FACULTY OF SCIENCES

Plant genome editing by CRISPR-Cpf1 is affected by temperature

Christophe Brugmans
Student number: 01203467

Promoters: Prof. Moritz Nowack and Dr. Thomas Jacobs
Scientific supervisor: Dr. Thomas Jacobs

Master's dissertation submitted to Ghent University to obtain the degree of Master of Science in Biochemistry and Biotechnology. Major Plant Biotechnology

Academic year: 2017 - 2018

UNIVERSITEIT GENT

Ghent University – Department of Plant Biotechnology and Bioinformatics
VIB – Center for Plant Systems Biology
Research Group: Plant genome editing

VIB-UGENT CENTER FOR PLANT SYSTEMS BIOLOGY

Preface

Je voudrais surtout remercier mes parents qui m'ont donné l'opportunité de poursuivre des études supérieures. Sans eux, je n'aurais jamais su atteindre ce stade. Merci pour votre soutien et confiance. J'espère que vous êtes fiers du résultat obtenu grâce à vos efforts.

Mijn beste vrienden Julie Andries, Jolien Bonte, Heleen Coreelman, Magali De Muynck en Simon Plovyt hebben gezorgd voor onvergetelijke tijden op de universiteit. Samen deelden wij de geur van gedissecteerde ratten, de douches door ontkoppelende waterleidingen in labos, de onverwachtse ondervragingen, de chaos voor de vele deadlines en zo verder. Teruglachend naar al de miserie van toen, staan we samen sterk voor de serieuzere problemen die ons te wachten staan.

As for the last phase of this journey, I would like to thank my promoter Thomas Jacobs for all his patient guiding efforts. I was pleased to receive my very own project and appreciated the independence that was given to me. Also I want to thank Debbie Rombaut, Ward Decaestecker, Mansour Karimi and Jonas Blomme not only for teaching me indispensable experimental procedures during my thesis, but also for the pleasant working atmosphere. I enjoyed being part of the group and I hope the group enjoyed my stay as much as I did.

Finally, I would like to thank all the professors, post-docs, assistants and students that contributed to my scientific education.

~ The plant genome editing group ~

Table of contents

List of abbreviations

DSB: double-stranded DNA break

NHEJ: non-homologous end-joining

HDR: homology-directed repair

ZFN: zinc finger nuclease

TALEN: transcription activator-like effector nuclease

CRISPR: clustered regularly interspaced short palindromic repeats

Cas: CRISPR-associated

PAM: protospacer adjacent motif

crRNA: CRISPR RNA

tracrRNA: trans-activating crRNA

gRNA: guide RNA

AsCpf1: Cpf1 from *Acidaminococcus sp. BV3L6*

FnCpf1: Cpf1 from *Lachnospiraceae bacterium ND2006*

LbCpf1: Cpf1 from *Francisella novicida U112*

PDS: *phytoene desaturase*

PPD: *peapod*

GG: golden gate

FAST: fluorescence-accumulating seed technology

gDNA: genomic DNA

RT-PCR: reverse transcription polymerase chain reaction

GFP: green fluorescent protein

NLS: nuclear localization signal

Abstract

In this project a CRISPR-Cpf1 system was evaluated for the induction of targeted mutations in *Arabidopsis* and tomato. Three Cpf1 homologues AsCpf1, FnCpf and LbCpf1 were used and different targets were chosen. Multiplex editing was tested by targeting the two *PPD* genes.

Expression vectors containing a crRNA, a Cpf1 and a FAST cassette were constructed and used for *Arabidopsis* and tomato transformation. Unexpectedly, the *Arabidopsis* T1 events, did not show any bleaching phenotypes consistent with the mutations in the *PDS* gene. Sanger sequencing of these *Arabidopsis* T1 seedlings confirmed there were no mutations present. Only one mutated plant was observed in the genotypically screened tomato plants. Troubleshooting pointed out that Cpf1 mRNA was present in tomato, however western blotting showed no signal of Cpf1 protein presence. BY-2 cells containing GFP-Cpf1 cassettes do not show consistent Cpf1 protein accumulation.

Recent reports indicated that AsCpf1 is temperature sensitive and that higher mutagenesis efficiencies can be obtained in *Arabidopsis* with Cas9 after temperature stress. Therefore, increased temperature intervals were tested during plant growth.

Arabidopsis T2 seeds were screened in high-throughput while implementing four 37°C growth intervals. This gave mutant PDS phenotypes in all T2 lines bearing the same construct with LbCpf1, but only in the heated condition and only for one target sequence. Genotyping results confirmed the functionality of CRISPR-Cpf1 in those lines. Deletions up to ten base pairs were dominating. Further experiments demonstrated that two 37°C growth intervals are sufficient to induce mutations and that 30°C is not.

Using the same heat treatment regime in tomato, plants did not show a convincing increase in mutagenesis efficiency. The dual *PPD* targeting strategy did not seem to work either.

From this project we can conclude that the capacity of LbCpf1 to generate mutations in *Arabidopsis* is affected by heat stress. This finding may be of interest in the plant research field as Cpf1-mediated mutagenesis was not yet reported in *Arabidopsis*, likely due to its relatively low optimal growth temperature. Enabling and optimizing the use of Cpf1 in *Arabidopsis* would not only broaden the CRISPR-Cas targeting range but also give the possibility to create staggered cleavages which could be of use in experiments studying homology-directed repair for genome rewriting.

1. Introduction

1.1 Plant research: gene mutagenesis may reveal its function

1.1.1. Random mutagenesis to create variation

A primitive form of genetics was already practiced over 9000 years ago, when humans first started to domesticate plants (Piperno *et al.*, 2009). Humans selected the plants with the most interesting traits to propagate. When plants are grown in equal conditions, usually (epi)genetic variation is required to be able to notice, compare and select such traits. This variation is naturally available as the existence of different alleles for a specific gene, where one allele may be better than the other depending on the cultivation purpose. Just as 9000 years ago, researchers now still compare plants carrying different alleles, but now not always for agricultural purposes. Plant research may help optimize crops, but also contribute to the discovery of potentially interesting bioactive molecules for medicine (*i.e.* artemisinin (Su and Miller, 2015)), molecular mechanisms that could serve as basis or inspiration for new (bio)technology tools (*i.e.* bioremediation and solar panels), and other possible values for society.

Plant researchers had to rely on natural alleles that resulted in a plant with a different phenotype. These natural alleles may be caused by incorrect replication, transposable elements (Bai *et al.*, 2007), etc. and could thus not directly be controlled by the researcher. As searching for plants with a remarkable phenotype in nature is time-consuming, researchers had to find different approaches to collect investigable plant material. This initiated the use of mutagens on plants. By exposing plants to physical or chemical mutagens, for example X-rays and ethylmethanesulfonate (EMS) respectively, random mutations are introduced in the plant genome (Neuffer and Coe, 1978; Stadler, 1928). These mutations can cause aberrant phenotypes to investigate using forward genetics to map the altered gene. Deficient genes showing no visible phenotype (e.g. altered metabolism pathways) require high-throughput molecular screening methods like RNA expression profiling, metabolome profiling etc., which have only recently become available, making changes in plant behavior much easier to assess today.

The fast and reliable new generation sequencing technology dramatically lowered the sequencing costs, which made researchers start to sequence many organism genomes. Now different genes can be predicted on these sequenced genomes, but the function of most of those genes remains unknown. However, this data unlocked the reverse genetic approach: a candidate gene is inactivated and the phenotype is investigated. Then arose the question how to inactivate a certain gene in a plant. T-DNA insertion mediated by *Agrobacterium* is, besides a way of introducing exogenous genes in plants (Schell and Van Montagu, 1977), also a way to inactivate endogenous genes by interrupting them. Unfortunately, T-DNA insertion is essentially random (Li *et al.*, 2006), making it difficult to get a desired knock-out. However, if the plant genome is known, the T-DNA insertion locus can be identified by sequencing outwards of the T-DNA borders. All these lines can be kept in a library for when a researcher desires to investigate a specific knock-out (Krysan *et al.*, 1999). If a researcher needs a knock-out that is not available, a problem arises.

1.1.2. Targeted mutagenesis

Specifically knocking out a gene can be accomplished using engineered endonucleases. Endonucleases recognize a specific sequence in the genome and induce a double-stranded DNA break (DSB). This DSB triggers the plant DNA repair system, briefly depicted in Figure 1.1.1. In plants, the primary means of repair is via the non-homologous end joining (NHEJ) mechanism, that ligates the two DNA ends back together (Rinehart *et al.*, 1997). But this often leaves indels behind due to the error-prone property of NHEJ (Gorbunova and Levy, 1997). If an indel is made in a gene it often alters the reading frame for protein translation, which often results in a premature translation stop codon creating non-functional proteins. Homology directed repair (HDR) is a different and rarer repair mechanism that occurs only when template is available homologous to the DNA ends. This may evoke the invasion of the homologous template by a strand of the DSB end, whereafter replication is done based on the homologous template to repair the break (Puchta and Hohn, 1991).

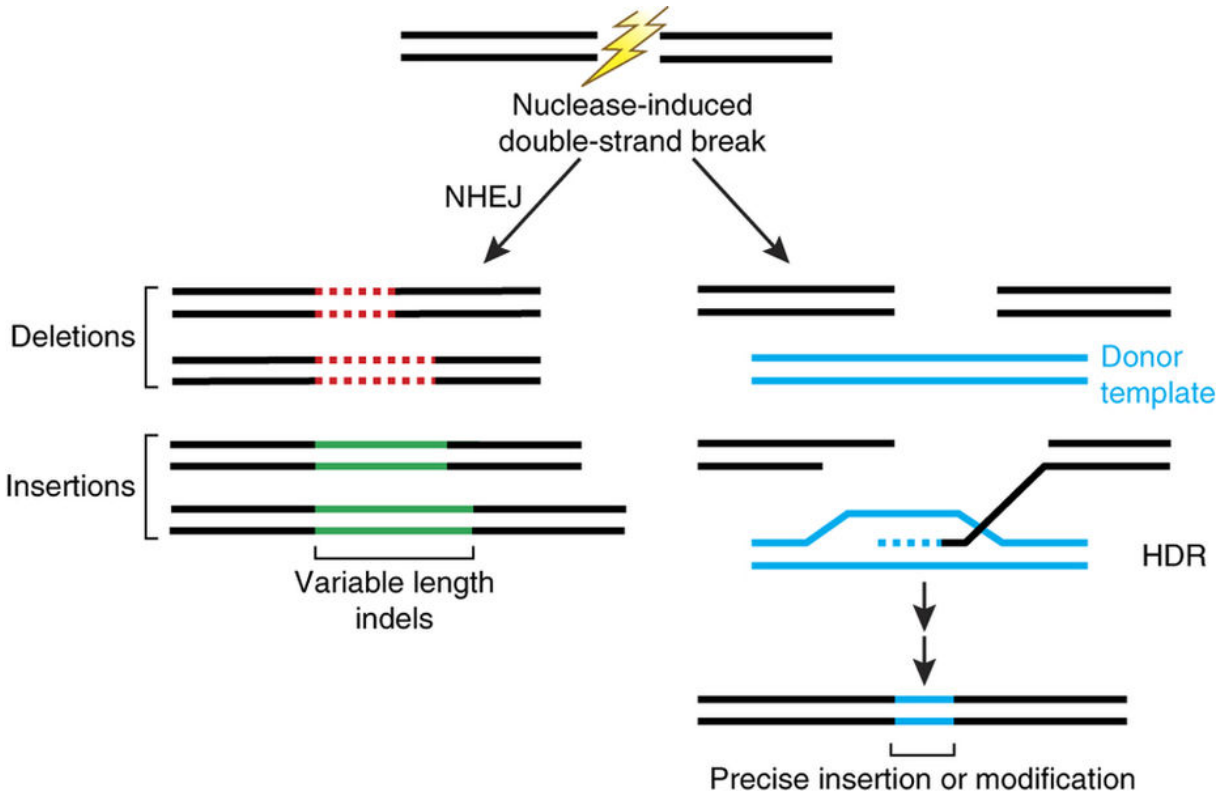

Figure 1.1.1: Repair mechanisms following a double stranded DNA break. Endonucleases induce a targeted double stranded break (DSB). In most cases the DSB is repaired by the nonhomologous end joining (NHEJ) pathway which is error-prone, leaving small indels of variable length at the DSB site. In some cases if a homologous donor template is provided, the homology-directed repair (HDR) mechanism makes one broken strand invade the provided DNA, using it as repair template. Figure taken from Sander & Joung, Nature biotechnology 2014.

Most endonucleases recognize a specific four to six base pair palindromic DNA sequence, and induce a DSB within or adjacent the recognition site. However, such small recognition sequences cannot be used genome wide, as the chance of finding such sequences by chance is high and would generate a fragmented genome, leading to lethality. Meganucleases recognize larger sequences (12 to 40 base pair long) and can be used to induce only one DSB in the genome (Puchta, 1999). Unfortunately, there are not enough meganucleases to cover all possible sequences. Some successful attempts were done to alter the recognition sequence (Seligman *et al.*, 2002), but the process is not straightforward and thus very time-consuming, limiting the possible targets.

Much more modular tools are the zinc-finger and the transcription activator-like effector nucleases (ZFNs and TALENs) (Kim *et al.*, 1996; Miller *et al.*, 2011). These tools consist of a sequence-specific DNA recognition protein linked to a FokI endonuclease as shown in Figure 1.1.2. The DNA recognition protein is made of different modules. In ZFNs, each protein module recognizes nucleotide triplets, while each module in TALENs recognizes single nucleotides. The DNA recognition protein binds the DNA sequence for which it was designed and brings the FokI nearby that sequence. However, the FokI endonuclease is only able to induce DSB when homodimerized. Therefore a second ZFN or TALEN must be designed to bind the DNA nearby the first FokI. Homodimerized FokI introduces a DSB between the two bound sequences (the spacer sequence). This makes it possible to induce DSB very specifically and showed very efficient in plants (Ma *et al.*, 2013; Zhang *et al.*, 2010).

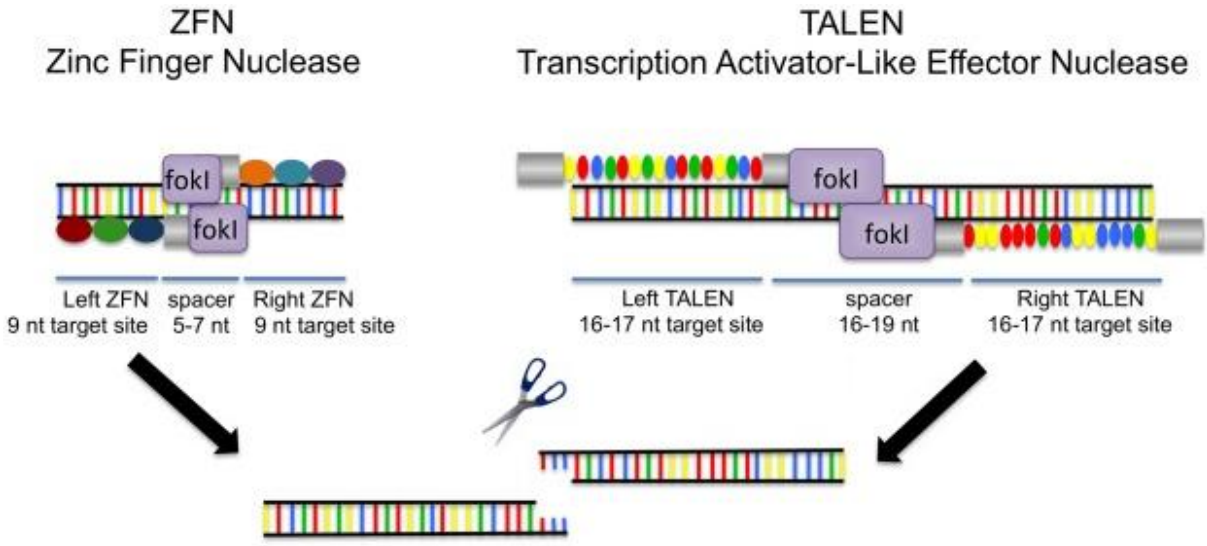

Figure 1.1.2: Schematic representation of zinc finger nuclease (ZFN) and transcription activator-like effector nuclease (TALEN) creating a double stranded DNA break (DSB). FokI nucleases are guided by a chain of DNA-binding protein modules. Each module recognizes three (ZFN) or one (TALEN) nucleotide. Dimerized FokI creates a DSB within the spacer region. Figure adapted from Moore *et al.*, PLoS One 2012.

An important breakthrough was reached in 2012, when researchers managed to program a prokaryotic adaptive immune system to induce targeted DNA breaks. As the mechanism of base-pair targeting and its development was much more simple compared to the DNA-recognizing proteins ZFNs and TALENs, it quickly gained popularity.

1.2. The CRISPR locus

1.2.1. From confusing sequence structures to an adaptive immune system

In 1987, Japanese researchers published the sequence of the *iap* gene and its flanks from *Escherichia coli* (Ishino *et al.*, 1987). The researchers remarked an unusual sequence structure at the 3'-end of this *iap* gene, described as five highly homologous sequences of 29 nucleotides arranged as direct repeats with 32 nucleotides as spacing. Unfortunately, they did not investigate it any further. The structure was not being reported further, until 1993, where such palindromic repeat sequences again caught attention. Dutch researchers found these repeat-spacer sequence motifs in *Mycobacterium tuberculosis*, and developed a spacer oligotyping method for *M. tuberculosis* strain differentiation, called spoligotyping (Groenen *et al.*, 1993; Kamerbeek *et al.*, 1997). While in the same year in Spain, similar genetic patterns were found by Mojica in the genome of archaea organisms and finally was decided to investigate the biological role of these structures (Mojica *et al.*, 1993). Knocking out the palindromic repeats, Mojica suggested that these had a role in correct DNA segregation during the cell cycle, but Mojica himself was not convinced (Mojica *et al.*, 1995). The following years, reports describing this structure accumulated and by the year 2000 Mojica *et al.* found these in 20 different prokaryotic species (Mojica *et al.*, 2000). To avoid confusion due to the different names given, these similar structures in phylogenetically distant prokaryotes were named CRISPRs, standing for the clustered regularly interspaced short palindromic repeats (Jansen *et al.*, 2002).

As CRISPRs were found so frequently in prokaryotes, it was reasoned that they must have an important biological role, which was still being questioned. The close linkage of CRISPR loci with sets of homologous genes, named *cas* genes (CRISPR associated) was remarked, but still no function could be attributed (Jansen *et al.*, 2002). In 2005, a key clue was unveiled by three different research groups: some spacer sequences matched with the sequence of bacteriophages and the spacers would have an extrachromosomal origin (Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Pourcel *et al.*, 2005). These groups hypothesized that CRISPR loci played a role in adaptive immunity in prokaryotes. Mojica *et al.* proposed that RNA transcripts of the CRISPR spacers recognized targets in the same way that eukaryotic cells do with the RNA interference system. Based on Mojica's proposal, numerous *cas* gene protein sequences were investigated and comparative analysis indeed revealed functional analogies between some Cas proteins and the proteins involved in the eukaryotic RNAi mechanism (Makarova *et al.*, 2006).

Experimental evidence confirmed that CRISPRs provided acquired resistance against viruses in prokaryotes (Barrangou *et al.*, 2007). This was done by exposing susceptible *Streptococcus thermophilus* bacteria to bacteriophages, generating phage-resistant mutants. Comparison of the CRISPR loci showed acquisition of spacer sequences in the CRISPR locus, matching parts of the genomes of the phages used. Afterwards, by adding and deleting spacers in the CRISPR locus of *S. thermophilus*, the changing survival outcomes confirmed the hypothesis that the CRISPR locus confers viral resistance in prokaryotes.

1.2.2. Elucidating the CRISPR-Cas mechanism

Bolotin *et al.* investigated some of the *cas* genes and found that Cas5 carried a HNH motif, present in various nucleases and thus suggesting its role as an endonuclease. They also described a five nucleotide conserved sequence which was always present near the extrachromosomal target sequence (Bolotin *et al.*, 2005). Three years later, this was demonstrated to be an essential element to localize the spacer in the phage genome to provide immunity (Deveau *et al.*, 2008). However, within this five nucleotide motif, it was found that a smaller conserved sequence was the real requirement to produce a spacer and this conserved sequence depends on the CRISPR locus, and thus on the CRISPR-Cas system (Horvath *et al.*, 2008). This CRISPR-type specific motif has been named the protospacer adjacent motif (PAM) and were mainly conserved di- or trinucleotides (Mojica *et al.*, 2009).

Kunin *et al.* compared the conserved CRISPR repeats from 195 bacterial genomes and saw that these repeats were not identical between genomes, but they could be clustered based on sequence similarity, suggesting the presence of CRISPR-Cas subtypes (Kunin *et al.*, 2007). Also they concluded that secondary structures in the repeats are evolutionary stable by remarking multiple compensatory base changes, which may indicate conserved binding sites.

A complex of five Cas proteins was identified in *E. coli*, called Cascade, which seemed necessary for the antiviral response (Brouns *et al.*, 2008). By cutting within each CRISPR repeat, the Cascade complex stands in for the production of individual CRISPR RNAs (crRNAs), which consists of spacers flanked by remainders of the cut CRISPR repeats. As suggested by Kunin *et al.*, these flanks indeed serve as binding sites for the Cascade complex. The bound crRNA then guides the complex towards viral nucleic acids, causing their cleavage. By testing an antisense spacer sequence, they showed that targeting the antisense strand was also possible to acquire resistance against the virus, suggesting that the Cascade-crRNA complex targeted double-stranded DNA and thus not the widely hypothesized mRNA. This was confirmed some months later by a Marraffini and Sontheimer that investigated a clinical isolate of *Staphylococcus epidermidis*, where a CRISPR spacer matched the *nickase* gene present in most staphylococcal conjugative plasmids and thus prevented conjugation (Marraffini and Sontheimer, 2008). Not only did they confirm that the CRISPR system targeted DNA, but they were the first researchers mentioning about using this system as a restriction enzyme that can be programmed by a suitable effector crRNA.

Further investigation on this system by Garneau *et al.* demonstrated that cleavage happened only if the *cas5* gene was functional and within the protospacer three nucleotides upstream the PAM sequence, generating a DSB with blunt ends (Garneau *et al.*, 2010). Also did he remark that cleavage happened at the same position when using a 29 nucleotide 5'-end truncated version of a protospacer comparing to the full 30 nucleotide protospacer, suggesting that the CRISPR-Cas system would use the 3'-end as anchoring point.

As the mechanism became clearer, it was still unclear how prokaryotes acquired spacers. Comparative genomics on the CRISPR locus pointed that the *cas1* and *cas2* genes were the signatures for CRISPR identification, indicating an important role for these genes. However, it made no part of the Cascade complex and had no role in CRISPR processing and interference (Brouns *et al.*, 2008). By studying the crystal structure of Cas1 and identifying it as a DNase, Wiedenheft *et al.* suggested a role in recognition, cleavage and integration of foreign DNA during the earlier steps of the immune system (Wiedenheft *et al.*, 2009).

Experimental evidence demonstrated later that Cas1 and Cas2 were the only Cas proteins necessary for spacer acquisition (Datsenko *et al.*, 2012; Yosef *et al.*, 2012). First functional insights were elucidated in 2014: the Cas1 complexes with Cas2, and is only then capable of recognizing the spacer integration site at the end of the leader sequence in the CRISPR locus (Nunez *et al.*, 2014). Furthermore, the complex provides correct orientation to catalyze the consecutive nucleophilic attacks of the protospacer on the flanks of the direct repeats as depicted in Figure 3 (Nunez *et al.*, 2015).

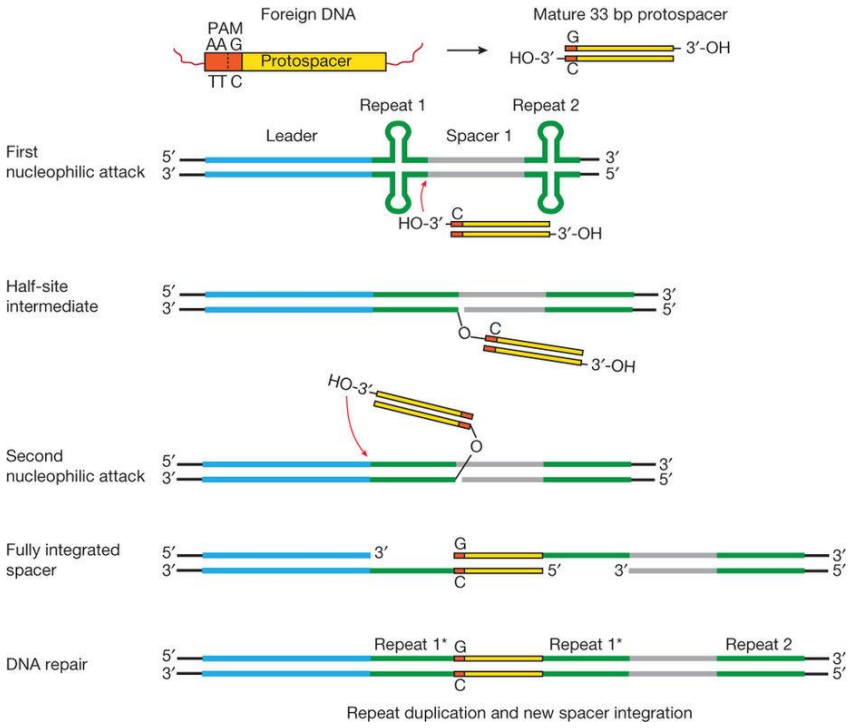

Figure 1.2.1: Model of molecular protospacer integration mediated by the Cas1-Cas2 complex during the CRISPR adaption phase. The 3'-OH end of the minus strand of the mature protospacer makes a nucleophilic attack on the CRISPR minus strand right after the first repeat, creating a half-site intermediate. After a second nucleophilic attack of the 3'OH end of the protospacer plus strand on the CRISPR plus strand right after the leader sequence, it creates ssDNA gaps which gets repaired. Spacer integration caused the duplication of the first repeat, marked with asterisks. Figure taken over from Nuñez *et al.*, 2015.

Spacer acquisition thus happens at the position directly downstream the CRISPR leader sequence independently of the amount of spacers. This confirms that the CRISPR locus represents a chronological record of prior infections, making the most downstream spacers the most ancient infection cicatrices (Sorek *et al.*, 2013).

1.2.3. Types of CRISPR-Cas systems

CRISPRs became more complex as research continued on increasing amounts of prokaryotes, requiring a CRISPR classification system. In 2002 the first four Cas protein families (homologues) were known (Jansen *et al.*, 2002). This number quickly increased to 45 (for only the first 200 completed prokaryotic genomes) and a phylogenetic-based subdivision was proposed (Haft *et al.*, 2005). A team of CRISPR pioneers decided to make a unified classification and nomenclature system for CRISPR-Cas, defined as polythetic due to the combined information from phylogenetic and comparative genomic analyses (Makarova *et al.*, 2011). This made the system divide into three major groups (Type I, II and III) but it was still excluding a significant amount of identified systems. A final revised classification was adopted by combining the analysis of signature *cas* proteins and CRISPR-Cas loci architecture (Makarova *et al.*, 2015). This is based on linking the presence of certain signature genes with certain CRISPR locus types.

First, it divided the CRISPR systems into two classes: multisubunit crRNA-effector complexes as class 1 and large single multidomain effectors as class 2 CRISPR-Cas systems. Second, there are five types of systems across these two classes. These types are compared on the required genes to fulfill the whole function of CRISPRs, subdivided into adaptation, expression, interference and eventual ancillary functions. A summary of the CRISPR types and structure is shown in Figure 1.2.2.

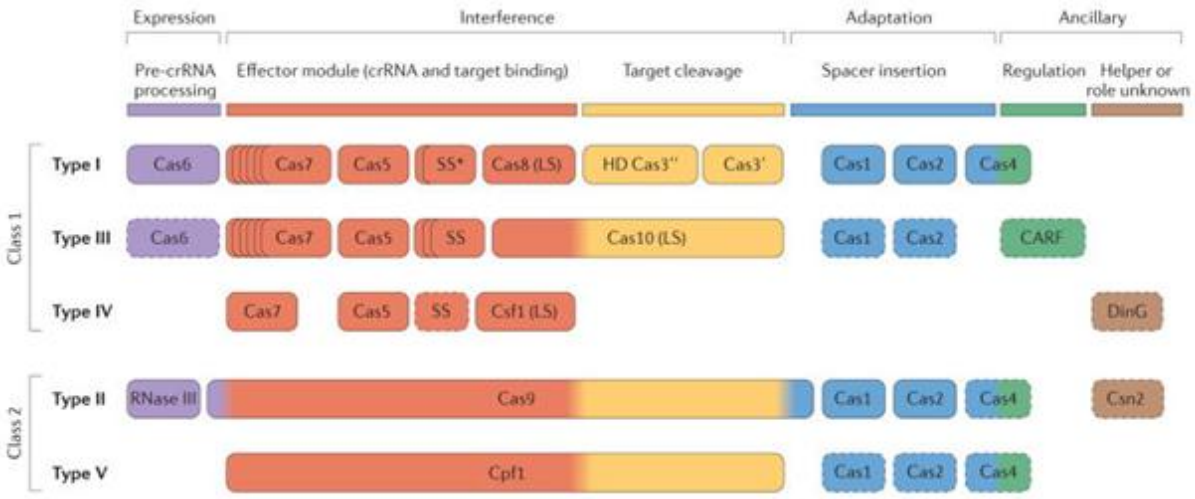

Figure 1.2.2: Functional classification of Cas proteins. Described earlier investigated Cas proteins and their contribution in the different stages of CRISPR-Cas immunity. Dashed outlines indicate dispensable components. Figure taken over from Makarova *et al.*, 2015.

As can be derived from Figure 1.2.3, the type I systems are the most ubiquitous and are defined by presence of the *cas3* gene, necessary for formation of the Cascade complex. Most type I systems have a single gene of all the *cas* gene families (*cas1 to cas8*). The less common type III systems have the *cas10* as signature gene. A putative type IV system was proposed that consisted of partially degraded *cas* genes less linked to CRISPRs. Type II systems have *cas9* as signature gene, which is a large multidomain protein capable of accomplishing multiple functions. Lastly, the type V system has the unique *cpf1* gene, but often unlinked with the CRISPR locus in genomes, rendering it also a putative system.

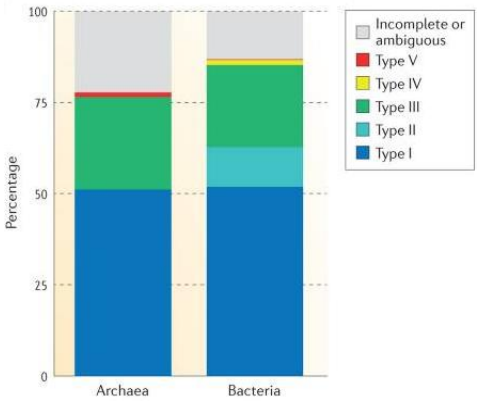

Figure 1.2.3: Distribution of CRISPR-Cas types amongst prokaryotic genomes. Figure taken over from Marakova *et al.*, 2015.

A computational pipeline has been constructed and used on all available prokaryotic genomes, which identified even more new (sub)types (Koonin *et al.*, 2017). One of these is a newly predicted type VI system that only targets RNA, in contrast with some type III systems that are capable of targeting both RNA and DNA (Samai *et al.*, 2015). This type VI interference was validated experimentally (Abudayyeh *et al.*, 2016; Smargon *et al.*, 2017).

It was still impossible to classify all identified CRISPR-Cas loci. But for now, unclassifiable systems would be limited to an insignificant amount.

1.3. Towards the genome editing tool CRISPR-Cas9

1.3.1. A type II CRISPR-Cas system

Further research was done on type II systems due to its simplicity compared to others. In 2010, while analyzing RNA of the human pathogen *Streptococcus pyogenes*, Deltcheva *et al.* found the tracrRNA which is a trans-encoded small RNA with 24 nucleotide complementarity to the CRISPR repeat regions (Deltcheva *et al.*, 2011). This tracrRNA seemed required for crRNA maturation with the help of RNaseIII and a protein Csn1, while no other reported Cas endoribonucleases (*cas6*) were detected in the genome. The use of a host factor RNaseIII which was not directly linked with the CRISPR locus suggested a new maturation mechanism which is now part of the type II CRISPR-Cas systems. They saw that Csn1 prevented RNA degradation and suggested that it helped duplexing between pre-crRNA and tracrRNA, facilitating cleavage by RNaseIII.

Around the same time, Sapranauskas *et al.* managed to express successfully a type II CRISPR-Cas system from *Streptococcus thermophilus* in *Escherichia coli*, providing heterologous protection against invading DNA (Sapranauskas *et al.*, 2011). They also concluded that their Cas9 protein was sufficient to achieve CRISPR interference. Grouped in the cluster of orthologues group COG3513, the described Csn1 and Cas9 proteins were actually orthologues of the earlier reported Cas5 protein by Barrangou *et al.* and was predicted to be a large multi-domain effector protein family required for CRISPR functioning (Barrangou *et al.*, 2007; Makarova *et al.*, 2006).

Jinek *et al.* further investigated the Cas9 protein from *Streptococcus pyogenes* and showed that SpCas9 cleaved the targets, guided by the both required tracrRNA and crRNA (Jinek *et al.*, 2012). They confirmed the blunt end-cutting property of SpCas9 in the target DNA, and added that each strand was nicked by a different nuclease domain: the HNH and RuvC-like domain nicked respectively the complementary and the noncomplementary crRNA sequence on the target DNA. They also reported the required PAM sequence for SpCas9, consisting of a NGG sequence directly downstream of the crRNA matching sequence on the target DNA. Finally, they proposed using a single guide RNA (gRNA) that replaced the tracrRNA-crRNA duplex, describing the gDNA as programmable to cleave any DNA of interest by just changing the DNA target-binding sequence. This reduced the CRISPR-Cas9 system into a simple two-component system, requiring only the Cas9 protein and the gRNA. However, these experiments were done *in vitro*. The replacement of the crRNA-tracrRNA duplex by a single guide RNA was the last major change to the CRISPR-Cas9 tool as we know now and is summarized in Figure 1.3.1.

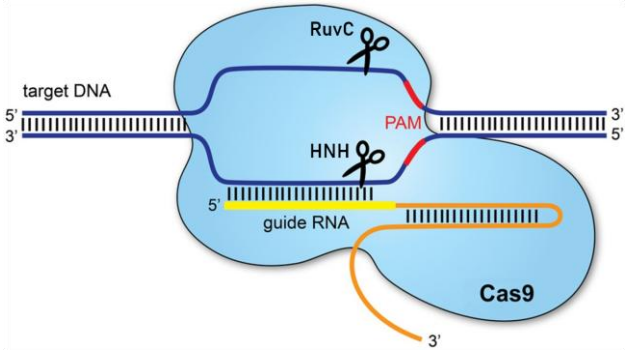

Figure 1.3.1: The CRISPR-Cas9 tool. The guide RNA is a fusion of the crRNA (yellow) and tracrRNA (orange) and can be bound to the Cas9 protein to guide the protein towards the genomic DNA (blue). Cas9 induces a double stranded blunt DNA break with its HNH and RuvC domains if the crRNA is matched with the genomic DNA and if this sequence match is directly followed by the PAM sequence (red). Figure adapted from Redman *et al.*, Education and practice 2016.

1.3.2. Genome editing capacities of CRISPR-Cas9

It did not take long to initiate the CRISPR report flood that started from 2013 on. In less than one year, successful targeted genome editing was reported *in vivo* in many eukaryotic organisms, including human cells, yeast, and plants (Cong *et al.*, 2013; DiCarlo *et al.*, 2013; Jiang *et al.*, 2013; Mali *et al.*, 2013). Different promoter systems for CRISPR-Cas9 expression were compare to optimize the efficiency of CRISPR-Cas9 and mutational frequencies reached up to 100% in plants (Tang *et al.*, 2016; Xie *et al.*, 2015).

One of the unique added values of the system was highlighted: the multiplexing capacity, giving the possibility to target multiple genes at the same time, being a natural feature of CRISPRs (Cong *et al.*, 2013). Multiplexing with ZFNs and TALENs require multiple huge protein constructs that need to be encoded by presumably more than one plasmid, which furthermore require restructuring of the protein module sequence for each new target which is time-consuming compared to the inexpensive purchase of a 20-24 nucleotide guide sequence. CRISPR-Cas9 also has a fixed cleaving site, while TALENs usually cleaves somewhere nonspecifically in the linker region between the TALEN monomers, making cleavage site prediction less precise (Miller *et al.*, 2011).

On the other hand, caution arose when a high frequency of CRISPR-Cas9 induced off-target mutations was described (Fu *et al.*, 2013). But this was caused by primitive imprudent choice of target sequences by ignoring the background genome and target mismatch tolerations. Therefore web-based software tools were developed for a better target prediction minimizing off-target mutagenesis (Hsu *et al.*, 2013). Later, further studies minimized undesired off-target effects of the native Cas9 protein, increasing the specificity of CRISPR-Cas9 (Bayat *et al.*, 2017; Doench *et al.*, 2016; Kleinstiver *et al.*, 2016a; Slaymaker *et al.*, 2016).

Initially used for targeted mutagenesis, CRISPR-Cas9 has been adapted to go beyond these borders. By mutating one nuclease domain the Cas9 becomes a nicking enzyme, which can be used with pairs of gDNA to generate a DSB with a much higher specificity, avoiding even more off-target mutations (Ran *et al.*, 2013). By mutating both nuclease domains of Cas9, dead Cas9 (dCas9) was created, still able to bind DNA in a targeted way, but could not cut the target anymore, giving the possibility to control the expression of target genes by recruiting, for instance, chromatin remodeling proteins (Bikard *et al.*, 2013). Furthermore, a cytidine deaminase can be coupled to dCas9 to change cytosines into thymidines in close proximity of the dCas9 target (Komor *et al.*, 2016). The latter has recently been tested in our group and confirmed the successful results. This demonstrates the polyvalence of the Cas9 protein for different genome editing purposes.

Lastly, true genome editing lays in the precise deletion, change or addition of sequences at a specific target. As described earlier, the usual outcome of CRISPR-Cas9 DSB is a deficient gene due to the frame-shifting NHEJ mechanism. But in some cases, HDR is done based on a provided template, giving the possibility to accurately integrate sequences of interest in a targeted manner. This was proved to work in plants, however the efficiency of such targeted integration is low, ranging from 0,1% to 10% depending on the methodology and plant species (Bortesi *et al.*, 2016). This does not limit the CRISPR-Cas9 tool for only research purposes, but could be expanded to medical therapies where precise genome editing may

be required. As this is also applicable on plants, the tool can furthermore be used for crop optimization in a cleaner way than *Agrobacterium*-mediated insertion of genes.

Unfortunately, CRISPR-Cas9 has an important shortcoming compared to TALENs when it comes to the targeting range. The required NGG sequence established a limit to the targeting range of Cas9. Alternative PAMs were tolerated by mutating the PAM recognition protein domain of Cas9 (Kleinstiver *et al.*, 2015). Recently xCas9, a new SpCas9 variant obtained by directed evolution, showed the capacity to recognize a broader range of PAMs including NG (Hu *et al.*, 2018). Due to these continuous improvements, researchers tend to focus on the Cas9 system. However, also other CRISPR-Cas mechanisms could have the potential to become a potent genome editing tool and could further increase the applications of CRISPR-Cas.

1.4. The debut of CRISPR-Cpf1

1.4.1. The type V CRISPR-Cas system has different features

A new Cas protein was found using bioinformatics analysis on the protein family database TIGRFAM (Haft *et al.*, 2003). TIGR04330 represented Cpf1 (also known as Cas12a), standing for the CRISPRs from *Prevotella* and *Francisella* 1, derived from the bacterial genomes in where it has been identified. Using CRISPRFinder (Grissa *et al.*, 2007) software, putative CRISPR systems were found in genomic sequences of *F. novicida* (Schunder *et al.*, 2013). Further *in silico* investigation of these loci suggested active acquisition of spacer sequences, indicating that *F. novicida* contains functional CRISPR-Cas systems.

The Cpf1 protein family consists of at least 16 reported members (Zetsche *et al.*, 2015). Three amongst them, the Cpf1 proteins from *Acidaminococcus sp. BV3L6* (AsCpf1), *Lachnospiraceae bacterium ND2006* (LbCpf1) and *Francisella novicida U112* (FnCpf1) are the main Cpf1 proteins being investigated. FnCpf1 is intuitively the most interesting Cpf1, as the PAM sequence is likely TTN resulting in broader targeting opportunities compared to the TTTV of AsCpf1 and LbCpf1.

Cpf1 distinguishes from Cas9 by some major characteristics summarized in Figure 1.4.1. First, the required PAM sequence to allow cutting by Cpf1 is (T)TTV, directly upstream of the target. This T-rich PAM highlights the importance of Cpf1 as genome editing tool complementary to Cas9, as AT-rich sequence are mostly found in intergenic regions, while intragenic regions are GC-favored (Pozzoli *et al.*, 2008). Secondly, no tracrRNA is required and Cpf1 has next to its DNase activity, also a RNase activity which makes it possible to process transcribed CRISPRs on its own (Fonfara *et al.*, 2016; Zetsche *et al.*, 2015). This could facilitate multiplex editing as only one transcript would be required for multiple targets, instead of using separate PolIII promoters for gRNA expression or gRNAs flanked by ribozyme sequences as used for Cas9 (Tang *et al.*, 2016). Lastly, Cpf1 is capable of inducing staggered DSBs, with 5' overhangs of 4-8 bases (Lei *et al.*, 2017). This last feature may attract a lot of interest as one could use Cpf1 as a very precise cloning enzyme to perform targeted integrations once it becomes very efficient.

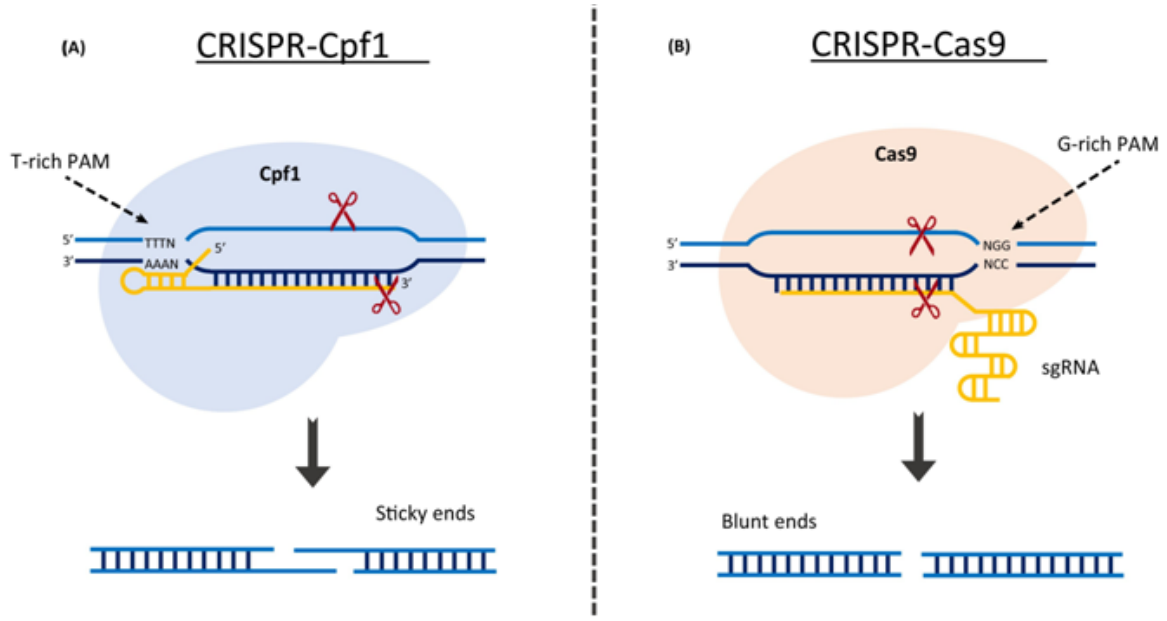

Figure 1.4.1: Key differences between CRISPR-Cpf1 and CRISPR-Cas9 systems. Cpf1 binds only self-processed crRNA, while Cas9 requires both crRNA and tracrRNA (fused together as longer single guide RNA (sgRNA). The PAM sequence to enable Cpf1 cut target DNA is thymidine-rich compared to the guanine-rich PAM for Cas9. Upon inducing a double stranded break, Cpf1 creates sticky ends compared to the blunt ends created by Cas9. Figure adapted from Zaidi *et al.*, Trends in plant science 2017.

The FnCpf1 wild-type spacer length is 24 nucleotides as shown by Zetsche. They demonstrated that detectable target cleavage was achieved for a 16 nucleotide crRNA. However, 18 nucleotide crRNA seemed required to achieve efficient target cleavage *in vitro*. Therefore 18-24 nucleotide crRNAs are suggested for use with Cpf1.

Possibly due to the different cleavage pattern induced by Cpf1 compared to Cas9, Cpf1 rarely causes insertions and instead induces larger deletions than Cas9 which could be interesting to increase the chance of a deficient gene product upon targeting (Kim *et al.*, 2016; Kim *et al.*, 2017). Also do Cpf1-induced DSBs have a more favorable HDR rate than Cas9 (Moreno-Mateos *et al.*, 2017; Toth *et al.*, 2016) which may be caused by the generated overhangs upon cleavage.

Interestingly, Moreno-Mateos reported that this increased HDR rate may be due to increased temperatures, possibly affecting the functionality of Cpf1. He remarked that AsCpf1 was less functional at 28°C while LbCpf1 showed more efficient. This made sense as reduced AsCpf1 activity was reported in plants earlier (Hu *et al.*, 2017). In the same year, papers claimed that Cas9 was temperature-dependent (LeBlanc *et al.*, 2018; Xiang *et al.*, 2017). Especially in plants this is a critical factor, as many plants are typically grown between 20°C and 24°C, while the efficiency of the nucleases seem to increase when growth temperature nears 37°C. LeBlanc quantitatively showed the increased efficiency in heat treated plants by targeting GFP reporter genes. They observed on average 77% more GFP-negative (indicative of mutations) nuclei in heat treated T1 plants compared to plants grown at 22°C. Heat treating the plants consisted of four heat cycles, each cycle consisting of 30 hours growth at 37°C followed by 48 hours recovery at 22°C. Yet no papers described increased Cpf1 activity in plants grown under comparable conditions. The temperature-dependency of these CRISPR nucleases does makes sense as these originate from human pathogenic bacteria.

The mechanism of target recognition by Cpf1 differs from that of Cas9 (Gao *et al.*, 2016). When mismatches are present, it prevents Cpf1 from binding properly to the target, which may prohibit DNA cleavage. Cpf1 was therefore suggested to be more precise due to this high sensitivity to mismatches. Indeed most predicted off-target sites did not produce mutations and use of preassembled recombinant Cpf1 ribonucleoproteins abolished completely off-target effects (Kim *et al.*, 2016; Kleinstiver *et al.*, 2016b).

There was some controversy on the dimerization requirement of Cpf1 to perform its activity as only a RuvC domain was detected for cleavage of the non-target strand but no HNH domain (Zetsche *et al.*, 2015). However, biochemical characterization and the crystal structure showed no traces of oligomerization (Dong *et al.*, 2016; Fonfara *et al.*, 2016). Not only is this finding straightforward as dimerization would then require tandem target sites, but this also might suggest a novel nicking domain (Yamano *et al.*, 2016). Based on this crystal structure of Cpf1, successful attempts to change the PAM sequence were achieved, contributing to broadening the target range of genome editing (Gao *et al.*, 2017; Nishimasu *et al.*, 2017).

1.4.2. Successful reports using CRISPR-Cpf1

In human cells, only AsCpf1 and LbCpf1 showed significantly higher capability to induce endogenous mutations compared to other Cpf1 proteins, with an efficiency up to 29% and 12% respectively, which was comparable with Cas9 induced mutation frequency (Zetsche *et al.*, 2015). However, later a different study on human cells showed that FnCpf1 is capable of robust targeted DNA cleavage, with an efficiency up to 34%, on exogenous GFP (Tu *et al.*, 2017). Also a triple target multiplexing study was done in human cells using AsCpf1 and resulted in a modification in all three targets. In yeast, single gene targeting efficiencies even reached 100% using FnCpf1, while 91% for three multiplexed targets in a single crRNA array (Swiat *et al.*, 2017).

The first reports in plants, showed in rice and tobacco respectively 28,2% and 47,2% of FnCpf1 induced average mutation frequency where the crRNA was expressed by U6 promoters (Endo *et al.*, 2016). Also LbCpf1 showed 41,2% efficiency in rice where a U3 promoter drove crRNA expression (Xu *et al.*, 2017). These first studies used RNA PolIII promoters for crRNA transcription, while in a later study, using a PolII expressed and ribozyme-processed crRNA system (adapted from Gao *et al.* (Gao and Zhao, 2014)) a 100% biallelic mutation frequency was achieved using LbCpf1 in T0 rice calli (Tang *et al.*, 2017). This ribozyme maturation system illustrated in Figure 1.4.2 uses ribozyme sequences that flank the mature crRNA on both sides. Upon translation, the ribozyme RNA sequences will fold and catalyze their own cleavage, yielding the mature crRNA that can bind Cpf1.

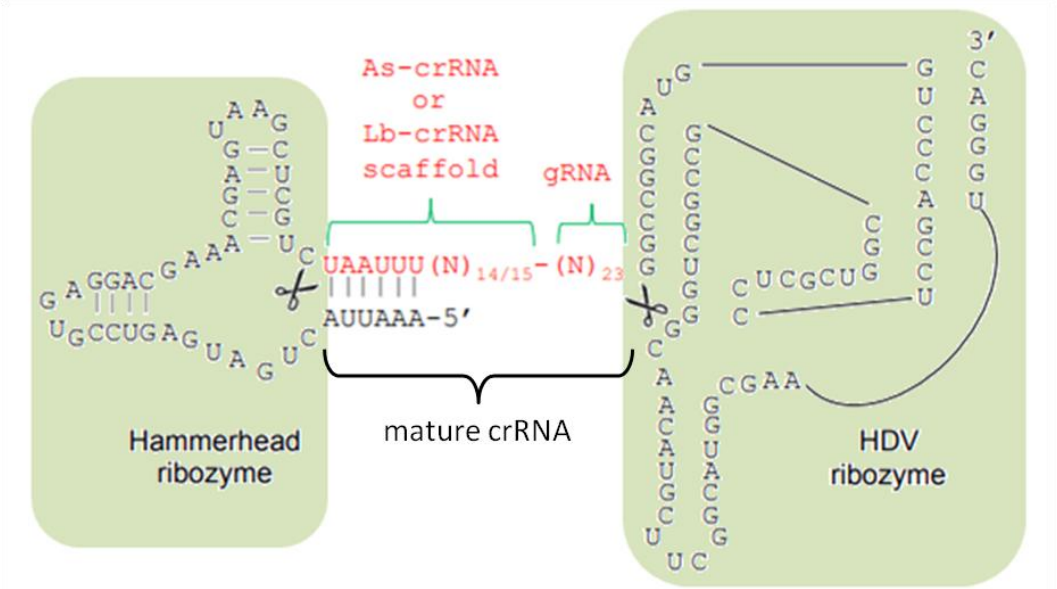

Figure 1.4.2: Self-processing ribozyme-flanked crRNA. Once transcribed, the desired mature crRNA sequence is flanked by a hammerhead and hepatitis delta virus (HDV) ribozyme at the 5'-end and 3'-end respectively. The ribozymes cleave between the nucleotides (indicated with scissors), yielding a mature crRNA able to bind Cpf1. Figure adapted from Tang *et al.*, Nature plants 2017.

In the same study, a dead version of Cpf1 (dCpf1) was used successfully for transcriptional repression in *Arabidopsis*. The expression of the targeted genes was lower than 10% of the wild-type expression when targeting the gene with dCpf1 fused to transcriptional repressors. Interestingly, yet no reports are available describing successful Cpf1-mediated mutagenesis in *Arabidopsis*. Lastly, DNA-free delivery of recombinant As/LbCpf1-crRNA complexes in protoplasts also demonstrated effective (Kim *et al.*, 2017). It must be noted that all successful Cpf1-targeting in plants was done using plant or species codon optimized Cpf1.

As can be seen, organism, vector design, delivery method, and many more factors may crucially affect the efficiency. Unfortunately, most CRISPR-Cpf1 studies were done in mammalian cells, while its utility in plant research can also be significant. Surprisingly, the targeted mutagenesis capacity of CRISPR-Cpf1 has not yet been reported in the important model plant *Arabidopsis*.

As a novel CRISPR-associated nuclease, still a lot could get optimized for its use in different organisms. Inspired by the new or different features of Cpf1, possibly new approaches for genome editing could get invented. Perfect plant genome editing lays in achieving a maximal precision together with a maximal efficiency, having the possibility to reach and rewrite any target in the genome.

2. Aim of the research project

CRISPR-Cas9 is a very efficient tool for targeted mutagenesis. However, Cas9 target choice is limited due to the required PAM sequence NGG, restricting its targeting range to more GC-rich coding sequences. Cpf1 is a different CRISPR associated nuclease and requires a (T)TTV PAM sequence, making Cpf1 complementary to Cas9 and expanding the targeting range of the CRISPR-Cas toolbox into AT-rich regions. Furthermore, Cpf1 differs in multiple aspects compared to Cas9 (like target cleavage and crRNA processing) and may therefore have different genome editing efficiencies and outcomes. In contrast to Cas9, the number of reports on the use of Cpf1 in plants is rather limited and therefore requires more investigation.

As using Cpf1 for plant genome editing is new to the institute, the main aim of this project is to demonstrate the functionality of the CRISPR-Cpf1 system at inducing targeted mutagenesis and quantifying the observed editing. We will use the different Cpf1 homologues AsCpf1, FnCpf and LbCpf1 as this will enable a comparison between them and these nucleases were shown to work well in mammalian cells and some plant species. The crRNA processing will be done by flanking the mature crRNA sequence by ribozymes.

In *Arabidopsis* we will target a standard genome editing target, the *PDS* (*phytoene desaturase*) gene, which should cause an easy-to-screen albino phenotype due to the deficient pigment pathway. Three different targets will be designed to *PDS* to compare efficiencies between the different target sequences.

We will also test CRISPR-Cpf1 in tomato by targeting *peapod* (*PPD*) genes which should result in a curved leaf surface (Gonzalez *et al.*, 2015). As tomato plants have two *PPD* genes, we will design two targets per gene, targeting both genes simultaneously to test the multiplexing capacity of CRISPR-Cpf1.

After plant transformation with the CRISPR-Cpf1 vectors, we will phenotypically screen the plants to find any expected mutagenic phenotypes. After this evaluation, relevant plant material will be collected to genotype the target sites. By sequencing these target regions, we will check if any mutations are induced, what kind of mutations are present and how efficient this mutagenesis mechanism is.

Depending on the remaining time, we want to try optimize the CRISPR-Cpf1 system by varying other factors. This could be done by increasing the temperature as the CRISPR nucleases appear to be temperature-dependent. We also want to build a vector system to clone up to four targets arranged as native CRISPR arrays as Cpf1 is able of processing its own CRISPRs and could therefore be a more efficient mechanism than the ribozyme-mediated crRNA maturation.

Summarized, in this project we want to test the mutagenesis capacity of CRISPR-Cpf1 in *Arabidopsis* and tomato. Mutagenesis efficiencies will be compared between Cpf1 homologues and between different targets. Finally, an attempt to optimize the efficiency of CRISPR-Cpf1 targeted mutagenesis can be done by varying the temperature and testing a different CRISPR processing method.

3. Results

In this project we evaluated CRISPR-Cpf1 to induce targeted mutations in *Arabidopsis* and tomato.

3.1. Vector cloning

3.1.1. Construction of CRISPR-Cpf1 expression vectors

The expression vectors were constructed according to Figure 3.1.1.

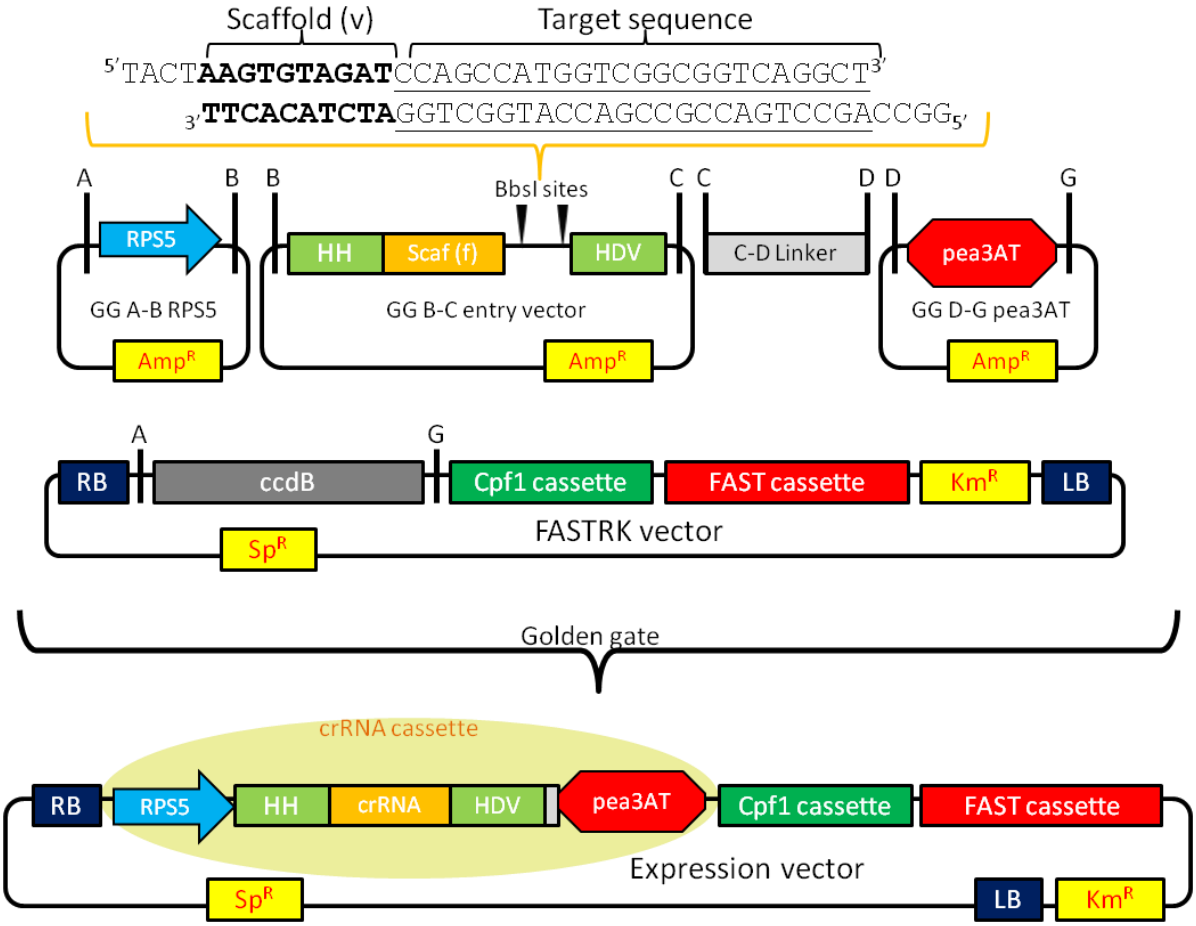

Figure 3.1.1: Cloning scheme for single targets. A forward and reverse oligo are annealed to form a dsDNA fragment containing the Cpf1 homologue-dependent (variable) scaffold part (Scaffold (v)) followed by the target sequence. This fragment is flanked by overhangs compatible with the overhangs of BbsI-digested golden gate (GG) B-C entry vectors enabling ligation. This will form, together with the fixed scaffold sequence part (Scaf (f)), the full mature crRNA sequence, which is flanked by hammerhead (HH) and hepatitis delta virus (HDV) ribozyme sequences. The created vector was then used in a golden gate reaction, where the ribozyme-flanked crRNA was provided by the RPS5 promoter and pea3AT terminator (C-D linker needed). During the golden gate reaction, this crRNA cassette is exchanged with the *ccdB* in the FASTRK vector (containing a Cpf1 expression cassette [PcUbiP-SpCas9-G7T] and a FAST (fluoresence accumulating seed technology) cassette)[OleP-mRuby-nosT], making the desired expression construct. (LB: left border; RB: right border; Amp^R: Ampicillin resistance; Km^R: Kanamycin resistance; Sp^R: Spectinomycin resistance).

Using the cloning strategy in Figure 3.1.1, 24-nucleotide targets were chosen as this corresponded to the wild-type crRNA target length reported (Zetsche *et al.*, 2015). As the crRNAs for the chosen Cpf1s homologues had to initiate with a uracil nucleotide, which cannot be achieved using U6 or U3 RNA polymerase III promoters, the ribozyme-mediated crRNA system was chosen as this was reported successful (Tang *et al.*, 2017). The RPS5 promoter was shown effective for crRNA expression in *Arabidopsis* plants in previous work. The Cpf1 cassettes contain the same promoter and terminator as Cas9 cassettes from previous work, minimizing varying factors between experiments. The FAST cassette simplifies selection of transgenic seeds (Shimada *et al.*, 2010), while the kanamycin resistance marker can be used for selection in cell cultures.

Three different *PDS* targets in *Arabidopsis* (PDS1, PDS2 and PDS3) and four targets in the tomato *PPD* genes divided as two targets (T1 and T2) per gene (G1 [Solyc09g065630] and G2 [Solyc06g084120]) were picked and oligos were designed (Table A.1). Following the cloning strategy illustrated in Figure 3.1.1, the oligo pairs were annealed and ligated inside BbsI-digested B-C entry vectors. Tomato T2 targets were also inserted in BbsI-digested C-D entry vectors, allowing assembly for multiplex targeting (replaces the C-D linker). Acquisition of vectors with correctly integrated targets was confirmed by sequencing and stocked (Table A.2).

Expression vectors were made via golden gate (GG) reactions using the components described in Table A.3 and was brought in *Agrobacterium* (Table A.4) to transform *Arabidopsis* and tomato.

3.1.2. Design and synthesis of vectors to construct native CRISPR-Cpf1 arrays

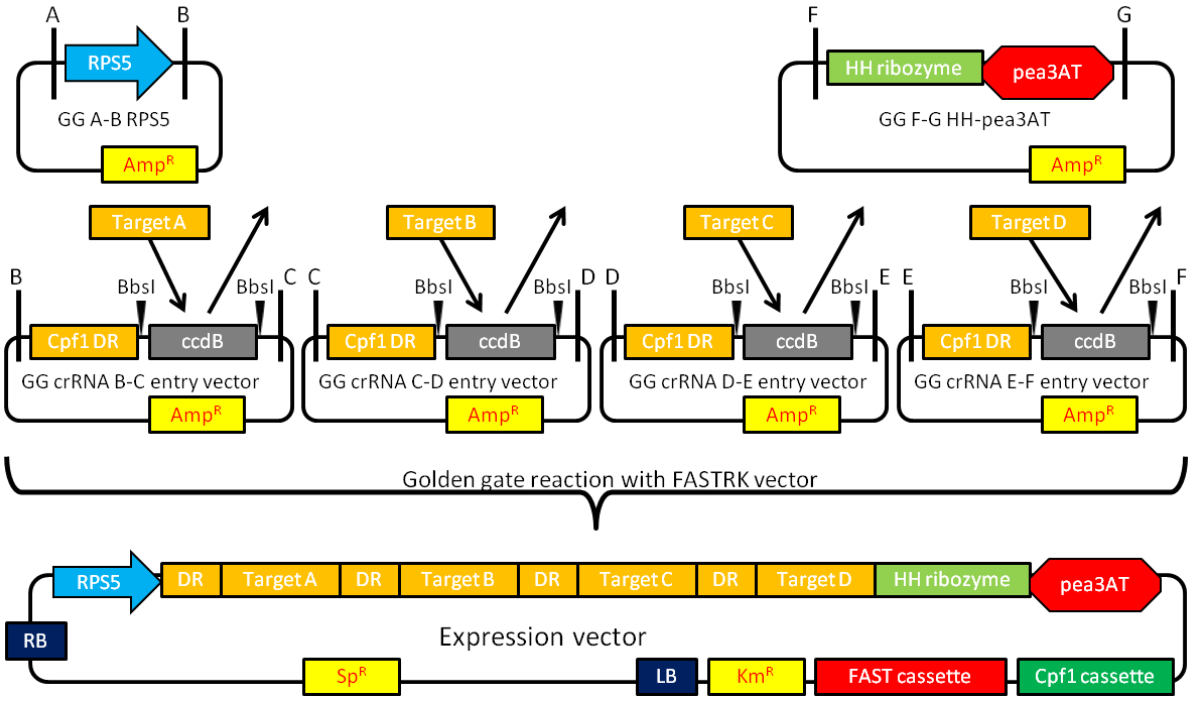

Figure 3.1.2: Desired cloning strategy to form vectors with native Cpf1 CRISPR arrays. In a first conventional cloning reaction, the Cpf1 targets replaces the *ccdB* in the golden gate (GG) crRNA entry vectors, forming a Cpf1 direct repeat (DR) followed by the target sequence. In a next golden gate reaction, four DR-Target fragments are ligated together, provided by a RPS5 promoter and pea3AT terminator, will replace the *ccdB* in the FASTRK vector (FASTRK vector not shown in figure, see Figure 3.1.1), forming the desired expression vector. (LB: left border; RB: right border; Amp^R: Ampicillin resistance; Km^R: Kanamycin resistance; Sp^R: Spectinomycin resistance)

As Cpf1 is able to process CRISPRs itself, a simple cloning system was devised to allow for up to four targets arranged as native CRISPRs. Future experiments will determine if the crRNA array system is more or less efficient than the ribozyme system used in this project.

Figure 3.1.2 shows the desired cloning strategy. As the GG crRNA entry vectors and the GG HH-pea3AT vector did not exist, these were constructed and the sequences can be found in addendum Sequences A.5 and A.6, respectively. Correctly sequenced vectors were stocked (Table A.7).

3.2. First transgenic generation do not show Cpf1-induced mutations

Red fluorescing seed was collected and sterilized. Up to 25 fluorescing seeds were stratified, placed on plates with growth medium and kept in the 21°C long day growth chamber.

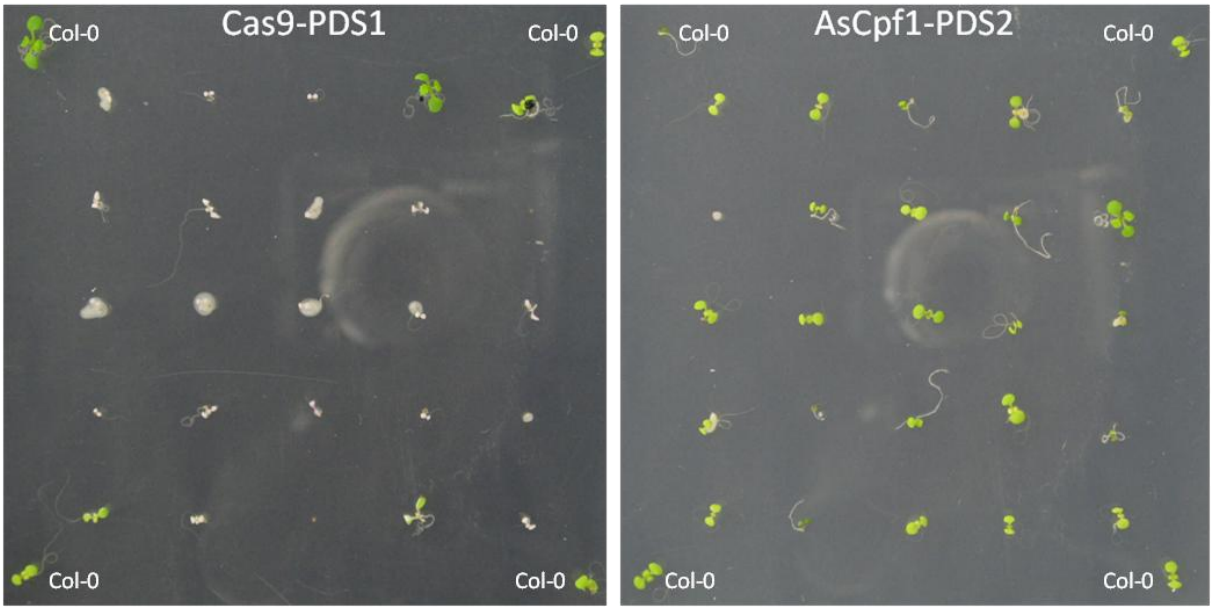

Figure 3.2.1: Comparison of *Arabidopsis* seedlings transformed with a Cas9 construct and a Cpf1 construct targeting *PDS*. Twenty-five *Arabidopsis* T1 seeds were placed on medium, harvested from *Arabidopsis* plants transformed with a Cas9 construct (left) and AsCpf1 construct (right) both targeting the *PDS* gene. Non-fluorescing seeds were indicated by a black dot. Photos were taken 11 days after placement in 21°C long day growth chamber. On each corner, one Col-0 seed was placed as negative control.

The contrast between Cas9 and Cpf1 transformants can be clearly seen (Figure 3.2.1). None of the 190 plants transformed with any of the nine Cpf1 constructs showed the expected bleaching phenotype (Table A.8). In contrast, positive-control seeds (Cas9-PDS) did show a high rate of albino seedlings as a 0:2:18 ratio was observed for the wild-type:chimera:albino phenotypes respectively. However, this rate was different between other Cas9 *PDS* targets as for the second and third target a 1:14:4 and 6:17:0 ratio was obtained respectively, showing that targeting efficiency is dependent on the chosen target (Table A.8).

For each Cpf1 plate, up to 10 seedlings were potted to let the T1 *Arabidopsis* plants grow in the greenhouse. Leaf samples were collected and the genomic DNA extracted to PCR-amplify the targeted region of *PDS* and purified amplicons were sent for sequencing.

All sequences were mapped to the reference *PDS* gene and no mutations were observed. Some sequences had an overall lower quality chromatogram (data not shown) indicative for gene editing and were used for TIDE analysis. However, the R-square values were low, making the model unreliable and these samples were likely just bad Sanger reads. These results were in agreement with the lack of phenotype; none of the T1 *Arabidopsis* plants were mutated at the target sites.

Leaf samples of the tomato plants were collected and the genomic DNA was extracted. Just as for the T1 *Arabidopsis*, no mutations were observed in the 30 genotyped tomato plants.

Together, these results indicated that our CRISPR-Cpf1 system was not functional. As the functionality of this system is critical for the project, troubleshooting began.

3.3. Troubleshooting of the used CRISPR-Cpf1 system

3.3.1. The CRISPR-Cpf1 transgenes are present in the genomic DNA

We began troubleshooting by confirming the presence of the T-DNAs in the transgenic lines. The *Arabidopsis* lines gave fluorescent seeds and the tomato T0 plants rooted in medium with kanamycin, therefore we expected these lines to contain the whole vector. We attempted to amplify the Cpf1 and the crRNA encoding sequences using the extracted genomic DNA of leaf samples as template.

Arabidopsis

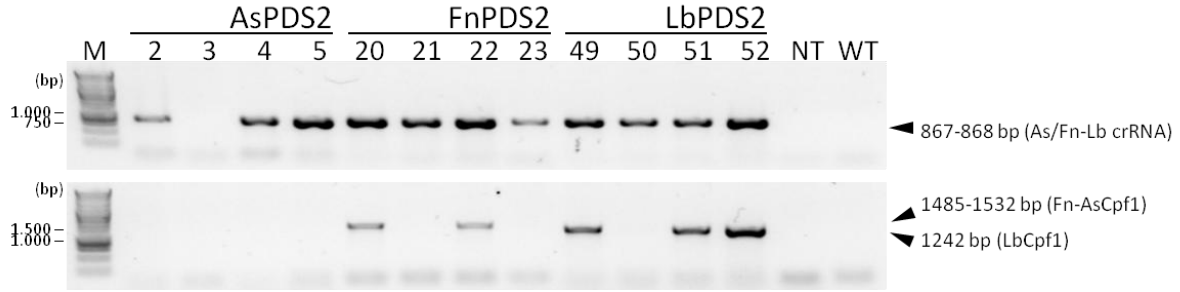

Figure 3.3.1: Amplification of crRNA and Cpf1 fragments from extracted genome of T1 Arabidopsis. Genomes of T1 *Arabidopsis* leaves were extracted and DNA fragments of the crRNA and Cpf1 were amplified. As/Fn/LbPDS2 indicates the construct with Cpf1 homologue targeting the second designed *PDS* target. NT = no template control; WT = Col-0 control. M = TopBench 1kb ladder.

The PCR results in Figure 3.3.1 indicate that the crRNA cassette is present in most plants. However, results from PCR amplification of a fragment of the Cpf1 cassette were inconsistent as there was sometimes a lack of the Cpf1 fragment between screened plants. This was especially the case for the AsCpf1 lines. By extending the PCR elongation time, we were able to see also inconsistencies between PCR results of the same samples (samples 5 and 20, Figure 3.3.2).

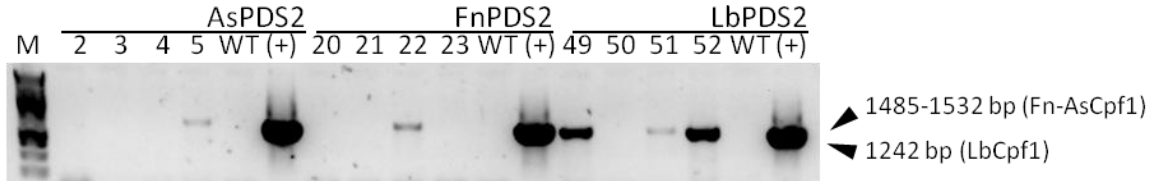

Figure 3.3.2: Amplification of crRNA and Cpf1 fragments from extracted genome of T1 Arabidopsis (2). Genomes of T1 *Arabidopsis* leaves were extracted and a DNA fragments of Cpf1 was amplified. As/Fn/LbPDS2 indicates the construct with Cpf1 homologue targeting the second designed *PDS* target. WT = Col-0 controls; (+) = Positive controls (the vector itself). M = TopBench 1kb ladder

An attempt was made to amplify the full Cpf1 coding sequence. No amplification was observed for the samples that were able to amplify the Cpf1 fragment earlier (Figure 3.3.3). We hypothesized that the quick *Arabidopsis* gDNA extraction protocol from Berendzen (Berendzen *et al.*, 2005) generated highly fragmented genomic DNA. To confirm this, we took *Arabidopsis* gDNA extracts from a different project that contains functional Cas9 constructs and had been extracted using the same protocol. Indeed, the full Cas9 coding sequence was not able to amplify (Figure 3.3.3).

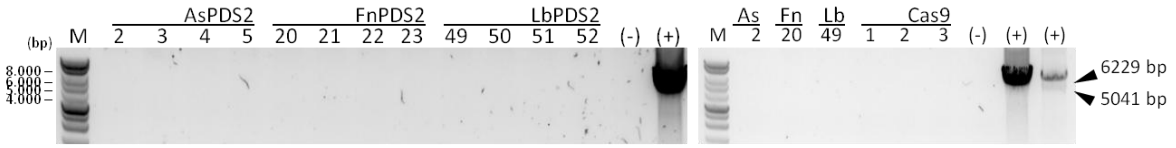

Figure 3.3.3: Amplification of full-size Cpf1 and Cas9 from extracted T1 *Arabidopsis* genomes. Genomes of T1 *Arabidopsis* leaves were extracted and the full DNA fragments of Cpf1 or Cas9 was amplified. As/Fn/Lb(PDS2) indicates the construct with Cpf1 homologue targeting the second designed *PDS* target. (-) = Negative control (Col-0); (+) = Positive controls (first two: construct with Cpf1, last one: construct with Cas9). M = TopBench 1kb ladder.

Tomato

Despite some inconsistencies between band intensities in both gels, it is clear that all tomato plants contain the transgenic construct except for the LbCpf1-T1G1/T2G2 samples 7 and 15 (Figure 3.3.4).

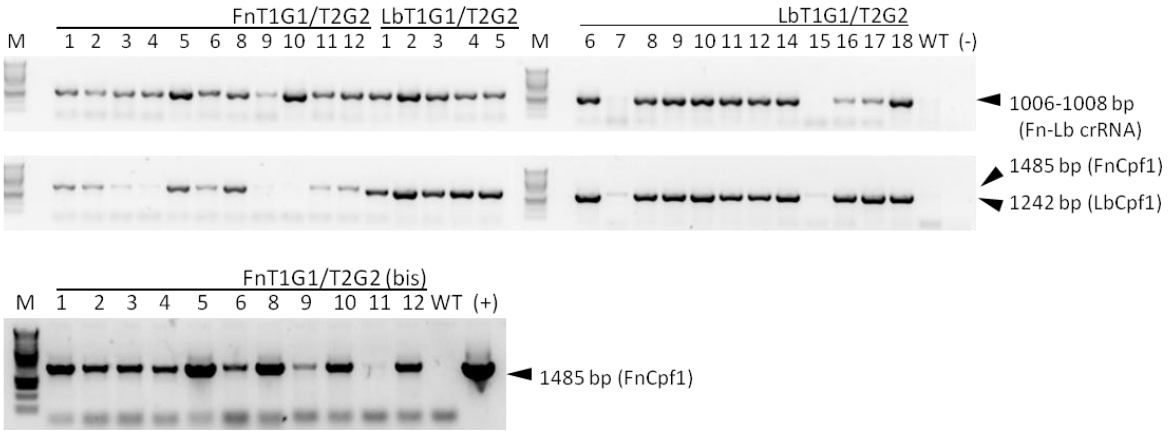

Figure 3.3.4: Amplification of crRNA and Cpf1 fragments from extracted genome of T0 tomato plant leaves. Genomes of T0 tomato plants were extracted and DNA fragments of the crRNA and Cpf1 were amplified. (-) = Negative control (MQ water); (+) = Positive control (FnT1G1/T2G2 construct itself); M = TopBench 1kb ladder.

It is not clear if the *Arabidopsis* lines carry the full-length Cpf1 gene as the negative results may have been due to the DNA extraction method used. However the crRNA cassettes were intact and the seeds were positive for the FAST selection marker.

In contrast, the data from the tomato plants, with DNA extracted using a superior method (Edwards *et al.*, 1991), showed that in 28 of the 30 tomato lines contained both the crRNA and Cpf1 cassettes. Therefore we reasoned that Cpf1 cassettes were also likely intact in the *Arabidopsis* lines and we continued with the troubleshooting.

3.3.2. Cpf1 is expressed in tomato

With the presence of the CRISPR-Cpf1 system in the genome confirmed, the next question was whether or not the Cpf1 and crRNAs were being expressed. This was only tested in tomato plants due to material availability.

RNA was extracted from tomato leaves for cDNA synthesis and used it as PCR template to detect Cpf1 expression. To ensure the entire Cpf1 gene was present, we amplified multiple overlapping Cpf1 cDNA fragments as depicted in Figure 3.3.5.

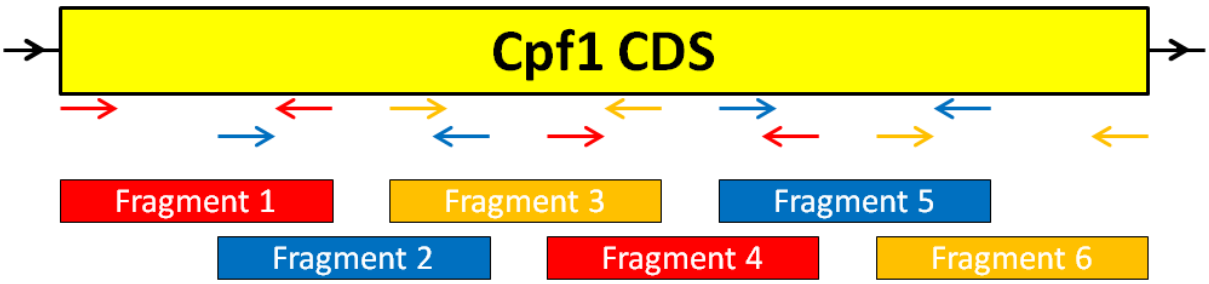

Figure 3.3.5: Amplification of six fragments covering the full Cpf1 sequence. Primers are presented by arrows underneath the Cpf1 CDS.

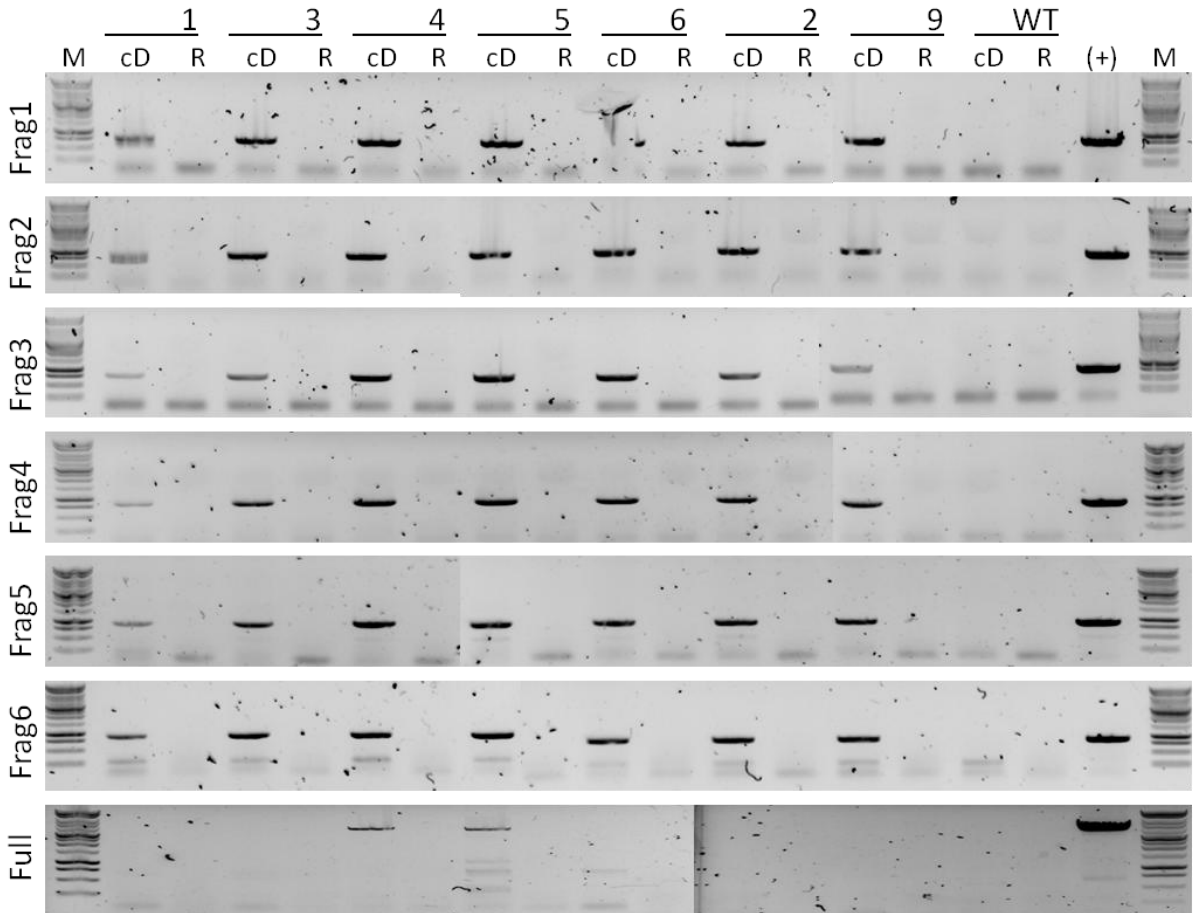

Figure 3.3.6: : Evaluating presence of all Cpf1 fragments in tomato leaf cDNA. cDNA made from T0 (A1 construct) tomato leaf mRNA extract was used as template to amplify fragments of Cpf1 cDNA, covering the complete Cpf1 cDNA. cD = cDNA; R = raw RNA sample; Numbers identify plants; (+) = positive control (A1 construct); M = TopBench 1kb.

Figure 3.3.6 shows that all individual Cpf1 cDNA fragments were present. No amplification was seen in the raw RNA sample indicating the samples were free of contaminating genomic DNA. However, the full length of Cpf1 was difficult to amplify; a band was only observed in samples 4 and 5 while all samples contain the individual fragments. cDNA synthesis is not as efficient for longer fragments, which may explain the lack of amplification here. Furthermore, we were later notified that the iScript used for RT-PCR was a bad batch, which could explain why we failed to get amplification of longer Cpf1 fragments. But here we can conclude that Cpf1 is getting expressed in the FnT1G1/T2G2 tomato plants.

3.3.3. The Cpf1 protein is undetectable on western blots

Transcribed Cpf1 does not mean it is translated and therefore western blots were performed to detect Cpf1 (via the hemaglutinin tag) in the transgenic tomato plants protein extracts.

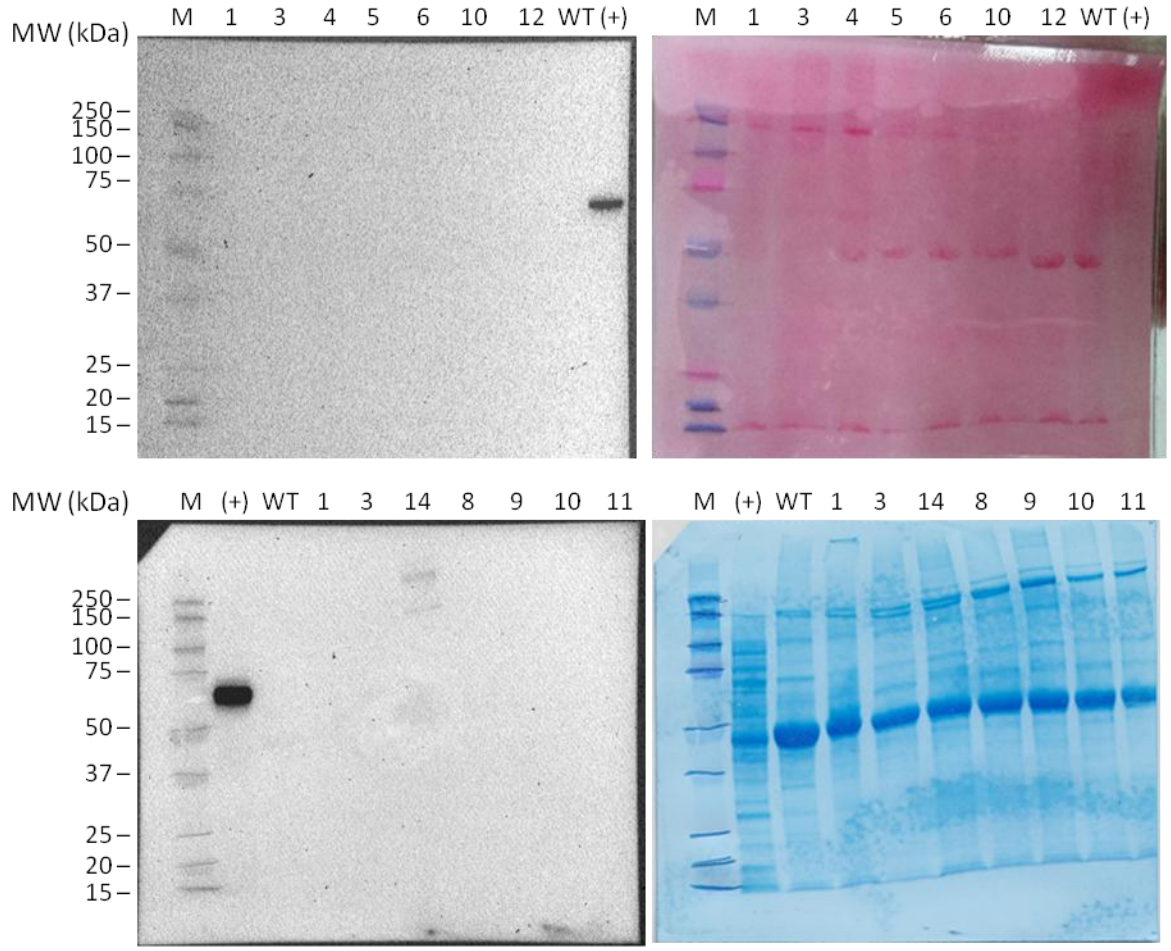

Figure 3.3.7: Western blotting of T0 tomato leaf protein extracts. Proteins were extracted from T0 tomato leaf samples and an equal amount of proteins were run on a PAGE gel. Upper part analyzed FnT1G1/T2G2 construct plants, with the western blot in the left and Ponceau staining in the right. Bottom half analyzed LbT1G1/T2G2 construct plants: western blot (left) and Comassie blue staining (right). Cpf1 contains a hemaglutinin (HA) tag which is targeted on the western blots. Numbers identify plants. (+) = Positive control (protein extract containing S6K-HA [62 kDa]) M = Precision Plus Protein Dual Colors Standards ladder.

There is no signal for Cpf1, having a molecular weight of 149 kDa (Figure 3.3.7). Sample 14 does show a very faint band around this size, however only after a doubled exposure time. Also a higher band beyond the protein ladder range can be noticed. This may indicate a very weak Cpf1 signal, but due to the other band on the same lane it may not be reliable as Cpf1 is not supposed to dimerize especially not after the denaturation steps.

3.3.4. BY-2 cells do not show consistent GFP-Cpf1 accumulation

An independent experiment to determine Cpf1 protein presence, including localization, is by tagging the protein with GFP. For this, GFP was N-terminally-fused to the three Cpf1 proteins and transformed into BY-2 cells.

As controls, a Cas9 construct with (positive control) and without (negative control) GFP-NLS cassette was implemented. On the positive control plate, all calli were fluorescent and showed a GFP signal in the nucleus as seen in Figure 3.3.8. All other plates did not show any fluorescence, except one callus out of 18 for AsCpf1 which had a chimeric fluorescence pattern. Zooming in on individual cells seem to show a weak fluorescence accumulation inside the nucleus, suggesting that if Cpf1 is being translated, it is transported inside the nucleus.

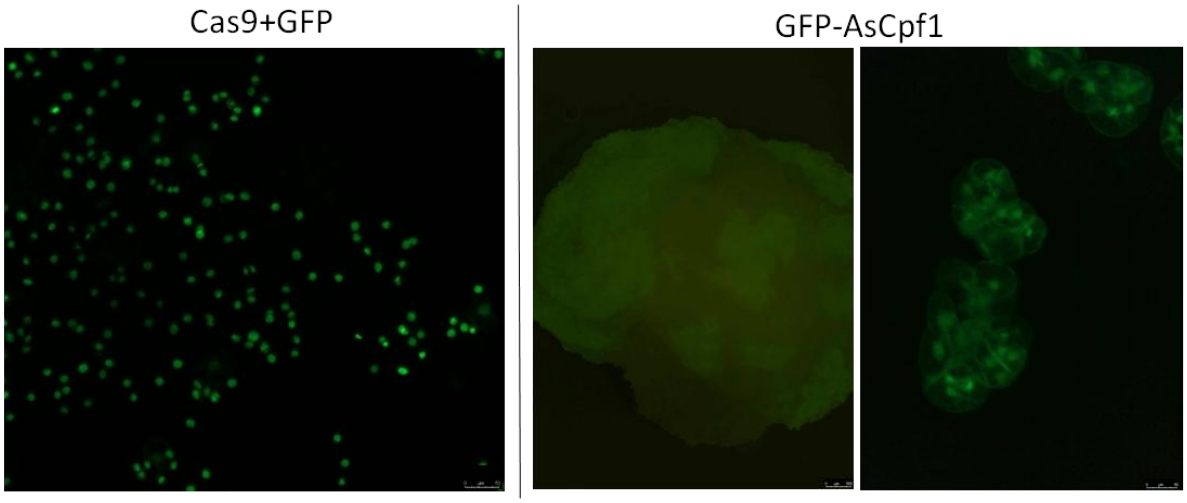

Figure 3.3.8: Fluorescent images of BY-2 cells bearing constructs containing GFP. All Cas9+GFP calli showed cells with a fluorescent nucleus. Only one GFP-AsCpf1 callus out of 18 showed chimeric pattern of fluoresence. Individual cells show a weak fluorescence signal in the nucleus.

After this section we concluded that if no mutations were made, the possible problem could be at translational level. The protein does not seem to be accumulating.

3.4. CRISPR-Cpf1 functionality is affected by heat in *Arabidopsis*

3.4.1. *Arabidopsis* T2 screen revealed functional CRISPR-Cpf1 lines at 37°C

As the first T1 seedling growth did not give any phenotype on 25 individuals, we wanted to make a bigger screen to see if any plants get mutagenized to see if CRISPR-Cpf1 is at least functional but with likely a lower efficiency. As not enough T1 seeds were available, we used the harvested T2 seeds from the T1 screen.

Inspired by the papers describing the heat-dependency of CRISPR nucleases (LeBlanc *et al.*, 2018; Moreno-Mateos *et al.*, 2017), we hypothesized that temperature would increase the rate of mutagenesis by CRISPR-Cpf1 and therefore we sowed 2 x 1000 seeds for each line. Eight days after sowing, heat treatments were initiated on one of the two sets consisting of 4 cycles of 72 hours, consisting of 30 hours at 37°C and 42 hours recovery in the *Arabidopsis* greenhouse where the other set remained permanently, as described by LeBlanc. After the last heat cycle, both sets are left in the greenhouse under observation. A simplified illustration of the experiment is shown in Figure 3.4.1.

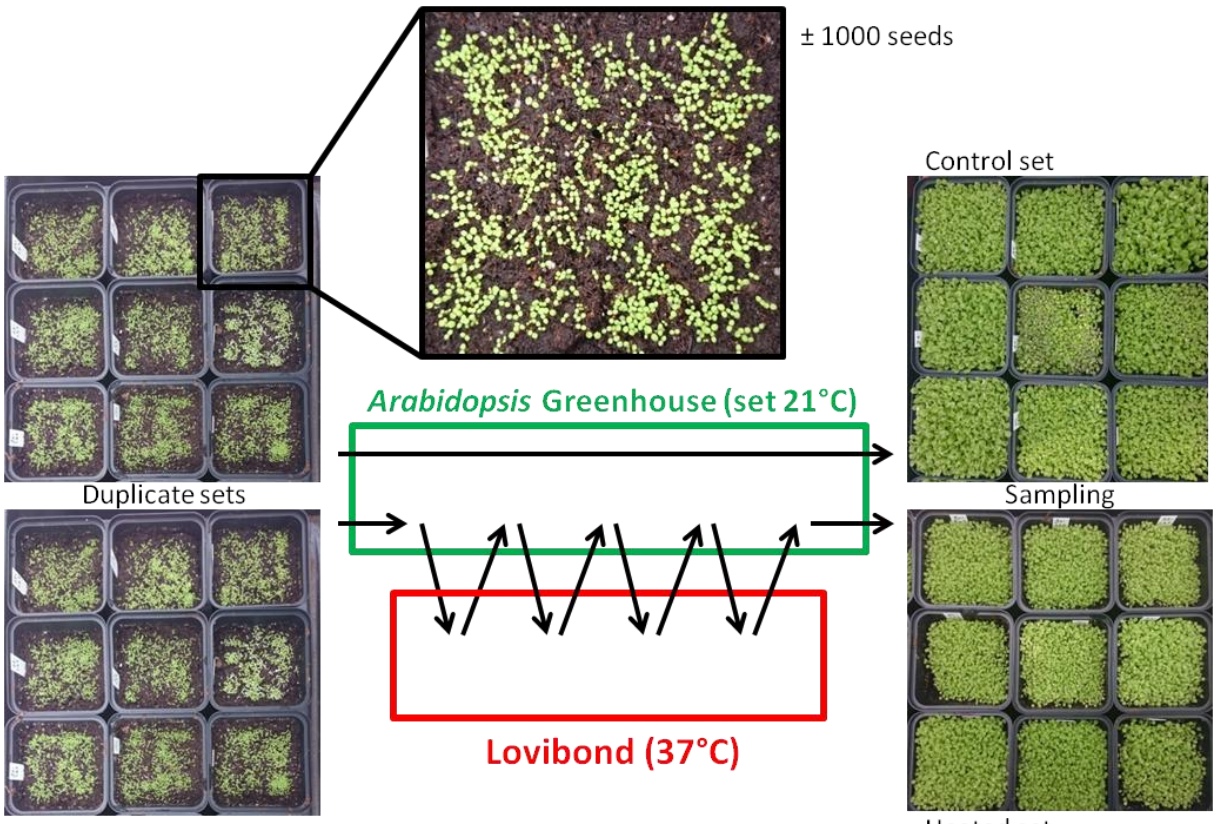

Figure 3.4.1: Illustration of the *Arabidopsis* T2 screening including a heat treatment condition. Around 1000 *Arabidopsis* T2 seeds were sown in soil-containing pots per line *in duplo* creating two equal sets. One set stayed permanently in the *Arabidopsis* greenhouse. The other set started four heat cycles from the eighth day on. Samples were taken four days after recovery from the last heat cycle.

Before any plants were subjected to heat, one seedling from a FnCpf1-PDS2 pot showed clear chimeric leaves. Sampling could have killed the plant and therefore the plant was placed in a separate pot for further growth without heat cycles.

Due to inconsistent greenhouse temperatures, the measured *Arabidopsis* greenhouse temperatures were graphed (Figure 3.4.2).

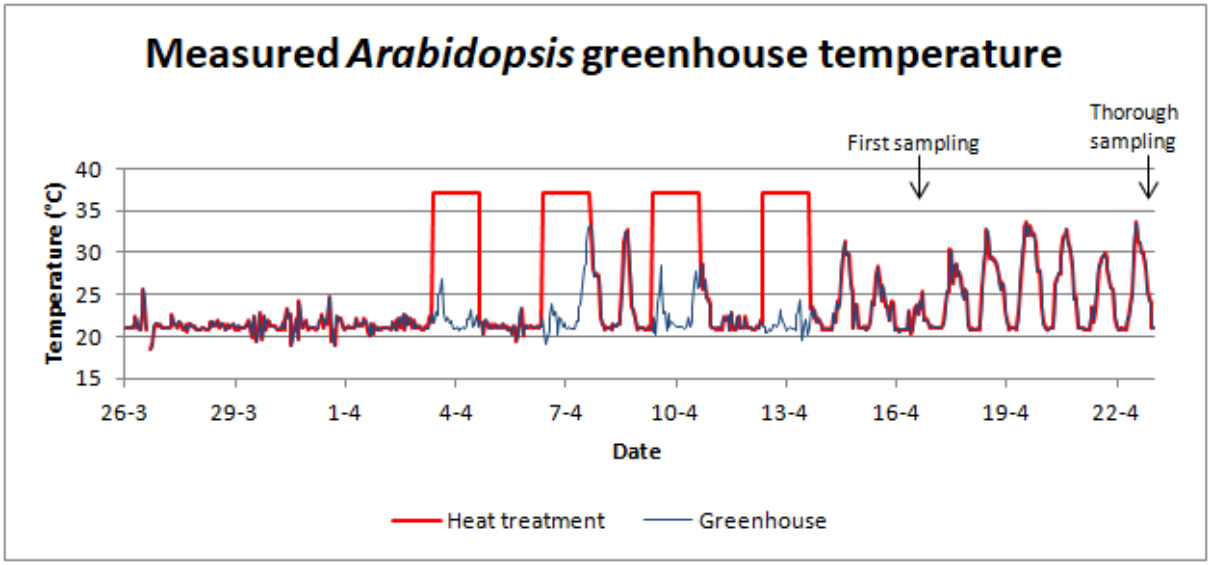

Figure 3.4.2: Measured *Arabidopsis* greenhouse temperatures from germination until sampling. T2 *Arabidopsis* seeds were sown on soil in duplicate, whereafter one set stayed permanently in the greenhouse while the other set was subjected to a heat treatment. The first sampling consisted of genotyping suspected plant material, while the thorough sampling genotyped the rosette leaves of all suspected lines (heated and non-heated) *in triplo*.

Four days after the fourth heat cycle, phenotypes were observed. However, only the LbCpf1-PDS2 lines showed the Cas9-PDS-like phenotype, even more pronounced as can be seen in Figure 3.4.3 (next page). Plants in AsCpf1-PDS3 pots showed bright green rosette leaves with even brighter green spots, together with bleached stems and cauline leaves. Plants from one FnCpf1-PDS3 pot had whitish leaves, possibly due to fungal growth. Up to three samples of the described suspected plant material (together with one unheated control of the same line) were collected for a first sequencing screen. Only all three of the heat-treated rosette leaves containing the LbCpf1-PDS2 construct showed clear mutations. A more thorough sampling was done on the rosette leaves of all suspected lines: for each of the suspected lines, exactly three non-heat-treated and three heat-treated rosettes were collected, choosing the most suspected rosette leaves if there were any. The sequencing results confirmed that only the LbCpf1-PDS2 construct successfully targeted *PDS* in the rosette leaves and only after the heat treatment. The sequencing data of the Cas9-PDS and the LbCpf1-PDS2 were analyzed using ICE (Hsiau *et al.*, 2018) (Table 3.4.1 and 3.4.2) and TIDE (Brinkman *et al.*, 2014) (Table A.9 and A.10), to roughly compare both tools.

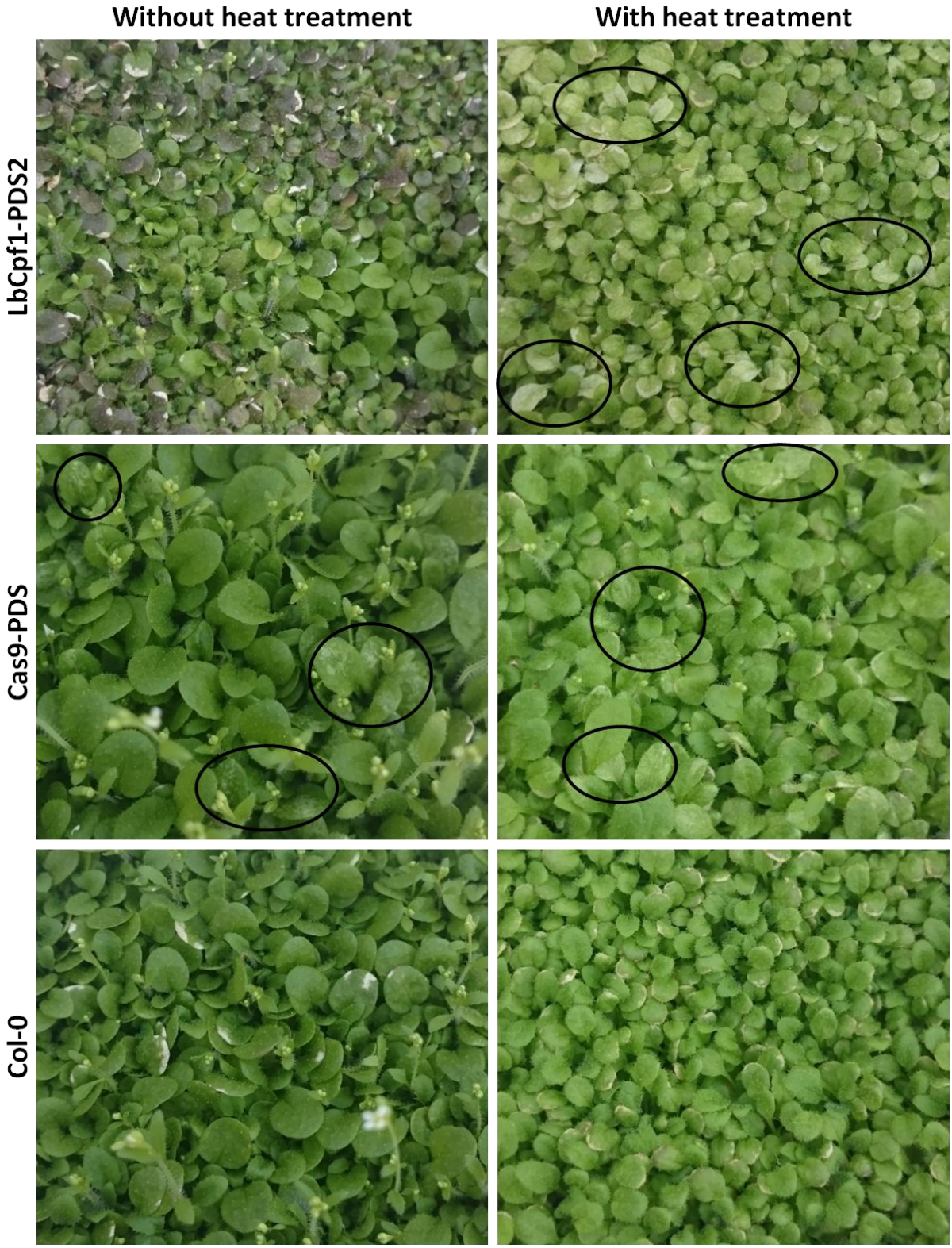

Figure 3.4.3: Three weeks old *Arabidopsis* T2 seedlings after growth in normal conditions and a heat treatment. *Arabidopsis* T2 seeds containing different CRISPR-Cpf1 constructs were sown in duplicate. One week later, one set was subjected to a heat treatment of four cycles while the other set was left in the greenhouse. After the fourth heat cycle, the seedlings were left in the greenhouse and pictures were taken four days later. Circles were drawn to indicate leaves with the expected bleached phenotype.

Table 3.4.1: Summarized ICE analysis for the sequences of LbCpf1-PDS2 rosette samples from heat-treated and non-heat-treated plants. 'Total' indicates the total indel frequency reported by the ICE tool. Only frequencies higher than 1% were taken.

		Total	R^2	-24	-23	…	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
Heated	Line 1	35	0,98		1,8		1,1				3,1	1,3			2,3	4,2	7,5			11	64
		16	0,99												1,2	1,9	5,7			6	83
		36	0,98		1,6		1,4				2,8				2,4	3,1	7,7			13	62
	Line 2	77	0,93		5,4		1,4				5,8	5,2			7,6	13	22			17	16
		73	0,95		4,4		2,2				6,4	4,1			7,7	11	22			13	22
		74	0,95		3,9		1,5				5	4,8			5,6	12	22			14	22
		70	0,93		5,4						4,6	3,1			4,1	12	24			15	24
		60	0,95		4,2		2,5				4,2	2,6			6	10	19			8,4	36
		52	0,97		3,4		1,3				3	1,5			3,7	8,6	13			13	45
	Line 3	24	0,98		1,8						1,2					3,4	8,7			7,8	74
		49	0,97		3,8		1,9				3,4	2,1			4,5	5,9	14			11	48
		68	0,95		3,6		3,1	2,5			6,7	2,7			6,4	11	18			12	28
	Line 4	40	0,96		1,7						2,1				2,1	4,1	14			15	56
		50	97		2,8		1,5	1,2			3,4	1,6			3,6	4	14			17	46
		45	0,96		1,6		1,5				3	1,2			2,5	5,7	11			15	51
Non-Heated	Line 1	1	1																		99
		1	1																		99
		3	1																	1,8	97
	Line 2	2	1																		98
		1	1																		99
		2	1																		98
		0	1																		100
	Line 3	1	1																		99
		0	1																		100
		1	1																		99
	Line 4	1	1																		99
		0	1																		100
		2	1																		98

Table 3.4.2: Summarized ICE analysis for the sequences of Cas9-PDS rosette samples from heat-treated and non-heat-treated plants. 'Total' indicates the total indel frequency reported by the ICE tool. Indels with lower than 2% frequency were left out for better visualization purposes.

	Total	R^2	-30	…	-25	-24	-23	…	-18	-17	-16	…	-4	-3	-2	-1	0	1	2	3	4
Heated	52	0,97									28					4	45	16			
	70	0,98														3,7	28	61			
	59	0,97															37	56			
	44	0,79	2,7		3,6												35	10			7,7
Non-Heated	53	0,92					2,7										39				46
	53	0,9											48				37				
	60	0,9							57								30				
	59	0,92								53							33				

The data clearly show that plants with the expected phenotype only arose after a heat treatment (Figure 3.4.3). As expected, the Cas9-PDS plants had the phenotypes in both conditions, however no conclusion can be given about eventually increased efficiency after heat treatment as the amount of germinated seeds and affected seedlings were not counted.

When looking at the measured temperatures in the greenhouse in Figure 3.4.2, we can see that the control set reached temperature peaks up to 33°C. However this does not appear to be sufficient to produce observable phenotypes, which could be due to not reaching a temperature threshold or because the warmer time interval was not long enough. Despite no visual phenotype was observed, the control plants could have subjected mutagenesis at lower rates and therefore also these were genotyped.

The mutation analysis summarized in Table 3.4.1 confirmed that the Cpf1 plants showing no phenotype indeed had no significant amount of mutations, as very low values given are likely background. The LbCpf1-PDS2 plants showed clearly mutations. Over 80% of these mutations appeared to be deletions of up to 10 base pairs.

Mutation analysis of the Cas9 plants (Table 3.4.2) showed a different indel pattern between plant samples of the heated and not-heated condition. However, the sample size is too small to generalize this conclusion.

3.4.2. 30°C is not sufficient for functional CRISPR-Cpf1 in the *Arabidopsis* lines

As the LbCpf1-PDS2 lines worked at 37°C, we investigated if a lower temperature of 30°C was sufficient for Cpf1-induced mutagenesis. Due to a lack of time, we shortened the heat treatment from four cycles to two cycles. One hundred seeds of each previously tested LbCpf1-PDS2 line were placed on medium and first grew under normal conditions. After one week, the shortened heat treatment started in dark incubators. After the second and last heat cycle, the plants were left in the growth chamber until the same timepoint was reached as when phenotypes were observed in the T2 screen after the fourth heat cycle, which was three weeks after germination. The plants were then scored based on their increasing bleached phenotype (example in Figure 3.4.3) and the data was summarized (Figure 3.4.4). The exact numbers can be found in Table A.11.

Figure 3.4.3: Example of scored *Arabidopsis* seedling phenotypes. (aff. cots: affected cotyledons but wild-type rosettes)

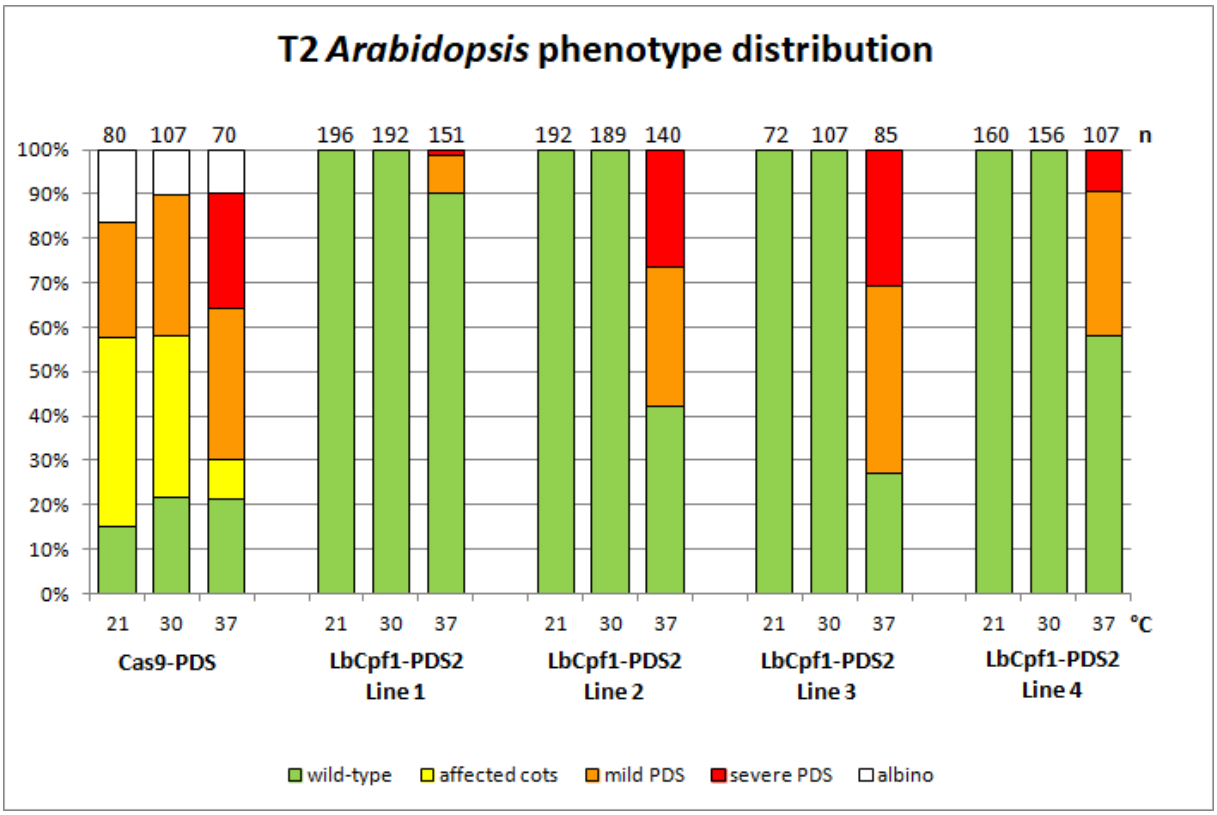

Figure 3.4.4: Visual distribution of the scored T2 *Arabidopsis* seedlings after different conditions. T2 *Arabidopsis* seedling plates were divided into three and each set was subjected to a different heat treatment temperature (21°C, 30°C or 37°C). The survived seedlings (n) were then scored according to their phenotype. Wild-type: no white spots were observed; affected cots: only the cotyledons were (partially) white while the emerging rosette leaves looked wild-type; mild PDS if white spots were present but covering less than 30% of total leaf surface; severe PDS if the white spots or parts covered over 30% of the leaf surface; albino if no pigment was observed in the seedling.

Bleached phenotypes only arose in LbCpf1-PDS2 heat-treated plants at 37°C and not at 30°C (Figure 3.4.4). The proportion of affected plants at 37°C varied between the lines. Line 3 is the most efficient line, but the sample size was reduced due to fungal growth.

Cas9-PDS plants showed different phenotypes at 21°C, including albinos. It was typical for the Cas9-PDS plants that there were seedlings showing affected cotyledons while emerging rosette leaves did not seem affected. This amount was strongly reduced for the seedlings with 37°C heat cycles, while severe PDS phenotypes emerged instead, which were previously not observed at the lower temperatures.

3.4.3. No clear hereditability of Cpf1-induced mutations in T2 *Arabidopsis*

To see if mutations induced in the T2 plants were heritable, the T3 seed was collected from the T2 population in the pots of each individual line and seed was sown for germination.

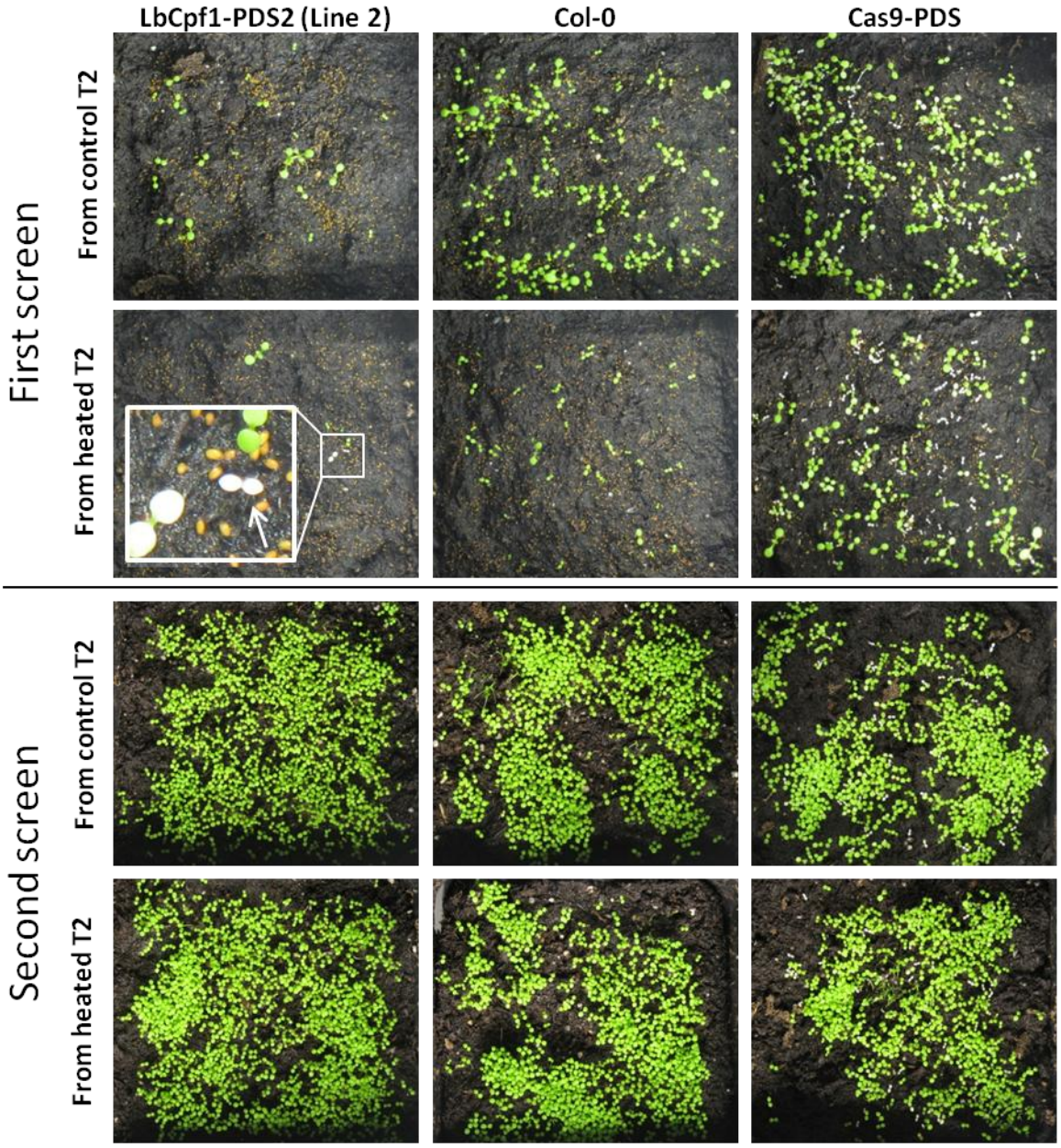

Figure 3.4.5: *Arabidopsis* T3 seedlings. *Arabidopsis* T3 seeds were harvested from heated and control T2 population and were sown on soil for germination. An albino seedling is indicated with a white arrow on a magnified window. A second T3 screen was done as the first screen had a low germination rate.

Our first *Arabidopsis* T3 screen had a low germination rate (Figure 3.4.5). Despite this, we were able to see albinos amongst the germinated seedlings, indicating heritability of the mutations. We expected to see more albino seedlings at higher germination rates. However after repeating the T3 screen, no albinos were found anymore despite the higher germination rate. Unfortunately no conclusion can be made based on this data.

3.5. Functionality of CRISPR-Cpf1 was not affected in tomato at 37°C

Encouraged by having found a heat-dependent, well-working line in *Arabidopsis*, we wondered if the same would be the case for the tomato plants. Therefore, we sampled T0 tomato plants at two timepoints including a set that was subjected to a heat treatment as illustrated in Figure 3.5.1. Four days after the last heat cycle, again young leaves were collected. The measured poplar greenhouse temperatures are graphed in Figure 3.5.2.

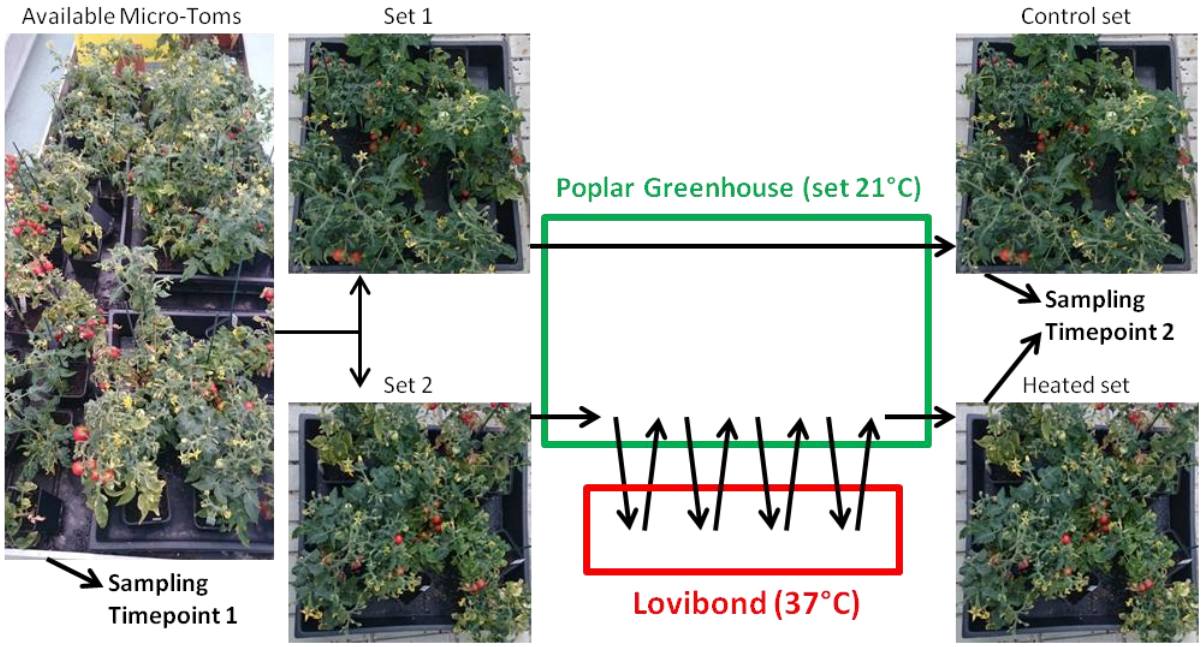

Figure 3.5.1: Illustration of the T0 tomato plants heat experiment. Leaf samples were taken from the available T0 tomato plants at timepoint 1. The plants were then split into two sets. One set stayed in the poplar greenhouse while the other set was subjected to four consecutive heat cycles. Four days after the last heat cycle (timepoint 2) again leaf samples were collected.

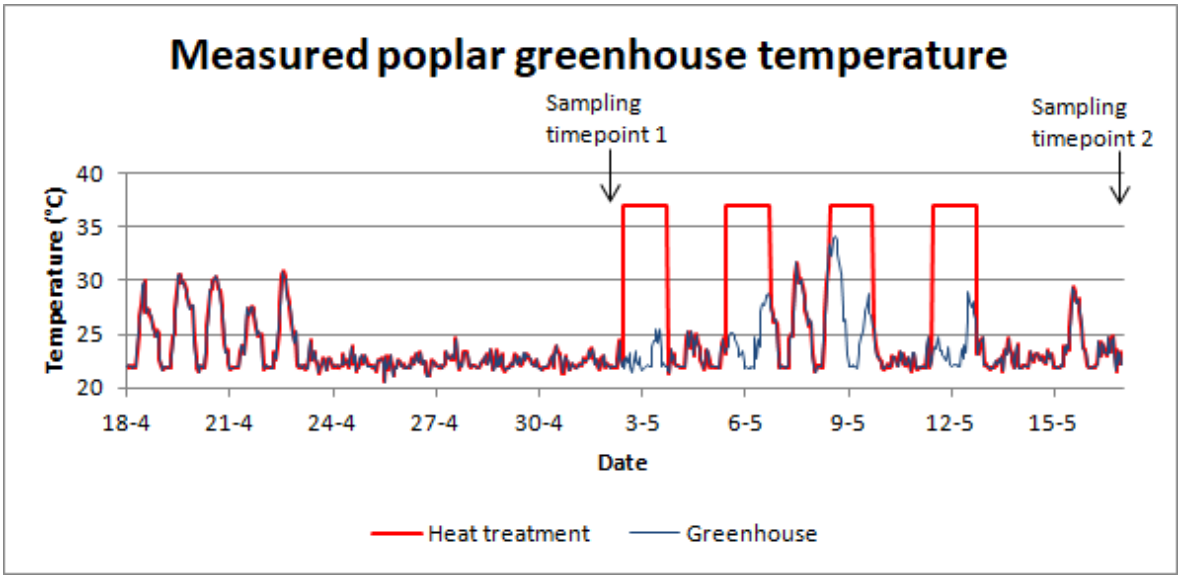

Figure 3.5.2: Measured poplar greenhouse temperature one month prior the last sampling. The tomato plants in the poplar greenhouse were divided into two: one set was subjected to a heat treatment while the other set stayed in the greenhouse. Young tomato leaf samples were collected before (Timepoint 1) and four days after the heat treatment (Timepoint 2).

The two targeted regions of all leaf samples were genotyped and total indel frequencies visualized in Figure 3.5.3 (numeric rates in Table A.12). The individual indel frequencies of samples with total indel frequency higher than 9% are summarized in Table 3.5.1.

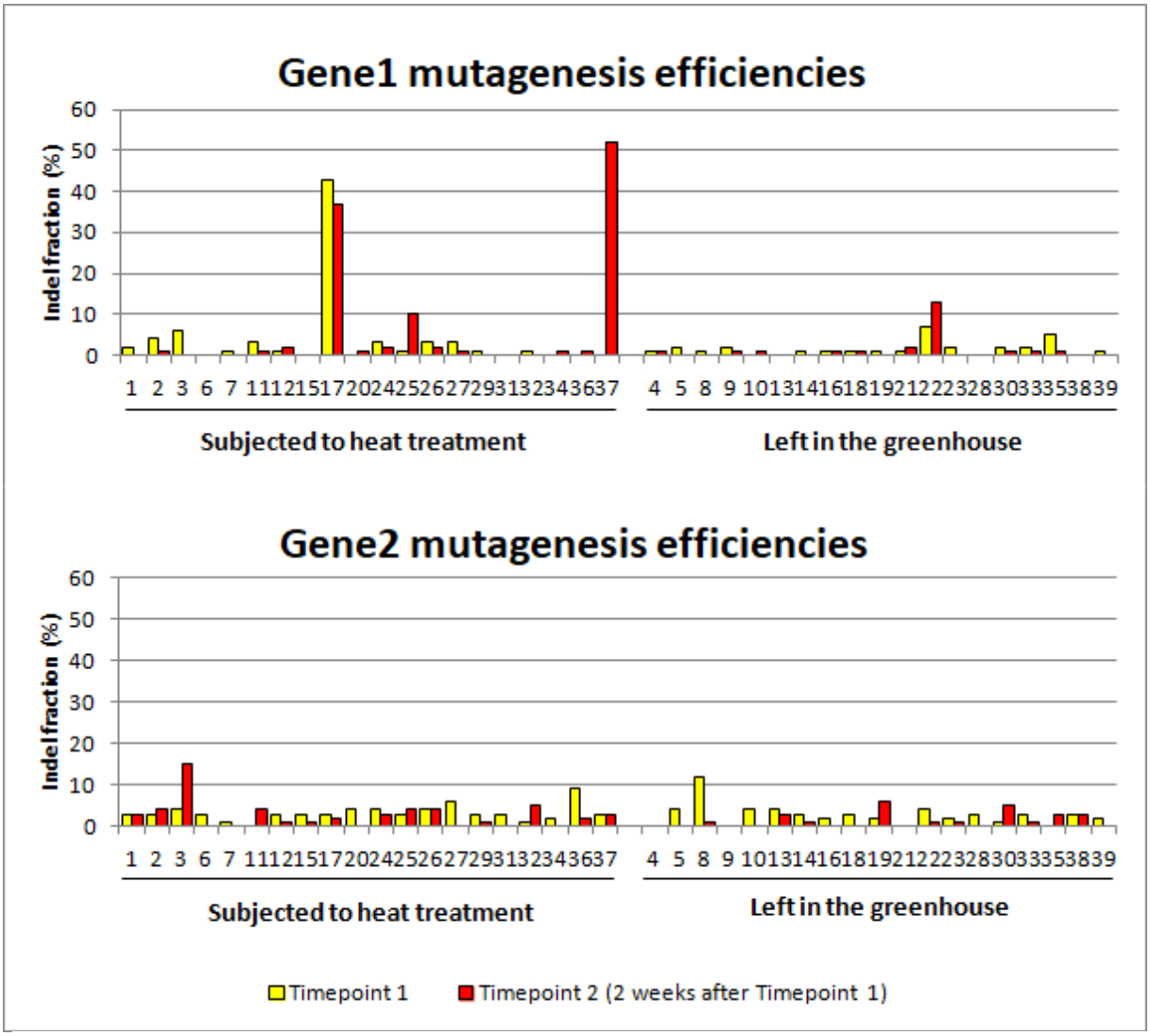

Figure 3.5.3: Total mutagenesis efficiencies (generated by ICE) of tomato leaf samples at two different timepoints. Young tomato leaves were collected (Timepoint 1). The tomato plants were then divided into two: one set stayed in the greenhouse while the other set was subjected to the heat treatment. Four days after the heat treatment, young tomato leaves were collected again (Timepoint 2). All samples were sequenced and analyzed using the ICE tool.

Table 3.5.1: Summarized ICE analysis for the genotyping to tomato leaf samples from heat-treated and non-heat-treated plants. Frequencies lower than 2% were left out.

	Sample	Total	R^2	-9	-8	-7	-6	-5	-4	-3	-2	-1	WT	1	2
Gene1	(17) FnCpf1	43	0,95				39						53		
	(17) FnCpf1 heated	37	0,96				36						59		
	(22) LbCpf1 heated	13	0,99						12				87		
	(25) LbCpf1 heated	10	1	2,7			3,8						90		
	(37) LbCpf1 heated	52	0,96				51						43		
Gene2	(8) FnCpf1	12	0,98										86		4,9
	(36) LbCpf1	9	0,99										90	5,9	
	(3) FnCpf1 heated	15	0,97						3,2	3,3			82		

One mutated tomato plant (Sample 17: plant 21 with FnT1G1/T2G2 construct) line was found prior to any heat treatment (Figure 3.5.3). The presence of these mutations is consistent after the heat treatment. Sample 37 (plant 20 with LbT1G1/T2G2 construct) is the only sample that showed mutations only after the heat treatment.

All other samples showed much lower indel frequencies even after a heat treatment. One could see a slight increase of indel frequencies in some samples upon heat treatment, but this is also shown for non-heated plants. Furthermore, there are also samples showing lower indel frequencies after heat treatment. The presence of this variability may be explained by varying qualities of sequencing chromatograms.

We expected that samples 17 and 37 would also have mutations in Gene2 as the plant lines showed to have a functional Cpf1 protein that targeted Gene1. Unfortunately, no convincing mutation frequencies were observed in Gene2, including samples 17 and 37 (Figure 3.5.3).

Before having started the heat treatments, red tomatoes were harvested and the seed stocked. By analyzing the seeds under the fluorescence microscope, red fluorescing tomato seeds can be seen (Figure 3.5.4). This proves the functionality of the FAST system, which was designed for use in *Arabidopsis*, also in tomato.

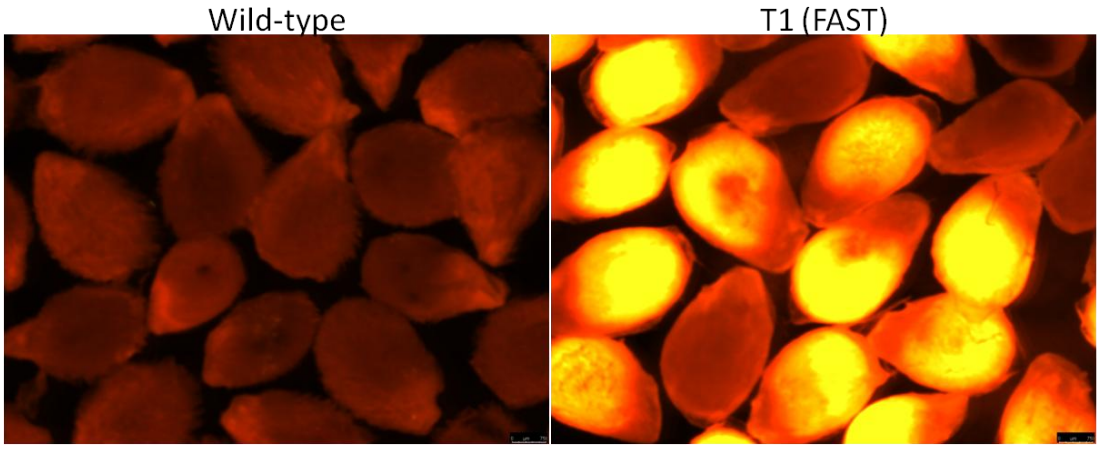

Figure 3.5.4: Tomato seeds containing the FAST (Fluorescence accumulating seed technology) cassette are fluorescent.

3.6. BY-2 cells show no GFP-Cpf1 accumulation at 37°C

Knowing that heat may be crucial for Cpf1 activity, we wondered if the GFP-Cpf1-containing BY-2 cells would show more fluorescence in heated conditions due to possibly a more stable Cpf1 protein accumulation. Therefore the previously used calli (see section 3.3.4) were brought on new medium in duplicate on two separate plates. A first set was kept at 25°C while the other set was kept at 37°C for 48 hours. The cells were analyzed with the fluorescence microscope (Figure 3.6.1).

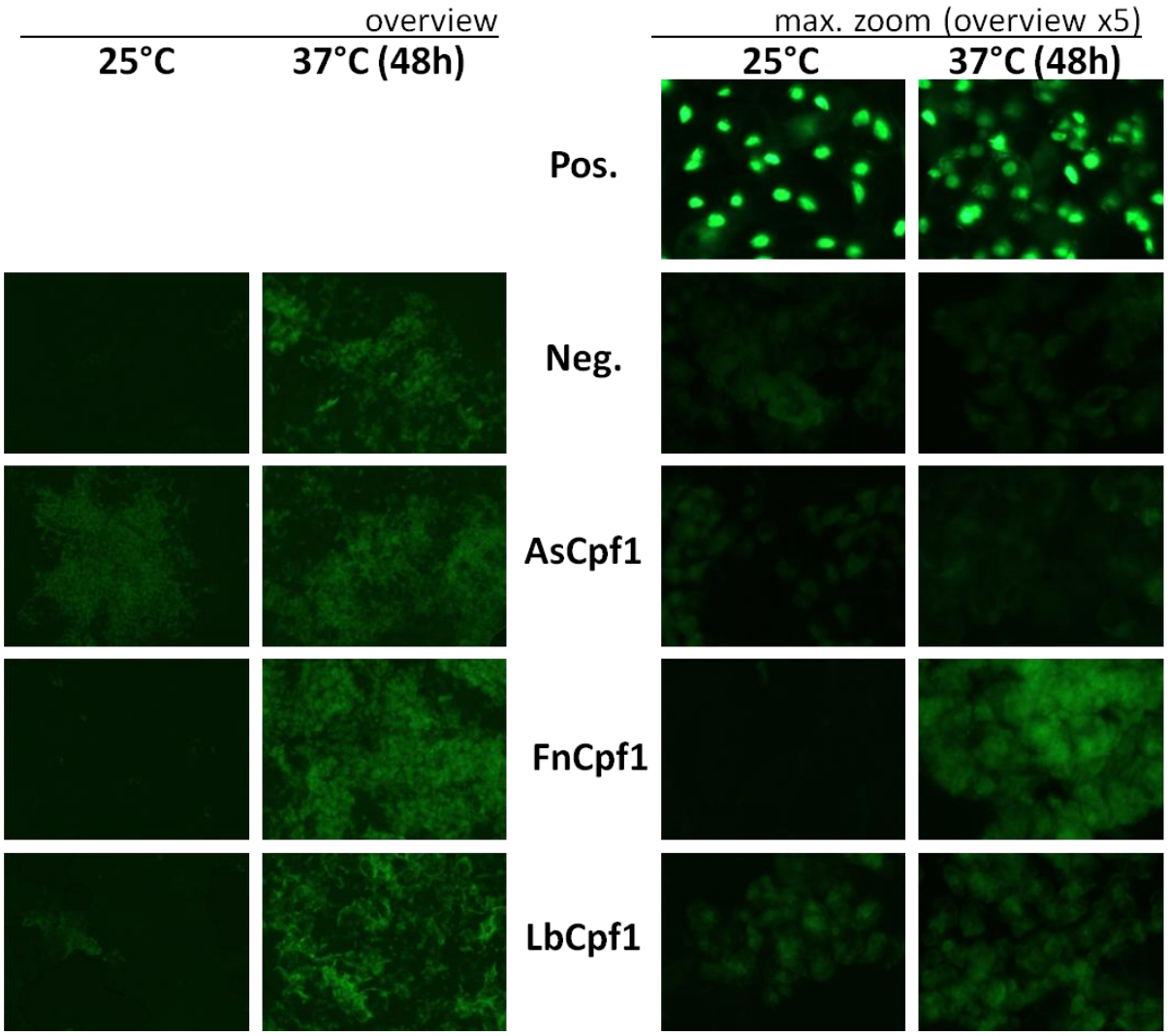

Figure 3.6.1: Fluorescence microscopy images of normal and heated BY-2 cells transformed with a GFP-Cpf1 expression construct. BY-2 cells containing the GFP-Cpf1 constructs were grown at 25°C or 37°C during 48 hours. Images were made at two different magnitudes to show overall fluorescence and cell-level fluorescence. Pos. = Positive control (Cas9 and GFP cassette); Neg. = Negative control (only Cas9 cassette)

A general increase in fluorescence can be observed upon heating BY-2 cells (Figure 3.6.1). Only the positive control (GFP-NLS cassette) shows a clear fluorescent signal in the nuclei.

4. Discussion

In this project we tested the CRISPR-Cpf1 system in *Arabidopsis* and tomato and created novel vectors to construct native crRNA arrays.

Vectors for crRNA array cloning system were successfully synthesized

As Cpf1 is able to process its own CRISPRs into mature crRNAs (Fonfara *et al.*, 2016) and multiplexing targets using single-transcript crRNA arrays proved efficient in yeast (Swiat *et al.*, 2017), we created a cloning system enabling the construction of crRNA array cassettes (Figure 3.1.2). This gave the possibility to compare the efficiency of ribozyme-mediated crRNA maturation using the same targets, however we were not able to do this comparison due to a lack of time. The entry vectors contain the *ccdB* gene between two BbsI recognition sites, which saves a lot of time as screening for target insertion would not be required anymore. Future experiments will be able to clone up to four ordered targets in native crRNA arrays to attempt optimize (multiplexing) efficiency of CRISPR-Cpf1.

No CRISPR-Cpf1 activity found in the first generation transformants

T1 *Arabidopsis* plants transformed with the nine Cpf1-PDS vectors did not show any phenotypes, while the plants with Cas9-PDS control clearly showed activity depending on the *PDS* gRNA/target (Figure 3.2.1 and Table A.8). It could have been that Cpf1 only generated in frame mutations or that the mutation rate was too low to observe a knockout phenotype. Unfortunately, DNA sequencing analysis confirmed the seedlings had no clear Cpf1-induced mutations. Lower quality chromatograms were used for TIDE analysis, showing indels with low frequencies, but also the R^2 value was quite low, indicating that the estimation may not be that accurate and thus not trustworthy. After genotyping leaves of 30 tomato plants, no convincing mutation frequencies were found either.

As we only had the outcome of 190 seedlings transformed with the Cpf1 vectors, we could assume it is not that efficient. However before assuming that our system is not functional, a larger screening population was needed. Unfortunately, due to the varying and sometimes low fluorescent seed amounts (possibly due to reported technical issues in the greenhouse), it was not possible to make a high-throughput screen for all Cpf1 constructs. Therefore the existing T1 seedlings were kept for growth until the next generation could be harvested.

Troubleshooting shows Cpf1 expression but no protein accumulation

First we investigated if the plants actually acquired the necessary transgenes to induce mutations. In *Arabidopsis* the crRNA was present in 92% of the cases, while the Cpf1 fragments seemed to be missing in multiple samples (Figure 3.3.1). By reamplifying the Cpf1 fragments, we started to obtain inconsistent results as new bands appeared while other bands disappeared (Figure 3.3.2). We then tried to amplify the full Cpf1 coding sequence and we saw no bands at all, while some samples did show a band for a smaller Cpf1 fragment previously (Figure 3.3.3). We then hypothesized that the quick *Arabidopsis* gDNA extraction protocol from Berendzen (Berendzen *et al.*, 2005) may have been too harsh and generated a quite fragmented gDNA extract. To confirm this, we tried to amplify the full Cas9 coding sequence from *Arabidopsis* samples with functional Cas9 that was extracted using the same protocol. As expected, the full Cas9 coding sequence was not able to amplify (Figure 3.3.3).

This may indicate that Berendzen extraction protocol generated a highly fragmented gDNA extract which may not be good for diagnostic purposes, making it not clear if our *Arabidopsis* samples contained the full Cpf1 coding sequence. As this protocol was specifically for *Arabidopsis*, we extracted the tomato plant leaves gDNA following the Edwards protocol (Edwards *et al.*, 1991) which is cleaner. This showed a better amplification for the Cpf1 fragment (Figure 3.3.4). Also the crRNA showed its presence in the genome.

As both transgenes were clearly present in tomato despite the lack of mutations, we thought the problem could be at the transcriptional level. We tried to show presence of the full Cpf1 cDNA by amplifying different overlapping cDNA fragments. Due to repetitive contamination of genomic DNA, we decided to repeat the RNA extraction using a different protocol which included the use of columns and DNase treatment. Using the cDNA made out of this second RNA extract, we successfully amplified all separate fragments while showing there was no gDNA contamination (Figure 3.3.6). However full-size Cpf1 cDNA still did not amplify for most samples, but did show amplification for two of them. We were later notified that we used a bad batch of iScript for RT-PCR, which could explain why longer fragments were not always able to amplify.

Despite the expression of Cpf1, no mutations were found in the plants and therefore a problem at the translational level could be the cause. As the western blots showed undetectable amounts of Cpf1 protein (Figure 3.3.7), there might be something wrong with translation. Maybe the protein is not stable and got degraded, but if Cpf1 is getting degraded then the western blot should show a signal but at lower protein sizes. One sample did seem to show a signal as a double band, however only after a doubled exposure time. The lower band may correspond to the LbCpf1 size of around 149 kDa, while the higher band seemed around the double in size. As Zetsche described possible dimerization of FnCpf1 at around 300 kDa (Zetsche *et al.*, 2015), this may also be the case for LbCpf1. However during the western blot the proteins are denatured, so dimerized proteins should not be seen on there. The signal is also quite weak still after an exposure time of over 30 minutes. If the Cpf1 protein is present, it would be in very low amounts.

As an independent experiment to detect protein expression and localization, we transformed BY-2 cells with GFP-Cpf1 translational fusion constructs. The positive control containing a separate GFP cassette shows a very clear signal inside the cell nucleus while no fluorescing GFP-Cpf1 transformed BY-2 calli were found except for one chimeric callus out of eighteen calli transformed with the GFP-AsCpf1 (Figure 3.3.8). Despite this low protein accumulation, we can see that fluorescence appears to accumulate inside the nuclei, showing that the NLS sequence is functional. Some fluorescence can be observed in the cytoplasm, but we assume this is due to autofluorescence which is not observable in the positive control due to a higher signal-to-noise contrast. We again see that the Cpf1 protein does not show a consistent and clear accumulation inside the cell.

Here, we use a human codon-optimized Cpf1, while all papers reporting efficient CRISPR-Cpf1 mutagenesis in plants used plant or even species codon-optimized Cpf1. Therefore we hypothesize that codon-optimization may be critical to obtain consistent Cpf1 activity. Therefore we suggest further experimentation with plant-optimized Cpf1. Although codon-optimization may not be sufficient for a significant increase in protein abundance, it could be

sufficient to make CRISPR-Cpf1 functional by reaching a certain protein abundance threshold.

***Arabidopsis* T2 screen revealed temperature-dependent CRISPR-Cpf1 activity**

In parallel with the troubleshooting experiments, *Arabidopsis* T2 seeds were harvested from the previously-phenotypically screened T1 plants. Over 1000 seeds per line were put on soil for germination. As by then papers were analyzed indicating that temperature affects Cpf1 and Cas9 (LeBlanc *et al.*, 2018; Moreno-Mateos *et al.*, 2017), we included a heat treatment in the experiment as described by LeBlanc.

One week after germination without any heat treatment, one single plant showed a chimeric phenotype with the FnPDS2 construct. We transplanted this in an individual pot for further growth. If the CRISPR-Cpf1 system appears to be functional once genotyped, the efficiency would be very low as only one plant out of thousands was found.

Four days after the heat treatment, aberrant phenotypes were observed in the heat-treated plants compared to the non-heat-treated plants. An important note to be made is that different reporter genes should be considered, as targeting *PDS* can give a difficult to distinguish phenotype especially when stressing the plants with heat. This could generate, when expecting *PDS*-deficient leaves, false positive phenotypes due to factors like heat stress or fungal growth. After a first genotyping screen on all four lines containing the LbCpf1-PDS2 construct showing whitish spots on the rosette leaves after heat treatment (Figure 3.4.3), mutations were found at the expected cutting site. We saw that no mutations were induced despite powdery whitish or light green leaves in other lines. A more thorough sampling was done for genotyping, confirming these results. The mutations induced were mainly deletions up to 10 nucleotides (Table 3.4.1), confirming that Cpf1 targeting tends to generate rather deletions than insertions (Kim *et al.*, 2017). This may be due to the staggered cleavage pattern of Cpf1, where the overhangs may get trimmed and directly causing deletions.

A limited comparison was done between the mutation analysis tools TIDE and ICE (Brinkman *et al.*, 2014; Hsiau *et al.*, 2018) (Tables 3.4.1, 3.4.2, A.9 and A.10). We compared some individual indel frequencies of the samples and saw that both methods appear to largely agree, although some minor frequency differences can be observed. Remarkably, when aligning the indel frequencies, we saw that for some samples TIDE gave results that made the sample seem to shift in indel frame, while ICE showed a perfect consistent pattern. Despite ICE is not designed for Cpf1, one may generally prefer to use ICE to analyze batches of samples as it is time-saving and gives a plausible estimation of indel distribution. However this was only tested for two targets, making it not possible to generalize our findings.

As there were hot days during the T2 growth and heat seemed to be a crucial factor for Cpf1 functionality, we collected the measured temperature data. During the heat treatment we can see that the control condition received warmer temperature peaks up to 33°C (Figure 3.4.2), however this did not seem sufficient to induce mutations probably due to the quite short time period of heat or heat temperatures still below a threshold to induce mutations.

The screening demonstrates that our cloning system works in *Arabidopsis*. However to function, a heat treatment was necessary. Different researchers reported about efficient use of Cpf1 in plants, but these were on other plants (tobacco and rice) using continuous growth

temperatures above 25°C up to even 30°C (Endo *et al.*, 2016). One report about using dCpf1 in *Arabidopsis* for transcriptional repression was published where the *Arabidopsis* were grown at 25°C (Tang *et al.*, 2017). No reports about Cpf1 inducing targeted mutations in *Arabidopsis* were found. The reason why there are so less reports about the use of Cpf1 in *Arabidopsis* may be due to the undetectable functionality at the growth temperature of 21°C, abandoning the subject. Eventually we could suggest further experiments where the *Arabidopsis* plants are grown in continuous higher growth temperatures, but of course there is a trade-off due to the subjected stress.

We wanted to get CRISPR-Cpf1 efficiency rates and compare the efficiency with an in-between heat treatment temperature of 30°C. As not much time was left, we had to shorten the heat treatment. Therefore we have sown a known number of T2 seeds on plate in triplicate for the lines showing mutations and treated each set with two cycles (instead of four) on set temperatures (21°C, 30°C and 37°C). The obtained data here was in line with previous results showing that 37°C heat cycles was required to get mutant phenotypes while the greenhouse temperatures were not sufficient. With the new data we show that two heat cycles were sufficient to get a good mutagenesis rate of Cpf1. More importantly, the data suggests that cycles set at 30°C is not sufficient to produce mutant phenotypes (Figure 3.4.4), suggesting that the temperature threshold is higher. This makes sense that there were still no mutants seen in the previous T2 screen despite some high temperature peaks. The temperature peaks may have not lasted long enough at a yet unknown threshold temperature above 30°C. Also did we note after the scoring of the Cas9-PDS phenotypes that more severe *PDS*-deficient plants were observed. This confirms that the nuclease may be more efficient at higher temperatures in plants as suggested by LeBlanc.

To investigate the hereditability of the mutations induced in the T2 generation, a T3 screen was performed. During a first screen, germination of T3 *Arabidopsis* seeds from heated T2 lines showed some albino seedlings (Figure 3.4.5), indicating that mutations in *PDS* were heritable. Unfortunately many seeds did not germinate, which may be caused by the too early seed harvesting that was done to be able to do this final experiment. When repeating the T3 screen, no albino seedlings were observed anymore despite the increased germination rate. No conclusion can be made about the mutation hereditability here.

CRISPR-Cpf1 in tomato does not seem affected by heat

As CRISPR-Cpf1 was functional in some *Arabidopsis* lines after the heat treatment, we were encouraged to also subject half of our growing tomato plants to the same heat treatment. We wanted to genotype each plant before and after the eventual heat treatment to investigate if indel frequencies would change in heat-treated plants. One plant showed clear Cpf1-induced mutations already before any heat treatment, and was maintained after the heat treatment (Figure 3.5.3). This indicated that our system can be functional in tomato without needing any heat treatment, seemingly with a low efficiency. This plant contained the FnCpf1, showing that also this homologue is functional. A larger population should have been screened earlier to have found the mutated plant. Unfortunately, only one plant containing LbCpf1 showed clear mutations after the heat treatment while no mutations were observed earlier. Small increases of total indel frequencies were observed after the heat treatment, but also many plants seemed to have a decreased total frequency. On the other hand, also some samples showed an increased total indel frequency without any heat

treatment at the second time point. This variation could be due to the varying quality of the sequencing chromatograms and we therefore cannot conclude that the heat treatment increased CRISPR-Cpf1 activity.

Interestingly, the clearly mutagenized plants using the first target generated mostly six nucleotide deletions (Table 3.5.1). As this results in a deletion of two amino acids and no frameshift, it is not an interesting target to generate knock-outs.

Our data also show that the double targeting system did not work as we wanted (Figure 3.5.1). As we thought that Cpf1 activity was the main issue, we expected that the samples containing mutations in Solyc09g065630 (*PPD* gene 1) would also have mutations in Solyc06g084120 (*PPD* gene 2) as the Cpf1 seemed to be functional in those plant cells. However this could be explained by an inefficient target choice and more targets should be tested in the future.

As the tomato plants were producing fruit, we harvested some T1 tomato seeds. Under the fluorescence microscope we clearly saw red fluorescing tomato seeds (Figure 3.5.4). This shows that, despite the FAST system was designed for *Arabidopsis*, it is also functional in tomato without requiring any adaptations as suggested by Shimada. This system is very handy as no kanamycin selection is required to grow T1 tomato plants bearing our vectors, saving time and resources.

Heated BY-2 cells do not show clear Cpf1 accumulation upon heating

As our *Arabidopsis* experiment showed that only mutations were induced in some lines after subjecting it to heat, we wondered if heat would have a stabilizing role for Cpf1. We therefore took our previously made BY-2 cells transformed with GFP-Cpf1 and brought cells on new plates to leave it during two days in the normal condition (25°C) or the heat condition (37°C). Under the fluorescent microscope we did again not see clear fluorescence accumulation inside GFP-Cpf1 BY-2 nuclei despite the heat treatment (Figure 3.6.1). We did see an overall increased fluorescence signal, but this can also be seen in the negative control (Cas9 without GFP). Therefore we think the increased fluorescence could be due to the accumulation of certain compounds caused by heat stress.

Conclusion

In this project we showed that 37°C heat stress was required to induce mutations by LbCpf1 in *Arabidopsis* while 30°C heat stress was not sufficient for mutagenesis. We showed that LbCpf1 and FnCpf1 were able to induce mutations in tomato plants, but mutagenized plants were very uncommon, even after 37°C heat stress. We would therefore tend to assume that Cpf1 functionality may be dependent on the status of the plant cell instead of the temperature. Assuming that the tomato plants were not as stressed as the *Arabidopsis* plants, we could increase the temperature for heat treatments on tomato plants. We further hypothesize that heat stress could be generalized as abiotic stress, suggesting further experiments testing functionality of CRISPR-Cpf1 in presence of other abiotic stresses like drought stress and DNA damage.

5. Materials and methods

Vector cloning

Cpf1 targets were identified according to the target motif $TTTN_{25}$ using the Geneious R11 software from Biomatters (Kearse *et al.*, 2012). The crRNA length of 24 nucleotides was based on the native crRNA length reported (Zetsche *et al.*, 2015). Three targets for the *Arabidopsis thaliana PDS3* gene (GeneID 827061) and two target for each tomato (*Solanum lycopersicum*) *PPD* (Gonzalez *et al.*, 2015)(Gene1: Solyc09g065630 and Gene2: Solyc06g084120) were chosen. Oligonucleotides with the N_{24} target sequences preceded by the relevant Cpf1 homologue-variable scaffold part (AsCpf1: CTTGTAGAT; FnCpf1: GTTGTAGAT; LbCpf1: AAGTGTAGAT (Zetsche *et al.*, 2015)) were ordered (via VIB online oligo request facility Oreol) in the way to be able to anneal these oligos and clone these inside BbsI-digested B-C and C-D (Lampropoulos *et al.*, 2013) entry vectors (PGE references p00295 and p00296) during a golden gate reaction (protocol A.13) to form crRNA vectors where the mature crRNA (consisting of Cpf1 scaffold followed by the target sequence) is flanked by the hammerhead and hepatitis delta virus ribozymes sequences (Gao and Zhao, 2014) at respectively the 5' and 3' flanks. The golden gate reaction products were transformed into competent *E. coli* DH5α cells via a heat shock (45 seconds at 42°C) and clones growing on Luria-Bertani (LB) medium with carbenicillin (Cb) (100 µg/mL) were picked for colony touch PCR to screen for the presence of the expected crRNA vectors. Positive clones were grown in liquid culture to be used for plasmid extraction using GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific). Plasmids were sent for sequencing to Eurofins Genomics (Overnight Mix2Seq Kit) and mapped in Geneious to the expected crRNA vector sequence. Clones bearing the correct crRNA vector sequences were stocked as glycerol stocks.

To create the expression vectors, golden gate reactions provided the crRNA by the RPS5 promoter and pea3AT terminator (obtained from Mansour Karimi) to form a crRNA cassette, replacing the *ccdB* in the available FASTRK vectors (PGE references p00379, p00380 and p00381). Once replaced, the binary FASTRK vectors will contain the crRNA cassette (as described above), a Cpf1 cassette (driven by PcUbi promoter and terminated by G7 terminator), a FAST cassette for fluorescence-accumulating seed technology (Shimada *et al.*, 2010) (mRuby3 driven by Ole1 promoter and terminated by NOS), and a kanamycin resistance marker. After transformation with this golden gate product in *E. coli* DH5α cells, individual clones growing on LB+Cb medium were picked for colony touch PCR and screened for the presence of golden gate reaction junctions to indicate the presence of the desired expression vectors. Positive clones were grown in liquid culture to be used for plasmid extraction as described earlier. Extracted plasmid was used for a diagnostic digest with NdeI (BioLabs for all restriction enzymes) to ensure absence of cloning errors. Clones showing the expected vector digest pattern were stocked as glycerol stocks.

The final expression vectors were transformed into *Agrobacterium tumefaciens* strains C58C1 (for *Arabidopsis* transformation) or EHA105 (for tomato transformation) via electroporation. Clones growing on YEB+Rifampicin+Gentamycin+Spectinomycin medium were picked for colony touch PCR amplifying the same golden gate junctions as for the DH5α cells to ensure presence of the correct expression vectors. Positive clones were stocked as glycerol stocks.

Geneious software was used to design new vectors for the golden gate B-C, C-D, D-E and E-F position that contain the full Cpf1 direct repeat (Zetsche *et al.*, 2015) followed by *ccdB* flanked by BbsI sites (Sequence A.5). Also was one F-G vector created that contained the HDV ribozyme sequence followed by the pea3AT terminator (Sequence A.6). The required unique fragments were synthesized by the BioXp 3200 system (SGI-DNA) and the fragment size were checked on agarose gels. When multiple bands were present, the correct band was excised from the gel and purified using the a gel purification kit (Gel DNA recovery kit, Zymoclean). The fragments were then ligated in BsaI-digested pGG vectors (PGE references p00013/14/15) via a NEBuilder HiFi DNA Assembly reaction (BioLabs). The reaction products were transformed in *E. coli ccdB*Survival2-T1 cells (Invitrogen) via a heat shock (45 seconds at 42°C). Colony touch PCR screen, plasmid extraction and sequencing was done as described earlier. The sequencing chromatograms were mapped to the designed sequences in Geneious and the clones with the correct vector sequence were stocked in glycerol.

Plant transformation and first generation growth

Floral dip transformation (Clough and Bent, 1998) of *Arabidopsis thaliana* Col-0 was done by the PSB facility and the plants were kept in the *Arabidopsis* greenhouse (set at 21°C, long day) until seeds could be harvested. Harvested seed was stocked and 25 seeds were picked, gas sterilized (chlorine), stratified (4 days at 4 °C) and placed on 1/2 MS medium plates. The seeds were analyzed with Leica stereo microscope and non-fluorescent seeds were indicated on plate. The plates were left in the long day 21°C growth chamber and pictures of the plates were taken after 11 days. Up to ten seedlings were placed in soil mixture (Potgrond voor zaaien en stekken, Saniflor) and left in the greenhouse for further growth.

Transformation of tomato variety Micro-Tom plants (Protocol A.18 from Michel Hernould group [Inra-Bordeaux, France]) was taken care by Jan Van Doorsselaere (VIVES institute, Roeselare) and the transformed plant calli placed on kanamycin-medium were gradually delivered and kept in the long day 24°C growth chamber. Rooting plants were transferred on soil mixture (Professioneel beroepspotgrond, Saniflor) and brought to the medicinal greenhouse (set at 24°C, long day) for further growth. Later the tomato plants were moved to the tobacco and poplar greenhouse (set at 21°C, long day).

Genotyping and mutation analysis

Rosette samples were taken during growth of the T1 *Arabidopsis* plants and used to extract genomic DNA (Berendzen *et al.*, 2005). The gDNA was used as template to PCR amplify the targeted region of *PDS* using forward GACGTCAGGAAGAACATGGTC and reverse AACACCTCGTCGGTCACGC primers. Gel electrophoresis analysis was done with a part of the PCR products to confirm amplification and visualize eventual large indels. The remaining part of the PCR products was purified following the HighPrep PCR Clean Up protocol (MagBio Genomics) and sent for sequencing to Eurofins Genomics using the forward primer. The obtained sequencing chromatograms were mapped to the *PDS* gene in Geneious to have a quick visualization of mutations. Suspected overlapping chromatograms were further analyzed using the online TIDE and ICE tools (Brinkman *et al.*, 2014; Hsiau *et al.*, 2018).

Genomic DNA was extracted from tomato leaf samples (Edwards *et al.*, 1991). The *PPD* targeting regions were PCR amplified using forward CAGCGGAAAAAGCGTCACAT and reverse ATTTGGTAGGGACGGCCAAC primers for Gene1 and forward TCACAGGCGATTCAGCAAGT and

reverse TGATAGGTAGAAGCTAGCCGC primers for Gene2. Downstream steps were identical as described for *Arabidopsis* except that sequencing chromatograms were mapped on the relevant *PPD* gene.

Troubleshooting

All used primers in this section can be found in addendum Table A.14.

The extracted gDNA from the *Arabidopsis* and tomato samples were used as template to PCR amplify fragments of the Cpf1 and crRNA transgenes. The PCR products were run on an electrophoresis gel to visualize the amplified products.

A first RNA extraction was done using the Shengben (2009) protocol (Protocol A.15) while a second RNA extraction was prepared following the 'RNAprep - Trizol combined with Columns' protocol from Untergasser A. (2008) (http://www.molbi.de/protocols/rna_prep_comb_trizol_v1_0.htm). Extracted RNA quality was visualized on agarose gel. DNase treatment (after the extraction protocol or during the protocol) was done following the RQ1 DNase treatment protocol (Promega). The DNase-treated RNA was used as template to synthesize cDNA using the iScript cDNA synthesis kit following manufacturer protocol (Bio-Rad). The cDNA was used as template to PCR amplify transgene and reference gene fragments**.** The PCR amplification was initially done with standard Taq polymerase (MyTaq Red Mix, Bioline) and later switched to the high-fidelity iProof kit (Bio-Rad).

Tomato leaf samples protein extraction, SDS-PAGE and western blot were done following protocol A.16.

BY-2 cell transformed with GFP-Cpf1

GFP-Cpf1 cassettes were built via golden gate reactions (PGE plasmids p00339, p00043, p00677, p00290/p00291/p00292, p00229, p00044 and p00045) providing exactly the same Cpf1 cassette of the FASTRK vectors except for the B-C position which now contains GFP. The reaction was used to transform *E. coli* DH5α cells and colonies were picked for colony touch PCR screening using forward GGACTTGTGCGACATGTCGTTTTC and reverse GTTTACGTCGCCGTCCAGCT primers. Diagnostic digests were done using NheI and NotI to check if no major cloning errors occurred. Correct constructs were transformed in *A. tumefaciens C58C1* and colonies were screened as described for *E. coli*. Positive clones were stocked in glycerol. BY-2 cell culturing and transformation were done following Protocol A.17. *Agrobacteria* stocks PGE00114 (Cas9 without GFP) and PGE00441 (Cas9 with GFP cassette) were used as negative and positive control respectively (obtained from Ward Decaestecker). Formed calli were brought on fresh medium *in duplo* for each construct. The first set was incubated at 25°C while the other set at 37°C. After 48 hours, parts of calli were brought on a glass slide with 2 drops of MQ water and covered with a thin glass cover. The slides were analyzed under the Leica stereo microscope and images were taken.

T2 *Arabidopsis* seedling screen including heat treatment

Arabidopsis T2 seeds were harvested and for each available line around 20 mg of seeds were brought in tubes, stratified during 4 days at 4°C and spread over pots with water-saturated soil mixture (Potgrond voor zaaien en stekken, Saniflor) for a total of two pots per line. The first set was left in the *Arabidopsis* greenhouse chamber, while the other set received four

consecutive cycles of heat treatment from the eighth day on. A heat treatment cycle consists of 30 hours growth at 37°C where after 42 hours in the greenhouse to allow recovery (LeBlanc *et al.*, 2018). After the fourth heat treatment cycle, the treated set was kept in the greenhouse. Four days after the last heat cycle, samples from suspected phenotypes were collected for genotyping and mutation analysis as described in the 'Genotyping and mutation analysis' section except that the gDNA was extracted following the Edwards protocol. Six days after the sample collection, a more thorough sampling was done by taking (suspected) 3 non-heated and heated rosette leaves from each suspected line and mutations were analyzed.

Determining Cpf1 targeting efficiency of T2 *Arabidopsis* seedlings

For the Cas9-PDS line and each LbCpf1-PDS2 line, one hundred sterilized and stratified T2 *Arabidopsis* seeds were placed on 1/2 MS medium plates containing timentin for a total of six plates per line. Using the fluorescence microscope, non-fluorescing seeds were indicated on the plate. All plates were kept in the long day 21°C growth chamber during eight days. From the eighth day on, two plates of each line was placed in the 21°C growth chamber wrapped with aluminium paper, in a 30°C incubator and a 37°C incubator. This treatment at different temperatures was in dark conditions and continued for 30 hours, afterwards all the plates were brought back in the long day 21°C growth chamber back exposed to light for a recovery of 42 hours. After the recovery, a second and last treatment cycle was done. After this second cycle, the plates were kept for recovery in the growth chamber for 9 more days. Seedlings were phenotypically scored according to the proportion of bleached area on total leaf surface.

T3 *Arabidopsis* screen

T3 *Arabidopsis* seeds were harvested from T2 lines showing mutations (seeds derived from the whole grown population in pot, each line and condition stocked separately). Approximately 20 mg of seeds of selected lines was brought in tubes, stratified during 3 days at 4°C, spread over pots with water-saturated soil mixture (Potgrond voor zaaien en stekken, Saniflor) and kept in the *Arabidopsis* greenhouse for germination. The second screen was done by stratifying the seeds during 4 days on water-saturated soil.

Tomato heat treatment

Young leaf samples were collected from the growing T0 tomato plants. The plants were then divided into two sets. One set was left in the tobacco and poplar greenhouse while the other set started immediately the equal heat treatment as described previously ('T2 Arabidopsis seedling screen including heat treatment'). Four days after the last heat cycle, again young leaf samples were collected from all plants of both sets. Genotyping and mutation analysis of both leaf samples from tomato plants were done as described above ('Genotyping and mutation analysis').

References

Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DB, Shmakov S, Makarova KS, Semenova E, Minakhin L, Severinov K, Regev A, Lander ES, Koonin EV, Zhang F (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353:** aaf5573

Bai L, Singh M, Pitt L, Sweeney M, Brutnell TP (2007) Generating novel allelic variation through Activator insertional mutagenesis in maize. *Genetics* **175:** 981-992

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

Bayat H, Omidi M, Rajabibazl M, Sabri S, Rahimpour A (2017) The CRISPR Growth Spurt: from Bench to Clinic on Versatile Small RNAs. *Journal of microbiology and biotechnology* **27:** 207-218

Berendzen K, Searle I, Ravenscroft D, Koncz C, Batschauer A, Coupland G, Somssich IE, Ulker B (2005) A rapid and versatile combined DNA/RNA extraction protocol and its application to the analysis of a novel DNA marker set polymorphic between Arabidopsis thaliana ecotypes Col-0 and Landsberg erecta. *Plant methods* **1:** 4

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research* **41:** 7429-7437

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151:** 2551-2561

Bortesi L, Zhu C, Zischewski J, Perez L, Bassie L, Nadi R, Forni G, Lade SB, Soto E, Jin X, Medina V, Villorbina G, Munoz P, Farre G, Fischer R, Twyman RM, Capell T, Christou P, Schillberg S (2016) Patterns of CRISPR/Cas9 activity in plants, animals and microbes. *Plant biotechnology journal* **14:** 2203-2216

Brinkman EK, Chen T, Amendola M, van Steensel B (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic acids research* **42:** e168

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321:** 960-964

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant journal : for cell and molecular biology* **16:** 735-743

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339:** 819-823

Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature communications* **3:** 945

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

Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of bacteriology* **190:** 1390-1400

DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic acids research* **41:** 4336-4343

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J, Root DE (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology* **34:** 184-191

Dong D, Ren K, Qiu X, Zheng J, Guo M, Guan X, Liu H, Li N, Zhang B, Yang D, Ma C, Wang S, Wu D, Ma Y, Fan S, Wang J, Gao N, Huang Z (2016) The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532:** 522-526

Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic acids research* **19:** 1349

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

Fonfara I, Richter H, Bratovic M, Le Rhun A, Charpentier E (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532:** 517-521

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31:** 822-826

Gao L, Cox DBT, Yan WX, Manteiga JC, Schneider MW, Yamano T, Nishimasu H, Nureki O, Crosetto N, Zhang F (2017) Engineered Cpf1 variants with altered PAM specificities. *Nature biotechnology* **35:** 789-792

Gao P, Yang H, Rajashankar KR, Huang Z, Patel DJ (2016) Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell research* **26:** 901-913

Gao Y, Zhao Y (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *Journal of integrative plant biology* **56:** 343-349

Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468:** 67-71

Gonzalez N, Pauwels L, Baekelandt A, De Milde L, Van Leene J, Besbrugge N, Heyndrickx KS, Cuellar Perez A, Durand AN, De Clercq R, Van De Slijke E, Vanden Bossche R, Eeckhout D, Gevaert K, Vandepoele K, De Jaeger G, Goossens A, Inze D (2015) A Repressor Protein Complex Regulates Leaf Growth in Arabidopsis. *The Plant cell* **27:** 2273-2287

Gorbunova V, Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic acids research* **25:** 4650-4657

Grissa I, Vergnaud G, Pourcel C (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic acids research* **35:** W52-57

Groenen PM, Bunschoten AE, van Soolingen D, van Embden JD (1993) Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. *Molecular microbiology* **10:** 1057-1065

Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology* **1:** e60

Haft DH, Selengut JD, White O (2003) The TIGRFAMs database of protein families. *Nucleic acids research* **31:** 371-373

Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R (2008) Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *Journal of bacteriology* **190:** 1401-1412

Hsiau T, Maures T, Waite K, Yang J, Kelso R, Holden K, Stoner R (2018) Inference of CRISPR Edits from Sanger Trace Data. *bioRxiv*

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31:** 827-832

Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, Zeina CM, Gao X, Rees HA, Lin Z, Liu DR (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556:** 57-63

Hu X, Wang C, Liu Q, Fu Y, Wang K (2017) Targeted mutagenesis in rice using CRISPR-Cpf1 system. *Journal of genetics and genomics = Yi chuan xue bao* **44:** 71-73

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

Jansen R, Embden JD, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* **43:** 1565-1575

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic acids research* **41:** e188

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

Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J (1997) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *Journal of clinical microbiology* **35:** 907-914

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28:** 1647-1649

Kim D, Kim J, Hur JK, Been KW, Yoon SH, Kim JS (2016) Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nature biotechnology* **34:** 863-868

Kim H, Kim ST, Ryu J, Kang BC, Kim JS, Kim SG (2017) CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nature communications* **8:** 14406

Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America* **93:** 1156-1160

Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK (2016a) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529:** 490-495

Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, Gonzales AP, Li Z, Peterson RT, Yeh JR, Aryee MJ, Joung JK (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523:** 481-485

Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM, Lopez JM, McCaw ZR, Aryee MJ, Joung JK (2016b) Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nature biotechnology* **34:** 869-874

Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533:** 420-424

Koonin EV, Makarova KS, Zhang F (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Current opinion in microbiology* **37:** 67-78

Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an insertional mutagen in Arabidopsis. *The Plant cell* **11:** 2283-2290

Kunin V, Sorek R, Hugenholtz P (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome biology* **8:** R61

Lampropoulos A, Sutikovic Z, Wenzl C, Maegele I, Lohmann JU, Forner J (2013) GreenGate---a novel, versatile, and efficient cloning system for plant transgenesis. *PloS one* **8:** e83043

LeBlanc C, Zhang F, Mendez J, Lozano Y, Chatpar K, Irish VF, Jacob Y (2018) Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. *The Plant journal : for cell and molecular biology* **93:** 377-386

Lei C, Li SY, Liu JK, Zheng X, Zhao GP, Wang J (2017) The CCTL (Cpf1-assisted Cutting and Taq DNA ligase-assisted Ligation) method for efficient editing of large DNA constructs in vitro. *Nucleic acids research* **45:** e74

Li Y, Rosso MG, Ulker B, Weisshaar B (2006) Analysis of T-DNA insertion site distribution patterns in Arabidopsis thaliana reveals special features of genes without insertions. *Genomics* **87:** 645-652

Ma N, Liao B, Zhang H, Wang L, Shan Y, Xue Y, Huang K, Chen S, Zhou X, Chen Y, Pei D, Pan G (2013) Transcription activator-like effector nuclease (TALEN)-mediated gene correction in integration-free beta-thalassemia induced pluripotent stem cells. *The Journal of biological chemistry* **288:** 34671-34679

Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (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:** 7

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV (2011) Evolution and classification of the CRISPR-Cas systems. *Nature reviews Microbiology* **9:** 467-477

Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJ, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews Microbiology* **13:** 722-736

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* **339:** 823-826

Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322:** 1843-1845

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD, Rebar EJ (2011) A TALE nuclease architecture for efficient genome editing. *Nature biotechnology* **29:** 143-148

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155:** 733-740

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* **60:** 174-182

Mojica FJ, Diez-Villasenor C, Soria E, Juez G (2000) Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology* **36:** 244-246

Mojica FJ, Ferrer C, Juez G, Rodriguez-Valera F (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:** 85-93

Mojica FJ, Juez G, Rodriguez-Valera F (1993) Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Molecular microbiology* **9:** 613-621

Moreno-Mateos MA, Fernandez JP, Rouet R, Vejnar CE, Lane MA, Mis E, Khokha MK, Doudna JA, Giraldez AJ (2017) CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nature communications* **8:** 2024

Neuffer MG, Coe EH (1978) Paraffin oil technique for treating mature corn pollen with chemical mutagens. *Maydica* **v. 23**

Nishimasu H, Yamano T, Gao L, Zhang F, Ishitani R, Nureki O (2017) Structural Basis for the Altered PAM Recognition by Engineered CRISPR-Cpf1. *Molecular cell* **67:** 139-147 e132

Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA (2014) Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature structural & molecular biology* **21:** 528-534

Nunez JK, Lee AS, Engelman A, Doudna JA (2015) Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519:** 193-198

Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R (2009) Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 5019-5024

Pourcel C, Salvignol G, Vergnaud G (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151:** 653-663

Pozzoli U, Menozzi G, Fumagalli M, Cereda M, Comi GP, Cagliani R, Bresolin N, Sironi M (2008) Both selective and neutral processes drive GC content evolution in the human genome. *BMC evolutionary biology* **8:** 99

Puchta H (1999) Use of I-Sce I to induce DNA double-strand breaks in Nicotiana. *Methods Mol Biol* **113:** 447-451

Puchta H, Hohn B (1991) The mechanism of extrachromosomal homologous DNA recombination in plant cells. *Molecular & general genetics : MGG* **230:** 1-7

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, Zhang F (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154:** 1380-1389

Rinehart TA, Dean C, Weil CF (1997) Comparative analysis of non-random DNA repair following Ac transposon excision in maize and Arabidopsis. *The Plant journal : for cell and molecular biology* **12:** 1419-1427

Samai P, Pyenson N, Jiang W, Goldberg GW, Hatoum-Aslan A, Marraffini LA (2015) Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **161:** 1164-1174

Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic acids research* **39:** 9275-9282

Schell J, Van Montagu M (1977) The Ti-plasmid of Agrobacterium tumefaciens, a natural vector for the introduction of nif genes in plants? *Basic life sciences* **9:** 159-179

Schunder E, Rydzewski K, Grunow R, Heuner K (2013) First indication for a functional CRISPR/Cas system in Francisella tularensis. *International journal of medical microbiology : IJMM* **303:** 51-60

Seligman LM, Chisholm KM, Chevalier BS, Chadsey MS, Edwards ST, Savage JH, Veillet AL (2002) Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic acids research* **30:** 3870-3879

Shimada TL, Shimada T, Hara-Nishimura I (2010) A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana. *The Plant journal : for cell and molecular biology* **61:** 519-528

Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science* **351:** 84-88

Smargon AA, Cox DBT, Pyzocha NK, Zheng K, Slaymaker IM, Gootenberg JS, Abudayyeh OA, Essletzbichler P, Shmakov S, Makarova KS, Koonin EV, Zhang F (2017) Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Molecular cell* **65:** 618-630 e617

Sorek R, Lawrence CM, Wiedenheft B (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual review of biochemistry* **82:** 237-266

Stadler LJ (1928) Genetic Effects of X-Rays in Maize. *Proceedings of the National Academy of Sciences of the United States of America* **14:** 69-75

Su XZ, Miller LH (2015) The discovery of artemisinin and the Nobel Prize in Physiology or Medicine. *Science China Life sciences* **58:** 1175-1179

Swiat MA, Dashko S, den Ridder M, Wijsman M, van der Oost J, Daran JM, Daran-Lapujade P (2017) FnCpf1: a novel and efficient genome editing tool for Saccharomyces cerevisiae. *Nucleic acids research* **45:** 12585-12598

Tang X, Lowder LG, Zhang T, Malzahn AA, Zheng X, Voytas DF, Zhong Z, Chen Y, Ren Q, Li Q, Kirkland ER, Zhang Y, Qi Y (2017) A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nature plants* **3:** 17103

Tang X, Zheng X, Qi Y, Zhang D, Cheng Y, Tang A, Voytas DF, Zhang Y (2016) A Single Transcript CRISPR-Cas9 System for Efficient Genome Editing in Plants. *Molecular plant* **9:** 1088-1091

Toth E, Weinhardt N, Bencsura P, Huszar K, Kulcsar PI, Talas A, Fodor E, Welker E (2016) Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells. *Biology direct* **11:** 46

Tu M, Lin L, Cheng Y, He X, Sun H, Xie H, Fu J, Liu C, Li J, Chen D, Xi H, Xue D, Liu Q, Zhao J, Gao C, Song Z, Qu J, Gu F (2017) A 'new lease of life': FnCpf1 possesses DNA cleavage activity for genome editing in human cells. *Nucleic acids research* **45:** 11295-11304

Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **17:** 904-912

Xiang G, Zhang X, An C, Cheng C, Wang H (2017) Temperature effect on CRISPR-Cas9 mediated genome editing. *Journal of genetics and genomics = Yi chuan xue bao* **44:** 199-205

Xie K, Minkenberg B, Yang Y (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences of the United States of America* **112:** 3570-3575

Xu R, Qin R, Li H, Li D, Li L, Wei P, Yang J (2017) Generation of targeted mutant rice using a CRISPR-Cpf1 system. *Plant biotechnology journal* **15:** 713-717

Yamano T, Nishimasu H, Zetsche B, Hirano H, Slaymaker IM, Li Y, Fedorova I, Nakane T, Makarova KS, Koonin EV, Ishitani R, Zhang F, Nureki O (2016) Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165:** 949-962

Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic acids research* **40:** 5569-5576

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163:** 759-771

Zhang F, Maeder ML, Unger-Wallace E, Hoshaw JP, Reyon D, Christian M, Li X, Pierick CJ, Dobbs D, Peterson T, Joung JK, Voytas DF (2010) High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 12028-12033

Addendum

Table A.1: Ordered oligos to target sequences of interest. In the row with ordered sequences: normal sequence style are the required overhangs (upon annealing) to ligate inside BbsI-digested entry vectors; bold sequence style represents the Cpf1-homologue-dependent scaffold part; underlined sequence style represents target sequence.

	Alias	Target Sequence (24 nt)	Ordered sequences (group stock number)
Arabidopsis *PDS*	LbPDS1	CCAGCCATGGTCGGCGGTCAGGCT	TACT**AAGTGTAGAT**CCAGCCATGGTCGGCGGTCAGGCT (316)
			GGCCAGCCTGACCGCCGACCATGGCTGG**ATCTACACTT** (325)
	LbPDS2	TCAGTCAAAGAATGGATGGAAAAG	TACT**AAGTGTAGAT**TCAGTCAAAGAATGGATGGAAAAG (317)
			GGCCCTTTTCCATCCATTCTTTGACTGA**ATCTACACTT** (326)
	LbPDS3	TCTGGCCATGTCAGCATCTCGTTG	TACT**AAGTGTAGAT**TCTGGCCATGTCAGCATCTCGTTG (318)
			GGCCCAACGAGATGCTGACATGGCCAGA**ATCTACACTT** (327)
	AsPDS1	CCAGCCATGGTCGGCGGTCAGGCT	TACT**CTTGTAGAT**CCAGCCATGGTCGGCGGTCAGGCT (319)
			GGCCAGCCTGACCGCCGACCATGGCTGG**ATCTACAAG** (401)
	AsPDS2	TCAGTCAAAGAATGGATGGAAAAG	TACT**CTTGTAGAT**TCAGTCAAAGAATGGATGGAAAAG (320)
			GGCCCTTTTCCATCCATTCTTTGACTGA**ATCTACAAG** (402)
	AsPDS3	TCTGGCCATGTCAGCATCTCGTTG	TACT**CTTGTAGAT**TCTGGCCATGTCAGCATCTCGTTG (321)
			GGCCCAACGAGATGCTGACATGGCCAGA**ATCTACAAG** (403)
	FnPDS1	CCAGCCATGGTCGGCGGTCAGGCT	TACT**GTTGTAGAT**CCAGCCATGGTCGGCGGTCAGGCT (322)
			GGCCAGCCTGACCGCCGACCATGGCTGG**ATCTACAAC** (404)
	FnPDS2	TCAGTCAAAGAATGGATGGAAAAG	TACT**GTTGTAGAT**TCAGTCAAAGAATGGATGGAAAAG (323)
			GGCCCTTTTCCATCCATTCTTTGACTGA**ATCTACAAC** (405)
	FnPDS3	TCTGGCCATGTCAGCATCTCGTTG	TACT**GTTGTAGAT**TCTGGCCATGTCAGCATCTCGTTG (324)
			GGCCCAACGAGATGCTGACATGGCCAGA**ATCTACAAC** (406)
Tomato Solyc09g065630 (Gene1)	LbT1G1	CGGCGACAAAGCTGCAAGTGCTGT	TACT**AAGTGTAGAT**CGGCGACAAAGCTGCAAGTGCTGT (415)
			GGCCACAGCACTTGCAGCTTTGTCGCCG**ATCTACACTT** (416)
	FnT1G1	CGGCGACAAAGCTGCAAGTGCTGT	TACT**GTTGTAGAT**CGGCGACAAAGCTGCAAGTGCTGT (417)
			GGCCACAGCACTTGCAGCTTTGTCGCCG**ATCTACAAC** (418)
	LbT2G1	TTGGGAGGTTTTCTACCGTACGGC	TACT**AAGTGTAGAT**TTGGGAGGTTTTCTACCGTACGGC (419)
			GGCCGCCGTACGGTAGAAAACCTCCCAA**ATCTACACTT** (420)
	FnT2G1	TTGGGAGGTTTTCTACCGTACGGC	TACT**GTTGTAGAT**TTGGGAGGTTTTCTACCGTACGGC (421)
			GGCCGCCGTACGGTAGAAAACCTCCCAA**ATCTACAAC** (422)
Tomato Solyc06g084120 (Gene2)	LbT1G2	CAATCATAGCACTGCCCTTCCGGT	TACT**AAGTGTAGAT**CAATCATAGCACTGCCCTTCCGGT (407)
			GGCCACCGGAAGGGCAGTGCTATGATTG**ATCTACACTT** (408)
	FnT1G2	CAATCATAGCACTGCCCTTCCGGT	TACT**GTTGTAGAT**CAATCATAGCACTGCCCTTCCGGT (409)
			GGCCACCGGAAGGGCAGTGCTATGATTG**ATCTACAAC** (410)
	LbT2G2	TCTTGCTGATTTCTTCTATCAGGC	TACT**AAGTGTAGAT**TCTTGCTGATTTCTTCTATCAGGC (411)
			GGCCGCCTGATAGAAGAAATCAGCAAGA**ATCTACACTT** (412)
	FnT2G2	TCTTGCTGATTTCTTCTATCAGGC	TACT**GTTGTAGAT**TCTTGCTGATTTCTTCTATCAGGC (413)
			GGCCGCCTGATAGAAGAAATCAGCAAGA**ATCTACAAC** (414)

Table A.2: Eurofins reference to the correctly sequenced construct. Stock number of plant genome editing group.

Construct (stock number)	Eurofins sequencing references
B-LbPDS1-C (p00376)	EF00317249
B-LbPDS2-C (p00377)	EF00317250
B-LbPDS3-C (p00378)	EF00317253
B-AsPDS1-C (p00384)	EF00317309
B-AsPDS2-C (p00385)	EF00317308
B-AsPDS3-C (p00386)	EF00317307
B-FnPDS1-C (p00387)	EF00317306
B-FnPDS2-C (p00388)	EF00317305
B-FnPDS3-C (p00389)	EF00317304
B-LbT1G2-C (p00394)	EF00317303
B-FnT1G2-C (p00396)	EF00317302
B-LbT2G2-C (p00395)	EF00317301
B-FnT2G2-C	EF00317320 (misses insert)
B-LbT1G1-C (p00390)	EF00317299
B-FnT1G1-C (p00392)	EF00317319
B-LbT2G1-C (p00391)	EF00317322
B-FnT2G1-C (p00393)	EF00317318
C-LbT2G2-D (p00442)	EF00321422
C-FnT2G2-D (p00397)	EF00317317
C-LbT2G1-D (p00443)	EF00321333
C-FnT2G1-D (p00444)	EF00321423

Table A.3: Components of the golden gate reaction to construct the desired expression constructs. Plasmids without reference were received from Mansour.

Alias	Used plasmid (Plant genome editing group plasmid stock numbers)				
	Destination	**A-B**	**B-C**	**C-D**	**D-G**
A1	pFASTRK-FnCpf1-AG (p00380)	Rps5	FnT1G1(p00392)	FnT2G2(p00397)	pea3AT
A2	pFASTRK-FnCpf1-AG (p00380)	Rps5	FnT2G1(p00393)	FnT2G2(p00397)	pea3AT
A3	pFASTRK-LbCpf1-AG (p00381)	Rps5	LbT1G1 (p00390)	LbT2G2 (p00442)	pea3AT
A4	pFASTRK-FnCpf1-AG (p00380)	Rps5	FnT1G2 (p00396)	FnT2G1 (p00444)	pea3AT
A5	pFASTRK-LbCpf1-AG (p00381)	Rps5	LbT1G2 (p00394)	LbT2G1 (p00443)	pea3AT
B1	pFASTRK-LbCpf1-AG (p00381)	Rps5	LbPDS1 (p00376)	Linker	pea3AT
B2	pFASTRK-LbCpf1-AG (p00381)	Rps5	LbPDS2 (p00377)	Linker	pea3AT
B3	pFASTRK-LbCpf1-AG (p00381)	Rps5	LbPDS3 (p00378)	Linker	pea3AT
B4	pFASTRK-AsCpf1-AG (p00379)	Rps5	AsPDS1 (p00384)	Linker	pea3AT
B5	pFASTRK-AsCpf1-AG (p00379)	Rps5	AsPDS2 (p00385)	Linker	pea3AT
B6	pFASTRK-AsCpf1-AG (p00379)	Rps5	AsPDS3 (p00386)	Linker	pea3AT
B7	pFASTRK-FnCpf1-AG (p00380)	Rps5	FnPDS1 (p00387)	Linker	pea3AT
B8	pFASTRK-FnCpf1-AG (p00380)	Rps5	FnPDS2 (p00388)	Linker	pea3AT
B9	pFASTRK-FnCpf1-AG (p00380)	Rps5	FnPDS3 (p00389)	Linker	pea3AT

Table A.4: Stocked (Plant genome editing group references) *Agrobacterium* colonies with its construct

Construct	in *Agrobacterium* screening colony
A1	Colony 2 (PGE00539)
A3	Colony 3 (PGE00576)
A4	Colony 4 (PGE00577)
A5	Colony 7 (PGE00578)
B1	Colony 3 (PGE00559)
B2	Colony 6 (PGE00560)
B3	Colony 8 (PGE00561)
B4	Colony 10 (PGE00562)
B5	Colony 15 (PGE00563)
B6	Colony 17 (PGE00564)
B7	Colony 19 (PGE00567)
B8	Colony 22 (PGE00565)
B9	Colony 27 (PGE00566)

Sequence A.5 Designed crRNA array entry vector sequence for AsCpf1 at position B-C:

GGTCTCAAACAGTCAAAAGACCTTTTTAATTTCTACTCTTGTAGATGGGTCTTCAAGCCAGATAACAGTATGCGTATTTGCG
CGCTGATTTTTGCGGTATAAGAATATATACTGATATGTATACCCGAAGTATGTCAAAAAGAGGTATGCTATGAAGCAGCGT
ATTACAGTGACAGTTGACAGCGACAGCTATCAGTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAA
CCATGCAGAATGAAGCCCGTCGTCTGCGTGCCGAACGCTGGAAAGCGGAAAATCAGGAAGGGATGGCTGAGGTCGCCCG
GTTTATTGAAATGAACGGCTCTTTTGCTGACGAGAACAGGGGCTGGTGAAATGCAGTTTAAGGTTTACACCTATAAAAGA
GAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGC
CAGTGCACGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGA
TGACCACCGATATGGCCAGTGTGCCGGTTTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGACATC
AAAAACGCCATTAACCTGATGTTCTGGGGAATATAAGAAGACAAGGCTTGAGACC

The outer gray pair of sequences are BsaI sites while the inner pair are BbsI sites.
The green sequences are the conventional greengate overhangs for position B-C
For C-D: replace by GGCT...TCAG
For D-E: replace by TCAG...CTGC
For E-F: replace by CTGC...ACTA
The purple sequence is the homologue-dependent Cpf1 direct repeat according to Zetsche (2015).
For FnCpf1: replace by GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT
For LbCpf1: replace by GTTTCAAAGATTAAATAATTTCTACTAAGTGTAGAT
The yellow sequence encodes the *ccdB* negative selection marker

Sequence A.6 Designed pea3AT terminator preceded by the HDV ribozyme for position F-G

GGTCTCAACTAGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACC
AGGCCTCCCAGCTTTCGTCCGTATCATCGGTTTCGACAACGTTCGTCAAGTTCAATGCATCAGTTTCATTGCCCACACACCA
GAATCCTACTAAGTTTGAGTATTATGGCATTGGAAAAGCTGTTTTCTTCTATCATTTGTTCTGCTTGTAATTTACTGTGTTCTT
TCAGTTTTTGTTTTCGGACATCAAAATGCAAATGGATGGATAAGAGTTAATAAATGATATGGTCCTTTTGTTCATTCTCAAA
TTATTATTATCTGTTGTTTTTACTTTAATGGGTTGAATTTAAGTAAGAAAGGAACTAACAGTGTGATATTAAGGTGCAATGT
TAGACATATAAAACAGTCTTTCACCTCTCTTTGGTTATGTCTTGAATTGGTTTGTTTCTTCACTTATCTGTGTAATCAAGTTTA
CTATGAGTCTATGATCAAGTAATTATGCAATCAAGTTAAGTACAGTATAGGCTTGTATTGAGACC

The grey sequences are BsaI sites
The green sequences are the conventional greengate overhangs for position F-G
The cyan sequence is the hepatitis delta virus ribozyme
The red sequence covers the pea3AT sequence

Table A.7: Summary of the synthesized, sequenced and stocked vectors for the Cpf1 crRNA array system.

	Position	Name	Plasmid	Glycerol	Sequence
AsCpf1	B-C	pGG-B-AsCpf1DR-ccdB-C	p00551	PGE00934	EF00388848
	C-D	pGG-C-AsCpf1DR-ccdB-D	p00552	PGE00935	EF00388857
	D-E	pGG-D-AsCpf1DR-ccdB-E	p00553	PGE00936	EF00388864
	E-F	pGG-E-AsCpf1DR-ccdB-F	p00554	PGE00937	EF00388871
FnCpf1	B-C	pGG-B-FnCpf1DR-ccdB-C	p00710	PGE01139	EF00591672
	C-D	pGG-C-FnCpf1DR-ccdB-D	p00730	PGE01156	EF00837481
	D-E	pGG-D-FnCpf1DR-ccdB-E	p00711	PGE01140	EF00591673
	E-F	pGG-E-FnCpf1DR-ccdB-F	p00712	PGE01141	EF00591674
LbCpf1	B-C	pGG-B-LbCpf1DR-ccdB-C	p00729	PGE01155	EF00837475
	C-D	pGG-C-LbCpf1DR-ccdB-D	p00708	PGE01137	EF00591668
	D-E	pGG-D-LbCpf1DR-ccdB-E	p00709	PGE01138	EF00591670
	E-F	pGG-E-LbCpf1DR-ccdB-F	p00728	PGE01154	EF00837474
HH-pea3AT	F-G	pGG-F-HDV-pea3AT-G	p00550	PGE00933	EF00388919

Table A.8: Phenotype rates of T1 *Arabidopsis* seedlings containing the CRISPR-Cas9/Cpf1 expression vectors

	Cas9			**AsCpf1**			**FnCpf1**			**LbCpf1**		
	PDS1	PDS2	PDS5	PDS1	PDS2	PDS3	PDS1	PDS2	PDS3	PDS1	PDS2	PDS3
Wild-type	0	**5%**	**26%**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**
Chimer	**10%**	**74%**	**74%**	0%	0%	0%	0%	0%	0%	0%	0%	0%
Albino	**90%**	**21%**	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Total amount	20	19	23	2	24	24	25	23	25	23	22	22

Table A.9: Summarized TIDE analysis for the sequences of LbCpf1-PDS2 rosette samples from heat-treated and non-heat-treated plants. 'Total' indicates the total indel frequency reported by the TIDE tool. Only significant indel frequencies (p-value < 0,05) were taken.

		Total	R^2	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1
Heated	Line 1	29,1	0,99						2,1	1			2,1	4,4	7,3			9,6	70	
		12,3	0,99											1,6	5,3			4,9	87	
		31,7	0,98						2				2	3,6	7,9			13	67	
	Line 2	71,6	0,87						4,9	4			6,8	14	24			17	15	
		70,7	0,93						5,3	3,5			8	13	25			14	23	
		68,3	0,9						4,9	4,1			6,2	13	25			13	22	
		67,1	0,81							2	2,5			3,6	12	24			14	21
		55,7	0,95						3,7	2,7			6	12	22			8,6	40	
		49,6	0,97						2,7	1,6			4	9,8	15			14	48	
	Line 3	17,9	0,95											2,6	7,5			6,5	77	
		42,3	0,98						1,8	1,5			4,6	6,7	15			11	55	
		64	0,93		1,5				6,7	1,8			7	13	20			14	29	
	Line 4	35,9	0,97										1,4	4,1	15			15	61	
		47,8	0,98						2,7				3,8	5	15			18	50	
		41,7	0,98					1	2	1,3			2,7	6,1	12			16	56	
Non-Heated	Line 1	0,1	0,99																99	
		0	0,99																99	
		8	0,82															3	74	
	Line 2	8,7	0,88															3,1	80	
		2,1	0,98																96	1,4
		0	0,99																99	
		3,6	0,96															1,6	92	1,8
	Line 3	4,8	0,92																87	
		3,1	0,89																86	
		0	0,99																99	
	Line 4	0,6	0,99																99	
		0,4	0,98																98	
		5,2	0,93																87	

Table A.10: Summarized TIDE analysis for the sequences of Cas9-PDS rosette samples from heat-treated and non-heat-treated plants. 'Total' indicated the total indel frequency reported by the TIDE tool. Only significant indel frequencies (p-value < 0,05) were taken.

	Total	R^2	-48	…	-18	-17	-16	-15	…	-6	-5	-4	-3	-2	-1	0	1	2	3	4
Heated	49,8	0,98					31								3,5	48	15			
	68,6	0,98													3,6	30	59			
	95,7	0,98														2,2	39	57		
	53,7	0,81	28													27	5,7	1,4		8,5
Non-Heated	50,8	0,98														47				49
	58,4	0,98										58				39				
	60	0,86			52						1,5					26				
	52,9	0,87				48										34				

Table A.11: Counted seedlings per scoring category. Scoring numbers were added per plate (two plates per temperature per construct). F+/- = FAST-positive/negative seedlings. NA = not available.

		21°C (1)		21°C (2)		30°C (1)		30°C (2)		37°C (1)		37°C (2)	
		F+	F-	F+	F-	F+	F-	F+	F-	F+	F-	F+	F-
Cas9-PDS	wild-type		**9**		**3**		**12**		**11**		**4**		**11**
	aff.cots		**22**		**12**		**23**		**16**		**3**		**3**
	mild PDS	NA	**14**	NA	**7**	NA	**15**	NA	**19**	NA	**14**	NA	**10**
	severe PDS		0		0		0		0		**10**		**8**
	albino		**10**		**3**		**7**		**4**		**2**		**5**
LbCpf1-PDS2 (L1)	wild-type	**97**	**2**	**99**	**1**	**98**	**1**	**94**	**1**	**73**	**1**	**63**	**1**
	aff.cots	0	0	0	0	0	0	0	0	0	0	0	0
	mild PDS	0	0	0	0	0	0	0	0	**7**	0	**6**	0
	severe PDS	0	0	0	0	0	0	0	0	**2**	0	0	0
	albino	0	0	0	0	0	0	0	0	0	0	0	0
LbCpf1-PDS2 (L2)	wild-type	**95**	**3**	**97**	**1**	**96**	**2**	**93**	**3**	**24**	**1**	**35**	**1**
	aff.cots	0	0	0	0	0	0	0	0	0	0	0	0
	mild PDS	0	0	0	0	0	0	0	0	**22**	0	**22**	0
	severe PDS	0	0	0	0	0	0	0	0	**19**	0	**18**	0
	albino	0	0	0	0	0	0	0	0	0	0	0	0
LbCpf1-PDS2 (L3)	wild-type	**72**	**25**	NA (fungal take-over)		**69**	**23**	**38**	**12**	**17**	**11**	**6**	**12**
	aff.cots	0	0			0	0	0	0	0	0	0	0
	mild PDS	0	0			0	0	0	0	**22**	0	**14**	0
	severe PDS	0	0			0	0	0	0	**19**	0	**7**	0
	albino	0	0			0	0	0	0	0	0	0	0
LbCpf1-PDS2 (L4)	wild-type	**82**	**18**	**78**	**22**	**83**	**17**	**73**	**26**	**28**	**17**	**34**	**11**
	aff.cots	0	0	0	0	0	0	0	0	0	0	0	0
	mild PDS	0	0	0	0	0	0	0	0	**21**	0	**14**	0
	severe PDS	0	0	0	0	0	0	0	0	**4**	0	**6**	0
	albino	0	0	0	0	0	0	0	0	0	0	0	0

Table A.12: ICE scores (total indel frequencies) for both *PPD* genes at timepoint 1 (T1) and timepoint 2 (T2).

				Efficiency (ICE-score = indel%)			
				Gene1		**Gene2**	
	Sample	**Construct**	**Plant**	**T1**	**T2**	**T1**	**T2**
heated	1	A1	1	2	0	3	3
	2	A1	2	4	1	3	4
	3	A1	3	6	0	4	15
	6	A1	6	0	0	3	0
	7	A1	9	1	0	1	0
	11	A1	14	3	1	0	4
	12	A1	15	1	2	3	1
	15	A1	18	0	0	3	1
	17	A1	21	43	37	3	2
	20	A1	24	0	1	4	0
	24	A3	14	3	2	4	3

	25	A3	16	1	10	3	4
	26	A3	9	3	2	4	4
	27	A3	28	3	1	6	0
	29	A3	30	1	0	3	1
	31	A3	15	0	0	3	0
	32	A3	17	1	0	1	5
	34	A3	25	0	1	2	0
	36	A3	22	0	1	9	2
	37	A3	20	0	52	3	3
control	4	A1	4	1	1	0	0
	5	A1	5	2	0	4	0
	8	A1	10	1	0	12	1
	9	A1	12	2	1	0	0
	10	A1	13	0	1	4	0
	13	A1	16	0	0	4	3
	14	A1	17	1	0	3	1
	16	A1	19	1	1	2	0
	18	A1	22	1	1	3	0
	19	A1	23	1	0	2	6
	21	A3	1	1	2	0	0
	22	A3	3	7	13	4	1
	23	A3	7	2	0	2	1
	28	A3	29	0	0	3	0
	30	A3	12	2	1	1	5
	33	A3	8	2	1	3	1
	35	A3	11	5	1	0	3
	38	A3	10	0	0	3	3
	39	A3	27	1	0	2	0

Protocol A.13: Golden Gate reaction for target insertion in entry vectors

Add 1 µL of forward and 1 µL of reverse primer (10 µM) in 48 µL of MQ water and run in thermocycler (5 minutes 95°C, 30 seconds 85°C - Ramp 2°C/s, 30 seconds 25°C - Ramp 0,1°C/s, hold at 4°C).

Use 1 µL of the dimer product with 1.5 10 µM ATP, 1.5 µL CutSmart buffer, 50 ng of GG entry vector, 200 U T4 Ligase, 0.5 µL BbsI-HF and add MQ water up to 15 µL. Run reaction in thermocycles (30 cycles of 2 minutes at 37°C and 2 minutes at 16°C, 5 minutes at 50°C, 5 minutes at 85°C, hold at 10 °C).

Table A.14: used oligo primers during troubleshooting PCRs. Ref. = Plant genome editing oligo stock number

Function	Ref.	Name	Sequence
T-DNA presence in gDNA	320	AsPDS2_FW	TACTCTTGTAGATTCAGTCAAAGAATGGATGGAAAAG
	323	FnPDS2_FW	TACTGTTGTAGATTCAGTCAAAGAATGGATGGAAAAG
	317	LbPDS2_FW	TACTAAGTGTAGATTCAGTCAAAGAATGGATGGAAAAG
	417	FnT1G1_FW	TACTGTTGTAGATCGGCGACAAAGCTGCAAGTGCTGT
	415	LbT1G1_FW	TACTAAGTGTAGATCGGCGACAAAGCTGCAAGTGCTGT
	343	PcUBIP_REV	GTGTTTGAGGCGGTGAAGGAAG
Cpf1 fragments cDNA	679	B-dummy5UTR_FW	GTCGACTGGTACCAACAGGC
	829	FnCpf1_REV1	CCTCGTCCAGGGAAAAGACC
	263	FnCpf1_FW1	CGACATTCCTACCTCCATCATCTAC
	830	FnCpf1_REV2	AGGATTTCCTCAAAGCGGCA
	264	FnCpf1_FW2	GCAGGTGTTTGATGACTATTCCGTG
	831	FnCpf1_REV3	ACTCTTGGCGCTGAAGAACA
	265	FnCpf1_FW3	CCTGTTCATCAAGGATGACAAGTAC
	832	FnCpf1_REV4	CTCCTCGGTCAATGCTCAGG
	266	FnCpf1_FW4	CACTCACCCAGCTAAGGAGGC
	833	FnCpf1_REV5	CGAAGGAGGCAATGGTCCAT
	267	FnCpf1_FW5	CCTACCAGCTGACAGCTCCC
	680	D-dummy3UTR_REV	ACAGGGAATGAAGGTAAAGGTT
GFP-Cpf1 screen	170	pGG4468_FW	GGACTTGTGCGACATGTCGTTTTC
	27	GFP_REV	GTTTACGTCGCCGTCCAGCT

Protocol A.15: Small RNA extraction (Shengben, Sep. 2009)

Grind tissue (~50 mg) in liquid nitrogen into fine powder, immediately add 0.5 mL TRI reagent and shake vigorously. Incubate 5 minutes at room temperature. Add 100 µL chloroform to each tube in hood, shake vigorously and put at room temperature for 5 minutes. Centrifuge at 10000 rpm in pre-cooled 4°C rotor during 15 minutes. Transfer aqueous phase to new tubes. Add 300 µL isopropanol, invert several times, let sit at room temperature for 10 minutes. Centrifuge at 10000 rpm at 4°C during 15 minutes. Discard supernatant and wash pellets with 0.5 mL 70% ethanol. Remove supernatant and air-dry the pellet in hood. Add 30 µL RNA-free water to dissolve the RNA. Determine concentration and keep at -80°C.

Protocol A.16: Protein extraction, SDS-PAGE and western blot

1) Protein extraction

Protein extraction buffer :

- 5% ethylene glycol
- 25 mM Tris-HCl pH 7.6
- 15 mM $MgCl_2$
- 5 mM EGTA pH 8
- 150 mM NaCl
- 15 mM para-nitrophenylphosphate
- 60 mM beta-glycerophosphate
- 1 mM dithiothreitol
- 0.1% NP-40
- 0.1 mM Na3VO4}
- 1 mM NaF
- 1 mM PMSF
- 10 µg/ml leupeptin
- 10 µg/ml aprotinin
- 10 µg/ml soybean trypsin inhibitor
- 0.1 mM benzamidin
- 5 µg/ml antipain
- 5 µg/ml pepstatin
- 5 µg/ml chymostatin
- 1 µM E64

OR ask to proteomics group, "BHB+" buffer

- Grind frozen sample with Retsch machine 30Hz for 30sec, freeze, 30Hz for 30sec, freeze. Keep samples very cold. Puncture the lid of the tubes with a needle.
- Add 50-100µL protein extraction buffer. The less, the better.
- Freeze sample in liquid N2.
- Vortex until sample becomes liquid.
- Freeze sample in liquid N2.
- Vortex until sample becomes liquid.
- Centrifuge @4°C for 15min @14000rpm.
- Transfer supernatans to new tube.
- Centrifuge @4°C for 15min @14000rpm.
- Transfer supernatans to new tube.
- Store at -70°C. It is NOT recommended to freeze protein samples, load the gel the same day if possible.

2) Protein quantification with BSA standard curve

3) SDS-PAGE

Sample preparation (on ice)

- Keeping the concentration in mind, add sample buffer [loading buffer + DDT (x5)] to the protein sample extract. Then cook the sample at 95°C for 5 min to denature the sample. Now the protein extract sample can be stored in the freezer -20°C.
- In a separate epp, pipet the calculated volume (20-50 µg protein) of sample and adjust with MQ water until 20µl
- Incubate again at 95°C for 5min.
- Put on ice.

Running the gel

- Needed: 2 (10%) Mini-PROTEAN TGX Precast Gels (Bio-Rad) (if only 1 gel is run, use a plastic Buffer Dam, provided with the system).
- Dilute 10X running buffer (from Bio-Rad) to 1X with pure water.
- Remove gels from packages in the sink. Discard buffer inside. Carefully remove combs and green strips (!) at the bottom.
- Place gels correctly in buffer tank and fill buffer tank with 1X running buffer.
- Load 10µL Precision Plus Protein Dual Color Standards (Bio-Rad, freezer).

- Load samples.
- Put on lid, connect to voltage generator and let run @180V for 45min until the dye reaches the bottom of the gel.

Prepare 3% semi-skimmed milk (1.5g Skim Milk (Difco) + add TBT-buffer to 50mL).
For 10X TBS (1L):

- 12,11g Tris
- 87,66g NaCl

adjust to pH8 with HCl
> filter sterilize & keep in the fridge.
For 1X TBT: 100mL 10X TBS + 900mL MQ-H2O + 1mL Triton (Sigma).

- Break' gel assembly with 'little green fork', carefully take gel out, cut off stacking and loading dye.

4) Western blot

Transfer of protein to PVDF-membrane

- Needed: 2 Trans-Blot Turbo Transfer Packs + PVDF (Mini-format, Bio-Rad).
- Open packages and assemble cassette in following order:
  - buffer-saturated Whatmann paper/membrane sandwich ('bottom')
    - remove air bubbles with clean rolling pin
    - Cut off a triangle at upper left corner
  - gel (don't move after first contact)
  - buffer-saturated Whatmann paper ('top')
    - remove air bubbles with clean rolling pin
- Put lid on cassettes, close them by turning (don't open again!!), put in voltage generator (Trans-Blot Turbo Transfer System) & let run @25V for 7min (turn on > TURBO > 1 mini gel per cassette > runA/runB).
- Dismantle construction, discard in BLUE barrel.

Ponceau staining

- Add Ponceau staining solution for about 1 minute (can be more for low amounts).
- Pour off staining solution and wash membrane with TBT-buffer.
- Scan membrane quickly as staining fades fast.
- Wash in TBT-buffer until staining has completely disappeared.

Blocking

- Add an excess (10mL) of 3% SM-milk in TBT to each membrane.
- Put on shaker ON @4°C.
- Wash membrane next day a few times with TBT-buffer.

Incubation with antibody

- Add 10 to 15mL of a 1:200* dilution of Anti-GFP antibody (dilute in 3% milk + TBT) to the membranes. *check Ab manual for concentration!
- Put on shaker for 1hour @RT. Incubation time depends on Ab (anti-HA, 1:10000 dilution)
- Pour off antibody dilution (back to falcon) and wash with TBT-buffer:
  - Add TBT-buffer, shake and pour off immediately.
  - Add TBT-buffer and put on shaker for 15min @RT.
  - Wash 3 more times for 5min while shaking @RT.
- Add 10ml of a 1:10000 dilution of secondary anti-rat (dilute in 3% milk + TBT).
- Put on shaker for 1hour @RT.
- Wash again as described before.
- Note: antibody dilutions can be used up to 3X when stored @-20°C. When using a PAP-antibody, use 10-15mL of a 1:2500 dilution and incubate for 1hour @RT, then wash as described.

5) Detection with the Bio-imager

- Add 1.5mL of oxidizing reagent + 1.5mL of luminol reagent.
- Put some drops of this mix on an open plastic sheet.
- Carefully pat membrane dry, using a tissue.
- Place membrane on the mix, protein side (side touching gel during blotting) downwards.
- Close the plastic sheet and remove air bubbles.
- Incubate for 1min.
- Cut a bigger and small piece of plastic foil (the one next to the qPCR-machine), put the bigger in the Bio-imager cassette.

- Place membrane on the foil and carefully cover with the smaller piece of foil, making sure no air bubbles are included.
- Close the detection cassette.
- Open *Image Lab* on the computer and take following steps:
    - *New protocol*
    - *Application* > *Select* > *Blots* > *Chemi*
    - *Image exposure* > *Signal accumulation set-up*
        - *First pic*: 10 (first picture will be taken at 10 seconds).
        - *Last pic*: 900 (=15min)
        - *Total # pics*: 30
    - *Position gel* (use *Camera zoom*)
    - *Filter position* > *No filter* (check on machine)
    - *Run protocol*
- Use *Image transform* at the left to change the contrast while running
- After the run:
    - Select image at the bottom
    - *Save all*, using right click
    - Select single picture > *Select pic and continue*
    - Take a pic of the marker, using *Chemi marker* as protocol
    - *Image tools* > *Merge*
    - *Invert data*
    - Export picture as a JPEG (give a new name!)

6) Eventual staining with coomassie brilliant blue

Wash again in TBT, take blot out, put coomassie brilliant blue on it, leave for 15 to 60 minutes. Take coomassie off, add clearing buffer (5:4:1 ratio of MQ water:MeOH:AceticAcid), take off after 15 minutes, dry with paper and scan.

Protocol A.17: BY-2 transformation

Plant medium:

for BY-2 liquid cultures	*1L*
MS salts (no additives Duchefa)	4.302 g
KH_2PO_4	0.2 g
Sucrose	30 g
pH with 1 M KOH	5.8

prepare 25 erlenmeyers 250 mL containing 40 mL medium and autoclave.

for solid medium	*1L*
add plant tissue culture agar	6.5 g

autoclave

prior to use, add BY-2 vitamines (1 mL/1L, or 40 uL in each erlenmeyer)

BY-2 vitamines	*50 mL (H_20)*
2,4D (auxin, dissolve in 1mL ethanol)	0.02 g
thiamine	0.05 g
myo-inositol	5 g

filter sterilize, aliquot 10 mL in 10 mL falcon tubes, store at –20 C
Prior to use, melt and dissolve

Growing the cultures

Use 250 mL erlenmeyers, red screw cap, 40 mL of medium.
Add vitamins prior to use.
Inoculate (the volume depends a bit on the history of the cells, after a while they grow faster) the cells using a 5 mL sterile Pasteur pipette. Release the cells directly into the medium.
Unscrew the cap half/one turn, and seal with surgical tape (permits airflow).
Incubate in a shaker, 28 C, 150 rpm.
Refresh each week, same time, same procedure, normally 1/40 dilution of a 7-day old culture.
Cells can be kept viable at least a week when stored at 4 C.
To maintain a backup culture, grow the cells on a solid plate. As such it can be grown at 25C up to one month. Refresh when browning of the callus starts to appear.

Transformation:

DAY 1 (by default we do this on Friday afternoon, during the weekend, the shaker remains rather closed, as only few people work then and cells do not stop shaking, which improves the quality)

1. dilute the BY-2 culture 2, 3 and 4 mL in 40mL, erlenmeyer 250 mL at 28 °C

The density of the BY-2 is by far the most important. If the density of the BY-2 is OK, the transformation will work with any volume of agrobacterium. To be sure about the density, we use the above 3 dilutions and at least one of these will give transformants with a certainty of over 90%.

2. Grow Agrobacterium (e.g. C58, LBA4404) in 5-10 mL YEB liquid medium with antibiotics in a 50ml Falcon tubes (NO RIFAMPYCIN !)
3. Dilute the growing agrobacterium culture 1/5 the afternoon before day 3 **(Sunday in our protocol)**

DAY3 (Monday)

Take 4 mL BY-2 culture and decant in a petridish plate (round and deep ones, NOT the ones used for bacteria but the same diameter). Mix gently with various concentrations (we normally only use 300ul) of bacteria (10 uL to 500 uL, ON culture). Incubate at max 25 °C to 28C without shaking or moving.

DAY5 (Wednesday)

Plate all cells on selective medium (BY2 medium containing 500 mg/L Carbenicilline and 200 mg/L Vancomycin, plus Kan = 100mg/L or Hygro = 30 mg/L).

If you put the plate upright, the cells should only slightly slide down or not at all, that is the optimal density to plate them out. If they are too few or too thick, the efficiency will drop. When plating, try to cover the entire plate by swirling the mix around after closing the plate.

After 14 days, calli should become visible. Pick them with a sterile tooth pick and transfer to a new selective plate (including the Vanc and Carb).

Protocol A.18: Genetic Transformation of Tomato (Micro-Tom)

Sterilize the seeds 2 min in 70% ethanol, 25 minutes in 10% bleach . Rinse with sterile water.

Sow the seeds (>100 seeds) on ½ MS (or MS) medium and germinate 3 days in the dark. Transfer to light. Friday – Monday or Monday – Monday.

Day D – 1:

- **Agrobacterium culture**:

Inoculate 2 tubes with 20 mL liquid LB + Antibiotics (eg 20 microl streptomycin). Incubate 24h at 28°C (200rpm). Strain EHA105.

- **Cotyledon preparation**:

Cut the cotyledons at both extremities and in the middle. Recover the two half cotyledons and lay them on solid KCMS (upper face on the medium). Do not allow the cotyledons to dry out. Place the pots in the dark at 25°C in growing chamber for 24h. Label the pots A – F.

COTYLEDONS are FRAGILE. Handle with CARE.

Day D:

- **Agrobacteria preparation**:

Centrifuge the bacterial suspension 10 min at 3000 rpm. Add acetosyringon to the liquid KCMS (0,1 ml in 100 ml). Resuspend each pellet in 20 ml liquid KCMS. Measure the OD (normally around 0,3).

Dilute the culture in a sterile falcon tube to an OD 0.1-0.2 (final volume of 30 mL). Homogenize before measuring OD. Let stand 1h.

- **Transformation**:

Collect the cotyledons from the KCMS. Keep the KCMS pots sterile. Soak the cotyledons in the bacterial suspension for 15 min with gentle shaking (50 rpm). Start with pots A, then pots B, etc…. Dry the cotyledons on sterile soft paper and lay them on solid KCMS. Pots are incubated in the dark for 48h at 25°C.

Alternatively, add drops of Agrobacterium on the cotyledons. Let stand 15 min. Remove bacteria from each pot.

Day D+2:

- **Regeneration**:

Transfer cotyledons to 2Z (max 10 explants per pot). Be gentle with cotyledons and drag on medium (in order to remove Agro, to prevent overgrowth). After one week, place the explants on another place in the same pot. Change the medium every 15 days.

When shoots are developed (>1 cm), they are picked on rooting medium. Plants that develop roots (and sideroots) are likely transgenic. Transgenic plants have a green appearance. If no callus on the cotyledons or shoots after 6 weeks, discard by autoclaving.
Acclimatization vitro to vivo: first week: high humidity; second week: gradually open transparent box.

MEDIUM

		solid KCMS	**liquid KCMS**	**2Z**	**Rooting medium (ENR)**	**MS ½ or MS**
Volume		1l	1l	1l	1l	1l
MS (Basal salt)					2,2 g	
MS (with vitamins)		4.4 g	4.4 g	4.4 g		2.2 or 4.4g
Sucrose		20 g	20 g	30 g	10 g	20 g
KH_2PO_4		0.2 g	0.2 g			
Nitsch vitamins				1 ml	1 ml	
Folic acid				1 ml	1 ml	
pH		5.7	5.7	5.7	5.7	5.7
Agar		6 g		6 g	6 g	6 g
		Autoclave	Autoclave	Autoclave	Autoclave	Autoclave
Thiamine (stock 0,5 mg/ml)		2 ml	2 ml			
Acetosyringone (200 mM stock)		1 mL	1 mL (add before use)			
2.4D (stock 0,5 mg/ml)		400 µl				
Kinetin (stock 0,5 mg/ml)		200µL				
Zeatin Ribosid (stock 0,5 mg/mL)				4ml		
Kanamycin (stock 50 mg/ml)				2 ml	2 ml	
Melatonine (stock 25mg/ml)				1 ml		
Timentin (stock 60 mg/ml)				5 mL mL	2,5 ml	

Acetosyringone: 200 mM = 0,4 g/10ml 70% ethanol
Melatonine: 0,1M; 0,25g/10ml 95% ethanol

Nitsch Vitamins 1000x

For 100 ml

Biotin	0.005 g
Glycine	0.2 g
Myo-inositol	10 g
Nicotinic Acid	0.5 g
Pyridoxine HCl	0.05 g
Thiamine HCl	0.05 g

Stock folic acid: 0,05g in 1ml 0,1M NaOH and then H20 till 100 ml

Biotin: dissolve 5 mg in 10ml
Modification Cermak et al
KCMS: 0,5 mg/L IAA; 0,1 mg/L kinetine;
Grow Agrobacterium on Monday (Day D-1)
No preculture of tomato;
Isolate cotyledons on Tuesday, inoculate 10min.
Day +2: put on 2Z; 2mg/L zeatin and 0,1 IAA mg/L (melatonin and timentin; no kan)
Day +6/7: transfer to 2Z; zeatin, IAA, melatonin en timentin and Kan)
Modification Dan et al
No preculture of tomato
Isolate cotelydons on Tuesday, make 3 pokes per cotyledon with fine needle, inoculate 10 min
Day +2: put on 2Z; 2 mg/L zeatin riboside, 0,1 IAA mg/L (melatonin and timentin and Kan)