr2-1	41.2382
MYB12-cr1-4	27.3085	MYB07-cr2-2	32.2117
MYB12-cr2-1	32.2136	MYB07-cr2-3	31.5237
MYB12-cr2-2	31.2932	MYB07-cr3-1	35.8842
MYB12-cr2-3	31.0247	MYB07-cr3-2	31.9255
MYB12-cr2-4	36.8061	MYB07-cr3-3	36.8161
MYB12-cr3-1	40.9773	MYB07-cr3-4	41.9974
MYB12-cr3-2	30.9681	MYB43-cr1-1	37.2099
MYB12-cr3-3	29.7421	MYB43-cr1-2	43.6297
MYB12-cr3-4	28.5455	MYB43-cr1-3	34.7946
MYB12-cr4-1	24.4977	MYB43-cr1-4	36.3889
MYB12-cr4-2	37.6635	ctrl-1-1	36.3938
MYB12-cr4-3	29.9819	ctrl-2-1	34.8549
MYB12-cr4-4	27.7692	ctrl-3-1	37.3556
MYB12-cr5-1	39.4130	ctrl-3-2	37.0277
MYB12-cr5-2	36.3907	ctrl-3-3	30.1829
MYB12-cr5-3	28.6736	Mb-ori-1	40.0604
MYB12-cr6-1	34.8070	Mb-ori-2	36.5136
MYB12-cr6-2	25.3992	Mb-ori-3	35.3090
MYB12-cr6-3	37.8370	Mb-self-1	33.7534
MYB12-cr6-4	33.8756	Mb-self-2	21.6997
MYB07-cr1-1	40.3575	Mb-self-3	34.1519
MYB07-cr1-2	32.0392		

Average coverage: 34.1, standard deviation: 4.7

Supplementary Table 2 - Number of predicted off-target sites containing a SNV, indel or BND in close proximity. SNVs, indels and BNDs in a window of 40 bp around predicted off-target sites (from 20 bp before until 20 bp after) were identified at these sites.

	MYB12			*MYB07*			*MYB43*		
PAM / mism	NGG	NAG	NGA	NGG	NAG	NGA	NGG	NAG	NGA
0	**1/1**	-	-	**1/1**	-	0/1	**1/1**	-	-
1	**1/1**	-	-	-	-	-	**1/2**	-	-
2	0/1	-	-	-	-	0/3	-	0/2	-
3	0/1	0/2	-	0/7	0/3	0/13	0/2	0/10	0/5
4	0/16	0/23	0/29	0/34	0/40	**1/85**	0/7	0/33	0/9
5	**1/149**	**2/560**	0/370	**1/536**	0/528	0/758	0/99	0/207	0/165
6	**6/1261**	**22/5225**	**5/2808**	**7/2772**	**2/4949**	**4/4818**	**1/818**	**3/1710**	0/1741

Supplementary Table 3 - Potential identified off-target mutations in MYB12-targeted plants. Plant: plant in which the variation was identified, chrom: chromosome at which the variation was identified, pos: chromosome position at which the variation was identified, original: original allele, new: alternative allele identified by sequencing, off-target sequence: the sequence of the potential off-target. The first number indicates the amount of mismatches to the target site, the – or + indicates the strand direction. Sequence including PAM is shown. Mismatches to the target site are indicated by lowercase.

plant	chrom	pos	original	new	off-target sequence
MYB12-cr1-1	MBtmv-ch12	64076107	CAAG	C	1_-_CTAATGGtGAAGGCTCTTGGAGG
MYB12-cr1-3	MBtmv-ch12	64076107	CAAG	C	1_-_CTAATGGtGAAGGCTCTTGGAGG
MYB12-cr1-4	MBtmv-ch12	64076107	CAAG	C	1_-_CTAATGGtGAAGGCTCTTGGAGG
MYB12-cr2-1	MBtmv-ch12	64076107	C	CA	1_-_CTAATGGtGAAGGCTCTTGGAGG
MYB12-cr2-2	MBtmv-ch12	64076107	C	CA	1_-_CTAATGGtGAAGGCTCTTGGAGG
MYB12-cr2-3	MBtmv-ch12	64076107	C	CA	1_-_CTAATGGtGAAGGCTCTTGGAGG
MYB12-cr2-4	MBtmv-ch12	64076107	C	CA	1_-_CTAATGGtGAAGGCTCTTGGAGG
MYB12-cr1-1	MBtmv-ch07	12715857	C	G	5_-_gTAgTGGAGcAGcaTCTTGGGAG
MYB12-cr1-2	MBtmv-ch07	12715857	C	G	5_-_gTAgTGGAGcAGcaTCTTGGGAG
MYB12-cr1-2	MBtmv-ch01	81659160	CG	C	5_+_gTAAaGGAGAAGGgTCggGGGGG
MYB12-cr1-3	MBtmv-ch01	81659160	CG	C	5_+_gTAAaGGAGAAGGgTCggGGGGG
MYB12-cr1-3	MBtmv-ch07	12715857	C	G	5_-_gTAgTGGAGcAGcaTCTTGGGAG
MYB12-cr1-4	MBtmv-ch07	12715857	C	G	5_-_gTAgTGGAGcAGcaTCTTGGGAG
MYB12-cr6-2	MBtmv-ch09	42543746	A	C	5_-_CgAtTGaAGAAGcCaCTTGGGAG
MYB12-cr4-4	MBtmv-ch01	45162100	A	C	6_+_CggtTGaAGAAGcCaCTTGGGAG
MYB12-cr1-3	MBtmv-ch09	8210505	C	G	6_-_CTAATaGAaAcGaCTCTcGcTGA
MYB12-cr2-1	MBtmv-ch07	62052567	T	G	6_-_CTAATGGAGcAGcaaCTcGaGGG
MYB12-cr2-1	MBtmv-ch08	8307012	TGG	T	6_+_CaAATaGAGAAGaaagTTGGGGG
MYB12-cr5-3	MBtmv-ch08	8307012	TGGG	T,TG	6_+_CaAATaGAGAAGaaagTTGGGGG
MYB12-cr1-1	MBtmv-ch09	36618645	T	TC	6_-_CaAATttAccAGaCTCTTGGGGG
MYB12-cr1-4	MBtmv-ch02	305095	A	C	6_+_CcgtTGaAGAAGcCaCTTGGGAG
MYB12-cr3-3	MBtmv-ch02	892860	A	C	6_+_aTttTGGAGAAGGtTtTgGGCAG
MYB12-cr1-3	MBtmv-ch03	13061218	A	C	6_+_CggtTGaAGAAGcCaCTTGGGAG
MYB12-cr1-4	MBtmv-ch04	14758569	A	C	6_+_CggtTGaAGAAGcCaCTTGGGAG
MYB12-cr1-3	MBtmv-ch09	36618645	T	TC	6_-_CaAATttAccAGaCTCTTGGGGG
MYB12-cr1-4	MBtmv-ch09	36618645	T	TC	6_-_CaAATttAccAGaCTCTTGGGGG

MYB12-cr1-2	MBtmv-ch09	54520118	TGG	T	6_+_tTttTGGAGAAGaCaaTTGGGGG
MYB12-cr1-4	MBtmv-ch09	54520118	TGG	T,TG	6_+_tTttTGGAGAAGaCaaTTGGGGG
MYB12-cr4-1	MBtmv-ch04	41890658	T	G	6_+_CggtTGaAGAAGcCaCTTGGGAG
MYB12-cr2-2	MBtmv-ch09	56807135	T	C	6_-_CTgAgGGgGAAGGgggTTGGAGG
MYB12-cr2-2	MBtmv-ch09	56807142	T	C	6_-_CTgAgGGgGAAGGgggTTGGAGG
MYB12-cr1-1	MBtmv-ch10	8646961	C	A	6_+_CaAgTGGAGgAaaCTCTTGaTGG
MYB12-cr1-2	MBtmv-ch10	8646961	C	A	6_+_CaAgTGGAGgAaaCTCTTGaTGG
MYB12-cr1-3	MBtmv-ch10	8646961	C	A	6_+_CaAgTGGAGgAaaCTCTTGaTGG
MYB12-cr6-1	MBtmv-ch04	32234263	C	A	6_-_CTtATGGAGgAGGtaCTTctAAG
MYB12-cr2-4	MBtmv-ch11	7641173	G	A	6_-_aTgtTtGAaAAGtCTCTTGGAGA
MYB12-cr2-4	MBtmv-ch11	7641178	A	C	6_-_aTgtTtGAaAAGtCTCTTGGAGA
MYB12-cr2-4	MBtmv-ch11	7641181	T	G	6_-_aTgtTtGAaAAGtCTCTTGGAGA
MYB12-cr2-4	MBtmv-ch11	7641185	T	G	6_-_aTgtTtGAaAAGtCTCTTGGAGA
MYB12-cr2-4	MBtmv-ch11	7641200	A	C	6_-_aTgtTtGAaAAGtCTCTTGGAGA
MYB12-cr2-4	MBtmv-ch11	7641203	G	T	6_-_aTgtTtGAaAAGtCTCTTGGAGA
MYB12-cr6-3	MBtmv-ch04	32234263	C	A	6_-_CTtATGGAGgAGGtaCTTctAAG
MYB12-cr4-1	MBtmv-ch05	22633658	A	C	6_+_CggtTGaAGAAGcCaCTTGGGAG
MYB12-cr3-1	MBtmv-ch02	29648578	G	T	6_-_CaAATGGgGAAGGaTtgTtGAGA
MYB12-cr2-2	MBtmv-ch06	41283587	C	CA	6_+_CTAAaGGAaAtGGCTCTcaaGAG
MYB12-cr2-4	MBtmv-ch06	29411462	C	T	6_+_CcgtTGaAGAAGcCaCTTGGGAG
MYB12-cr3-2	MBtmv-ch02	29648578	G	T	6_-_CaAATGGgGAAGGaTtgTtGAGA
MYB12-cr3-1	MBtmv-ch07	48420817	C	A	6_-_CcgtTGaAGAAGcCaCTTGGGAG
MYB12-cr3-2	MBtmv-ch07	48420817	C	A	6_-_CcgtTGaAGAAGcCaCTTGGGAG
MYB12-cr3-3	MBtmv-ch07	48420817	C	A	6_-_CcgtTGaAGAAGcCaCTTGGGAG
MYB12-cr3-4	MBtmv-ch07	48420817	C	A	6_-_CcgtTGaAGAAGcCaCTTGGGAG
MYB12-cr3-3	MBtmv-ch02	29648578	G	T	6_-_CaAATGGgGAAGGaTtgTtGAGA
MYB12-cr4-1	MBtmv-ch07	42412203	A	C	6_+_CggtTGaAGAAGcCaCTTGGGAG
MYB12-cr1-4	MBtmv-ch08	25356533	A	C	6_+_gggAaGGAGgAGGCTgTTGGGAG
MYB12-cr3-4	MBtmv-ch02	29648578	G	T	6_-_CaAATGGgGAAGGaTtgTtGAGA
MYB12-cr3-1	MBtmv-ch08	13624959	GT	G	6_-_CTAgTtGAGAAaaaaCTTGGTAG
MYB12-cr3-2	MBtmv-ch08	13624959	GT	G	6_-_CTAgTtGAGAAaaaaCTTGGTAG
MYB12-cr3-3	MBtmv-ch08	13624959	GT	G	6_-_CTAgTtGAGAAaaaaCTTGGTAG
MYB12-cr3-4	MBtmv-ch08	13624959	GT	G	6_-_CTAgTtGAGAAaaaaCTTGGTAG
MYB12-cr4-3	MBtmv-ch08	64170858	GT	G	6_+_tTAATtGAGcAGttTCTTtGGAG
MYB12-cr6-1	MBtmv-ch08	25530402	C	A	6_+_CTAcTatAGAAGatTtTTGGAAG
MYB12-cr6-2	MBtmv-ch08	25530402	C	A	6_+_CTAcTatAGAAGatTtTTGGAAG
MYB12-cr6-3	MBtmv-ch08	25530402	C	A	6_+_CTAcTatAGAAGatTtTTGGAAG
MYB12-cr4-4	MBtmv-ch01	90984340	A	C	6_-_CaAATtGgGAAGGtcCcTGGTGA
MYB12-cr1-4	MBtmv-ch10	8646961	C	A	6_+_CaAgTGGAGgAaaCTCTTGaTGG
MYB12-cr6-4	MBtmv-ch08	25530402	C	A	6_+_CTAcTatAGAAGatTtTTGGAAG
MYB12-cr1-4	MBtmv-ch09	14541598	T	G	6_+_aTtATGagGAAtGCTCTTGaAAG
MYB12-cr3-3	MBtmv-ch09	55450674	A	C	6_+_CggtTGaAGAAGcCaCTTGGGAG
MYB12-cr6-1	MBtmv-ch07	35121831	G	T	6_-_CTAATGaAGAAGGagCgTctGGA
MYB12-cr4-1	MBtmv-ch10	26526465	A	T	6_+_CcgtTGaAGAAGcCaCTTGGGAG
MYB12-cr4-2	MBtmv-ch10	26526465	A	T	6_+_CcgtTGaAGAAGcCaCTTGGGAG
MYB12-cr6-2	MBtmv-ch07	35121831	G	T	6_-_CTAATGaAGAAGGagCgTctGGA
MYB12-cr6-2	MBtmv-ch10	27013197	A	C	6_+_CggtTGaAGAAGcCaCTTGGGAG
MYB12-cr6-1	MBtmv-ch11	37229460	C	T	6_-_aTtATGagGAAtGCTCTTGaAAG

MYB12-cr6-3	MBtmv-ch11	37229460	C	T	6_-_aTtATGagGAAtGCTCTTGaAAG
MYB12-cr6-3	MBtmv-ch07	35121831	G	T	6_-_CTAATGaAGAAGGagCgTctGGA
MYB12-cr6-4	MBtmv-ch11	37229460	C	T	6_-_aTtATGagGAAtGCTCTTGaAAG
MYB12-cr4-1	MBtmv-ch12	49638567	A	C	6_+_CggtTGaAGAAGcCaCTTGGGAG

Supplementary Table 4 - Potential identified off-target mutations in MYB07-targeted plants. Plant: plant for which the variation was identified, chrom: chromosome at which the variation was identified, pos: chromosome position at which the variation was identified, original: original allele, new: alternative allele identified by sequencing, off-target sequence: the sequence of the potential off-target. The first number indicates the amount of mismatches to the target site, the – or + indicates the strand direction. Sequence including PAM is shown. Mismatches to the target site are indicated by lowercase.

sample	chrom	pos	Original	new	off_target_sequence
MYB07-cr2-2	MBtmv-ch12	18814582	G	GC	4_-_tGTAtTTAATCCAagGGAGATGA
MYB07-cr2-2	MBtmv-ch12	18814586	TC	T	4_-_tGTAtTTAATCCAagGGAGATGA
MYB07-cr2-2	MBtmv-ch12	18814598	GATTA	G	4_-_tGTAtTTAATCCAagGGAGATGA
MYB07-cr2-2	MBtmv-ch12	18814605	TACACTTTGACA	T	4_-_tGTAtTTAATCCAagGGAGATGA
MYB07-cr2-1	MBtmv-ch05	35769058	C	T	5_+_AGTtATcAATCCgggGGAGAGGG
MYB07-cr2-2	MBtmv-ch05	35769058	C	T	5_+_AGTtATcAATCCgggGGAGAGGG
MYB07-cr2-4	MBtmv-ch05	35769058	C	T	5_+_AGTtATcAATCCgggGGAGAGGG
MYB07-cr3-4	MBtmv-ch03	43215781	G	T	6_+_tGTAATTAcTCCtTCaGtGgTAG
MYB07-cr2-2	MBtmv-ch03	62387413	G	A	6_-_AGTAtggAATaCATtGaAGAGGA
MYB07-cr2-1	MBtmv-ch04	23364417	G	T	6_-_AGTAtcTtAcCCcTtGGAGATGA
MYB07-cr2-2	MBtmv-ch04	23364417	G	T	6_-_AGTAtcTtAcCCcTtGGAGATGA
MYB07-cr1-2	MBtmv-ch04	65010373	C	T	6_+_AGTAATcAcTtCtTCcGgGAAAG
MYB07-cr1-3	MBtmv-ch04	65010373	C	T	6_+_AGTAATcAcTtCtTCcGgGAAAG
MYB07-cr1-4	MBtmv-ch04	65010373	C	T	6_+_AGTAATcAcTtCtTCcGgGAAAG
MYB07-cr2-3	MBtmv-ch06	44351840	TTGGGAGG	T	6_+_AGTAgTTgAgCCtTgGGAGgGGG
MYB07-cr1-3	MBtmv-ch06	44351841	TGGGAGG	T	6_+_AGTAgTTgAgCCtTgGGAGgGGG
MYB07-cr1-4	MBtmv-ch07	36791943	TA	T	6_-_AaTAATTAtTCCAatatAGATGG
MYB07-cr3-1	MBtmv-ch07	41338468	G	A	6_+_AtcAAgaAATtCtTCGGAGAAGG
MYB07-cr3-2	MBtmv-ch07	41338468	G	A	6_+_AtcAAgaAATtCtTCGGAGAAGG
MYB07-cr3-3	MBtmv-ch07	41338468	G	A	6_+_AtcAAgaAATtCtTCGGAGAAGG
MYB07-cr3-4	MBtmv-ch07	41338468	G	A	6_+_AtcAAgaAATtCtTCGGAGAAGG
MYB07-cr1-4	MBtmv-ch08	1168305	TTAA	T	6_-_taTtATTAATCCATCaaAcATGA
MYB07-cr2-2	MBtmv-ch08	15847866	AC	A	6_-_AGTgAaTAATCaAaCGGgGgGGG
MYB07-cr1-3	MBtmv-ch09	27102458	TG	T	6_+_AGgcATaAATCCATtGGgGgGGG
MYB07-cr1-1	MBtmv-ch09	55981319	ACCCCCC	A	6_-_AGTAATTAAaCgggCGGgGgGGG
MYB07-cr1-2	MBtmv-ch09	55981319	ACCCCCC	A	6_-_AGTAATTAAaCgggCGGgGgGGG
MYB07-cr1-3	MBtmv-ch09	55981319	ACCCCCC	A	6_-_AGTAATTAAaCgggCGGgGgGGG
MYB07-cr3-2	MBtmv-ch09	55981319	ACCCCCC	A	6_-_AGTAATTAAaCgggCGGgGgGGG
MYB07-cr3-3	MBtmv-ch09	55981319	ACCCCCC	A	6_-_AGTAATTAAaCgggCGGgGgGGG
MYB07-cr3-4	MBtmv-ch09	55981319	ACCCCCCC	A,AC	6_-_AGTAATTAAaCgggCGGgGgGGG
MYB07-cr2-1	MBtmv-ch10	5850940	CA	C	6_-_AGTAATTgATaCcTaaGAcATGA
MYB07-cr1-1	MBtmv-ch11	7765437	A	T	6_+_AtTAATTAATCgATaGtttATGG
MYB07-cr1-2	MBtmv-ch11	7765437	A	T	6_+_AtTAATTAATCgATaGtttATGG
MYB07-cr1-3	MBtmv-ch11	7765437	A	T	6_+_AtTAATTAATCgATaGtttATGG
MYB07-cr1-4	MBtmv-ch11	7765437	A	T	6_+_AtTAATTAATCgATaGtttATGG

Supplementary Table 5 - Potential identified off-target mutations in MYB43-targeted plants. Plant: plant for which the variation was identified, chrom: chromosome at which the variation was identified, pos: chromosome position at which the variation was identified, original: original allele, new: alternative allele identified by sequencing, off-target sequence: the sequence of the potential off-target. The first number indicates the amount of mismatches to the target site, the – or + indicates the strand direction. Sequence including PAM is shown. Mismatches to the target site are indicated by lowercase.

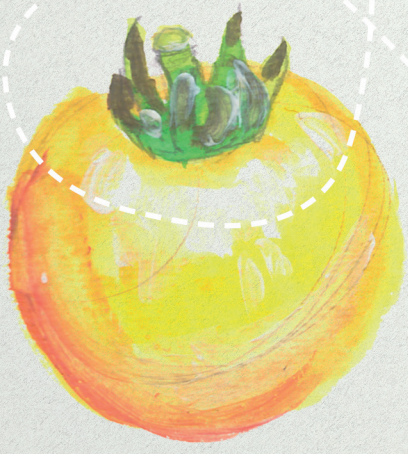
sample	chrom	pos	original	new	off-target sequence
MYB43-cr1-2	MBtmv-ch08	65078349	GGTCCATGGACA	G	1_+_aAAAGGTCCATGGACAGCTGAGG
MYB43-cr1-3	MBtmv-ch08	65078349	GGTCCATGGACA	G	1_+_aAAAGGTCCATGGACAGCTGAGG
MYB43-cr1-4	MBtmv-ch08	65078349	GGTCCATGGACA	G	1_+_aAAAGGTCCATGGACAGCTGAGG
MYB43-cr1-1	MBtmv-ch08	65078353	CATGGACAG	C	1_+_aAAAGGTCCATGGACAGCTGAGG
MYB43-cr1-3	MBtmv-ch08	65078353	CATGGACAG	*,C	1_+_aAAAGGTCCATGGACAGCTGAGG
MYB43-cr1-4	MBtmv-ch08	65078353	CATGGACAG	*,C	1_+_aAAAGGTCCATGGACAGCTGAGG
MYB43-cr1-2	MBtmv-ch04	9818217	G	A	6_-_GAAAGGTaCcTGGACtctTcAAG
MYB43-cr1-4	MBtmv-ch04	9818217	G	A	6_-_GAAAGGTaCcTGGACtctTcAAG
MYB43-cr1-1	MBtmv-ch05	66840992	A	T	6_-_GAAAGGTCaAacGAgccCTGAAG
MYB43-cr1-4	MBtmv-ch09	5291626	C	A	6_-_GgggGGTttATGGACAGgTGCAG
MYB43-cr1-2	MBtmv-ch10	64180108	TGG	T,TG	6_+_GgggGGTagtTGGACAGCTGAGG

Supplementary Table 6 - Primers used for genotyping. Sanger: genotyping using Sanger sequencing, amplicon: genotyping using amplicon sequencing. For amplicon sequencing, and additional 8-bp barcode was attached to the 5' end of both the forward and reverse primer to allow pooling of different samples into a single sequencing library.

site	purpose	F/R	sequence (5'-3')
SlMYB12-target	Sanger	F	GCGGGGTCCATTTATTTGTGTTGT
SlMYB12-target	Sanger	R	GCTCGAATCCATTACACTATGTTATC
SlMYB12-offtarget	amplicon	F	TTTTGAATGAGATGATTATGCGTTA
SlMYB12-offtarget	amplicon	R	TGGTAAGAACACCTTGTTGTGAG
SlMYB07-target	Sanger	F	CAAGTGCCGGCTTTAATTAATTTC
SlMYB07-target	Sanger	R	GTCCAAAACACCATTACTCTTTAATCC
SlMYB07-offtarget	amplicon	F	CATACTTGTTTCCAAGGAGGCTA
SlMYB07-offtarget	amplicon	R	TTAGAGCTCATGGTGAAGGTTGT
SlMYB43-target	Sanger	F	GTGGTACACCACTAGGGTTAGTGG
SlMYB43-target	Sanger	R	GACGCAACAGGTTCTGTCATAACC
SlMYB43-offtarget1	amplicon	F	TAATAGGGATGGGAAGACAACCT
SlMYB43-offtarget1	amplicon	R	ACTCTTACCGCAACGCCTAAG
SlMYB43-offtarget2	amplicon	F	AGAGAAGAAACCTGCAAGATGTG
SlMYB43-offtarget2	amplicon	R	GTGTATGTGTGTGTAAGGGATGG
TCcontrol_T-DNA	T-DNA	F	ATGGGAAGAGGAAGAGTTGAG
TCcontrol_T-DNA	T-DNA	R	CAGGATTGTACCCAAAATGTGGAG

Supplementary Datasets

A Supplementary Dataset containing IGV snapshots of potential off-target sites is available from the author upon request. Access to the raw sequencing data is available from the author upon request.



Chapter 6

General Discussion

Genome editing

The ability to modify organisms' genomes as desired has been a long-standing goal in molecular biology research. Over the years, several tools working towards this goal have become available, such as zinc-finger nucleases and transcription-activator-like effector nucleases (Urnov et al. 2010; Miller et al. 2011). However, none of these systems have the ease of use and programmability that the CRISPR-Cas (an abbreviation for Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated protein) system boasts (Jinek et al. 2012; Zhu et al. 2020; Jaganathan et al. 2018).

Most applications of the systems above for genome editing function by inducing double-strand breaks (DSBs) at specific, user-defined sites in the target genome. DSBs threaten the cell's survival; therefore, all organisms – from bacteria to fungi to plants to animals – have developed pathways to repair these breaks. These repair pathways are not always perfect. In plants, the most frequently employed pathway for DSB repair is non-homologous end joining (NHEJ) (Puchta 2005). This pathway can result in small insertions and deletions (indels) that may disrupt the function or regulation of genes. Because of this, the CRISPR-Cas system is frequently used to knock out genes to study the effect of these genes on the plant's phenotype. For precise genome editing, the homology-directed repair (HDR) pathway offers great opportunities: by providing the plant with a repair template, specific changes or insertions of genetic material can be accomplished. Unfortunately, HDR occurs at low frequencies in plants' somatic cells, preventing its routine application (Puchta 2005; 2002; Steinert et al. 2016).

As DSBs are mutagenic, genome editing systems should induce their DSBs as precisely as possible. DSBs at sites other than the intended target site can result in undesirable mutations. As CRISPR-Cas has great potential for therapeutical applications in humans, the specificity of CRISPR-Cas mediated genome editing has been studied in-depth in mammalian cells (Doudna 2020; D. Kim et al. 2019). In plants, research on specificity, especially using genome-wide techniques, has been more limited (Sturme et al. 2022).

In this thesis, I have studied the specificity of CRISPR-Cas mediated genome editing in tomato in depth. To do this, I have applied both targeted and untargeted approaches. In this chapter, I will discuss my main findings, their implications, and avenues for further research.

The value of high-throughput techniques in genome editing research

One of the biggest limitations to studying plant genome editing is the comparative lack of high-throughput systems. In the mammalian research field, the use of cell lines is highly standardized. A wide range of available cell types can be transfected, selected, and cultured. Although these cell lines have their limitations – for example, they are frequently derived from cancerous cells that have impaired DNA repair mechanisms – they are great tools for high-throughput studies of genome editing. Cell pools can be transfected with CRISPR-Cas machinery, selected for successful transfection or even editing events, and subsequently cultured (Shalem et al. 2014; T. Wang et al. 2014). In every cell, different mutations may occur. An overview of all possible mutational outcomes, including low-frequency on- and off-target mutations, can be obtained by high-throughput sequencing of edited cell pools. These cell-based approaches are the main pillars upon which our understanding of the specificity of CRISPR-Cas mediated genome editing was built. Popular algorithms that predict gRNA efficiency and specificity, such as the Azimuth score and the Moreno-Mateos score, have all been developed using this type of high-throughput data (Doench et al. 2014; Moreno-Mateos et al. 2015).

In plants, mutant lines are frequently obtained through *Agrobacterium*-mediated transformation. For tomato and many other plant species, this procedure includes tissue culture, which is a laborious and time-consuming process. In this research, I aimed to investigate many nuclease-gRNA complexes. To identify low-frequency on- and off-target mutations, large numbers of mutant lines are needed for every gRNA. Acquiring these through stable transformation is challenging, bordering on practically impossible.

An approach more similar to mammalian cell lines, and thus more amenable to high-throughput research is by applying protoplasts (Reed and Bargmann 2021; Lin et al. 2018). Protoplasts are plant cells that are stripped of their cell walls. The components necessary for genome editing can subsequently be delivered to the cell encoded on a plasmid or as a preassembled ribonucleoprotein (RNP). In Chapter 2, I set up a system to transfect tomato protoplasts in a 96-well format. By coupling this to next-generation amplicon sequencing, I could investigate on- and off-target mutations induced by CRISPR-Cas9 (Chapters 2 and 3) and CRISPR-Cas12a (Chapter 3) in-depth. Additionally, protoplasts are suitable for optimizing genome editing systems due to the relatively fast turnaround time of protoplast experiments and the sensitivity of next-generation sequencing. I have performed this, on a small scale, for Cas12a in Chapter 3. These systems could also be used to test and optimize new genome editing applications, such as prime editing, before continuing to more time-

consuming stable transformation. Recently, a study was published in which a combination of protoplasts and high-throughput imaging and cell sorting techniques was used to further optimize Cas12a-mediated genome editing. Additionally, they reported strongly improved Cas12a-mediated base editing in wheat and maize (Gaillochet et al. 2023). This observation emphasizes the strength and promise of high-throughput approaches. Furthermore, the availability of more high-throughput data on on- and off-target mutations in plants could be used to improve gRNA prediction algorithms for use in plants, increasing their still limited predictive value (Chapter 2).

Major downsides to the use of protoplast systems, as compared to mammalian cell lines, are that protoplasts need to be isolated freshly each time and cannot be cultured. Neither can they easily be regrown into complete plants (Reed and Bargmann 2021). The use of cell suspension cultures may overcome these hurdles (Moscatiello et al. 2013). In industrial settings, these cultures are used to produce secondary metabolites and other high-value substances (Kieran et al. 1997; Yue et al. 2016; Seki et al. 1997). When established, cell cultures can provide a large amount of homogeneous material, allowing research on a wide variety of topics. Cultured cells can be protoplasted, infected by *Agrobacterium*, and/or regenerated into complete plants. It should however be kept in mind that whereas protoplasts are primary cells still (partly) with the same identity as in the plant, suspension cultures are maintained and passaged possibly *ad infinitum*. Over time, this may affect different pathways, including pathways for DNA repair, changing DNA repair outcomes as compared to whole plants.

Other high-throughput alternatives to protoplasts are transient *Agrobacterium*-mediated assays such as agroinfiltration or calli resulting from *Agrobacterium*-mediated transformation or particle bombardment. However, these approaches are more limited than protoplast assays regarding cargo that can be delivered (*Agrobacterium*-based approaches do not allow for the use of RNPs) and speed.

To level the playing field for genome editing in mammalian and plant cells, the continuing development of methods to perform high-throughput research in plants is important (Gaillochet et al. 2021; Pan et al. 2023).

On-target genome editing

Predicting sgRNA or crRNA efficiency

Using the protoplast-amplicon sequencing approach, I characterized on-target mutations induced by 89 sgRNAs (Chapter 2) and 35 guides overlapping for Cas9 and Cas12a (Chapter 3). Critical for a successful genome editing experiment is selecting an efficient sgRNA or crRNA. For plants – where creating a mutant line is frequently a time-consuming, tissue-culture endeavour – it is desirable to minimize the risk of selecting an inefficient target site-gRNA combination as much as possible. Several algorithms have been developed that can predict sgRNA or crRNA efficiencies (Doench et al. 2014; Fusi et al. 2015; Moreno-Mateos et al. 2015; H. K. Kim et al. 2018; Luo et al. 2019). However, these algorithms have all been trained using data acquired from experiments with mammalian cells. Although this software is heavily used in the plant field (Lei et al. 2014; Liu et al. 2017), it is largely unknown how actual sgRNA efficiency in plants correlates to predicted efficiency (Naim et al. 2020). In Chapter 2, I correlated actual efficiencies to those predicted by two algorithms. I found that the Azimuth score (Doench et al. 2016) – as opposed to the Moreno-Mateos score (Moreno-Mateos et al. 2015)– indeed had some predictive value. In Chapter 3, I performed a similar correlation between the DeepCpf1 score (Luo et al. 2019) and Cas12a efficiency. Although top and bottom performers could be somewhat predicted, predictability was generally even lower. These low correlations could indicate that key determinants affecting sgRNA efficacy, such as the epigenetic landscape at target sites (Konstantakos et al. 2022), are not readily transferable from mammalian to plant cells.

Like in mammalian cells, selecting a high-scoring sgRNA is no guarantee for efficient mutagenesis, and a low-scoring sgRNA will not necessarily fail to induce any mutations. In practice, selecting multiple sgRNAs for a knockout mutagenesis experiment is advisable, if possible. Another potential advantage of using multiple sgRNAs for creating a knockout mutant is that guide-to-guide deletions may remove the complete coding sequence, eliminating the risk of creating dominant-negative alleles or partially active proteins (R. Wang et al. 2020). If selecting multiple guides is not possible, testing guide activity prior to transformation by protoplast or agroinfiltration assays (for example, as outlined in Chapter 4) is advisable. Finally, the high-throughput approaches used in this thesis could be applied on an even larger scale to gain more data on Cas9 and Cas12a efficiency. Data obtained in that way can be used to train prediction algorithms specifically suited to genome editing in tomato.

Mutation patterns and their implications for nuclease selection and DNA repair pathway choice

In our studies, I obtained thousands of mutant alleles distributed over all tested sgRNAs and crRNAs. This allowed us to determine mutation patterns for both Cas9 and Cas12a (Chapters 2 and 3). These patterns show interesting differences. Perhaps the most notable difference is that whereas Cas9 frequently induces insertions and deletions, Cas12a induces almost exclusively deletions. These deletions are generally larger than Cas9-induced deletions. The large majority of Cas9-induced insertions are one bp insertions. The nature of these insertions is, to some extent, predictable (Lemos et al. 2018; van Overbeek et al. 2016). They seem to result from the duplication of the 4th nucleotide before the PAM, likely due to staggered cuts with a 1 bp 5' overhang Cas9 sometimes introduces (Zuo and Liu 2016; Shou et al. 2018). Interestingly, Cas12a also induces 4 or 5 bp 5' overhang (Zetsche et al. 2015), but this does not result in a high frequency of 4 or 5 bp insertions (Chapter 3).

The differences in the mutation pattern of these two enzymes have multiple potential implications. As a result of the bias towards the characteristic one bp insertion, Cas9-induced DSBs have a higher chance of resulting in a frameshift mutation than Cas12a-induced DSBs (Chapter 3). Therefore, the use of Cas9 over Cas12a might be preferred when a knockout mutation is desired. Additionally, the predictability of + 1 bp insertions induced by Cas9 could be harnessed to induce specific mutations or repair alleles. The deletions induced by Cas12a, in combination with the A/T-rich PAM, could make Cas12a preferable over Cas9 for editing cis-regulatory elements in the promoter of a gene.

Another implication might be that different DSBP repair pathways are employed to repair Cas9-induced breaks than to repair Cas12a-induced breaks. It is important to remember that it is not the nucleases that realize mutations – the nucleases only induce DSBs. Mutations arise due to the subsequent non-perfect DNA repair pathways employed by the host cell. A difference in mutation pattern may therefore indicate a difference in DNA repair pathway choice.

The first major decision made in DSB repair is whether to repair the break using canonical non-homologous end joining (cNHEJ) or to proceed to end resectioning, which can result in a multitude of different outcomes: alternative end joining (a-EJ, also known as microhomology-mediated end joining), single strand annealing (SSA), or homologous recombination (HR) (Puchta 2005; Vu et al. 2021). In our high-throughput experiments, I could not detect HR in the absence of an exogenous template, because of the highly homozygous nature of the used cv. Moneyberg background.

Additionally, large deletions resulting from SSA will result in the removal of primer binding sites and are not detected. However, we can detect the smaller, a-EJ-induced deletions – and it is precisely these deletions that may explain the difference between the observed mutation patterns of Cas9 and Cas12a-induced DSBs (Bernabé-Orts et al. 2019). This difference is most likely a result of the staggered DSBs induced by Cas12a.

The most straightforward explanation is that staggered overhangs preferentially proceed to end-resectioning and subsequent a-EJ, HR, or SSA, whereas blunt ends preferentially proceed to NHEJ. This difference could also explain why we often find that dsODN integrations at target sites for Cas12a are often accompanied by deletions, which is not the case for Cas9 (Chapter 4). We should remember that most of the induced DSBs are likely repaired perfectly without the formation of mutations (Bétermier et al. 2014; Ben-Tov et al. 2023). These sequences will be cut repeatedly until an error is made, a mutation is introduced, and the target site is lost, which is the visible end-point. Therefore, an alternative explanation is that simple re-ligation of the free DNA ends is often perfect for nucleases that induce staggered breaks and that most visible outcomes result from the intrinsically error-prone a-EJ or SSA. This shift in thinking also has implications for target efficiency: differences in the perceived efficiency of sgRNAs might not only result from the efficacy of the sgRNA-nuclease complex but also from the tendency of a specific target site to be repaired by a more perfect or more error-prone process. Both scenarios could also explain why CRISPR-Cas12a is reported to be more successful in inducing HDR and similar approaches, for which end-resectioning is a necessary first step (Wolter and Puchta 2019; Van Vu et al. 2020).

In conclusion, site-specific nucleases are valuable tools for genome editing. They can also open avenues to study the fundamental processes of DSB repair and pathway choice, which have shaped evolution and will continue to do so. More insight into and higher predictability of DSB repair can, in turn, inform genome editing experiments.

Off-target activity in genome editing

Biased off-target detection

Off-target detection methods can broadly be divided into biased and unbiased categories. In biased or targeted approaches, off-target sites are predicted based on homology to the target site and then screened for mutations. In unbiased approaches, the whole genome is screened for potential off-target activity without being limited by predictions. Using a targeted approach, I first investigated the specificity of CRISPR-Cas-mediated genome editing in tomato cells. In Chapters 2 and 3, I

selected predicted off-target sites and investigated the presence of mutations at those sites using highly sensitive amplicon sequencing. In these two chapters combined, I screened 324 potential off-target sites for Cas9, and 57 for Cas12a. As I aimed to better understand when off-target mutations do happen and when they do not, I selected predicted off-target sites with varying numbers of mismatches to the target site and included some that I expected to be mutagenized. This purposeful inclusion of genuine off-target sites contrasts with many studies that report on CRISPR-Cas specificity in plants (Nekrasov et al. 2017; N. Zhang et al. 2020; X. Wang et al. 2021). Generally, studying specificity and off-target mutations of genome editing comes second to studying the phenotype of the genome-edited plant, and usually, sgRNAs are chosen that are already predicted to be very specific. Although practical and pragmatic, the latter approach does not serve to establish the limits of specificity.

Mutations could only be detected at a few predicted off-target sites - 24 for Cas9, and 10 for Cas12a. The majority of off-target mutations (19 out of 24 for Cas9, and 7 out of 10 for Cas12a) occurred at sites that only had 1 mismatch compared to the target site. In our large Cas9 screen (Chapter 2), I could not identify any off-target mutations at sites with three or more mismatches to the target site, although 138 out of 224 screened sites had three or four mismatches. I could identify Cas9-induced mutations at sites with three mismatches in Chapter 3. However, these occurred at very low frequencies. For Cas12a, I could not identify any off-target mutations at loci that had 3 or more mismatches to the target site, although 32 out of 57 tested sites had three mismatches.

Seed sequence: more applicable to Cas12a than Cas9

The position of the mismatch in the spacer seems to have a more significant impact on the likelihood of an off-target mutation for Cas12a than for Cas9. For Cas9, a single mismatch at the position most distal to the PAM makes a high-risk off-target site, with the potential to be cleaved at the same frequency as the on-target (Chapter 2, Chapter 3). However, the relationships between off-target activity risk and mismatch positions are less straightforward for the other positions. I found off-target mutations at varying frequencies for single mismatches scattered over the complete length of the spacer, including positions close to the PAM – often designated the “seed sequence” (Chapter 2, Chapter 3, Chapter 4). Earlier work has stated that mismatches in this seed sequence will abolish Cas9 cleavage activity (Jinek et al. 2012; Semenova et al. 2011). Although mutation frequencies are generally low at such positions, I and others have shown that off-target activity is still a risk (Pattanayak et al. 2013; Hsu et al. 2013).

For Cas12a, the positional range of mismatches allowing near on-target activity seems wider than for Cas9. The presence of mismatches in the last four nucleotides of the spacer can still result in near-on-target mutation efficiency (Chapter 3). This agrees with studies that show that spacers with a length of 19 nucleotides retain almost complete activity (H. Kim et al. 2017; Zetsche et al. 2015). I used spacers with the standard length of 23 nucleotides. Off-target mutations were not found at sites that had mismatches within the first fourteen nucleotides following the PAM. Therefore, a "seed sequence" that does not tolerate mismatches may hold more true for Cas12a than for Cas9. However, we must remember that in this thesis, I tested significantly more off-target sites for Cas9 than for Cas12a.

Off-target mutations

I calculated relative mutation frequencies for identified off-target sites to gain more insight into the chance that an on-target mutation and off-target mutation occur together in a single genome. Based on this, I draw the important conclusion that in most cases – even if the guide RNA used tends to induce off-target mutations – it will be possible to obtain plants that contain only on-target mutations. Potential exceptions are Cas9 sgRNAs with off-target sites with a single mismatch at high-risk positions and Cas12a crRNAs with off-target sites with up to two mismatches at high-risk positions – the positions furthest from the PAM. This is also reflected by the results obtained from the stable transformation of three promiscuous sgRNAs in tomato (Chapter 5). For sgRNA *MYB06*, targeting *Solyc01g079620* (*MYB12*), heterozygous off-target mutations at a known off-target site were identified at four out of six lines, indicating that off-target mutations can be avoided by obtaining enough transformed lines or by subsequent outcrossing of the off-target mutation. The relative frequency of the off-target mutation in protoplasts was approximately 5% for this sgRNA. The plant line obtained for sgRNA *MYB54* targeting *Solyc11g011050* contained biallelic mutations at one off-target site. The relative mutation frequency at this position, 54-2 in Chapter 2, was approximately 60% in protoplast pools. Notably, the other known off-target site of this sgRNA, 54-3 in Chapter 2, which had a much lower relative frequency of only around 1%, was not found to be mutagenized in the obtained line. However, better comparing mutation frequencies in protoplast pools with mutation frequencies obtained in stable transformation, requires the regeneration of many more mutant lines.

The ability of CRISPR-Cas systems to induce mutations at sites that are highly homologous to the target site is not always undesirable. Plant genomes are often highly duplicated and contain

multiple paralogs of a gene. The promiscuity of some sgRNAs may allow for knocking out multiple genes with a single sgRNA (Jouanin et al. 2018).

Taken together, Chapters 2 and 3 tell us that genome editing by CRISPR-Cas9 and CRISPR-Cas12a is generally specific. The large majority of predicted off-target sites are not mutagenized. “High-risk” sgRNAs can be avoided by using off-target prediction software, such as Cas-OFFinder, CRISPOR, and CRISPR-P – although these depend on reliable genome assemblies, which are not yet available for every crop – and certainly not for every cultivar. I and others have shown differences in predicted off-target site sequences between widely-used reference genomes and used laboratory cultivars (Chapter 5, X. Wang et al., 2021). Even if particular gRNAs can induce off-target mutations, those are generally induced at much lower frequencies, allowing for the selection of mutant plant lines that do contain on-target mutations but no off-target mutations.

Whole Genome Sequencing: strengths and weaknesses

A limitation of the studies in Chapters 2 and 3 is that they were biased towards predictable sites with some homology to the on-target site. Off-target mutations at unpredicted, and therefore, untested positions in the genome will not be identified. Therefore, I additionally applied unbiased approaches to off-target screening in Chapters 4 and 5.

Although protoplasts offer significant advantages, such as their applicability to high-throughput research and the possibility to easily deliver cargo to the cell, mutant tomato lines are generally obtained through *Agrobacterium*–mediated stable transformation (Brooks et al. 2014). Hence, in Chapter 5, I performed whole genome resequencing on 38 mutant T_1 tomato lines, five T_1 transformation and tissue culture control lines, and six seed-grown wildtype lines from two generations. This unbiased method yields an in-depth overview of all mutations – both on-target, off-target, and germ-line or somatic – that have arisen in one individual but is less suitable to map all potential off-target mutations of a single sgRNA. Although prices have decreased rapidly, WGS remains costly for hundreds of lines, so that low-frequency off-target mutations may be missed.

Off-target mutation detection using WGS presents several difficulties. First, reliable mapping, particularly for indels (the most common CRISPR-induced mutation) and SVs, is difficult to achieve (Nakamura et al. 2011; Minoche et al. 2011). This point is illustrated by whole genome sequencing results presented for rice (Tang et al. 2018). Three different software packages (Pindel, MuTect, and VarScan2) were used to call indels in resequenced lines. Each software individually identified hundreds to tens of thousands of indels relative to the reference genome. However, only a handful

of these would overlap for all three, indicating that false positive and false negative calls occur. Secondly, except for target sites, highly homologous off-target sites, or sites previously identified using a method such as GUIDE-seq, interpretation of found variation is difficult as it cannot be determined with certainty whether the variation has arisen naturally or through CRISPR-Cas9. Finally, when standing variation within a cultivar is not entirely captured by sequenced negative controls, remaining natural variation could result in the false calling of off-target mutations. These difficulties made it impossible for us to perform true unbiased off-target detection (Chapter 5).

Nevertheless, I was able to conclude that CRISPR-Cas9 mediated genome editing did not result in the identification of significantly more indels or SVs than in either transformation and tissue culture control plants or seed-grown wild type plants. As Cas9-induced mutations are mostly indels following a particular size distribution (Chapter 2, Chapter 3) I compared the patterns of called indels between these groups. No differences were found, and neither did the identified indel patterns show characteristics of Cas9-induced mutations. Additionally, WGS allowed me to inspect over 28.000 predicted off-target sites for the presence of mutations. Apart from two previously verified off-target sites (Chapter 2), at only one site with five mismatches to the target could off-target activity of CRISPR-Cas9 not be ruled out. All these results agree with previous WGS-based studies on CRISPR-Cas specificity in plants and indicate that Cas9 is specific.

GUIDE-seq: an attractive alternative for unbiased off-target site detection

In Chapter 4, I employed a method for unbiased off-target detection based on capturing small pieces of DNA in DSBs. This technique, called GUIDE-seq, is suitable for gaining an overview of the location of all DSBs a gRNA-nuclease pair may induce and all DSBs from other causes– and thus where mutations may occur (Malinin et al. 2021; S Q Tsai et al. 2015). I applied GUIDE-seq to six sgRNAs – five that I knew could induce off-target mutations and one for which no off-target sites were predicted. I used this technique to identify all on-target sites and known off-target sites for these six sgRNAs. The detection of a DSB at one of these sites (MYB19-OffT) gave independent evidence for the tolerance of Cas9 for mismatches in the so-called seed sequence. I could not identify any additional unpredicted off-target effects. This observation indicates that they do not occur or occur at frequencies below our detection limit, which is currently ~ 0.7%. Reported detection limits in mammalian cells are lower (~0.1%) (Kleinstiver, Tsai, et al. 2016; S Q Tsai et al. 2015), and I have proposed several methods to improve sensitivity.

This method is an important addition to the toolbox of available off-target detection methods in plants. It offers a way to validate the predictions of widely used off-target identification algorithms.

It is an *in vivo* method, offering advantages over other methods of unbiased off-target detection, such as Digenome-seq (D. Kim et al. 2015) or CIRCLE-seq (Shengdar Q Tsai et al. 2017), which often result in false positives. Arguably, it is a more reliable method for the detection of off-target sites than whole genome resequencing, as every site identified using GUIDE-seq originates from a DSB. In contrast, variants detected using WGS may also simply be a result of natural variation or imprecise variant calling. Therefore, the availability of a perfect reference genome is less critical for GUIDE-seq than for WGS. Thus, GUIDE-seq offers an unbiased way to compile a "shortlist" of putative off-target sites that could be checked for mutations after stable transformation.

Furthermore, the integration of small dsODNs in DSBs in a targeted manner may have exciting applications for research. This method may integrate small regulatory motifs or protein tags into the genome, facilitating, for example, research on transcriptional regulation or protein interaction while the genes involved remain in their native environment.

Limitations: large deletions, loss-of-heterozygosity, undesired insertions

Altogether, I have performed in-depth research on the specificity of, particularly CRISPR-Cas9, and to a lesser extent, CRISPR-Cas12a mediated genome editing in tomato. In my research, I have focused on the identification of genuine off-target sites and the frequency of their mutations. However, apart from off-target mutations, CRISPR-mediated genome editing can also result in undesirable effects at on-target locations. Studies have shown that Cas9 can induce large deletions, spanning up to hundreds of kilobases, as well as chromosome rearrangements (Cullot et al. 2019; Kosicki et al. 2018; Zhou et al. 2014). These large deletions result in loss-of-heterozygosity (LOH), which poses a considerable risk in human genome editing.

Many approaches I have applied in my research would not have identified such large-scale effects. They would be undetectable using the amplicon sequencing I applied in Chapters 2 and 3, as they disrupt primer binding sites. Although GUIDE-seq could identify the genomic break sites at which these mutations might occur, this technique is unsuitable for detecting such mutations and determining their frequencies. It is possible to identify this type of structural variation using whole genome resequencing, as applied in Chapter 5. In the 38 resequenced genome-edited lines, I did not find evidence of these effects at target sites and predicted off-target sites. As we find that larger deletions and insertions are increasingly rare (Chapter 2, Chapter 3), these large structural variations are not likely to occur frequently. To get a better overview of the frequency of these types of events, long-read sequencing such as Nanopore sequencing or PacBio sequencing of target or off-target sites could be applied. Additionally, specific PCR techniques such as anchored multiplex

PCR followed by sequencing (AMP-seq) or high-throughput genome-wide translocation sequencing (HTGTS) can be used to detect these events (Hu et al. 2016; S Q Tsai et al. 2015).

LOH can also be caused by allele conversion. This phenomenon can occur when a DSB induced in one allele on a chromosome is repaired in an HDR-involving process using the allele on another chromosome, resulting in the loss of one allele. This undesirable effect has been shown to occur in human embryos, hematopoietic cells, and yeast (Alanis-Lobato et al. 2021; Boutin et al. 2021; Gorter De Vries et al. 2019). In my experiments, I would not have been able to detect allele conversion, as the Moneyberg genome is highly homozygous and therefore does not contain the heterozygous alleles necessary to distinguish the two chromosomes. This pre-existing homozygosity also illustrates that risks associated with LOH are likely low for tomato cultivars such as Moneyberg. Additionally, the occurrence of LOH – either due to large deletions or as a result of allele conversion – at target sites is relatively simple to avoid in mutant tomato plants by ensuring that two different alleles are found.

In Chapter 2, I have shown that pieces of vector-derived DNA can integrate at DSB sites. As plasmid preparations can sometimes be contaminated by genomic DNA of *Escherichia coli*, it is not unlikely that *E. coli* DNA might also integrate at low frequencies at DSBs. Previous studies have additionally shown that incomplete (and therefore, difficult to detect) fragments of T-DNA, as well as parts of the *Agrobacterium*-genome itself, might integrate into plant genomes (Salomon and Puchta 1998; Schouten et al. 2017; Ülker et al. 2008). Such integrations would result in a transgenic plant and might have regulatory consequences. RNP transfection is often proposed to rule out this risk (J. Kim and Kim 2016). However, I have found that even using RNP transfection, short DNA stretches derived from the dsDNA template used to *in vitro* produce sgRNAs may integrate into DSBs (Chapter 2). It is important to note that this occurred at low frequencies in our experiments. Additionally, using techniques such as GUIDE-seq, the most likely integration sites for such DNA sequences can be predicted and subsequently screened for the presence of foreign DNA. To further avoid these integrations, sgRNAs might also be entirely synthetically produced without using a DNA template. Nevertheless, the risk of foreign DNA integration is never zero (Bertheau 2021).

Strategies for minimizing the off-target mutation frequency

Especially in mammalian cells, much research has been done to minimize the frequency of off-target mutations. Of course, developing off-target prediction algorithms allowed for easy selection of sgRNAs with a minimum number of putative off-target sites. Additionally, researchers have tried to improve specificity by making use of two Cas9 nickases, the action of which will result in a DSB

only when both enzymes bind at the correct position (Ran et al. 2013; Mali et al. 2013). In a variation on this theme, completely inactive dCas9s have been fused to *Fok*I endonuclease domains to induce DSBs similarly to ZFNs and TALENs (Guilinger et al. 2014; Shengdar Q. Tsai et al. 2014). Furthermore, Cas enzymes have been engineered to be more specific, for example, by introducing amino acid mutations that weaken non-target DNA binding (Slaymaker et al. 2016), reduce non-specific DNA contacts (Kleinstiver, Pattanayak, et al. 2016) or modify the REC3 domain, which plays a role in the activation of the nuclease domain (J. S. Chen et al. 2017). These nucleases have also been successfully applied in plants, albeit on a limited scale (D. Zhang et al. 2017; Q. Zhang et al. 2018; Xu et al. 2019). In this research, I have focused on the unmodified, original SpCas9 and LbCas12a. By using these more specific nucleases, off-target mutation frequencies, as reported in this thesis, could likely be further reduced.

Regulation of CRISPR-bred crops

The current regulatory status of CRISPR-bred crops

In Chapter 1, I mentioned that the Japanese "Sicilian Rouge High GABA" tomato is the first CRISPR-edited crop available on the market. Unlike in Japan, it is currently not possible to buy CRISPR-bred fruits or vegetables in a Dutch supermarket – or any other EU supermarket, for that matter. The current European legislation on genetically modified organisms (GMOs) was put in place in 2001 (Directive 2001/18/EC, amended by Regulations in 2003: No 1829/2003 and No 1830/2003), before site-directed nucleases (ZFNs, TALENs) were widely applied and over a decade before CRISPR-Cas9 was first used as a genome editing tool. Existing GM crops were then mostly developed by integrating foreign genes in the crop genome and are thus transgenic crops. Examples of the latter are Golden Rice (and its successor, Golden Rice 2) and Bt corn (Ostlie et al. 1997; Ye and Beyer 2000).

Crops obtained through mutation breeding, the random introduction of mutations in the genome using techniques such as irradiation and ethyl methane sulfonate (EMS) treatment (Jankowicz-Cieslak & Till, 2015), are also classified as GMOs. The legal, European definition of a GMO is: "organisms in which the genetic material has been altered in a way that does not occur naturally by mating or natural recombination". However, crops derived from mutation breeding were exempted from GMO regulation, with the motivation that the techniques had a long history of safe use. Currently, over 3400 food and ornamental crops bred using mutagenesis techniques are grown worldwide (Mutant Varieties Database).

Although not an explicit criterium in current GMO regulation, there is one practical difference between GMOs regulated as such and GMOs exempted from this regulation: the presence – or absence – of foreign DNA. This difference also reflects the enforceability of the legislation. Foreign DNA, such as the bacterial genes in Golden Rice and Bt corn, frequently-used genetic elements such as the CaMV35S promoter, NOS terminator, or different antibiotic resistance genes can readily be detected with techniques such as (q)PCR or microarrays (Marmiroli et al. 2008). Alternatively, though with more difficulty, sequencing approaches such as whole genome sequencing can be used and the foreign origin of integrated genes can be identified using existing databases. The mutations induced by irradiation, chemical treatments, or mutagenic CRISPR-Cas approaches can be detected by sequencing methods, but their origin cannot be determined with certainty, as these induced mutations are indistinguishable from naturally occurring mutations. I encountered similar difficulties in Chapter 5.

With the advent of site-specific nucleases, and their ability to perform targeted mutagenesis without the integration of foreign DNA, it became unclear whether crops bred using these techniques should be regulated as GMOs or be exempted from existing GMO regulations. In 2016, the Confédération Paysanne, amongst other groups, asked the Court of Justice of the European Union for clarification on this subject. The scientific community assumed that at least targeted mutagenesis would be exempted, especially when no foreign DNA was introduced to the plant, as is illustrated by the papers published around that time, highlighting phrases such as "DNA-free transformation of plant cells" (H. Kim et al. 2017; Woo et al. 2015; Y. Zhang et al. 2016). This expectation was reinforced by the legal opinion of the appointed Advocate General of the Court of Justice of the EU, which stated that "the exemption […] covers all organisms obtained by any technique of mutagenesis, […] on the condition that they do not involve the use of recombinant nucleic acid molecules or genetically modified organisms […]". However, in July 2018, the European Court of Justice ruled that all crops bred using mutagenesis techniques "are GMOs and […] subject to the obligations laid down by the GMO Directive".

Based on the precautionary principle, this conservative decision was met with both surprise and discontent by the scientific community (Kupferschmidt 2018; Christiansen et al. 2019; Callaway 2018). This sentiment got stronger when it later became clear that other leading countries would allow CRISPR-Cas-bred crops to be brought to the market. In the USA, crops are regulated from a product-centred approach rather than an approach based on the technology that was used to create them. Thus, CRISPR-edited crops are not regulated as long as they contain minor changes that could have been made through traditional breeding (USA Department of Agriculture 2020).

Similar regulations exist in Canada, much of South America, Australia, and Japan. China is still in the process of setting up regulations but is also expected to only moderately regulate CRISPR-edited crops (Global Gene Editing Regulation Tracker). Currently, the European Commission is working on a new legislative proposal for plants produced by certain new genomic techniques, including CRISPR-Cas (European Commission 2023).

Challenges in regulating CRISPR-bred crops

A difficulty in regulating genome-edited crops is that site-directed nucleases (SDNs), such as ZFNs, TALENs, and CRISPR-Cas, do not easily fit into the current European framework of process-based regulation. Whereas in the pre-genome editing era, one procedure would typically result in one type of outcome, the application of SDNs can result in many different outcomes, ranging from simple indels to nucleotide substitutions and to the targeted integration of foreign genes, depending on the experimental setup. These outcomes were previously defined in three categories: SDN-1, SDN-2, and SDN-3 applications (Sturme et al. 2022). In SDN-1 applications, targeted mutations (indels) are made by inducing DSBs, which are subsequently repaired using the NHEJ or alt-EJ pathways. In SDN-2 applications, targeted point mutations or small indels are achieved through HDR, base-editing or prime editing. In SDN-3 applications, a complete gene or other genomic feature is integrated at a target site, using either HDR and a large DNA template, or NHEJ. If the integrated DNA is derived from the same or a crossable species, this will result in a cisgenic plant; otherwise, the plant will be transgenic. The outcomes of SDN-1 and SDN-2 approaches are similar to – and sometimes indistinguishable from – traditional breeding and mutagenesis techniques. SDN-3 outcomes are more similar to the "classic" GMO, with the difference being that the integration site is not random as with T-DNA insertion.

Specificity and Safety

The specificity and the safety of genome editing are often mentioned in the same breath. However, it is good to define what is meant by "safe", and how specificity – or lack thereof – can affect these aspects. Furthermore, it is essential to note that risks may differ per SDN application.

When it comes to GM crops, safety comprises two things: first, that the GM crop is as safe for human consumption as the non-edited counterpart ("assessment of substantial equivalence"), and second, that the crop is not harmful to the environment – either directly, or indirectly (Bawa and Anilakumar 2013). Several potential hazards can be identified in both categories. Hazards to consumer safety include introducing or creating new allergens, introducing or increasing toxin

production, and losing nutritional value (Bawa and Anilakumar 2013). Environmental hazards are mostly associated with introducing pesticide- or herbicide-related traits to plants. These introductions may lead to the evolution of resistance in both pests and weeds (Schütte et al. 2017; Bates et al. 2005). Additionally, traits may "escape" to wild or weedy relatives of the genome-edited crop (Sudianto et al. 2013). Several crops (Bt corn, Bt cotton) have been bred to be resistant to lepidopteran pests by producing Bt toxin. Although this toxin is harmless to humans and most other insects, there is still a chance of it accumulating higher up in the food chain (M. Chen et al. 2009; Obrist et al. 2006). Furthermore, the loss of one pest may open up a niche for other, previously minor pests, which can result in a loss of the advantages of the genome-edited crop (such as increased yield and reduced pesticide usage) over the non-edited crop (Krishna and Qaim 2012; J. H. Zhao et al. 2011).

Generally speaking, the risks of SDN-1 and SDN-2 approaches concerning consumer and environmental safety are considered low (Lassoued et al. 2019; Naegeli et al. 2020). As no new genetic material is introduced to the genome, the risk of introducing new allergens or toxins is low, as are environmental risks. Risks associated with these outcomes are comparable to those associated with classic breeding and mutagenesis methods. SDN-3 approaches, in which foreign DNA is integrated, similar to traditional GMOs, potentially carry higher risks. In all cases, however, risks depend on the exact application of the SDN and aspects such as the altered trait and the nature of the integrated DNA (Mullins et al. 2022; Lassoued et al. 2019; Kawall 2021).

Risks posed by off-target mutations

Using multiple approaches, I have shown that CRISPR-Cas9-mediated genome editing in tomato is highly specific. Furthermore, I have shown that off-target mutations are predictable and thus avoidable. Using whole genome sequencing, I have shown that genome-edited plants do not contain significantly more indels and SVs to the reference genome than non-genome-edited plants. Therefore, we could say that the range of genetic variation found in genome-edited plants falls within the range of genetic variation found within the cultivar. Although I have not applied such methods in this thesis, the frequency of off-target mutations could be further reduced using more specific, engineered Cas-nucleases (Xu et al., 2019; D. Zhang et al., 2017; Q. Zhang et al., 2018).

If off-target mutations are found in regenerated plants, these plants can either be discarded or, if possible, the off-target mutation can be removed by backcrossing (Graham et al. 2020). Therefore, safety risks of off-target genome editing will be associated with unpredictable off-target sites – although I have not found evidence of those - or large or complex structural variation that is not as

easily detected as the more frequently occurring indels. It is worth noting that even if such an unintended genomic change occurs, it will not necessarily negatively affect consumer or environmental safety. The risks associated with off-target mutations are not a result of these mutations themselves but of the potential undesired effect of these mutations on plant traits (H. Zhao and Wolt 2017). Mutation breeding, using EMS or irradiation, is known to induce mutations all over the genome – the vast majority of which can be considered unintended and even undesired (Pacher and Puchta 2016). However, products resulting from these procedures have a long history of safe use, indicating that these unintended mutations have not affected consumers.

Therefore, I propose that off-target mutations pose limited additional risk to consumer and environmental safety as compared to on-target mutations. However, as for other mutagenesis approaches, risks can never be completely ruled out. Thus, it should be determined what level of risk is acceptable and if it should trigger safety assessment regulation. Additionally, as SDN approaches can result in such a wide range of changes, individual assessment of genome-edited products, instead of regulating them by the process used to create them, might be the method of choice.

Conclusions

In this thesis, I have thoroughly studied the specificity of CRISPR-Cas9-mediated genome editing in tomato. Targeted amplicon sequencing allowed us to characterize high-risk off-target sites and conclude that the large majority of predicted off-target sites are not altered. Using GUIDE-seq, I show that unpredicted off-target DSBs either do not occur or only at very low frequencies, reinforcing the value of off-target prediction software used in biased approaches. Additionally, using whole genome resequencing, I show that CRISPR-mediated genome editing does not result in an overall increase of indels or SVs in the genome. Combining these results, I conclude that CRISPR-Cas9-mediated genome editing in tomato is highly specific. Additionally, I conclude that off-target sites are predictable and, therefore, that off-target mutations in products for farmers and consumers are avoidable. Furthermore, I demonstrated the value of high-throughput research techniques in plants. I showed that genome editing is not only a useful tool for reverse genetics but can also be applied to study more fundamental processes such as DSB repair. I discussed the regulatory status of CRISPR-bred crops, the difficulties surrounding regulation, and touched upon the link between specificity and safety. Overall, the research presented in this thesis will provide valuable information to both users of CRISPR-Cas mediated genome editing and professionals involved in the regulation and legislation surrounding genome-edited crops.

References

Alanis-Lobato, G., J. Zohren, A. McCarthy, N.M.E. Fogarty, N. Kubikova, E. Hardman, M. Greco, D. Wells, J.M.A. Turner, and K.K. Niakan. 2021. "Frequent Loss of Heterozygosity in CRISPR-Cas9-Edited Early Human Embryos." *Proceedings of the National Academy of Sciences of the United States of America* 118 (22). https://doi.org/10.1073/pnas.2004832117.

Bates, S.L., J.Z. Zhao, R.T. Roush, and A.M. Shelton. 2005. "Insect Resistance Management in GM Crops: Past, Present and Future." *Nature Biotechnology*. https://doi.org/10.1038/nbt1056.

Bawa, A.S., and K.R. Anilakumar. 2013. "Genetically Modified Foods: Safety, Risks and Public Concerns - A Review." *Journal of Food Science and Technology*. https://doi.org/10.1007/s13197-012-0899-1.

Ben-Tov, D., F. Mafessoni, A. Cucuy, A. Honig, C. Melamed-Bessudo, and A.A. Levy. 2023. "Uncovering the Dynamics of Precise Repair at CRISPR/Cas9-Induced Double-Strand Breaks." *BioRxiv*. https://doi.org/10.1101/2023.01.10.523377.

Bernabé-Orts, J.M., I. Casas-Rodrigo, E.G. Minguet, V. Landolfi, V. Garcia-Carpintero, S. Gianoglio, M. Vázquez-Vilar, … D. Orzaez. 2019. "Assessment of Cas12a-Mediated Gene Editing Efficiency in Plants." *Plant Biotechnology Journal* 17 (10): 1971–84. https://doi.org/10.1111/pbi.13113.

Bertheau, Y. 2021. "Advances in Identifying GM Plants: Toward the Routine Detection of 'Hidden' and 'New' GMOs." In , 87–150. https://doi.org/10.19103/AS.2021.0097.22.

Bétermier, M., P. Bertrand, and B.S. Lopez. 2014. "Is Non-Homologous End-Joining Really an Inherently Error-Prone Process?" *PLoS Genetics* 10 (1). https://doi.org/10.1371/journal.pgen.1004086.

Boutin, J., J. Rosier, D. Cappellen, F. Prat, J. Toutain, P. Pennamen, J. Bouron, … A. Bedel. 2021. "CRISPR-Cas9 Globin Editing Can Induce Megabase-Scale Copy-Neutral Losses of Heterozygosity in Hematopoietic Cells." *Nature Communications* 12 (1). https://doi.org/10.1038/s41467-021-25190-6.

Brooks, C., V. Nekrasov, Z.B. Lippman, and J. Van Eck. 2014. "Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/ CRISPR-Associated9 System." *Plant Physiology* 166 (3): 1292–1297. https://doi.org/10.1104/pp.114.247577.

Callaway, E. 2018. "CRISPR Plants Now Subject to Tough GM Laws in European Union." *Nature* 560 (7716): 16. https://doi.org/10.1038/d41586-018-05814-6.

Chen, J.S., Y.S. Dagdas, B.P. Kleinstiver, M.M. Welch, A.A. Sousa, L.B. Harrington, S.H. Sternberg, J.K. Joung, A. Yildiz, and J.A. Doudna. 2017. "Enhanced Proofreading Governs CRISPR–Cas9 Targeting Accuracy." *Nature* 550 (7676): 407–10. https://doi.org/10.1038/nature24268.

Chen, M., G.Y. Ye, Z.C. Liu, Q. Fang, C. Hu, Y.F. Peng, and A.M. Shelton. 2009. "Analysis of Cry1Ab Toxin Bioaccumulation in a Food Chain of Bt Rice, an Herbivore and a Predator." *Ecotoxicology* 18 (2): 230–38. https://doi.org/10.1007/s10646-008-0276-z.

Christiansen, A.T., M.M. Andersen, and K. Kappel. 2019. "Are Current EU Policies on GMOs Justified?" *Transgenic Research* 28 (2): 267–86. https://doi.org/10.1007/s11248-019-00120-x.

Cullot, G., J. Boutin, J. Toutain, F. Prat, P. Pennamen, C. Rooryck, M. Teichmann, … A. Bedel. 2019. "CRISPR-Cas9 Genome Editing Induces Megabase-Scale Chromosomal Truncations." *Nature Communications* 10 (1): 1136. https://doi.org/10.1038/s41467-019-09006-2.

Doench, J.G., N. Fusi, M. Sullender, M. Hegde, E.W. Vaimberg, K.F. Donovan, I. Smith, … D.E. Root. 2016. "Optimized SgRNA Design to Maximize Activity and Minimize Off-Target Effects of CRISPR-Cas9." *Nature Biotechnology* 34 (2): 184–91. https://doi.org/10.1038/nbt.3437.

Doench, J.G., E. Hartenian, D.B. Graham, Z. Tothova, M. Hegde, I. Smith, M. Sullender, B.L. Ebert, R.J. Xavier, and D.E. Root. 2014. "Rational Design of Highly Active SgRNAs for CRISPR-Cas9-Mediated Gene Inactivation." *Nature Biotechnology* 32 (12): 1262–67. https://doi.org/10.1038/nbt.3026.

Doudna, J.A. 2020. "The Promise and Challenge of Therapeutic Genome Editing." *Nature*. Nature Research. https://doi.org/10.1038/s41586-020-1978-5.

European Commission. 2023. "European Commision Initiative: Legislation for Plants Produced by Certain Genomic Techniques." 2023. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13119-Legislation-for-plants-produced-by-certain-new-genomic-techniques_en.

Fusi, N., I. Smith, J. Doench, and J. Listgarten. 2015. "In Silico Predictive Modeling of CRISPR/Cas9 Guide Efficiency." *BioRxiv*, 021568. https://doi.org/10.1101/021568.

Gaillochet, C., W. Develtere, and T.B. Jacobs. 2021. "CRISPR Screens in Plants: Approaches, Guidelines, and Future Prospects." *Plant Cell*. American Society of Plant Biologists. https://doi.org/10.1093/plcell/koab099.

Gaillochet, C., A. Peña Fernández, V. Goossens, K. D'Halluin, A. Drozdzecki, M. Shafie, J. Van Duyse, … T.B. Jacobs. 2023. "Systematic Optimization of Cas12a Base Editors in Wheat and Maize Using the ITER Platform." *Genome Biology* 24 (1). https://doi.org/10.1186/s13059-022-02836-2.

Genetic Literacy Project. n.d. "Global Gene Editing Regulation Tracker." Accessed May 4, 2023. https://crispr-gene-editing-regs-tracker.geneticliteracyproject.org/.

Gorter De Vries, A.R., L.G.F. Couwenberg, M. Van Den Broek, P. De La Torre Cortés, J. Ter Horst, J.T. Pronk, and J.M.G. Daran. 2019. "Allele-Specific Genome Editing Using CRISPR-Cas9 Is Associated with Loss of Heterozygosity in Diploid Yeast." *Nucleic Acids Research* 47 (3): 1362–72. https://doi.org/10.1093/nar/gky1216.

Graham, N., G.B. Patil, D.M. Bubeck, R.C. Dobert, K.C. Glenn, A.T. Gutsche, S. Kumar, … P.L. Morrell. 2020. "Plant Genome Editing and the Relevance of Off-Target Changes1[OPEN]." *Plant Physiology* 183 (4): 1453–71. https://doi.org/10.1104/pp.19.01194.

Guilinger, J.P., D.B. Thompson, and D.R. Liu. 2014. "Fusion of Catalytically Inactive Cas9 to FokI Nuclease Improves the Specificity of Genome Modification." *Nature Biotechnology* 32 (6): 577–82. https://doi.org/10.1038/nbt.2909.

Hsu, P.D., D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, Y. Li, … F. Zhang. 2013. "DNA Targeting Specificity of RNA-Guided Cas9 Nucleases." *Nature Biotechnology* 31 (9): 827–32. https://doi.org/10.1038/nbt.2647.

Hu, J., R.M. Meyers, J. Dong, R.A. Panchakshari, F.W. Alt, and R.L. Frock. 2016. "Detecting DNA Double-Stranded Breaks in Mammalian Genomes by Linear Amplification-Mediated High-Throughput Genome-Wide Translocation Sequencing." *Nature Protocols* 11 (5): 853–71. https://doi.org/10.1038/nprot.2016.043.

IAEA. n.d. "Mutant Varieties Database." Accessed May 4, 2023. https://www.iaea.org/resources/databases/mutant-varieties-database.

Jaganathan, D., K. Ramasamy, G. Sellamuthu, S. Jayabalan, and G. Venkataraman. 2018. "CRISPR for Crop Improvement: An Update Review." *Frontiers in Plant Science* 9 (July): 1–17. https://doi.org/10.3389/fpls.2018.00985.

Jankowicz-Cieslak, J., and B.J. Till. 2015. "Forward and Reverse Genetics in Crop Breeding." In *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*, 1:215–40. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-22521-0_8.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Jouanin, A., L.J.W.J. Gilissen, L.A. Boyd, J. Cockram, F.J. Leigh, E.J. Wallington, H.C. van den Broeck, … M.J.M. Smulders. 2018. "Food Processing and Breeding Strategies for Coeliac-Safe and Healthy Wheat Products." *Food Research International* 110 (August): 11–21. https://doi.org/10.1016/j.foodres.2017.04.025.

Kawall, K. 2021. "The Generic Risks and the Potential of SDN-1 Applications in Crop Plants." *Plants*. MDPI. https://doi.org/10.3390/plants10112259.

Kieran, P.M., P.F. Macloughlin, and D.M. Malone. 1997. "Plant Cell Suspension Cultures: Some Engineering Considerations." *Journal of Biotechnology*. Vol. 59.

Kim, D., S. Bae, J. Park, E. Kim, S. Kim, H.R. Yu, J. Hwang, J.-I. Kim, and J.-S. Kim. 2015. "Digenome-Seq: Genome-Wide Profiling of CRISPR-Cas9 off-Target Effects in Human Cells." *Nature Methods* 12 (3): 237–44. https://doi.org/10.1038/nmeth.3284.

Kim, D., K. Luk, S.A. Wolfe, and J.-S. Kim. 2019. "Evaluating and Enhancing Target Specificity of Gene-Editing Nucleases and Deaminases." *Annual Reviews in Biochemistry*. https://doi.org/10.1146/annurev-biochem-013118.

Kim, H., S.T. Kim, J. Ryu, B.C. Kang, J.S. Kim, and S.G. Kim. 2017. "CRISPR/Cpf1-Mediated DNA-Free Plant Genome Editing." *Nature Communications* 8: 1–7. https://doi.org/10.1038/ncomms14406.

Kim, H.K., S. Min, M. Song, S. Jung, J.W. Choi, Y. Kim, S. Lee, S. Yoon, and H. (Henry) Kim. 2018. "Deep Learning Improves Prediction of CRISPR–Cpf1 Guide RNA Activity." *Nature Biotechnology* 36 (3): 239-241 3. https://doi.org/10.1038/nbt.4061.

Kim, J., and J.-S. Kim. 2016. "Bypassing GMO Regulations with CRISPR Gene Editing." *Nature Biotechnology* 34 (10): 1014–15. https://doi.org/10.1038/nbt.3680.

Kleinstiver, B.P., V. Pattanayak, M.S. Prew, S.Q. Tsai, N.T. Nguyen, Z. Zheng, and J. Keith Joung. 2016. "High-Fidelity CRISPR–Cas9 Nucleases with No Detectable Genome-Wide off-Target Effects." *Nature* 529 (7587): 490–95. https://doi.org/10.1038/nature16526.

Kleinstiver, B.P., S.Q. Tsai, M.S. Prew, N.T. Nguyen, M.M. Welch, J.M. Lopez, Z.R. McCaw, M.J. Aryee, and J.K. Joung. 2016. "Genome-Wide Specificities of CRISPR-Cas Cpf1 Nucleases in Human Cells." *Nature Biotechnology* 34 (8): 869–74. https://doi.org/10.1038/nbt.3620.

Konstantakos, V., A. Nentidis, A. Krithara, and G. Paliouras. 2022. "CRISPR-Cas9 GRNA Efficiency Prediction: An Overview of Predictive Tools and the Role of Deep Learning." *Nucleic Acids Research*. Oxford University Press. https://doi.org/10.1093/nar/gkac192.

Kosicki, M., K. Tomberg, and A. Bradley. 2018. "Repair of Double-Strand Breaks Induced by CRISPR–Cas9 Leads to Large Deletions and Complex Rearrangements." *Nature Biotechnology* 36 (8): 765–71. https://doi.org/10.1038/nbt.4192.

Krishna, V. V., and M. Qaim. 2012. "Bt Cotton and Sustainability of Pesticide Reductions in India." *Agricultural Systems* 107 (March): 47–55. https://doi.org/10.1016/j.agsy.2011.11.005.

Kupferschmidt, K. 2018. "EU Verdict on CRISPR Crops Dismays Scientists." *Science*. American Association for the Advancement of Science. https://doi.org/10.1126/science.361.6401.435.

Lassoued, R., D.M. Macall, S.J. Smyth, P.W.B. Phillips, and H. Hesseln. 2019. "Risk and Safety Considerations of Genome Edited Crops: Expert Opinion." *Current Research in Biotechnology* 1 (November): 11–21. https://doi.org/10.1016/j.crbiot.2019.08.001.

Lei, Y., L. Lu, H.Y. Liu, S. Li, F. Xing, and L.L. Chen. 2014. "CRISPR-P: A Web Tool for Synthetic Single-Guide RNA Design of CRISPR-System in Plants." *Molecular Plant* 7 (9): 1494–96. https://doi.org/10.1093/mp/ssu044.

Lemos, B.R., A.C. Kaplan, J.E. Bae, A.E. Ferrazzoli, J. Kuo, R.P. Anand, D.P. Waterman, and J.E. Haber. 2018. "CRISPR/Cas9 Cleavages in Budding Yeast Reveal Templated Insertions and Strand-Specific Insertion/Deletion

Profiles." *Proceedings of the National Academy of Sciences of the United States of America* 115 (9): E2010–47. https://doi.org/10.1073/pnas.1716855115.

Lin, C.S., C.T. Hsu, L.H. Yang, L.Y. Lee, J.Y. Fu, Q.W. Cheng, F.H. Wu, … M.C. Shih. 2018. "Application of Protoplast Technology to CRISPR/Cas9 Mutagenesis: From Single-Cell Mutation Detection to Mutant Plant Regeneration." *Plant Biotechnology Journal* 16 (7): 1295–1310. https://doi.org/10.1111/pbi.12870.

Liu, H., Y. Ding, Y. Zhou, W. Jin, K. Xie, and L.L. Chen. 2017. "CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants." *Molecular Plant* 10 (3): 530–32. https://doi.org/10.1016/j.molp.2017.01.003.

Luo, J., W. Chen, L. Xue, and B. Tang. 2019. "Prediction of Activity and Specificity of CRISPR-Cpf1 Using Convolutional Deep Learning Neural Networks." *BMC Bioinformatics* 20 (1): 332. https://doi.org/10.1186/s12859-019-2939-6.

Mali, P., J. Aach, P.B. Stranges, K.M. Esvelt, M. Moosburner, S. Kosuri, L. Yang, and G.M. Church. 2013. "CAS9 Transcriptional Activators for Target Specificity Screening and Paired Nickases for Cooperative Genome Engineering." *Nature Biotechnology* 31 (9): 833–38. https://doi.org/10.1038/nbt.2675.

Malinin, N.L., G. Lee, C.R. Lazzarotto, Y. Li, Z. Zheng, N.T. Nguyen, M. Liebers, … S.Q. Tsai. 2021. "Defining Genome-Wide CRISPR–Cas Genome-Editing Nuclease Activity with GUIDE-Seq." *Nature Protocols* 16 (December). https://doi.org/10.1038/s41596-021-00626-x.

Marmiroli, N., E. Maestri, M. Gullì, A. Malcevschi, C. Peano, R. Bordoni, and G. De Bellis. 2008. "Methods for Detection of GMOs in Food and Feed." *Analytical and Bioanalytical Chemistry*. https://doi.org/10.1007/s00216-008-2303-6.

Miller, J.C., S. Tan, G. Qiao, K.A. Barlow, J. Wang, D.F. Xia, X. Meng, … E.J. Rebar. 2011. "A TALE Nuclease Architecture for Efficient Genome Editing." *Nature Biotechnology* 29 (2): 143–48. https://doi.org/10.1038/nbt.1755.

Minoche, A.E., J.C. Dohm, and H. Himmelbauer. 2011. "Evaluation of Genomic High-Throughput Sequencing Data Generated on Illumina HiSeq and Genome Analyzer Systems." *Genome Biology* 12 (11). https://doi.org/10.1186/gb-2011-12-11-r112.

Moreno-Mateos, M.A., C.E. Vejnar, J.D. Beaudoin, J.P. Fernandez, E.K. Mis, M.K. Khokha, and A.J. Giraldez. 2015. "CRISPRscan: Designing Highly Efficient SgRNAs for CRISPR-Cas9 Targeting in Vivo." *Nature Methods* 12 (10): 982–88. https://doi.org/10.1038/nmeth.3543.

Moscatiello, R., B. Baldan, and L. Navazio. 2013. "Plant Cell Suspension Cultures." In *Plant Mineral Nutrients*, 77–93. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-152-3_5.

Mullins, E., J.L. Bresson, T. Dalmay, I.C. Dewhurst, M.M. Epstein, L.G. Firbank, P. Guerche, … R. Schoonjans. 2022. "Criteria for Risk Assessment of Plants Produced by Targeted Mutagenesis, Cisgenesis and Intragenesis." *EFSA Journal* 20 (10). https://doi.org/10.2903/j.efsa.2022.7618.

Naegeli, H., J.L. Bresson, T. Dalmay, I.C. Dewhurst, M.M. Epstein, L.G. Firbank, P. Guerche, … N. Rostoks. 2020. "Applicability of the EFSA Opinion on Site-Directed Nucleases Type 3 for the Safety Assessment of Plants Developed Using Site-Directed Nucleases Type 1 and 2 and Oligonucleotide-Directed Mutagenesis." *EFSA Journal* 18 (11): 6299. https://doi.org/10.2903/j.efsa.2020.6299.

Naim, F., K. Shand, S. Hayashi, M. O'Brien, J. McGree, A.A.T. Johnson, B. Dugdale, and P.M. Waterhouse. 2020. "Are the Current GRNA Ranking Prediction Algorithms Useful for Genome Editing in Plants?" *PLoS ONE* 15 (1): 1–12. https://doi.org/10.1371/journal.pone.0227994.

Nakamura, K., T. Oshima, T. Morimoto, S. Ikeda, H. Yoshikawa, Y. Shiwa, S. Ishikawa, … S. Kanaya. 2011. "Sequence-Specific Error Profile of Illumina Sequencers." *Nucleic Acids Research* 39 (13). https://doi.org/10.1093/nar/gkr344.

Nekrasov, V., C. Wang, J. Win, C. Lanz, D. Weigel, and S. Kamoun. 2017. "Rapid Generation of a Transgene-Free Powdery Mildew Resistant Tomato by Genome Deletion." *Scientific Reports* 7 (1): 482. https://doi.org/10.1038/s41598-017-00578-x.

Obrist, L.B., A. Dutton, R. Albajes, and F. Bigler. 2006. "Exposure of Arthropod Predators to Cry1Ab Toxin in Bt Maize Fields." *Ecological Entomology* 31 (2): 143–54. https://doi.org/10.1111/j.0307-6946.2006.00762.x.

Ostlie, K.R., W.D. Hutchison, and R.L. Hellmich. 1997. "Bt Corn and European Corn Borer." *North Central Region Publication* 602. https://digitalcommons.unl.edu/entomologyfacpub/597.

Overbeek, M. van, D. Capurso, M.M. Carter, M.S. Thompson, E. Frias, C. Russ, J.S. Reece-Hoyes, … A.P. May. 2016. "DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks." *Molecular Cell* 63 (4): 633–46. https://doi.org/10.1016/j.molcel.2016.06.037.

Pacher, M., and H. Puchta. 2016. "From Classical Mutagenesis to Nuclease-Based Breeding - Directing Natural DNA Repair for a Natural End-Product." *The Plant Journal : For Cell and Molecular Biology* 38 (1): 42–49. https://doi.org/10.1111/tpj.13469.

Pan, C., G. Li, A. Bandyopadhyay, and Y. Qi. 2023. "Guide RNA Library-Based CRISPR Screens in Plants: Opportunities and Challenges." *Current Opinion in Biotechnology*. Elsevier Ltd. https://doi.org/10.1016/j.copbio.2022.102883.

Pattanayak, V., S. Lin, J.P. Guilinger, E. Ma, J.A. Doudna, and D.R. Liu. 2013. "High-Throughput Profiling of off-Target DNA Cleavage Reveals RNA-Programmed Cas9 Nuclease Specificity." *Nature Biotechnology* 31 (9): 839–43. https://doi.org/10.1038/nbt.2673.

Puchta, H. 2002. "Gene Replacement by Homologous Recombination in Plants." *Plant Molecular Biology* 48: 173–82. https://search-proquest-com.ezproxy.library.wur.nl/docview/768110637?OpenUrlRefId=info:xri/sid:wcdiscovery&accountid=27871.

Puchta, H. 2005. "The Repair of Double-Strand Breaks in Plants: Mechanisms and Consequences for Genome Evolution." *Journal of Experimental Botany* 56 (409): 1–14. https://doi.org/10.1093/jxb/eri025.

Ran, F.A., P.D. Hsu, C.-Y. Lin, J.S. Gootenberg, S. Konermann, A.E. Trevino, D.A. Scott, … F. Zhang. 2013. "Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity." *Cell* 154 (6): 1380–89. https://doi.org/10.1016/j.cell.2013.08.021.

Reed, K.M., and B.O.R. Bargmann. 2021. "Protoplast Regeneration and Its Use in New Plant Breeding Technologies." *Frontiers in Genome Editing* 3 (September). https://doi.org/10.3389/fgeed.2021.734951.

Salomon, S., and H. Puchta. 1998. "Capture of Genomic and T-DNA Sequences during Double-Strand Break Repair in Somatic Plant Cells." *EMBO Journal* 17 (20): 6086–95. https://doi.org/10.1093/emboj/17.20.6086.

Schouten, H.J., H. vande Geest, S. Papadimitriou, M. Bemer, J.G. Schaart, M.J.M. Smulders, G.S. Perez, and E. Schijlen. 2017. "Re-Sequencing Transgenic Plants Revealed Rearrangements at T-DNA Inserts, and Integration of a Short T-DNA Fragment, but No Increase of Small Mutations Elsewhere." *Plant Cell Reports* 0 (0): 1–12. https://doi.org/10.1007/s00299-017-2098-z.

Schütte, G., M. Eckerstorfer, V. Rastelli, W. Reichenbecher, S. Restrepo-Vassalli, M. Ruohonen-Lehto, A.G.W. Saucy, and M. Mertens. 2017. "Herbicide Resistance and Biodiversity: Agronomic and Environmental Aspects of Genetically Modified Herbicide-Resistant Plants." *Environmental Sciences Europe*. Springer Verlag. https://doi.org/10.1186/s12302-016-0100-y.

Seki, M., C. Ohzora, M. Takeda, and S. Furusaki. 1997. "Taxol (Paclitaxel) Production Using Free and Immobilized Cells of Taxus Cuspidata." *Biotechnology and Bioengineering* 53 (2): 214–19. https://doi.org/10.1002/(SICI)1097-0290(19970120)53:2<214::AID-BIT12>3.0.CO;2-K.

Semenova, E., M.M. Jore, K.A. Datsenko, A. Semenova, E.R. Westra, B. Wanner, J. Van Der Oost, S.J.J. Brouns, and K. Severinov. 2011. "Interference by Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) RNA Is Governed by a Seed Sequence." *Proceedings of the National Academy of Sciences of the United States of America* 108 (25): 10098–103. https://doi.org/10.1073/pnas.1104144108.

Shalem, O., N.E. Sanjana, E. Hartenian, X. Shi, D.A. Scott, T.S. Mikkelsen, D. Heckl, … F. Zhang. 2014. "Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells." *Science* 343 (6166): 84–87. https://doi.org/10.1126/science.1247005.

Shou, J., J. Li, Y. Liu, and Q. Wu. 2018. "Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion." *Molecular Cell* 71 (4): 498-509.e4. https://doi.org/10.1016/j.molcel.2018.06.021.

Slaymaker, I.M., L. Gao, B. Zetsche, D.A. Scott, W.X. Yan, and F. Zhang. 2016. "Rationally Engineered Cas9 Nucleases with Improved Specificity." *Science* 351 (6268): 84–88. https://doi.org/10.1126/science.aad5227.

Steinert, J., S. Schiml, and H. Puchta. 2016. "Homology-Based Double-Strand Break-Induced Genome Engineering in Plants." *Plant Cell Reports* 35 (7): 1429–38. https://doi.org/10.1007/s00299-016-1981-3.

Sturme, M.H.J., J.P. Van Der Berg, L.M.S. Bouwman, A. De Schrijver, R.A. De Maagd, G.A. Kleter, and E. Battaglia-De Wilde. 2022. "Occurrence and Nature of Off-Target Modifications by CRISPR-Cas Genome Editing in Plants." *ACS Agricultural Science and Technology* 2 (2): 192–201. https://doi.org/10.1021/acsagscitech.1c00270.

Sudianto, E., S. Beng-Kah, N. Ting-Xiang, N.E. Saldain, R.C. Scott, and N.R. Burgos. 2013. "Clearfield® Rice: Its Development, Success, and Key Challenges on a Global Perspective." *Crop Protection*. https://doi.org/10.1016/j.cropro.2013.02.013.

Tang, X., G. Liu, J. Zhou, Q. Ren, Q. You, L. Tian, X. Xin, … Y. Zhang. 2018. "A Large-Scale Whole-Genome Sequencing Analysis Reveals Highly Specific Genome Editing by Both Cas9 and Cpf1 (Cas12a) Nucleases in Rice." *Genome Biology* 19 (1): 84. https://doi.org/10.1186/s13059-018-1458-5.

Tsai, S Q, Z. Zheng, N.T. Nguyen, M. Liebers, V. V Topkar, V. Thapar, N. Wyvekens, … J.K. Joung. 2015. "GUIDE-Seq Enables Genome-Wide Profiling of off-Target Cleavage by CRISPR-Cas Nucleases." *Nat Biotechnol* 33 (2): 187–97. https://doi.org/10.1038/nbt.3117.

Tsai, Shengdar Q, N.T. Nguyen, J. Malagon-Lopez, V. V Topkar, M.J. Aryee, and J.K. Joung. 2017. "CIRCLE-Seq: A Highly Sensitive in Vitro Screen for Genome-Wide CRISPR–Cas9 Nuclease off-Targets." *Nature Methods*, no. May. https://doi.org/10.1038/nmeth.4278.

Tsai, Shengdar Q., N. Wyvekens, C. Khayter, J.A. Foden, V. Thapar, D. Reyon, M.J. Goodwin, M.J. Aryee, and J.K. Joung. 2014. "Dimeric CRISPR RNA-Guided FokI Nucleases for Highly Specific Genome Editing." *Nature Biotechnology* 32 (6): 569–76. https://doi.org/10.1038/nbt.2908.

Ülker, B., Y. Li, M.G. Rosso, E. Logemann, I.E. Somssich, and B. Weisshaar. 2008. "T-DNA-Mediated Transfer of Agrobacterium Tumefaciens Chromosomal DNA into Plants." *Nature Biotechnology* 26 (9): 1015–17. https://doi.org/10.1038/nbt.1491.

Urnov, F.D., E.J. Rebar, M.C. Holmes, H.S. Zhang, and P.D. Gregory. 2010. "Genome Editing with Engineered Zinc Finger Nucleases." *Nature Reviews Genetics* 11 (9): 636–46. https://doi.org/10.1038/nrg2842.

USA Department of Agriculture. 2020. Ruling on the Movement of Certain Genetically Engineered Organisms. Docket No. APHIS-2018-0034, 85 29790–838.

Vu, T., D. Doan, J. Kim, Y.W. Sung, M. Tran, Y.J. Song, S. Das, and J. Kim. 2021. "CRISPR/Cas-based Precision Genome Editing via Microhomology-mediated End Joining." *Plant Biotechnology Journal* 19 (2): 230–39. https://doi.org/10.1111/pbi.13490.

Vu, T. Van, V. Sivankalyani, E. Kim, D.T.H. Doan, M.T. Tran, J. Kim, Y.W. Sung, M. Park, Y.J. Kang, and J. Kim. 2020. "Highly Efficient Homology-directed Repair Using CRISPR/Cpf1-geminiviral Replicon in Tomato." *Plant Biotechnology Journal* 18 (10): 2133–43. https://doi.org/10.1111/pbi.13373.

Wang, R., G.C. Angenent, G. Seymour, and R.A. de Maagd. 2020. "Revisiting the Role of Master Regulators in Tomato Ripening." *Trends in Plant Science* 25 (3): 291–301. https://doi.org/10.1016/j.tplants.2019.11.005.

Wang, T., J.J. Wei, D.M. Sabatini, and E.S. Lander. 2014. "Genetic Screens in Human Cells Using the CRISPR-Cas9 System." *Science* 343 (6166): 80–84. https://doi.org/10.1126/science.1246981.

Wang, X., M. Tu, Y. Wang, W. Yin, Y. Zhang, H. Wu, Y. Gu, Z. Li, Z. Xi, and X. Wang. 2021. "Whole-Genome Sequencing Reveals Rare off-Target Mutations in CRISPR/Cas9-Edited Grapevine." *Horticulture Research* 8 (1). https://doi.org/10.1038/s41438-021-00549-4.

Wolter, F., and H. Puchta. 2019. "In Planta Gene Targeting Can Be Enhanced by the Use of CRISPR /Cas12a." *The Plant Journal*, tpj.14488. https://doi.org/10.1111/tpj.14488.

Woo, J.W., J.-S.J. Kim, S. Il Kwon, C. Corvalán, S.W. Cho, H. Kim, S.-G.S.-T. Kim, S.-G.S.-T. Kim, S. Choe, and J.-S.J. Kim. 2015. "DNA-Free Genome Editing in Plants with Preassembled CRISPR-Cas9 Ribonucleoproteins." *Nature Biotechnology* 33 (11): 1162–64. https://doi.org/10.1038/nbt.3389.

Xu, W., W. Song, Y. Yang, Y. Wu, X. Lv, S. Yuan, Y. Liu, and J. Yang. 2019. "Multiplex Nucleotide Editing by High-Fidelity Cas9 Variants with Improved Efficiency in Rice." *BMC Plant Biology* 19 (1): 1–10. https://doi.org/10.1186/s12870-019-2131-1.

Ye, X., and P. Beyer. 2000. "Engineering the Provitamin A (β-Carotene) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm." *Science* 287 (5451): 303–5. https://doi.org/10.1126/science.287.5451.303.

Yue, W., Q.L. Ming, B. Lin, K. Rahman, C.J. Zheng, T. Han, and L.P. Qin. 2016. "Medicinal Plant Cell Suspension Cultures: Pharmaceutical Applications and High-Yielding Strategies for the Desired Secondary Metabolites." *Critical Reviews in Biotechnology*. Taylor and Francis Ltd. https://doi.org/10.3109/07388551.2014.923986.

Zetsche, B., J.S. Gootenberg, O.O. Abudayyeh, I.M. Slaymaker, K.S. Makarova, P. Essletzbichler, S.E. Volz, … F. Zhang. 2015. "Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System." *Cell* 163 (3): 759–71. https://doi.org/10.1016/j.cell.2015.09.038.

Zhang, D., H. Zhang, T. Li, K. Chen, J.L. Qiu, and C. Gao. 2017. "Perfectly Matched 20-Nucleotide Guide RNA Sequences Enable Robust Genome Editing Using High-Fidelity SpCas9 Nucleases." *Genome Biology* 18 (1). https://doi.org/10.1186/s13059-017-1325-9.

Zhang, N., H.M. Roberts, J. Van Eck, and G.B. Martin. 2020. "Generation and Molecular Characterization of CRISPR/Cas9-Induced Mutations in 63 Immunity-Associated Genes in Tomato Reveals Specificity and a Range of Gene Modifications." *Frontiers in Plant Science* 11 (February): 1–13. https://doi.org/10.3389/fpls.2020.00010.

Zhang, Q., H.L. Xing, Z.P. Wang, H.Y. Zhang, F. Yang, X.C. Wang, and Q.J. Chen. 2018. "Potential High-Frequency off-Target Mutagenesis Induced by CRISPR/Cas9 in Arabidopsis and Its Prevention." *Plant Molecular Biology* 96 (4–5): 445–56. https://doi.org/10.1007/s11103-018-0709-x.

Zhang, Y., Z. Liang, Y. Zong, Y. Wang, J. Liu, K. Chen, J.L. Qiu, and C. Gao. 2016. "Efficient and Transgene-Free Genome Editing in Wheat through Transient Expression of CRISPR/Cas9 DNA or RNA." *Nature Communications* 7 (1): 1–8. https://doi.org/10.1038/ncomms12617.

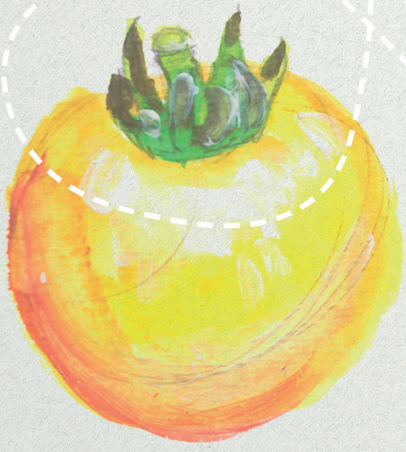
Zhao, H., and J.D. Wolt. 2017. "Risk Associated with Off-Target Plant Genome Editing and Methods for Its Limitation." *Emerging Topics in Life Sciences*. Portland Press Ltd. https://doi.org/10.1042/ETLS20170037.

Zhao, J.H., P. Ho, and H. Azadi. 2011. "Benefits of Bt Cotton Counterbalanced by Secondary Pests? Perceptions of Ecological Change in China." *Environmental Monitoring and Assessment* 173 (1–4): 985–94. https://doi.org/10.1007/s10661-010-1439-y.

Zhou, H., B. Liu, D.P. Weeks, M.H. Spalding, and B. Yang. 2014. "Large Chromosomal Deletions and Heritable Small Genetic Changes Induced by CRISPR/Cas9 in Rice." *Nucleic Acids Research* 42 (17): 10903–14. https://doi.org/10.1093/nar/gku806.

Zhu, H., C. Li, and C. Gao. 2020. "Applications of CRISPR–Cas in Agriculture and Plant Biotechnology." *Nature Reviews Molecular Cell Biology* 21 (11): 661–77. https://doi.org/10.1038/s41580-020-00288-9.

Zuo, Z., and J. Liu. 2016. “Cas9-Catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations.” *Scientific Reports* 5 (October): 1–9. https://doi.org/10.1038/srep37584.



Summary

Summary

CRISPR-Cas mediated genome editing is a powerful tool to induce targeted mutations in plant genomes and can potentially speed up plant breeding significantly. However, studies on the off-target effects of this technique in plants are limited. Therefore, I set out to acquire more data about the specificity of CRISPR-Cas-mediated mutagenesis. I performed this research in tomato, which has all the benefits of a model species, yet is also an important food crop.

In **Chapter 1**, I review the importance, history, and application of genome editing, with a focus on genome editing in plants, ending with an outline of this thesis. In **Chapter 2**, I describe a high-throughput study on CRISPR-Cas9 specificity in tomato. I set up a system for 96 well protoplast transfections, coupled this to next-generation amplicon sequencing, and used this system to map on- and off-target mutations for 89 gRNAs. This high-throughput approach allowed me to explore Cas9 on-target mutations in plants on a larger scale than previously performed. This experiment allowed me to make statements about the predictability of guide efficiency, the mutation pattern of Cas9, and the predictability of the characteristic and prevalent Cas9-induced one-base pair insertion. Additionally, I investigated Cas9 specificity by looking for mutations at 224 predicted off-target sites. At 17 sites, off-target mutations could indeed be identified. Most of these off-target sites contained only a one bp mismatch to the target. These mismatches were found all over the spacer, indicating that a mismatch in the so-called "seed region" proximal to the PAM does not necessarily preclude Cas9 activity. I also found that the nucleotide most distal from the PAM added little to Cas9 specificity. Off-target sites with a mismatch at this position could be mutated at frequencies similar to the target site. No off-target mutations were found at sites that had three or more mismatches to the target site. Importantly, I found that there is also a risk of undesirable mutations at on-target sites: at low frequencies, integrations of plasmid fragments can be found. The use of pre-assembled nuclease-sgRNA complexes (ribonucleoproteins or RNPs) removes this risk but may instead result in the integration of the DNA template used for sgRNA synthesis at DSBs. Although this occurred at very low frequencies, these results indicate that even using RNPs, the risk of integrating foreign DNA in the plant genome is probably never wholly absent. Additionally, although RNPs offer an efficient alternative to plasmids for genome editing in tomato protoplasts, their use will not eliminate off-target mutations.

In **Chapter 3**, I studied a different Cas effector: Cas12a. I first optimized Cas12a crRNA expression systems for use in tomato protoplasts. Then, I aimed to compare Cas12a and Cas9 in terms of efficiency, specificity, and mutagenic spectrum. I performed this by selecting 35 sites with

overlapping Cas12a and Cas9 targets. Again, I used the protoplast-amplicon sequencing combination to screen target sites and 157 predicted off-target sites – 100 for Cas9, and 57 for Cas12a - for mutations. I found Cas12a to be an efficient alternative for Cas9, with a tendency to produce more and larger deletions than Cas9. However, Cas12a induces fewer frameshift mutations than Cas9. These differences and the A/T rich "TTTV" PAM utilized by Cas12a might make Cas12a the more suitable choice for cis-regulatory element editing, whereas Cas9 might be more suited for generating gene knockouts. Although off-target mutations occur for Cas12a, these are predictable and happen mostly when mismatches to the target site occur distal from the PAM. To avoid off-target mutations, the selection of crRNAs that have predicted off-target sites with mismatches in four nucleotides most distal from the PAM should be avoided.

In **Chapters 2 and 3**, I applied prediction-based methods to identify off-target mutations: I predicted off-target sites and inspected only these sites for mutations. In **Chapter 4**, I stepped away from biased screening methods and applied GUIDE-seq to identify off-target double-stranded breaks (DSBs) in tomato protoplasts on a genome-wide scale. This method is based on integrating a short piece of double-stranded DNA of known sequence (a so-called dsODN) in DSBs. By using specific next-generation sequencing techniques, one can retrace where these dsODNs have integrated and, thus, where DSBs have occurred. I first optimized the dsODN integration in tomato protoplasts at known target sites and then continued with unbiased off-target site detection. I did not find any unpredicted off-target DSBs, indicating that these either do not occur or at very low frequencies below 0.7% in protoplast pools. This important conclusion validates off-target prediction software as used in Chapters 2 and 3.

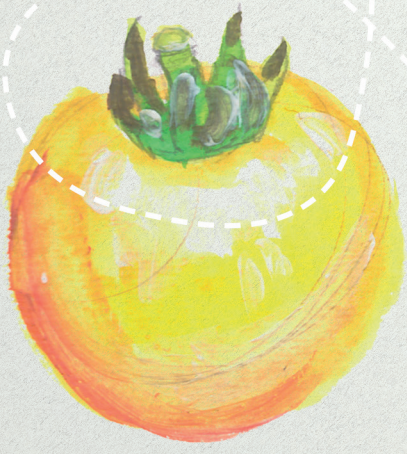
In **Chapter 5**, I moved to unbiased off-target detection in regenerated plants instead of protoplasts. I performed whole genome sequencing (WGS) of 49 tomato plants at ~34x read coverage. Of these, 38 plants were Cas9 genome-edited T_1 plants. These plants were generated using sgRNAs that I had previously identified as being able to induce off-target mutations in protoplasts (**Chapters 2, 4**). As controls, I sequenced six seed-grown wild type cv. Moneyberg plants from two consecutive generations (three plants each) and five T_1 plants that had been through the transformation and tissue culture procedure but had never been exposed to a CRISPR nuclease.

I determined the number of single nucleotide variants (SNVs), indels, and structural variants (SVs) in all plants compared to the reference genome. No significant differences for indel and SV numbers were found between the genome-edited and control groups. No difference in indel size pattern was identified either, indicating that Cas9 did not induce dramatic numbers of *de novo* variation. In the

genome-edited lines, we inspected 29,803 predicted off-target sites with up to six mismatches to the target site for the presence of mutations. This inspection confirmed the presence of mutations at two previously verified off-target sites (**Chapters 2, 4**), both having a single mismatch to the target site.

Additionally, a single potential off-target mutation was found in two plants at a site that had five mismatches to the target site. However, this identified -1 bp deletion might also have resulted from sequencing errors and subsequent erroneous indel calling, as it was present in a homopolymer stretch. This observation illustrates one of the limitations of WGS for unbiased off-target mutation detection. Overall, I conclude that CRISPR-Cas9 genome editing is specific. However, performing truly unbiased off-target mutation detection using WGS data is complicated by the difficulties of assembling repetitive regions, aligning reads to such regions, reliably calling SNVs and indels in such regions, and pre-existing variation within cultivars.

Finally, in **Chapter 6**, I integrated all findings of this thesis and draw conclusions about the specificity of CRISPR-Cas mediated genome editing in tomato. In addition, I discuss the value and importance of (developing) high-throughput research methods in plants, describe the current regulatory status of CRISPR-bred crops in the European Union, and comment on the link between specificity and safety.



Samenvatting

Samenvatting

CRISPR-Cas is een krachtig middel om gerichte mutaties in plantengenomen te induceren en heeft de mogelijkheid om plantenveredeling aanzienlijk te versnellen. Studies naar de off-target effecten van deze techniek bij planten zijn echter beperkt. Daarom wilde ik meer gegevens verzamelen over de specificiteit van mutagenese met behulp van CRISPR-Cas. Ik heb dit onderzoek uitgevoerd in tomaat, wat alle voordelen heeft van een modelsoort, maar ook een belangrijk voedselgewas is.

In **Hoofdstuk 1** geef ik een overzicht van het belang, de geschiedenis en toepassingen van genoombewerking, waarbij ik vooral inga op genoombewerking in planten, en eindig met een outline van deze thesis. In **Hoofdstuk 2** beschrijf ik een high-throughput studie naar de specificiteit van CRISPR-Cas9 in tomaat. Ik zette een systeem op om tomatenprotoplasten te kunnen transfecteren in 96-well formaat, koppelde dit aan amplicon sequencing, en gebruikte deze combinatie om alle on-target en off-target mutaties voor 89 gRNAs in kaart te brengen. Dankzij deze high-throughput aanpak kon ik Cas9-geïnduceerde on-target mutaties op grotere schaal onderzoeken dan eerder voor planten was gedaan. Hierdoor kon ik uitspraken doen over de voorspelbaarheid van de efficiëntie van gRNAs, het mutatiepatroon van Cas9, en de voorspelbaarheid van de karakteristieke, veelvoorkomende één-basepaar inserties geïnduceerd door Cas9. Bovendien onderzocht ik de specificiteit van Cas9 door op 224 voorspelde off-target locaties te kijken of er mutaties hadden plaatsgevonden. Op 17 plaatsen was dit daadwerkelijk het geval. De meerderheid van deze posities verschilde slechts één base met de target site. Deze zogenaamde “mismatches” waren verdeeld over de gehele spacer aanwezig, wat aangeeft dat een mismatch in de zogenaamde “seed region”, dichtbij de PAM, Cas9-activiteit niet onmogelijk maakt. Ik ontdekte ook dat de nucleotide die het verst van de PAM verwijderd is weinig toevoegt aan Cas9 specificiteit. Off-target sites die hier een mismatch hadden, konden gemutageniseerd worden met frequenties die vergelijkbaar waren met de target site. Er werden geen off-target mutaties gevonden op plaatsen die drie of meer baseparen verschilden met de target site. Een belangrijke bevinding was dat er ook een risico op ongewenste DNA-veranderingen bestaat op on-target sites: met lage frequenties kunnen integraties van plasmidefragmenten gevonden worden. Het gebruik van reeds geassembleerde nuclease-gRNA complexen (ribonucleoproteins of RNPs) elimineert dit risico, maar kan in plaats daarvan leiden tot de integratie van het DNA dat gebruikt is voor de synthese van sgRNAs op de plek van de DNA breuk. Alhoewel dit slechts met lage frequenties gebeurde, is dit een indicatie dat zelfs wanneer men RNPs gebruikt, het risico op de integratie van vreemd DNA waarschijnlijk nooit geheel afwezig is. Alhoewel RNPs een efficient alternatief voor plasmiden zijn

om genoombewerking in tomatenprotoplasten uit te voeren, elimineert hun gebruik off-target mutaties niet volledig.

In **Hoofdstuk 3** heb ik een andere Cas-effector bestudeerd: Cas12a. Eerst heb ik Cas12a crRNA-expressiesystemen geoptimaliseerd voor gebruik in tomatenprotoplasten. Vervolgens heb ik Cas12a en Cas9 met elkaar vergeleken wat betreft efficiëntie, specificiteit en mutatiespectrum. Dit heb ik gedaan door 35 locaties te selecteren met overlappende Cas12a- en Cas9-doelwitten. Opnieuw heb ik de combinatie van protoplasttransfectie en amplicon sequencing gebruikt om doelwitsites en 157 voorspelde off-target sites - 100 voor Cas9 en 57 voor Cas12a - te screenen op mutaties. Ik ontdekte dat Cas12a een efficiënt alternatief is voor Cas9, met de neiging om meer en grotere deleties te veroorzaken dan Cas9. Echter, Cas12a veroorzaakt minder frameshift-mutaties dan Cas9. Deze verschillen en de A/T-rijke "TTTV" PAM die Cas12a gebruikt, maken Cas12a mogelijk een meer geschikte keuze voor het bewerken van cis-regulerende elementen, terwijl Cas9 meer geschikt kan zijn voor het genereren van gen knock-outs. Hoewel off-target mutaties voorkomen bij Cas12a, zijn deze voorspelbaar en treden ze voornamelijk op wanneer er mismatches aanwezig zijn ver van de PAM-sequentie. Om off-target mutaties te vermijden, moeten crRNAs die voorspelde off-target sites hebben met mismatched basenparen in de vier nucleotiden het verst van de PAM, worden vermeden.

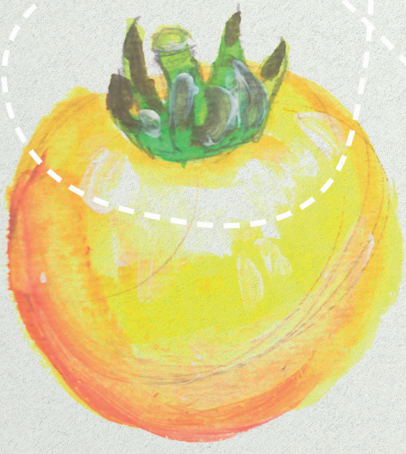
In **Hoofdstukken 2 en 3** heb ik voorspellingsgebaseerde methoden toegepast om off-target mutaties te identificeren: ik voorspelde off-target sites en inspecteerde alleen deze sites op mutaties. In **Hoofdstuk 4** stap ik af van dit soort vooringenomen screeningmethoden en heb ik GUIDE-seq toegepast om off-target dubbelstrengs DNA breuken (DSB's) op genoomwijde schaal in kaart te brengen in tomatenprotoplasten. GUIDE-seq is gebaseerd op het integreren van een kort stukje dubbelstrengs DNA met een bekende sequentie (een zogenaamde dsODN) in DSB's. Door gebruik te maken van specifieke sequencingtechnieken kan men traceren waar deze dsODN's zijn geïntegreerd en dus waar DSB's hebben plaatsgevonden. Ik heb eerst de integratie van deze dsODNs op bekende doelwitsites geoptimaliseerd in tomatenprotoplasten en ben vervolgens doorgegaan met de detectie van off-target sites. Ik heb geen onverwachte off-target DSB's gevonden, wat aangeeft dat deze ofwel niet voorkomen ofwel met zeer lage frequenties onder 0,7% in protoplastenpools. Deze belangrijke conclusie bevestigt de validiteit van off-target voorspellingssoftware zoals toegepast in Hoofdstukken 2 en 3.

In **Hoofdstuk 5** stap ik over op onbevooroordeelde detectie van off-targets in geregenereerde planten in plaats van protoplasten. Ik heb whole genome sequencing (WGS) uitgevoerd op 49 tomatenplanten met een dekking van ~34x. Van deze planten waren er 38 Cas9-bewerkte T1-planten. Deze planten waren gegenereerd met behulp van sgRNA's die ik eerder had geïdentificeerd als in staat om off-target mutaties te veroorzaken in protoplasten (**Hoofdstukken 2, 4**). Als controlegroep heb ik zes zaadgegroeide wildtype cv. Moneyberg-planten uit twee opeenvolgende generaties (drie planten per generatie) gesequenced, en vijf T1-planten die het transformatie- en weefselkweekproces hadden ondergaan maar nooit waren blootgesteld aan een CRISPR-nuclease.

Ik heb het aantal single nucleotide variants (SNV's), indels en structurele varianten (SV's) bepaald in alle planten in vergelijking met het referentiegenoom. Er werden geen significante verschillen gevonden in het aantal indels en SV's tussen de genoombewerkte en controlegroepen. Er werd ook geen verschil in het patroon van indelgroottes geïdentificeerd, wat aangeeft dat Cas9 geen dramatische aantallen *de novo* variaties veroorzaakte. In de genoombewerkte lijnen hebben we 29.803 voorspelde off-target sites met maximaal zes mismatches met de doelwitsite onderzocht op aanwezigheid van mutaties. Deze inspectie bevestigde de aanwezigheid van mutaties op twee eerder geverifieerde off-target sites (**Hoofdstukken 2, 4**), beide met één mismatch met de doelwitsite.

Daarbovenop werd één mogelijke off-target mutatie gevonden in twee planten op een plaats die vijf mismatches had met de doelwitsite. Deze geïdentificeerde -1 bp deletie kan echter ook het gevolg zijn van sequentiefouten en daaropvolgende foutieve indel-aanduiding, omdat deze aanwezig was in een homopolymeerreeks. Dit illustreert één van de beperkingen van WGS voor het onbevooroordeeld detecteren van off-target mutaties. Over het algemeen concludeer ik dat CRISPR-Cas9 genoombewerking specifiek is. Het uitvoeren van echt onbevooroordeelde detectie van off-target mutaties met behulp van WGS-data is echter gecompliceerd. Dit komt door de moeilijkheden bij het assembleren van regio's met herhalende sequenties, het mappen van reads naar dergelijke regio's, het betrouwbaar aangeven van SNV's en indels in deze regio's, en reeds bestaande variatie binnen cultivars.

Tot slot heb ik in **Hoofdstuk 6** alle bevindingen van dit proefschrift geïntegreerd en conclusies getrokken over de specificiteit van CRISPR-Cas-gemedieerde genoombewerking in tomaten. Daarnaast bespreek ik de waarde en het belang van (het ontwikkelen van) high-throughput onderzoeksmethoden in planten, beschrijf ik de huidige regelgevende status van CRISPR-gekweekte gewassen in de Europese Unie en ga ik in op de relatie tussen specificiteit en veiligheid.



Appendices

Acknowledgements

Man oh man, these past five-and-a-half years have been *a ride*. There have been good times, and unfortunately there have also been very bad times, and I'm fairly sure I wouldn't have made it to the end without the support of many amazing, wonderful, talented people that were there for me along the way. Here comes the shoutout you've all been waiting for.

First of all, as is tradition, the supervising team: **Ruud** and **Gerco**. **Ruud**, thank you for all the support, discussions, brainstorming, feedback, *everything*. I have grown a lot these past few years, and for a large part, that's thanks to you. **Gerco**, I think you've heard this before, but apart from being a great scientist, you're also a great group leader. I have always loved the Friday Five O'Clock Beers spirit and I hope you'll keep it up! Importantly, I want to thank **both of you** for being patient with me, giving me time and space when I needed it, and not pushing me at times when work couldn't be my top priority. It really meant a lot to me and made all the difference.

My two paranymphs, **William** and **Xinping**. **William**, you started out as my student, back when I was still a first-year PhD student who didn't really know what the heck she was doing. I'm happy to see that the student has now long surpassed the master. Over time, we became friends, and additionally you became the CRISPR buddy I could always count on to discuss my experiments, marvel about the wonders of Golden Gate cloning, and accompany me to conferences. Lecce was a nice trial run, but the real highlight was of course the CRISPR Frontiers one at CSHL. I truly enjoyed my time with you in New York and I even came up with a mediocre proposition based on that trip (*"visiting a museum without a gift shop is like cooking a meal without salt")*. I can't thank you enough for all the 15:00 o'clock coffee breaks and walkie-talkies - especially during the times of the Big Sad after my mom passed away. I can't believe that these days, we even have a ball for kicking! It's amazing to see you (my former student! I'm like a proud mom) finish your PhD thesis and get ready to start your postdoc. Even though we're in separate countries now (…barely), I'm happy that we can at least still visit conferences together: Porto, here we come. **Xinping**, I appreciate your kind and welcoming personality so, so much. You're very accepting of who people are and I could always just be myself with you. I really loved our Koningsdag trip, when you found the perfect Xinping-sized chair and I the perfect red typewriter that was missing from my collection of… *stuff*. Thank you for the amazing food you cooked for me, the painting we did together, and for allowing me to help you build your IKEA furniture. You have a talent for making people feel at home and at ease and there have been a few times when I really needed that. I'm happy to have you at my side during my defense.

Then, of course, there are all the people that worked with me on this project. To start with: **Michiel**. You're the whole package: the WD40 that keeps the small lab running smoothly, the ducttape that keeps everything from falling apart, and the performer that keeps everyone entertained. You're always willing to help out, troubleshoot experiments, or just listen. Thank you for checking in with me from time to time, thank you for transforming tomato plants for me, thank you for all the minipreps. I hope I chose the right colours for my graphs so that you, too, can enjoy the fruits of our labour. **Jacqueline**, you started on my project later but did an impressive amount of work. Thank you so much for all the tissue culture work, genotyping, and diving into the GUIDE-seq project with me. It's been a pleasure working with you. **Martijn**, you're the reason I still haven't ever sown an *Arabidopsis* in my life. Thank you for taking over the Leiden part of this project, for organising the amazing lab outings together with **Tjitske**, and of course – the fries. **Sven**, I've leaned heavily upon you for the last experimental chapter. Thank you so much for working with me on the whole genome sequencing project – I had fun! **Elio, Linda**, and **Femke**, thank you for helping me out with my sequencing- and bioinformatics-related questions. **Bert, Reza** and **Melissa** – the Leiden part of the project – thank you for the useful and fun discussions we've had over the years.

Then, there were all the students I had the privilege of supervising: **Jop, William** (yes you again), **Jonna, Remi, Rick** and **Lisanne**. Some of you worked on side projects that eventually didn't end up in this thesis – nevertheless, I'm grateful for all the work you did. Also, I'm very proud to see that three of you have found positions within Wageningen University – William and Lisanne as PhDs, and Rick as a technician.

Then, of course, there were all the other colleagues that made my (long) time at PDS a very pleasant one. Starting from the Extended Tomato Group, there's **Vera**, our organizing, scheduling, movie-directing superhero. I started my Master's thesis when you started your PhD and it was great to see your progress. Tomato Group Meetings weren't the same without you. I've enjoyed visiting you in Rotterdam and it's wonderful to see you and Joost with little Ayla. **Rufang**, you were also really there for me from the beginning. Thank you for teaching me how to do tissue culture, thank you for all your support and kindness. I very much enjoyed our walks around the lake behind Lumen and would still love to come visit you in China one day. **Xiaobing**, you always brought a certain *je-ne-sais-quoi* to tomato group meetings that I now miss. Thank you for being my small-lab neighbour. **Julia** and **Jin**, it was very nice to have people working on the same type of experiments as me in the group while you were there – thank you for that. **Iris**, thank you for always inspiring me to give sports another try – your enthusiasm really is contagious. You still might not believe it, but I did actually enjoy those few times we went for a run – although perhaps more for the company than for the

running, I'll admit it. Additionally, thank you very much for organising the Indoor Trampoline event with me and always cheerfully saying "Goeiemorgen Ellen!" when I would arrive. You've been the best "overbuurvrouw" I could ask for. **Gül**, I love the way your brain works and how we can have a lot of fun about nothing in particular. Thank you for the dinners (so happy that we eventually made it to MyAsia!), thank you for the walks, thank you for the 3 AM plant tour. I will call you on banana phone later – don't stay in the lab for too long. **Kai**, thank you for being amused (or at least convincingly pretending to be so) by all my dumb puns, teaching me some Flemish which, in a twist of faith, turned out to be very useful, and – of course – the fries with frikandel speciaal speciaal. But most of all – mercikes for the Googly Eyes. **Victor**, you really have an amazing sense of humour and I'm very happy that we got to see more of it over time. **Xiaowei**, I'm always impressed by the design of your presentations and I can't wait to see what your PhD thesis will look like! **Kailash**, thank you for being my neighbour for approximately three whole days, haha. I hope you've started to settle in the Netherlands by now and will enjoy your time in our group. **Indira**, so nice to see you return as a PhD student! I'm sure you'll be a great addition to the team. **Marian**, you're an amazing scientist and a very compassionate person. Thank you for being a great example. **Chris,** more expertise present during tomato group meetings is always a good thing (especially for finicky RNA extractions from ripe fruits), thanks for that. **Annemarie**, thank you so much for letting me hold your baby Greek tortoises, it really made my day, week, perhaps even month. **Kim**, you're very good at what you do. It always puzzles me how people can keep all those complicated plant hormone pathways straight – but you manage. Thanks for keeping us on our toes. **Lena**, you're impressingly ambitious and always manage to get what you want. Thank you for the amazing trip to Serbia together with **Mengfan**, **Rufang** and **Vera** – I have to admit it feels like ages ago! And, of course – the fries. **Baojian**, you're perhaps the nicest person I've ever met – always willing to help out. Seeing you back at PDS just before submitting my thesis was a nice last-minute surprise! **Charlotte**, we started our MSc theses at PDS around the same time, and later our PhDs. Both your kindness and your level of experimental organization are inspiring. **Mengran**, thank you so much for telling me stories about and showing me pictures of your cats. Life is better with them. **Tjitske**, apart from the lab outings, thank you for helping me out with the confocal. And thanks for organizing the traditional Sinterklaarborrel together with **Mieke**! Mieke, additionally, thank you for bringing the Brabantse gezelligheid to the lab. **Judit**, I don't know where you get all your energy from – but it's much appreciated! The Open Workspace truly is much quieter when you're away. **Amalia**, lady, thank you so much for the pleasant talks. I'm very sorry that the quinceañera never came to be, but feel free to dress up anyway for my defense party. **Suze**, thank you for picking fun technical papers for

Literature Discussion and your enthusiasm for (Friday) beers! **Francesca**, you've been a great neighbour over the years and it was a wonderful experience to wrap up our theses at the same time. Thank you so much for all the pleasant chats. I hope to see you in Leuven! **Marco**, you're part of the team that makes the lab run smoothly, and it's much appreciated. Thank you for always being willing to help out with questions and problems. **Richard**, it's impressive how well aware you are of *everything* that relates to your research – and everyone else's. Thank you for bringing this to our group meetings. **Jan**, thank you for the eLabjournal support and helping me out when I needed 30°C growth cabinets (and immediately short-circuited one when I used it). **Steven**, thank you for being our amazing in-house photographer. Our Brazilian guests: **Tati** and **Ricardo**, thank you for bringing a bit of sunshine, warmth and interesting cachaça to the lab! Of course, I would also like to thank all the other students and visitors who were a part of PDS over the years for making this amazing group what it is.

Although I'm very lucky to have had such amazing colleagues, I was also lucky to have people outside of work that had no idea what I was doing, apart from "something with tomatoes" ("well, you know, actually I don't *really* work with tomatoes, more *tomato cells*…" "yeah whatevs").

Of course, there was my jaarclub (you know, a group of elf, now twelve, people): **Annelou, Lisa, Anouk, Ellen (Kunst), Gregory, Albus, Milco, Nick, Nienke, Sander,** en **Daan**. "Unfortunately", it's not like the old days anymore – we don't meet up every Monday to drink Soup of the Day in Nienke's chicken. I still love going on weekends with you guys – to the Ardennes, to Fryslân -, celebrating our second lustrum (we are old), celebrating 5th of May. The Pur Schnee skiing trip last winter with **Lisa, Annelou, Milco, Marijn, Keje** and **Jaap** really was a highlight for me and made me completely forget about the PhD thesis in quite a stressful time – it was wonderful. Vielleicht kunnen we in die zukunft noch einmal enkele pizza's pakken. All in all, it's amazing to see you grow, branch out, build families, buy houses, have careers. You guys are great.

Then, there are the Tropikallies: **Marijn, Daan, Anouk, Karina, Eeltje, Naomi** (like in the paragraph before, it's very strange to actually use your proper names). We don't see each other very frequently these days, but whenever we're together, it's immediately like we're back in 2014/2015, trying (and succeeding) to convince the members of our student association that the best investment we could possibly do at that moment is to buy a huge palm tree to Tropikallify the garden. And I love it.

Special shoutout to **the Beaver Boys**: Marijn & Milco, aka Albus & Agterberg. I'm very lucky to have deux amis qui sont aussi mes… never mind actually. In any case: avec vous, le temps est toujours bon, except when we go to look at spotted garden eels in Nancy/check out the Schadow-Arkaden in

Düsseldorf/perform sick carves in Austria/do basically anything else et il y a trois jours de neige. Thank you so, so much for all the times you made me feel like I could actually be a fun person when I was deep in the PhD trenches and was feeling like anything but. I hope we'll sing Ameezing karaoke again soon. (Still can't believe I won that karaoke box!)

Of course, there's also my family. First of all, **mama**. I had never expected you not to be there at the finish line – but sometimes, things happen and you have no choice but to deal with it. I know you would have been so proud. I am extremely grateful for your unconditional love and support and will carry that with me for the rest of my life. I will miss you at my defense. **Papa**, it's also your love and support that I can always count on. Thank you for always taking me seriously – for teaching me how to solve equations with two variables in primary school, for building me contraptions to catch bugs, for putting up shelves, for (secretly) fixing my bike and pumping up the tires, for helping me move house (once again). The solid foundation that you and mama provided allowed me to grow into the person I am now. **Luuk**, we've always been very similar – not only in hair- and eye colour, but also in our way of thinking. I love the humour that we share. I've really enjoyed our trips together, remember the one when we went to buy your bonsai tree and we met that whacko employee at the Hornbach? Sometimes I'm scared that I'll never *truly* fit in in this world – but then I'm home again and I realize that at least, I fit in seamlessly with my family. You play a big part in that. **Marjolein**, you're one of the strongest women I know. You're so amazingly creatively talented – she is the one who drew the tomatoes for this thesis, y'all! Apart from drawing, you play guitar (and a couple other instruments) like no other, write your own music, play in a band, 3D animate stuff, *build* your own electric guitars, etc, etc. In other words, you're way cooler than me and I am so, so proud of the person you grew up to be.

And finally, as is also tradition, my partner – **Jaap**. You've been with me for over a decade now and have seen me at both my highest and my lowest – and still decided to stick by me. You truly are my rock – you are the level-headed person I can always count on. You bring me down safely when I fly too high, and you drag me back up when I've dug myself into a deep dark hole. Thank you for always having my back, yet never holding me back. Without you, there would have been many more microwaved meals in my life. I can't wait to see how the next chapter of our lives will turn out.

About the Author

Ellen Slaman (30-05-1993) was born and raised in Zaandam, The Netherlands. One of the things she finds very awkward is having to write an "About the Author" about herself. However, she would probably find it slightly more awkward to have someone else write an "About the Author" about her. So, here we are.

Already from a young age, Ellen has had an interest in biology, which mostly manifested itself as a desire to research and keep exotic pets. She has owned a multitude of stick insects, giant African land snails, garden snails, shrimp, a musk turtle, and regular cats. She has also thought about acquiring tree frogs, dart frogs, hermit crabs, turtles of the land, a dwarf puffer fish, several Coleoptera, and giant millipedes, but has thus far managed to suppress the impulse. Unsurprisingly, she was very pleased with the family subscription to Artis, where visits to the reptile house, aquarium and insectarium were, of course, the main highlights.

Apart from biology, Ellen also enjoyed stories. At four years old, she demanded that her parents taught her how to read because she had decided that this would be an important skill for the rest of her life. (For similar reasons, she insisted on teaching her siblings how to read at ages at which they were perhaps still more interested in drawing on the walls with crayons, and was sorely disappointed whenever they didn't want to "play school" with her.) At five years of age, she wrote her first illustrated flash fiction called "De maakulaar", in which a person buys a house for which they, as it turns out, already own the key. Speaking of plot twists!

After finishing primary school and getting considerably better at spelling, she moved on to Het Zaans Lyceum. Here, she learnt that she enjoys learning about physics, chemistry, mathematics and languages, but most of all biology and especially DNA, and sucks at any sport that involves a ball. When it was time to move on, she briefly considered becoming a medical doctor or studying French, but finally settled on studying Molecular Life Sciences at Wageningen University.

Oh, Wageningen! As it turns out, she loved the University and the place so much, she would remain connected to it for the next decade or so. She spent three years on her Bachelor's Degree, figuring out organic chemistry, physical chemistry, molecular biology, thermodynamics, quantum mechanics (unsuccessfully) and how to combine a full-time course load with a full-time party load. After that, she took a small break to focus on her soft skills and run her student association for a

year, together with six amazing people. In the capacity of "Penningmeester Sociëteit" she was responsible for the most important jobs in keeping such a society running: ordering the beer and other booze, and filing the taxes. She still has fond memories of counting wads of sticky, dirty, sweaty, liquor-drenched cash money (yes, we still used cash money in those days) at 10 am the day after Open Parties. What a time to be alive.

Then, it was time to become a bit more Serious and focus on her Master's studies. Although she started out doing Molecular Life Sciences, she quickly switched to Biotechnology, with a focus on Molecular Biotechnology, as she enjoyed genetic engineering far more than hardcore physical chemistry. She nurtured her love for DNA and genome editing, which culminated in an MSc thesis at the Plant Developmental Biology group, working on developing CRISPR-Cas mediated genome editing in tomato. After an internship at SciENZA in Amsterdam, also working on genome editing, she graduated cum laude – most likely as a result of her newfound Seriousness – and returned to Wageningen and the wonderful research group at PDS to pursue a PhD on the specificity of CRISPR-Cas mediated genome editing in tomato.

The PhD was quite a challenging time for her, with a lot of life happening. At times, she was convinced she would quit the PhD, at other times, she was convinced she would at least quit academia, and then there were times while presenting her research that she felt she could do this for the rest of her life. Funny how that works, isn't it?

After a bit over five years, she eventually managed to finish her dissertation and got to write this "About the Author" statement that is turning out rather long. This dissertation that you have before you right now, a combination of her love of biological sciences and her love of stories, would certainly have made young Ellen very proud. And isn't that what life is all about? Impressing your younger self?

Now, she is a postdoctoral researcher at VIB in Belgium, where she can further combine her love for all things DNA with her love for all things language. You can expect a publication on the difference in codon optimization between Flemish Dutch and Dutch Dutch, as well as insight into their phylogenetic relationships, any time now. Saluutjes!

ellen.slaman@gmail.com
ellen.slaman@psb.vib-ugent.be

Education Statement of the Graduate School EPS

Issued to: Ellen Slaman

Date: 30 August 2023

Group: Plant Developmental Systems (Bioscience)

University: Wageningen University

1) Start- Up Phase	date	cp
▶ **First presentation of your project**		
CRISPR-mediated genome editing in tomato: opportunities and safety aspects	24 May 2018	1.5
▶ **Writing or rewriting a project proposal**		
▶ **MSc courses**		
	Subtotal Start-Up Phase	1.5

2) Scientific Exposure	date	cp
▶ **EPS PhD Student Days**		
Get2Gether	15-16 Feb 2018	0.6
Get2Gether	1-2 Feb 2021	0.4
▶ **EPS Theme Symposia**		
EPS theme 1 symposium, Wageningen	31 Jan 2018	0.3
EPS theme 1 symposium, Leiden	31 Jan 2019	0.3
EPS theme 4 symposium, online	11 Dec 2020	0.2
EPS theme 1 symposium, online	28 Jan 2021	0.2
▶ **Lunteren Days and other national platforms**		
Annual meeting "Experimental Plant Sciences", Lunteren	9-10 Apr 2018	0.6
Annual meeting "Experimental Plant Sciences", Lunteren	8-9 Apr 2019	0.6
Annual meeting "Experimental Plant Sciences", online	12-13 Apr 2021	0.5
Annual meeting "Experimental Plant Sciences", Lunteren	11-12 Apr 2022	0.6
▶ **Seminars, series, workshops and symposia**		
Seminar: "Modern domestication of Pelargonium in a commercial environment", Ronald Snijder, Syngenta	5 May 2018	0.1
Seminar: "What we learnt about auxin transcriptional regulation from meta-analysis of whole genome data", dr. Victoria Mironova	27 Jun 2018	0.1
Seminar: "Transcriptional and epigenetic regulation of STK during flower development in Arabidopsis", Rosanna Petrella	13 Nov 2018	0.1
Seminar: "The role of ALOG family genes in inflorescence patterning in rice and Arabidopsis", Emanuela Franchini	13 Nov 2018	0.1

	Seminar: "Celebrating five years of bioinformatics collaboration at EPS"	10 Jul 2019	0.2
	Seminar: "Reconstructing the ancestral role of flowering pathway genes", dr. Veronica di Stilio	29 Aug 2019	0.1
	Seminar: "Effect of Salt Stress on Root Development", prof. dr. Christa Testerink	29 Sep 2020	0.1
	Seminar: "Seed germination, the role of translational regulation", prof. dr. Leónie Bentsink	1 Dec 2020	0.1
	Seminar: "Cis-regulatory control of quantitative trait variation and pleiotropy", prof. dr. Zach Lippman	9 Dec 2020	0.1
	Seminar: "Accelerating hypothesis testing using RNA viral vectors to deploy synthetic transcription factors in plants", dr. Arjun Khakhar	9 Dec 2020	0.1
	EPS Flying Seminar: "Challenges as a PhD and PostDoc in science and tips to overcome those", Dr Stefan Geisen	20 Jan 2021	0.1
	Workshop: NWO-TTW project workshop "Specificity and side-effects of mutagenesis by CRISPR-Cas in plants"	31 Aug 2022	0.3
	Symposium: "Editing the genome - challenges and opportunities"	14 Mar 2018	0.2
	Symposium: "Food for Future", session Technological Food Innovations	22 Jun 2018	0.2
	Symposium: "Gene Edited Crops; Global Perspectives and Regulation"	10 Oct 2019	0.3
	Resource Debate: "Is CRISPR Safe?"	21 May 2019	0.1
►	**Seminar Plus**		
►	**International Symposia and Conferences**		
	European Plant Science Retreat, Utrecht, NL	3-6 Jul 2018	0.9
	CRISPRcon, Wageningen, NL	20-21 Jun 2019	0.6
	1st PlantEd Conference, Novi Sad, RS	5-7 Nov 2019	0.8
	Solanacea International Online Meeting	9-11 Nov 2020	0.6
	2nd PlantEd Conference, Lecce, IT	20-22 Sep 2021	0.9
	Genome Engineering: CRISPR Frontiers, Cold Spring Harbor Laboratory, New York, US	24-27 Aug 2022	0.9
	Workshop "New Breeding Techniques", Brussels, BE	13 Sep 2022	0.2
	The Road To Sustainability Conference, Brussels, BE	16 May 2023	0.2
►	**Presentations**		
	Poster: "CRISPR-Cpf1 mediated mutagenesis in tomato", European Plant Science Retreat	4 Jul 2018	1.0

Poster: "Optimizing Cas12a-mediated genome editing in tomato", 1st PlantEd Conference	6 Nov 2019	1.0
Poster: "DNA repair outcomes and specificity of SpCas9 and LbCas12a activity in tomato cells", CRISPR Frontiers	26 Aug 2022	1.0
Talk: "High-Throughput Survey of Cas9 off-targets in Tomato", EPS Get2Gether	2 Feb 2021	1.0
Talk: "High-Throughput Survey of Cas9 off-target mutations in Tomato", Annual meeting "Experimental Plant Sciences"	13 Apr 2021	1.0
Talk: "Applying high-throughput technology to identify CRISPR-Cas9 induced off-target mutation in tomato", Lecce, Italy	21 Sep 2021	1.0
Talk: "Applying high-throughput technology to identify CRISPR-Cas induced off-target mutation in tomato", NWO-TTW project Workshop	31 Aug 2022	1.0
Talk: "CRISPR specificity in tomato", VIB, Ghent, BE	4 Apr 2023	1.0
► **3rd year interview**		
► **Excursions**		
Dekker Chrysanten	March 5, 2019	0.2
Online company visit Rijkzwaan	June 16, 2021	0.2
	Subtotal Scientific Exposure	20.1

3) In-Depth Studies	**date**	**cp**
► **Advanced Scientific Courses and Workshops**		
EPS course “Transcription Factors and Transcriptional Regulation”	10-12 Dec 2018	1.0
PE&ERC course “Introduction to R for statistical analysis”	17-18 May 2018	0.6
The Data Carpentries Workshop: Genomics Data	5-6 Feb 2019	0.6
Python for Biologists	Mar 2020	1.0
► **Journal Club**		
Journal club of Bioscience, cluster Plant Developmental Systems	Jan 2018 - 2022	3.0
► **Individual Research Training**		
	Subtotal In-Depth Studies	6.2

4) Personal Development	**date**	**cp**
► **General Skills Training Courses**		
EPS Introduction Course	27 Mar 2018	0.3
WGS course “Scientific Artwork - Vector Graphics and Images”	2-3 Oct 2018	0.6
YoungWUR Workshop “Creating Impactful Infographics & Data Visuals”	24 Nov 2020	0.3
WGS course “Scientific Writing”	Apr - Jun 2021	1.8
WGS Career Assessment	Jun 2021	0.3

▶	**Organisation of meetings, PhD courses or outreach activities**		3.0
	Talk: "CRISPR in Plant Breeding", Pint of Science	22 May 2019	
	Guest lecture: "Learning from Deep Sequencing", Hogeschool Arnhem-Nijmegen	28 Oct 2019	
	Organization of a practical at the EPS CRISPR-Cas workshop	14 Sep 2021	
	Invited panelist at workshop on "New Breeding Techniques", CEJA, Brussels, Belgium	13 Sep 2022	
	Talk: "CRISPR and sustainability", The Road To Sustainability Conference, Brussels, Belgium	16 May 2023	
▶	**Membership of EPS PhD Council**		
		Subtotal Personal Development	6.3

5) Teaching and Supervision Duties		**date**	**cp**
▶	**Courses**		
▶	**Supervision of BSc/MSc students**		3.0
	BSc Project "GFP to BFP conversion in tomato protoplasts", Jop Rang	Apr - Jul 2018	
	BSc Project "The role of methylation on the CNR phenotype of tomato", Jonna van der Putten	Mar - Jul 2019	
	MSc Project "Developing novel crRNA expression systems for CRISPR Cpf1 in tomato protoplasts", William de Martines	Jun - Oct 2018	
	MSc Project: "Comparing Cas9 and Cas12a in terms of efficiency, specificity and mutagenic spectrum", Lisanne Kottenhagen	Jun - Nov 2021	
	HBO Project: "Occurence of Cas9 off-targets in tomato", Rick Dekker	Sep 2020 - May 2021	
		Subtotal Teaching and Supervision Duties	3.0

	TOTAL NUMBER OF CREDIT POINTS*	**37.1**

Herewith, the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS with a minimum total of 30 ECTS credits.

*A credit represents a normative study load of 28 hours of study.

This research was financially supported by the Ministry of Infrastructure and Water Management through the NWO Research Programme Biotechnology and Safety (project number: 15792).

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover Design: Marjolein Slaman and the author

Thesis Layout: the author

Printed by: Proefschriftmaken.nl