

An effective and inexpensive strategy to identify multiplex CRISPR-edited plants

Renyu Li  
Jiaying, Zhejiang, China

Bachelor of Science, University of Washington, 2015

A Thesis presented to the Graduate Faculty  
of the University of Virginia in Candidacy for the Degree of  
Master of Science

Department of Biology,

University of Virginia, Charlottesville, Virginia  
May, 2019

Thesis Committee Members:  
Dr. Cristian H. Danna (First reader)  
Dr. Keith Kozminski  
Dr. Michael Timko

---

**Abstract**

Since its first introduction in 2013, CRISPR-Cas9 has become the preferred gene targeting tool to produce loss-of-function mutants in plants. In spite of the high specificity and ease of use, the identification of CRISPR-edited plants has remained a time consuming and onerous process. I have developed and tested an easy-to-use and inexpensive strategy to select for multiplex CRISPR mutagenized Arabidopsis plants. This strategy is based on targeting the gene/s of interest simultaneously with a proxy for CRISPR-Cas9 activity: an endogenous gene that produces an easy-to-detect visible phenotype. To test this strategy, I have chosen Arabidopsis gene *JAR1*, *GL1*, *EIN2* as the candidate proxies. I have tested the T2 progeny of independent T1 plants harboring CRISPR/Cas9 and successfully identified plants where the visible marker and the genes of interest were simultaneously edited at a high frequency. The co-editing frequency ranged from 55.6% to 93.75% for two genes, and 14.3% to 50% for three genes, depending on the T1 progeny tested and the proxy gene of choice. The visual phenotype selection provides a narrow pool of plants to analyze, hence increasing the recovery frequency while decreasing the cost of identifying mutants. This selection strategy also offers a framework to similarly facilitate the identification of CRISPR-edited plants in other plant species with more complex polyploid genomes where multiplex mutants are essential for studying gene function.

## Dedication

*To my grandfather, Yueqing Li, who is my torchbearer of  
biology life science*

## Acknowledgements

I would like to thank Dr. Cristian H. Danna, who is my mentor and the first reader for this MS thesis. This work would not have been possible without his support and guidance. I also need to thank Dr. Keith Kozminski and Dr. Michael Timko for their contributions as my thesis committee members. Thanks to their insightful opinions, I could see problems from different angles.

I wish to express my thank to Dr. Anthony Spano for his knowledge and help with western blots. I am grateful to all the members of the Danna lab for their help. I want to specially thank Charles Vavrik for his hard work during the DNA analysis process.

Lastly and more importantly, I want to thank my parents and my wife. Without their support, I will lose an important piece of energy to keep me on the track and finish this work.

---

## **Introduction**

### **Targeted mutagenesis technologies**

Prior to the development of CRISPR-Cas9 as a gene targeting tool in plants, most gene function studies relied on mutants generated via random mutagenesis. The main sources of mutants for *Arabidopsis thaliana* were created through EMS mutagenesis or T-DNA insertion. Both methods use random mutagenesis and require subsequent mapping of the mutations/insertions. By 2012, only two gene-targeting methods, the Zinc Finger Nuclease (ZFN) and the Transcription Activator-Like Effector Nuclease (TALEN), were developed and applied successfully in *Arabidopsis* and other plants (Miller et al., 2007&2011; Christian et al., 2010). The specificity of ZFN is based on protein binding to the target DNA, which introduces serious limitations to the design and customization. (Ramirez et al., 2008). Sub-optimal protein design of ZFN can result in poor binding activity or low specificity that causes high toxicity (Cornu et al., 2008). In 2009, a distinctive set of naturally occurring Type-3 Secretion Effectors proteins from the bacterium *Xanthomonas* sp. were found to have specific DNA binding activity towards DNA promoter sequences of plant genes (Moscou and Bogdanove, 2009). The various amino acid repeats found in these Transcription Activator-Like effectors (TAL effectors) can be modified to change the domain's DNA binding specificity. When the binding domain of TAL effectors is fused to a nuclease, the chimeric protein binds to the target DNA and produces double strand breaks (DSBs) that lead to loss-of-function mutations in the targeted gene (Christian et al., 2010; Bogdanove et al., 2011). However, the binding specificity of TALEN requires several repeating sequences that together encode the modular DNA-binding domain, making the cloning of such repeats hard to attain. In addition, the binding activity of TALEN is highly sensitive to epigenetic modifications in the target DNA (Malzahn et al., 2017).

In 2012, a new DNA editing tool was developed based on the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and the CRISPR-Associated System protein-9 (Cas9) of the bacterium *Streptococcus pyogenes*. In CRISPR/Cas9 system, a small guide RNA (sgRNA) directs the Cas9 nuclease to a specific gene locus. The sgRNA is comprised of a crRNA (CRISPR RNA) sequence and a tracrRNA sequence (trans-activating crRNA). The crRNA is a 20nt RNA sequence complementary to the targeted DNA sequence. In this case, Watson-Crick base pairing rules make CRISPR/Cas9 highly predictable and easy to implement. The tracrRNA of the sgRNA has a unique secondary structure that activates Cas9 upon binding (Karvelis et al., 2013). Once activated, Cas9 will introduce a DSB next to the PAM sequence. The designing of the sgRNA only requires the identification of a 20nt gene-specific sequence upstream of a Protospacer Adjacent Motif (PAM) site, typically a 5'-NGG-3' sequence in the gene of interest (Osakabe et al., 2016; Peterson et al., 2016). Once the specific 20nt DNA region upstream of the PAM site is identified, changing the 20nt in the crRNA region of the sgRNA only involves introducing the 20mer specific primer into the new construct via PCR. Both the Cas9 enzyme and the tracrRNA region of the sgRNA remain the same across different targeted genes. This makes the customization of sgRNAs to target specific genes significantly less expensive and easier than designing and cloning gene-specific DNA-binding proteins as in the case of ZFN and TALEN (Mahfouz et al., 2014). The endogenous DNA repair proteins of the plant will unfaithfully repair the DSB via Non-Homologous End Joining (NHEJ) recombination, and the function of the targeted gene will be lost.

---

## Hindrances at identifying CRISPR-mutated plants

Forward genetic screens in *Arabidopsis* have been very successful in isolating mutants that are impaired for specific gene functions (Koorneef and Meinke, 2010). However, gene redundancy often hinders the efforts to link gene function to phenotypes in reverse genetics screenings of single loss-of-function mutants (Bowers et al., 2003). Like most plant's genomes, *Arabidopsis*'s contains a large number of gene duplications, which often translates into completely or partially overlapping gene functions that mask the effect of loss-of-function mutations in individual genes (*Arabidopsis* Genome Initiative, 2000; Vision et al., 2000). Therefore, several months of crossing and selection are needed to produce multiplex mutant backgrounds before any conclusion can be drawn about the function of the gene/s under study. Owing to its *in-trans* activity, CRISPR-Cas9 can be used to simultaneously target several genes with potential redundant functions. (Barrangou et al., 2016).

However, the mutagenesis frequency and inheritability of each mutation is typically low. This is thought to be in part due to the method used for delivering CRISPR-Cas9 constructs into the plant's genome, which involves the production of transgenic plants harboring CRISPR-Cas9. In *Arabidopsis*, DNA fragment encoding sgRNAs and Cas9 is randomly transferred to the plant's genome via *Agrobacterium*-mediated transformation of immature flower cells. This transformation method is easy to use and does not require tissue culture and plant regeneration as it would in other plant species; the method only involves dipping the immature flowers in a suspension of *Agrobacterium tumefaciens* harboring the CRISPR-Cas9 construct in a binary plasmid (Hood et al., 1993; Clough and Bent 1998). The random insertion happens only in a few cells, most of which are somatic cells that do not give origin to ovules or pollen. Therefore, the inserted foreign DNA will not pass to the progeny. Because few germline cells will be transformed with a T-DNA and pass the transgene to the next generation (Clough and Bent, 1998), the capacity of delivering sgRNA and Cas9 enzyme to the germline is limited by the low frequency at which the T-DNA passes to the next generation (Belhaj et al., 2015).

To select for transgenic plants in the next generation, antibiotic/herbicide resistance is commonly used. Thus, an antibiotic/herbicide resistance gene is intergraded into this binary plasmid and the expression of the resistance gene is controlled by a constitutively active promoter. But such a selection method can only inform of the successful insertion of the T-DNA rather than the functionality of CRISPR-Cas9. Each randomly generated transgenic plant expresses diverse levels of Cas9 and sgRNAs depending on the chromosomal location and the chromatin landscape where the construct was randomly inserted (Cong et al., 2013; Nekrasov et al., 2013). Another complication is that, in those T1 generation transgenic plants where Cas9 and the sgRNAs are expressed, most of the editing of the targeted genes happens throughout the lifetime of the plant in somatic cells (Feng et al., 2014). Nevertheless, some editing will occur in the germline and the CRISPR induced mutations will pass on to the next generation. Most often, homozygous plants are identified in the T2 generation, where mendelian genetics and self-fertilization plays out to combine mutated gametes into one single embryo.

Lastly, if the gene under study is not known to produce a phenotype, DNA genotyping techniques will have to be used to identify plants bearing the CRISPR-induced mutation. PCR followed by regular sanger sequencing/restriction enzyme digestion are often used to genotype the mutation (Kim et al. 2014). Other methods applying high resolution melting curve or qRT-PCR are newly developed for high-

---

throughput genotyping, but they are still onerous and time consuming. (Thomas et al., 2014; Peng et al., 2018). These hindrances make the identification of mutants a slow and expensive process.

### **The use of trackable markers**

A potential solution to overcome the complications described in the previous section would be the use of visible markers to aid in the process of identifying and tracking plants bearing CRISPR targeted genes. For example, in animal cells, which typically repair CRISPR induced DSB of DNA via homologous recombination (HR), CRISPR-Cas9 can be used to splice-in a marker gene (i.e. GFP) thereby knocking out an endogenous gene. This allows for the identification of cells, or eventually entire organisms, where GFP is expressed, and hence, the endogenous gene of interest was successfully knocked out (Mashiko et al. 2013; Tálas et al. 2017). In plants, HR is rarely used to repair DSB of DNA making it hard to replace endogenous genes with easily trackable markers that would facilitate the selection of CRISPR mutants (Li et al., 2013; Schiml and Puchta, 2016). Therefore, the identification of mutants mostly relies on a repetitive process of trial and error. A few alternatives have been pursued to alleviate the cost and to reduce the time invested in the identification of CRISPR-edited plants. Cas9-GFP fusions can be used to identify T1 plants that express CRISPR/Cas9 (Osakabe et al. 2016). This strategy yields information about Cas9 expression but does not address the problem of identifying CRISPR-edited plants. Some improvements in the frequency of mutation and inheritability of CRISPR induced mutations have recently been made by using egg- or meristem-specific promoters to drive Cas9 expression. This approach has allowed for the editing of single nucleotides in an endogenous Arabidopsis gene via gene replacement (Wolter et al., 2018). In addition, a similar approach allowed for the in-frame GFP splice-in at specific Arabidopsis locus using donor sequences that provided DNA homology to drive HR-mediated DNA repair (Miki et al., 2018). Yet, the frequency at which these events occur, and entire plants harboring the markers are recovered, is still low and in most cases impractical.

A couple of studies in Arabidopsis and rice have demonstrated that co-editing, the editing of several CRISPR targeted genes simultaneously in the same plant, happens at a high frequency in the somatic tissues of T1 plants (Ma et al. 2015; Minkenberg et al. 2017; Zhang et al. 2016; Yan et al. 2016). Hence, I hypothesized that I could make use of this co-editing phenomenon to aid in the identification of plants bearing CRISPR mutated genes: if I target a gene that produces a visible and easy to identify phenotype, I could use this gene as a proxy for identifying T2 plants where several genes of interest might have been simultaneously mutated. Here, I present an inexpensive strategy to identify multiplex CRISPR-Cas9 mutagenized Arabidopsis plants.

This strategy is based on targeting gene/s of interest simultaneously with a proxy for CRISPR-Cas9 activity, an endogenous gene that produces an easy-to-detect visible phenotype when its function is eliminated by CRISPR-Cas9. For this study, I choose three genes with independent functions and located in different chromosomes. In Arabidopsis, the formation of leaf trichomes is contingent on the function of the *GLABRA-1* (*GLI*) gene. Loss-of-function mutants of *GLI* do not produce trichomes, a phenotype that is easily observable as these plants produce smooth leaves (Herman and Marks 1989; Marksa and Feldmann 1989; Hahn et. al, 2017). This well documented phenotype mitigates the concern that the strategy could interfere with the basal physiology of the studied plant. For the second and third targeted gene, I have chosen *Jasmonic Acid Resistant-1* (*JARI*) and *Ethylene Insensitive-2* (*EIN2*). Loss-of-function

mutations in *JAR1* and *EIN2* produce insensitivity to Methyl-Jasmonate (JA) and Ethylene (ET), respectively. The responses to both JA and ET can be monitored in seedlings exposed to JA or ET in tissue culture plants within a few days upon germination (Costigan et. al, 2011; Alonso et al, 1999). Exposure to JA causes root growth inhibition in plants harboring a wild type allele of *JAR1*(Costigan et. al, 2011). ET exposure under etiolation conditions causes the hypocotyl to bend downwards in seedling harboring a wild type allele of *EIN2*(Alonso et al, 1999). Targeting *GL1*, *JAR1* or *EIN2* with CRIPR-Cas9 will produce plants that have no trichomes or that are no longer sensitive to JA or ET. I have tested the feasibility of using these genes as proxies via assessing the frequency of co-editing at these loci. The co-editing frequency ranged from 55.6% to 93.75% for two genes, and 14.3% to 50% for three genes, depending on the T1 progeny tested and the proxy gene of choice. More importantly, among plants that lacked trichomes (*gl1* mutants) up to 30.8% were homozygous for either *JAR1* or *EIN2* mutant alleles and 3.85% were double homozygous mutants. The selection strategy laid out in this study will facilitate the identification of CRISPR-edited plants not only in Arabidopsis but also in other plant species where CRISPR-edited individual cells cannot be selected for the regeneration of entire plants in *in-vitro* culture. Importantly, the selection strategy laid out here will accelerate the identification of multiplex mutants where several genes with potentially overlapping functions need to be mutated simultaneously.

## **Material and Methods**

### **Design and synthesis of sgRNA expression cassettes.**

The sgRNA were designed with the online web tool at the Zhang lab at MIT (crispr.mit.edu). The retrieved sequences were verified by PCR on genomic DNA from wild-type (Col-0) plants and Sanger sequencing of the PCR amplicon. Each candidate target site was evaluated based on the calculated specificity score and the number of off-target sites. The chosen target site was inserted into an in-silico cloning construct template between the AtU6P promoter sequence and the tracrRNA sequence. The AtU6 promoter, crRNA, tracrRNA and a poly-T (“TTTT”) tail together forms a complete sgRNA expression cassette (Peterson et al., 2016). Each individual cassette was assembled into stackable arrays. Upstream the 5’ sequence of the first AtU6 promoter region, there was a 32bp sequence inserted for the convenience of future cloning. Also, a 17bp sequence was inserted downstream the 3’ sequence of the last poly-T tail for the same purpose (Supp Fig. 4). The in-silico construct is stored in an Ape format file. The final DNA sequence was synthesized through Genscript™ Custom Gene Synthesis services (Cat#SC1010).

### **T-DNA construct and bacteria preparation**

The DNA fragment of the synthesized sgRNA expression cassettes was amplified through PCR reaction (NEB Phusion® Cat#M0530S, Forward Primer : 5’aggctcccgggtgctgcgacggtctcaggtcagagcttg3’, Reverse Primer 2: 5’gaaagctgggtgattcaagcttggtctcatcaggatccaaag3’). The PCR amplified DNA fragment was then assembled with a pDONR vector which was flanked by restriction enzyme Sall (NEB, Sall-HF® Cat#R3138S) and HindIII (NEB, HindIII-HF® Cat#R3104S) through In-Fusion reaction (Takara® In-Fusion® HD Eco-Dry™ Cloning Plus Cat#638915) to form a donor vector (pDONR-CE). The pDONR-CE vector contains the Gateway (GW) cloning sites AttL1 and AttL2. The sgRNA expression cassettes stacking was inserted between the two GW cloning sites after the In- Fusion

assembling. Further through GW LR cloning reaction (ThermoFisher Scientific<sup>®</sup>, Gateway<sup>™</sup> LR Clonase<sup>™</sup> Enzyme mix Cat#11791019), the entire sgRNA expression cassettes fragment was transferred into a binary vector (pCUT3) which also encodes Cas9 enzyme conjugated with nuclear localization signal (NLS) and an epitope tag (HA) under the control of a UBQ10 promoter.

*E. coli* cell Top10 and DH5Alpha were used for propagate vectors pDONR-CE and pCUT3-CE respectively. The finalized pCUT3-CE vector was also transformed into *Agrobacterium* (GV3101) for plant transformation. *E. coli* cells and *Agrobacterium* cells were grown in Lysogeny broth (LB) and YEP medium respectively. Bacteria selection was based on the resistance gene carried on the vector and using 50 µg/mL Zeocin (Thermo Fisher Scientific<sup>®</sup>, Cat#R25001) and 100µg/mL Spectinomycin (Millipore Sigma<sup>®</sup>, SKU#S4014).

### **Plant transformation, selection and handling**

All the plants used in this study were Columbia-0 (Col-0) background. All transgenic plants were transformed with the finalized binary vector (pCUT3-CE) through standard *Agrobacterium*-mediated flora dipping process. To select transgenic seeds, T1 seeds were sown on sterilized plate made from the mixture of Murashige and Skoog (MS) medium (Murashige and Skoog, 1962), 0.7% phyto agar (w/v, Plant Media Cat#40100072-2) and 50µg/mL Kanamycin (fisher scientific<sup>®</sup> CAS#25389-94-0). Seeds were surface sterilized by 10%(v/v) bleach and 0.1% Tween 20 (v/v). After 14 days of growth on the sterilized agar plate, resistant seedlings were count and moved into soil for further growing. Green leaf tissue was collected from each independent transgenic plant after 4 weeks growing and preserved in liquid Nitrogen and stored under -80°C.

To visual select the glabrous plants, T2 seeds were stratified in 0.1% phyto agar mixture at 4°C for 3 days. Then, each individual seed was sown 1cm apart from each other to facilitate visual inspection. After 3 weeks of 16hr light photoperiod growth, glabrous plants were visually identified by the lack of trichomes. The visual identification of *jar1* and *ein2* homozygous mutants was carried out in MS plates supplemented with Methyl-Jasmonate (Millipore Sigma<sup>®</sup> SKU#W341002) or ACC (Millipore Sigma<sup>®</sup>, SKU#A3903) as previous described (Alonso et al, 1999, Costigan et. al, 2011) in tissue culture plates. All the identified seedlings were transferred out carefully from the selection plates into an independent box with soil for further growth, sampling and seeds propagation.

### **Protein and Western Blot assay**

To probe the expression of Cas9 enzyme using leaf tissue, total denatured protein content was extracted from 4-weeks old green leaf tissue. Green leaf tissue was ground in liquid nitrogen mixed with protein loading buffer (Tris-HCl, pH:8.8 ) and heat to 100°C and boil for 5min. The extracts were then centrifuged at 17,000x g for 5min at 25°C and the supernatant was used for gel blot analysis. Protein separation was done by poly-acrylamide gel (0.375M Tris-HCl pH=8.8, 8% Acrylamide, 0.05% APS, 0.1% SDS). After gel separation, proteins were transferred into PVDF membrane (Thermo Scientific<sup>™</sup>, Cat#88520) with 70V under 4°C for 90min. Monoclonal rabbit anti-HA antibody (Cell Signaling Technology<sup>™</sup>, mAb#3724, 1:4000 dilution) and Monoclonal mouse anti-β-Actin antibody (Millipore Sigma<sup>™</sup>, Cat#MAB1501) were used as primary antibody. Secondary HRP conjugated anti-rabbit IgG antibody (Jackson Immuno Research Laboratories, Inc. Code#111-0350144) and fluorescent (LiCor<sup>™</sup> IRDye<sup>®</sup> 800CW Goat anti-rabbit IgG P/N#925-32210) conjugated anti-mouse antibody

---

were used for probing the primary antibody and blot visualization respectively. Chemiluminescence was activated by ECL substrate (Bio-Rad®, Cat#1705060). Whole blot was detected by Bio-Rad® ChemiDoc MP® system and image result was analyzed by Bio-Rad Image Lab™ software.

### **Gene sequencing and mutation detection**

To detect the mutation created via CRISPR/Cas9, primers that are ~300bp upstream the target site and ~200bp downstream the target site were used to amplify such DNA fragment from identified plant genomic DNA material as template with PCR (New England BioLab® Inc., Phusion®, Cat#M0530S) reaction. After removing the excessive PCR primers (Fisher Scientific® ExoSap-IT™ Cat#78-201-1ML), the PCR product was sent to sanger sequencing (Eurofin Genomic). The sequencing result in FASTA format was analyzed using BLASTA web tool from NCBI website, aligned against wild-type (Col-0) genomic sequence to identify wild-type (no change) gene, and homozygous edited gene. In order to further discriminate heterozygous, raw chromatograph reading result in ab1 format was input into an online web tool (Synthego®, ICE® Analysis; <https://ice.synthego.com>). This process was repeated with all the three targeted genes in this study.

### **Analysis of sgRNA nucleotide composition and secondary structure**

The calculation of G/C content of all sgRNAs that were used in the experiment were done by a Python script. The secondary structure was predicted by input the sgRNA FASTA sequence into the Mfold web server (Zuker et al., 2003)

## **Results**

### **Identification of mutant plants using visually identifiable phenotypes to report CRISPR activity**

To deliver each component of the multiplex gene targeting CRISPR/Cas9 system (sgRNAs and Cas9 nuclease), I used an all-in-one vector (pCUT3-CE) that contains the sgRNA expression cassette stacking, and the Cas9 protein with HA fusion downstream of the UBQ10 promoter (Supp Fig. 1A). Arabidopsis U6 promoter (AtU6) and poly-T (“TTTT”) tail were used to control the expression of each sgRNA (Supp Fig. 1A; Supp Fig. 4). To provide similar expression level across the sgRNAs, each sgRNA is transcribed from an independent AtU6 promoter and each one has its own transcriptional termination. The RNA polymerase III transcription start site “G” was incorporated between the last nucleotide of the AtU6 promoter sequence and the first nucleotide of the sgRNA. The AtU6 promoter, the sgRNAs and the transcriptional termination sequences together constitute one expression cassette. (Supp Fig. 1A; Supp Fig. 4).

To visually identify CRISPR induced mutants I chose to target *JASOMNIC ACID RESISTANCE-1* (*JAR1*), *GLABROUS-1*(*GL1*) and *ETHYLENE INSENSITIVE-2* (*EIN2*). These genes are well characterized and their loss-of-function mutants are easily identifiable. The 20nt sgRNA sequence (target site) of each gene was designed to anneal the DNA sequence of either 1<sup>st</sup> or 2<sup>nd</sup> exon (Supp Fig. 1, B-C-D) to provide the highest likelihood that the DNA repair will destroy the translational frame of all predicted isoforms encoded by *JAR1*, *GL1* and *EIN2*. Three individual expression cassettes with its own 20nt crRNA sequence corresponding to each targeted gene were assembled as a stack (Supp Fig. 1A).

After agrobacteria-mediated transformation of immature flowers, I recovered a total of 26 independent T1 transgenic plants. After self-fertilization, I randomly choose the progeny of 4 independent T1 lines (#1, #3, #4 and # 25) for visual phenotype screening. For each independent T1 line, ~1000 T2 seeds were used and screened for each one of the 3 visual phenotypes linked to loss-of-function mutations in either *JAR1*, *GL1* or *EIN2*. I identified T2 plants in the progeny of T1 line #1, #3 and #25 that showed either jasmonate insensitivity (*jar1*), loss of trichome (*gl1*) or ethylene insensitivity (*ein2*) (Fig 1; Supp Fig. 2 A-F). As the screening of visual phenotypical mutants are sampled from and grouped by different independent transgenic plants, each screening dataset represents an independent experiment.

Cas9 expression for all four T1 lines analyzed was monitored via western blotting. The results showed Cas9 expression in lines #1, #3 and #25, but no in T1 line #4, which explains the lack of observable phenotypes in its T2 progeny (Supp Fig. 3). The percentage of visually detected T2 plants varied across independent T1 lines and did not correlate Cas9 expression (Fig 1; Supp Fig. 3). To increase the probability of selecting for CRISPR/Cas9 germline mutagenized plants, I only selected plants that had no trichomes anywhere in the leaves, or showed long roots or long hypocotyl (Supp Fig 2 D-E-F); I removed from the screening pipeline those plants that showed mosaic phenotypes, where, for instance, only patches of somatic cells showed the expected phenotype. In addition, the visually identified T2 mutants were allowed to self-pollinate and produce seeds for progeny tests. All the T2 plants that show *jar1*, *gl1* or *ein2* phenotypes produced 100% mutant T3 progenies, which confirmed T2 plants were loss-of-function mutants, as indicated by their phenotypes.

---

## Assessment of co-editing

To weigh the power of each proxy to predict co-editing, that is, to predict the occurrence of CRISPR-induced mutations in other targeted genes of the same plant, I PCR amplified and Sanger sequenced 600bp of DNA surrounding the sgRNA annealing site of each targeted gene. The Sanger sequencing results were analyzed using two software tools, NCBI BLAST and ICE online software (<https://ice.synthego.com>), and yielded three different indicative results: (1) elucidation of mutation types; (2) general co-editing frequency; (3) detailed mutant gene allelic state.

Among T2 visually selected plants, NCBI BLAST detected five different alleles of *jar1*, three different alleles for *gll* and four alleles for *eni2* (Fig 2). The mutations detected were consistent with previous studies reporting insertions of 1 or 2 nucleotides (+1 or +2) and deletions of 1, 2 or 3 nucleotides (-1, -2 and -3) within the six nucleotides upstream (5'-) of the PAM sites, which are the hallmarks of Non-Homologous End-Joining DNA repair in Arabidopsis (Jinek et al., 2012; van der Oost et al., 2013; Peterson et al., 2016 & Hahn et al., 2017).

To assess co-editing frequency in visually selected plants, and to determine whether the mutations in the second and third gene tested would also produce loss-of-function mutants, I analyzed the DNA sequences with NCBI BLAST alignment tools. Here, frequencies were determined by counting the number of double and triple co-edited plants and then dividing by the total number of T2 plants identified with each proxy (either smooth leaves, JA insensitivity or ET insensitivity). Among plants without trichomes (smooth leaves), the co-editing frequency varied from 61.1% to 93.75% for double co-edited plants where either *JAR1* or *EIN2* were targeted simultaneously with *GLI*, and 18.5% to 50% for triple co-edited plants, where both *JAR1* and *EIN2* were co-targeted with *GLI* (Fig. 3). Among ethylene insensitive plants, double gene co-editing frequency varied from 55.6% to 81.25%, while triple gene co-editing frequency ranged from 0% to 56.25% depending on the T2 progeny (Fig. 3). Although JA insensitive plants (*jar1*) were also observed in *in plate* selection, the numbers were too low to run co-editing analysis (these data were not included in Fig. 3). The low recovery of JA-insensitive mutants in the T2 progenies of independent T1 plants was not entirely unexpected. Again, as the assessment of co-edited mutants are sampled from and grouped by different independent transgenic plants, each assessment dataset is an independent experiment

ICE software from Synthego® (<https://ice.synthego.com>) was used to weigh the DNA sequence quality of the Sanger sequencing data, as well as to determine the relative contribution of each nucleotide to each given position in the DNA sequence of each PCR product (Hsiao et al., 2018). This detailed analysis of each T2 mutant plant's second and third CRISPR-targeted genes' allelic state, recovered from the visual selection screening, revealed the frequency at which double or triple mutants would be indirectly selected for based on the phenotype produced by each proxy. It will be expected that the PCR products obtained from somatic tissues (leaf) of each T2 mutant plant will contain a mix of alleles that do not align with the reference wild type allele, particularly within the 6 nucleotides 5' of the PAM site. This mix of alleles stems from two factors: (1) CRISPR/Cas9 in T1 plants could produce mutated ovules that may be fertilized by wild type or mutated pollen; in turn, the pollen may bear a similar or a different allele compared to the ovule's allele; (2) the *in-trans* activity of CRISPR-Cas9 will continue to target each gene in the germline of T1 plants as well as in the somatic cells of each T2 plant for as long as Cas9 and the target genes remain in the same

background. Because of these two factors, the PCR products obtained from each T2 selected mutant will reveal each plant's genetic constituency (allelism) in the form of DNA sequencing fluorescence chromatograms where a given fluorescence peak (nucleotide) in the wild type reference allele will be occupied by one or more nucleotides. These alternative nucleotides would be revealed as overlapping smaller peaks that correspond to mutant allele present in the PCR mix. Thus, addressing the allelic composition in the DNA sequence of each target site will reveal if the plant under study is either wild type, homozygote mutant, heterozygote or Bi-allelic (or higher order allelism) for each gene under study (Feng et al., 2014).

Among plants that lacked trichomes (*g1l* mutants) up to 30.8% were also homozygous for either *JAR1* or *EIN2* mutant alleles. Although most plants tested had a least one wild type allele for one of the three genes tested, double homozygous mutants were found in the T2 progeny of T1 line #25, where no wild type alleles for *JAR1* and *EIN2* were detected in 3.85% of plants analyzed. Importantly, up to 30.8% of the T2 plants were homozygous mutants for a second gene and heterozygous for a third targeted gene (Table I). Among ET insensitive plants, I found up to 46.2 % of plants (in T2 progeny of T1 #25) that were also homozygous for either *JAR1* or *GLI* mutant alleles. Most plants in all of the three T2 progenies analyzed were heterozygous for either *JAR* or *EIN2* mutant alleles (Table I). Interestingly, neither the percentage of visually selected plants (Fig. 1) nor the percentage of mutant alleles in the second or a third gene within the visually selected T2 plants seem to correlate with Cas9 expression (Fig. 3; Supp Fig. 3).

## **Discussion**

### **Differential efficiency of CRISPR/Cas9 mediated mutagenesis across *GLI*, *EIN2* and *JAR1***

Although the abundance of a 20nt site (crRNA) with NGG adjacent downstream is usually not a limiting factor within the topology of a gene, the efficiency among these crRNA target sites are always different. I posited three causes of the low editing efficiency in *JAR1*. These included: (1) a potentially higher number of off-targets which could have led to low specificity and therefore competition between the target sites; (2) a lower G/C content which might have resulted in a low target site binding efficiency; (3) the 20nt crRNA sequence could contain specific nucleotides that interfere with the sgRNA secondary structure folding, which would have inhibited Cas9-sgRN complex activity.

The first factor under consideration, the specificity score, which is reported as a whole number out of 100, among the three 20nt crRNA sequences were similar (cr*JAR1*: 98/100, cr*GLI*: 98/100, cr*EIN2*: 99/100). Yet, the number of potential off-targets, which are reported as a whole number, for *JAR1* crRNA was higher than for *GLI* and *EIN2* crRNAs (cr*JAR1*: 9, cr*GLI*: 5, cr*EIN2*: 3). Assuming that there is competition for sgRNAs across potential off-targets, I will expect to see that targeting efficiency of *EIN2* is the highest. However, the off-targets numbers do not seem to explain the high mutagenesis of *GLI* and low mutagenesis of *JAR1* (Doench et al., 2014&2016). Therefore, I should consider other explanations.

The second factor under consideration is the lower annealing stability of the sgRNA to the target due to a low G/C content. The G/C content of cr*JAR1* is 35%, while the G/C content of cr*GLI* and cr*EIN2* is much higher (cr*GLI*: 50%, cr*EIN2*: 55%). In plant cells, G/C content between 30% and 80% is thought to be a key factor

affecting CRISPR/Cas9 mutagenesis efficiency because appropriate G/C content can improve the binding stability between the crRNA sequence and the target site sequence on the genome (Liang et al, 2016). Hence, albeit still within normal values, the low G/C content of crJAR1 could explain its low mutagenesis efficiency.

A third factor to consider is secondary structure of the entire sgRNA. A study of the secondary structure of sgRNAs revealed that 3 stem loops (hairpins) are necessary for effectively forming a sgRNA-Cas9 complex (Liang et al., 2016). Among these, stem loop #1 is crucial for the formation of a functional Cas9-sgRNA-DNA complex, while stem loop #2 is critical to improve complex stability and *in vivo* activity. I performed in-silicon analysis of the *JAR1*, *GL1* and *EIN2* sgRNAs sequences using the online tool Mfold (Zuker et al., 2003). The secondary structure predictions of *JAR1* sgRNA showed that stem loop #1 and #2 are missing in all three predicted results (Supp Fig. 5, A-B-C). A single in-silicon prediction of *GL1* sgRNA revealed that all 3 stem loops are intact (Supp Fig. 5, D). The *EIN2* sgRNA received two predictions (Supp Fig. 5, E-F). One of the *EIN2* sgRNA predictions showed all 3 stem loops are intact (Supp Fig. E), while the 2<sup>nd</sup> predicted structure missed stem loop #1 (Supp Fig. F). The design of *EIN2* sgRNA may not be optimal and this could explain the result of the recovery percentage of ET insensitive plants, which is lower than trichome less plants and higher than JA insensitive plants.

The missing loops result from undesired base pairings between crRNA and tracrRNA nucleotides within each sgRNA. The crRNA sequences should have no more than 12 nucleotides that can pair within the sgRNA sequence. Within these 12 nucleotides, the number of consecutive base pairs (CBPs) should be less than 7 and no more than 6 internal base pairs (IBPs) (Liang et al., 2016). Unfortunately, there were 16 CBPs between the *JAR1* crRNA and the rest of the sgRNA in all *JAR1* sgRNA predicted structures (Supp Fig. 5, A-B-C). While predictions of *GL1* sgRNA showed 7 CBPs on the crRNA region and no IBPs (Supp Fig. 5, D), one *EIN2* sgRNA prediction showed 5 IBPs (less than 6) and a second prediction rendered 10 IBPs on its crRNA region (Supp Fig. 5, E-F).

In summary, the differential efficiency of mutagenesis should be explained by one of the three issues delineated above—although I hypothesize that suboptimal RNA secondary structure is the likely cause. Yet, there still remains the possibility of competition for limiting transcription proteins or physical exclusion at neighboring promoters. This possibility has been ruled out. As each sgRNA was clustered in the same insert and expressed from individual U6 promoters, I did not find compelling reasons to think that dissimilar sgRNA expression across the three sgRNAs could contribute to differences in mutagenesis frequencies across the three targets (Zhang et al., 2016). In addition, the position of each sgRNA in the cluster does not correlate with the observed co-targeting frequencies. Considering all the above, I can conclude that CG content and secondary structure of the sgRNA is critical for mutagenesis efficiency. These factors, together with target specificity and the existence of PAM site near a unique gene specific target sequence, are major factors that limit CRISPR-mediated mutagenesis efficiency.

### **The use of proxies to aid in the identification of CRISPR mutagenized plants.**

There are no precedents in the literature of attempts to deal with improvements in the identification of CRISPR-Cas9 mutagenized plants. Previous studies focused on identifying T1 plants where Cas9 is expressed. Most of these studies use Cas9 fusions to fluorescent proteins or an epitope tag (eg. HA/FLAG), and require taking samples

---

leaves and observing them under a fluorescence microscope or performing western blots to assess Cas9 expression (Osakabe et al. 2016; Cody et al. 2017); after that initial identification, brute force is needed to collect samples of T2 progenies of those T1 plants and proceed with the molecular DNA fingerprinting needed to identify plants bearing mutated genes of interest. My approach of targeting a proxy to identify plants where other genes of interest were mutated, allowed the identification of *gll* T2 plants where more than 3/100 plants were homozygous for mutations in a second gene and a third gene, while as much as 30/100 *gll* plants were homozygous for a second and heterozygous for a third gene of interest. The results showed that *GLI* mutagenesis is very efficient and that the identification of *gll* mutant plants is inexpensive and easy to implement without the need of any sophisticated piece of equipment: it only involves growing T2 plants in soil and visually identifying plants with smooth leaves that are easily observable among the majority of hairy plants. As *GLI* is conserved in many plant species and its expression dictates the development of trichomes, I hypothesize that a similar strategy could be efficiently applied to the identification of CRISPR mutagenized plants in crops with complex, often polyploid, genomes, where the search for individual plants bearing multiplex mutations in genes of interest could be substantially more onerous and time consuming than it is in *Arabidopsis*. The selection of the appropriate proxy gene to target will depend on the expected phenotype produced by the genes of interest. Whenever possible, *GLI* targeting will serve as a good proxy. In other cases, and other species, perhaps other proxies will be more suitable.

### **Ethical concerns about the use of CRISPR**

Most of the ethical concerns about the use of CRISPR-mediated genome editing relate to the use of this technology to genetically modify human germline cells. For the first time in human history, it is relatively easy and affordable to make changes to the genome of a human cell *in vitro*. Coupled to cell cloning techniques, CRISPR could open the door to regenerate genetically modified human embryos that will stably inherit the desired genetic modification. Also, new technologies have opened the possibility to manipulate sperm or ovules *in vivo*, which would allow germline inheritable modification without the need for expensive *in vitro* culture of cells and embryo regeneration. Although the potential for these technologies to cure diseases is evident, there are ethical concerns based on two potential drawbacks: (1) the changes are inheritable and potentially irreversible; (2) the technology is not safe enough yet to use without the risk of editing non-targeted genes (i.e. tumor suppressor genes or oncogenes), which could derive in unintended health consequences. There is an even stronger concern: the use of virus-delivered CRISPR to create biological weapons (Clapper, National Intelligence, 2016).

As CRISPR/Cas9 mediated mutagenesis and editing techniques developed, several research groups in academic and industry settings started to use CRISPR/Cas9 to produce GM crops (Wang et al., 2014; Khatodia et al., 2016). The use of CRISPR in plants have raised the old concerns about genetically modified food and environmental pollution. GM crops were introduced in the field for the first time 1996. The debate on human health risk posed by GM crops is still ongoing, but it seems clear by now that all the major threats feared by environmental activists did not materialized. From an environmental pollution perspective, not having had made use of more genetically modified crops has actually played against the environment, as the use of polluting chemicals (pesticides, fertilizer, herbicides, etc.) could have been cut to a greater extent over the past 25 years if more genetically modified crops would have been developed

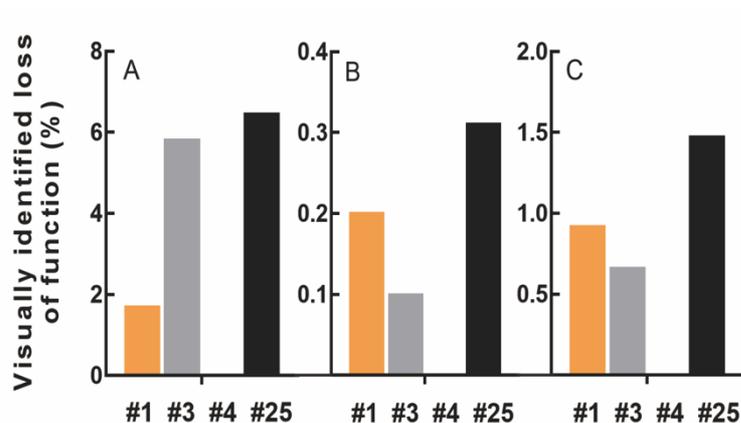
and commercialized. A major concern about GM crops has always been the presence of foreign DNA encoding antibiotic or herbicide resistance that could spread to other species in the environment. There have been documented cases of GM plants of canola and turfgrass (among others) growing in the wild near farms and roads, but the extent to which these plants can harm the environment via displacing or outcompeting non-GM species is unknown, and presumably very low. Indeed, the genetic modification that they bear (i.e. herbicide resistance) will not provide these plants with any advantage out of the farm where the herbicides are not applied. For insect resistant GM crops (like corn, cotton and soybeans) the consequences for the environment are higher if crosspollination with wild relatives happens, as the spreading of insect resistant genes could potentially shift an entire ecosystem balance in unpredictable ways. Two factors alleviate this risk: (1) cross-pollination in these species is very unlikely; (2) insect resistant genes typically have a very narrow range of target species. Cross-pollination is major concern, particularly for canola, as these plants can interchange genes with several related weeds in the US and Canada. It is hard to predict how evolution will play, but assuming that the genetic modifications will only give an advantage over any non-GM plants under selective condition, the problem relates more to a practical issue than an ethical issue: how to control herbicide resistant weeds. There are several alternatives that the biotech industries can use to mitigate scape of GM genes into the ecosystem: (1) limit the use of transgene to only self-fertilizing species; (2) turn out-crossing species into self-fertilizing species; (3) use positive selection to maintain crops alive (plants will die in the absence of the selection out of the farm), etc. But all these alternatives to make GM crops safer require genetic manipulation. Hence, the stronger the opposition to the use of GM crops, the longer it will take to see the benefits.

As CRISPR/Cas9 targets DNA sequences *in-trans*, the transgenic piece of DNA where CRISPR-Cas9 and the selection markers (antibiotic/herbicide resistance genes) reside, could be segregated apart from the CRISPR-targeted genes once the *in-trans* editing event/s has happened. By crossing mutants created by CRISPR/Cas9 with wildtype crops and actively selecting plants that *do not* contain the transformed DNA fragment but *do contain* the mutated alleles, gene drive can be avoided. As these crops would not harbor foreign DNA in its genetic makeup, the genetic modifications introduced with CRISPR would be no different from any other DNA polymorphism that naturally exists across different cultivars of the crop. However, one important caveat that hinders the use of CRISPR to produce GM crops is that the technology is still inadequate to make precise single base editing and gene replacement, or splicing in foreign DNA sequences (Komor et al., 2016). These limitations stem from the low frequency at which plants repair DNA via HR. Recently, two new studies reported the use of meristematic cell and embryo/germline specific promoters to drive Cas9 expression for gene replacement in Arabidopsis (Wang et al., 2015; Wolter et al., 2018; Miki et al., 2018). The authors reported a significant increase in the recovery of successfully replaced genes and successfully spliced-in foreign DNA sequences (GFP) attributed to the use of egg or meristem-specific promoters to drive the activity of Cas9. Importantly, Miki et al., in a 2018 *Nature Communication* article, reported the use of CRISPR to splice in a GFP marker in frame with the DNA coding sequence of two endogenous Arabidopsis genes without the use of a selection marker. The selection strategy presented in this study shows a second option of convenient tracking/selection of CRISPR modified plants: targeting *GL1* as the selection marker allows for the growth of plant in a stress-free environment during the selection pipeline. In addition, the labor work for selecting trichome-less plants is minimum. These discoveries have set the

stage for the easy and relatively inexpensive use of CRISPR technology to produce GM crops.

### Figures and Legends

Figure 1



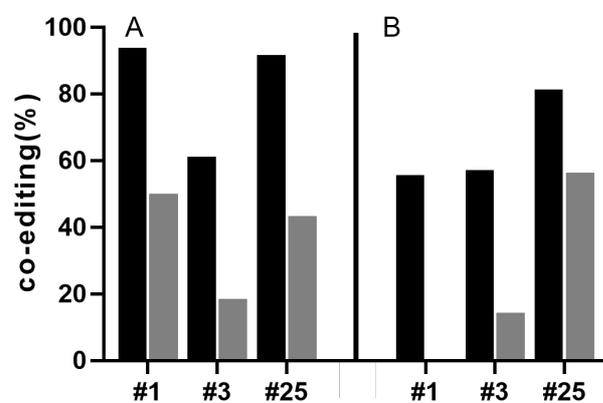
**Figure 1.** Frequency analysis of edited genes in T2 plants. Percentage of plants that showed the corresponding mutant phenotypes calculated as mutant/total number of T2 plants obtained from four independent T1 plants (CE-1, CE-3, CE-4 and CE-25). N = 130 *gll* plants; 5 *jar1* plants; 32 *ein2* plants

Figure 2

A	<i>JAR1</i>	
WT	5'-ACCACAGT <b>TTTGTAAGTAAATGGCGGATTGG</b> TTCT-3'	In/Del
M1	5'-ACCACAGT <b>TTTGTAAGTAAATGGCG</b> <b>G</b> ATTGGTTCT-3'	+1bp
M2	5'-ACCACAGT <b>TTTGTAAGTAAATGGCG</b> <b>A</b> GATTGGTTCT-3'	+1bp
M3	5'-ACCACAGT <b>TTTGTAAGTAAATGGCG</b> <b>T</b> GATTGGTTCT-3'	+1bp
M4	5'-ACCACAGT <b>TTTGTAAGTAAATGGCG</b> -ATTGGTTCT-3'	-1bp
M5	5'-ACCACAGT <b>TTTGTAAGTAAATGG</b> <b>---</b> ATTGGTTCT-3'	-3bp
	<b>sgRNA target spacer</b> <b>PAM</b>	
B	<i>GL1</i>	
WT	5'-CAGTTTTTCTGACGATGCGGTTCCATTGGCCAGTG-3'	In/Del
M1	5'-CAGTTTTTCTGACGATGCGGTT <b>C</b> ATTGGCCAGTG-3'	+1bp
M2	5'-CAGTTTTTCTGACGATGCGGTT <b>CC</b> ATTGGCCAGTG-3'	+1bp
M3	5'-CAGTTTTTCTGACGATGCGG <b>T</b> -ATTGGCCAGTG-3'	-2bp
	<b>sgRNA target spacer</b> <b>PAM</b>	
C	<i>EIN2</i>	
WT	5'-TGTTGCAGCTCGCATAAGCGTTGTGACTGGTAAAC-3'	In/Del
M1	5'-TGTTGCAGCTCGCATAAGCGTT <b>G</b> TGACTGGTAAAC-3'	+1bp
M2	5'-TGTTGCAGCTCGCATAAGCGTT <b>GT</b> GACTGGTAAAC-3'	+2bp
M3	5'-TGTTGCAGCTCGCATAAGCGTT <b>G</b> -ACTGGTAAAC-3'	-1bp
M4	5'-TGTTGCAGCTCGCATAAGCGTT <b>G</b> <b>---</b> ACTGGTAAAC-3'	-2bp
	<b>sgRNA target spacer</b> <b>PAM</b>	

**Figure 2.** Sequence analysis of CRISPR/Cas9 edited mutants. **A-C)** CRISPR/Cas9 induced mutations detected via PCR and sanger sequencing for *JAR1* (A), *GL1* (B) and *EIN2* (C) in T2 plants. In green, blue and red are depicted the sgRNA target/spacer site, the PAM site and insertions and deletions detected in the DNA sequence, respectively. WT: wild type reference DNA sequence. M1-M5: mutation types.

Figure 3



**Figure 3.** Scoring of T2 plants that bearing two or three genes edited simultaneously. A: selection based on lack of trichomes (*g11*). B: selection based on ethylene insensitivity (*ein2*). Black bars indicate that mutations were detected in one gene in addition to the selectable visual marker. Grey bars indicate that mutations were detected in two gene in addition to the selectable visual marker.

Table I

		<i>jar1</i> or <i>ein2</i>			<i>jar1</i> and <i>ein2</i>			
		HM	HET	Bi	Double-HM	HM/HET	Double-HET	Bi/HET
Lack of Trichome	#1	20.0	73.3	6.67	0	25	75	0
	#3	0	100	0	0	0	100	0
	#25	30.9	63.6	5.45	3.85	30.8	57.7	7.69

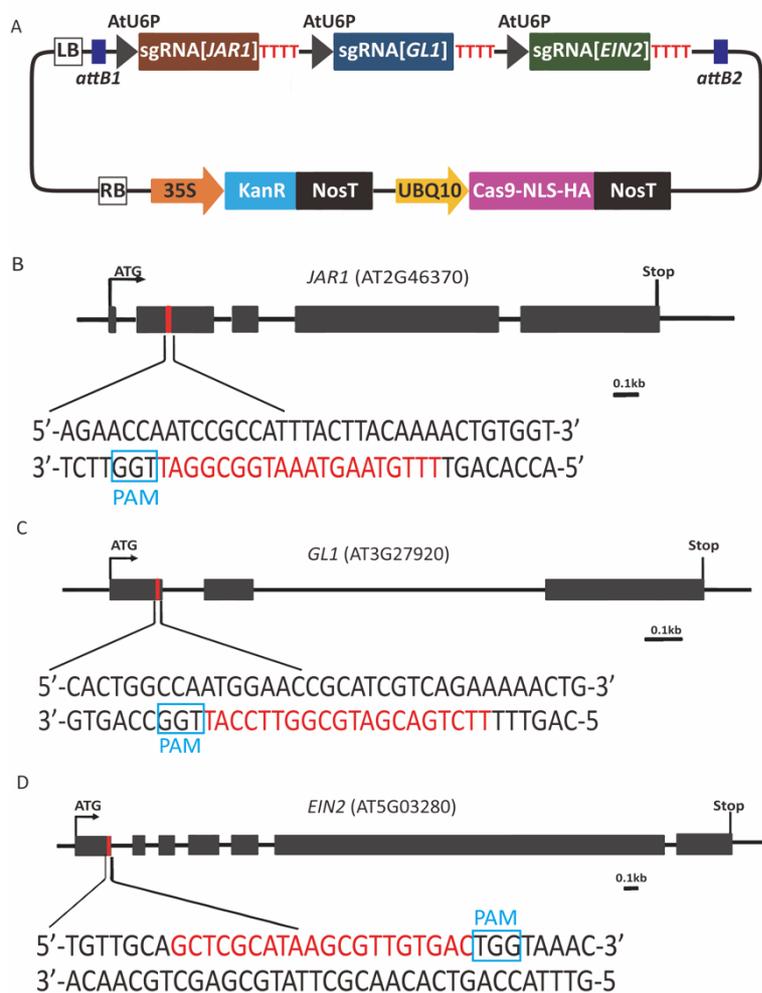
---

		<i>gl1</i> or <i>jar1</i>			<i>gl1</i> and <i>jar1</i>			
		HM	HET	Bi	Double-HM	HM/HET	Double-HET	Bi/HET
Ethylene insensitivity	#1	40.0	60.0	0	0	0	0	0
	#3	0	100	0	0	0	100	0
	#25	46.2	53.8	0	0	44.4	55.6	0

**Table 1.** Genetic constituency of co-edited genes. Numbers indicate the percentage of co-edited genes and their allelism in T2 plants organized by alleles detected within each group. HM: homozygous. HET: heterozygous. BI: bi-allelic. Double-HM: similar or different mutant alleles for both genes under study. HM/HET: similar or different mutant alleles for one gene and at least one wild type allele for the second gene under study. Double-HET: at least one wild type allele for each gene under study. Bi-HET: multiple alleles in one gene and at least one wild type allele for the second gene under study.

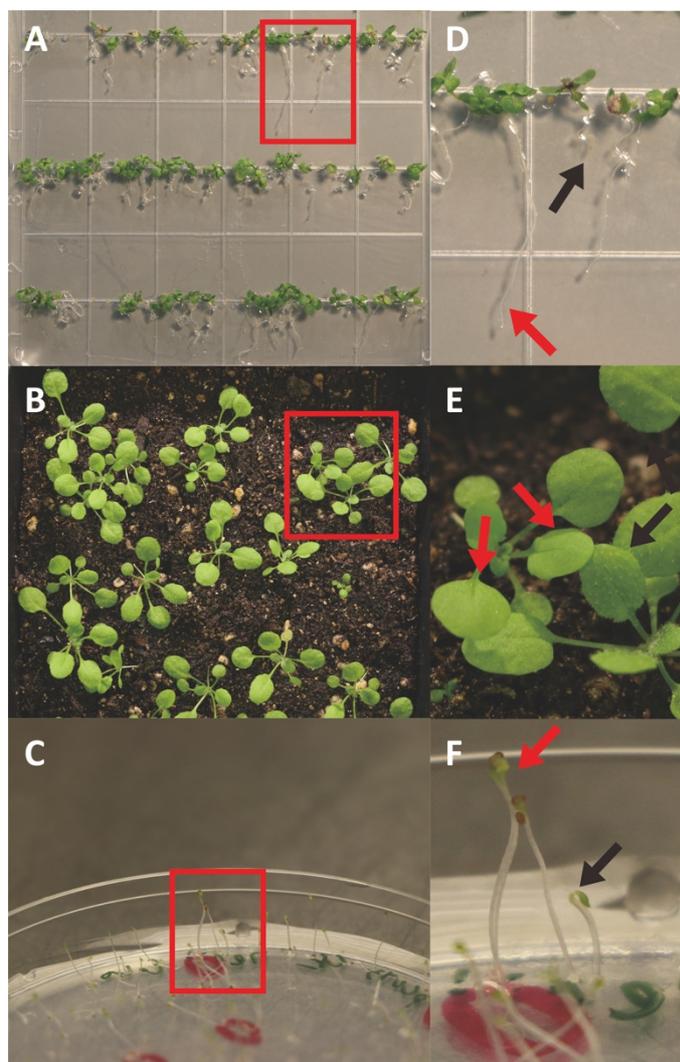
## Supplementary Figures

## Supplementary Figure 1



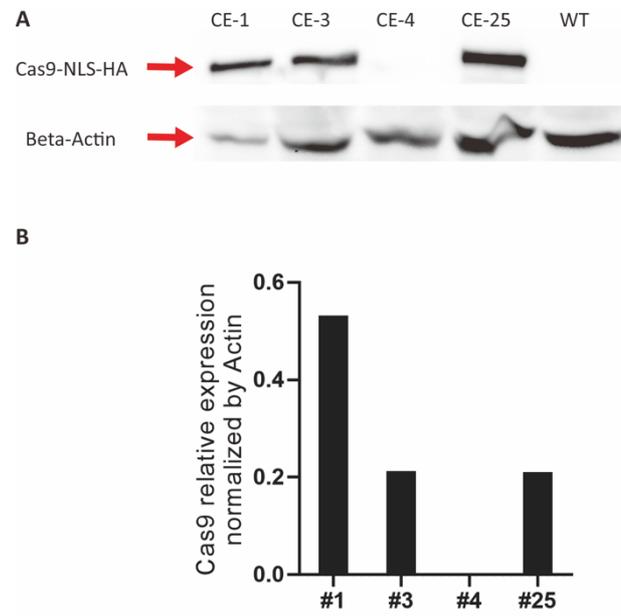
**Supplementary Figure 1.** Schematic illustration of the CRISPR/Cas9 construct and the sgRNA target. A) Schematic illustration of the CRISPR/Cas9 construct. The construct contains 3 individual cassettes that can generate single-strand guide RNA targeting *JAR1*, *GL1* and *EIN2*. The expression of each sgRNA is controlled by individual *AtU6* promoters (U6P) and Individual poly-T terminator (TTTT). B-D) *JAR1*, *GL1* and *EIN2* target sites. Sequence highlighted in red denote the 20nt crRNA target site. The blue rectangular highlights the PAM site. Scale bar=0.1kb

## Supplementary Figure 2



**Supplementary Figure 2.** Identification of visual phenotypes. A-C) Mutant phenotypes identified through visual observation for *jar1* (A), *gl1* (B) and *ein2* (C). The red box highlights the edited loss-of-function mutant plants side-by-side with wild type non-edited plants. D-F) Enlarged picture area in red box in A (*jar1*), B (*gl1*) and C (*ein2*). The red arrow points out the mutant phenotype. The black arrow points out the wild type (WT) phenotype. Each mutant plant was selected as described in Material and Methods.

## Supplementary Figure 3



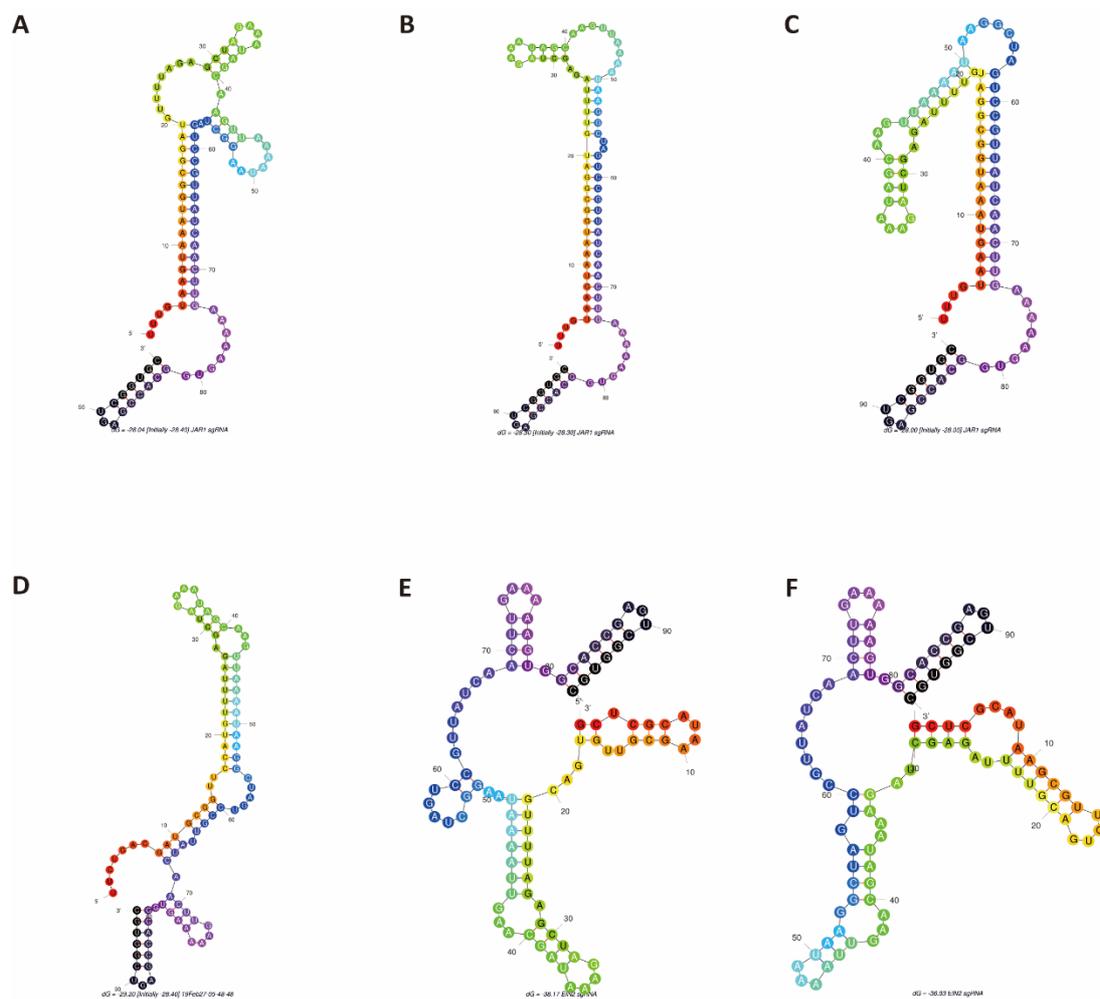
**Supplementary Figure 3.** Cas9 expression. A) Cas9 expression in 4 independent transgenic lines were tested via western blot. B) Quantification of Cas9 expression normalized by Beta-Actin. WT; wild-type Arabidopsis. CE; CRISPR-editing.

## Supplementary Figure 4

GGTCTCAGGTCAGAGCTTGTTTCAGGACTCGAGcattctcattcttaagatatg  
aagataatcttcaaaagcccctgggaatctgaaagaagagaagcaggcccatttatatgggaaagaaca  
atagtatttcttatatagcccatttaagtgaaaacaatcttcaaaagcccacatcgcttagataagaaaacg  
aagctgagttatatacagctagagtcgaagtagtgatt**gTTTGTAAAGTAAATGGCGGA**  
**TGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT**  
**ATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT**GAGCTCcatc  
tcattcttaagatatgaagataatcttcaaaagcccctgggaatctgaaagaagagaagcaggcccattta  
tatgggaaagaacaatagtatttcttatatagcccatttaagtgaaaacaatcttcaaaagcccacatcg  
ttagataagaaaacgaagctgagttatatacagctagagtcgaagtagtgatt**gTTCTGACGAT**  
**GCGGTTCCAT**GTTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG  
CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT**TTT**  
**T**GAGCTCcatcttcattcttaagatatgaagataatcttcaaaagcccctgggaatctgaaagaagag  
aagcaggcccatttatatgggaaagaacaatagtatttcttatatagcccatttaagtgaaaacaatcttca  
aagcccacatcgcttagataagaaaacgaagctgagttatatacagctagagtcgaagtagtgatt**gGC**  
**TCGCATAAGCGTTGTGAC**GTTTTTAGAGCTAGAAATAGCAAGTTA  
AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT  
CGGTGCT**TTT**GGATCCCTGATGAGACC

**Supplementary Figure 4.** In-silico designed and *in vitro* synthesized sgRNA expression cassettes stacks. Each AtU6 promoter sequence is depicted in lowercase. Each sgRNA sequence is depicted in uppercase brown (*JAR1*), blue(*GLI*) and green(*EIN2*). Each crRNA sequence is depicted in bold font. The RNA polymerase III transcription start site “g” is depicted in lower case red, and the stop site (poly T tail) as upper case red. The 32 and 17 extra nucleotides at 5’ and 3’ of the synthetic DNA fragment are shown in back uppercase font.

## Supplementary Figure 5



**Supplementary Figure 5.** sgRNA secondary structure prediction. Illustration of in-silico prediction of *JARI*(A-C), *GLI*(D) and *EIN2*(E-F) sgRNA secondary structure by Mfold (Zuker et al. 2003). Nucleotides are colored coded as red and black for the 5' and 3' ends, respectively

## Citation list

- Alonso, J. M., Hirayama, T., Roman, G., Nourizadeh, S., & Ecker, J. R. (1999). EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. *Science*, 284(5423), 2148–2152.
- Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, 408(6814), 796–815.
- Barrangou, R., & Doudna, J. A. (2016). Applications of CRISPR technologies in research and beyond. *Nature Biotechnology*, 34(9), 933–941.
- Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N. J., & Nekrasov, V. (2015). Editing plant genomes with CRISPR/Cas9. *Current Opinion in Biotechnology*, 32, 76–84.
- Bogdanove, A. J., & Voytas, D. F. (2011). TAL effectors: customizable proteins for DNA targeting. *Science*, 333(6051), 1843–1846.
- Bowers, J. E., Chapman, B. A., Rong, J., & Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, 422(6930), 433–438.
- Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A. J., & Voytas, D. F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 186(2), 757–761.
- Clapper J.R., National Intelligence. (2016). Worldwide Threat Assessment of the US Intelligence Community. Senate Armed Services Committee, US.
- Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant Journal: For Cell and Molecular Biology*, 16(6), 735–743.
- Cody, W. B., Scholthof, H. B., & Mirkov, T. E. (2017). Multiplexed Gene Editing and Protein Overexpression Using a Tobacco mosaic virus Viral Vector. *Plant Physiology*, 175(1), 23–35.
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121), 819–823.
- Cornu, T. I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichtinger, M., Joung, J. K., & Cathomen, T. (2008). DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Molecular Therapy*, 16(2), 352–358.
- Costigan, S. E., Warnasooriya, S. N., Humphries, B. A., & Montgomery, B. L. (2011). Root-localized phytochrome chromophore synthesis is required for photoregulation of root elongation and impacts root sensitivity to jasmonic acid in Arabidopsis. *Plant Physiology*, 157(3), 1138–1150.
- Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., ... Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology*, 32(12), 1262–1267.
- Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., ... Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2), 184–191.
- Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.-L., ... Zhu, J.-K. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *PNAS*, 111(12), 4632–4637.
- Hahn, F., Mantegazza, O., Greiner, A., Hegemann, P., Eisenhut, M., & Weber, A. P. M. (2017). An Efficient Visual Screen for CRISPR/Cas9 Activity in Arabidopsis thaliana. *Frontiers in Plant Science*, 8, 39.

- 
- Herman, P. L., & Marks, M. D. (1989). Trichome Development in *Arabidopsis thaliana*. II. Isolation and Complementation of the GLABROUS1 Gene. *The Plant Cell*, 1(11), 1051–1055.
- Hood, E. E., Gelvin, S. B., Melchers, L. S., & Hoekema, A. (1993). New *Agrobacterium* helper plasmids for gene transfer to plants. *Transgenic Research*, 2(4), 208–218.
- Hsiao, T., Maures, T., Waite, K., Yang, J., Kelso, R., Holden, K., & Stoner, R. (2018). Inference of CRISPR Edits from Sanger Trace Data. *BioRxiv*.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), 816–821.
- Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P., & Siksnys, V. (2013). crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biology*, 10(5), 841–851.
- Khatodia, S., Bhatotia, K., Passricha, N., Khurana, S. M. P., & Tuteja, N. (2016). The CRISPR/Cas Genome-Editing Tool: Application in Improvement of Crops. *Frontiers in Plant Science*, 7, 506.
- Kim, J. M., Kim, D., Kim, S., & Kim, J.-S. (2014). Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nature Communications*, 5, 3157.
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 533(7603), 420–424.
- Koornneef, M., & Meinke, D. (2010). The development of *Arabidopsis* as a model plant. *The Plant Journal: For Cell and Molecular Biology*, 61(6), 909–921.
- Liang, G., Zhang, H., Lou, D., & Yu, D. (2016). Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Scientific Reports*, 6, 21451.
- Li, J., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., ... Sheen, J. (2013). Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nature Biotechnology*, 31(8), 688–691.
- Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., ... Liu, Y.-G. (2015). A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. *Molecular Plant*, 8(8), 1274–1284.
- Mahfouz, M. M., Piatek, A., & Stewart, C. N. (2014). Genome engineering via TALENs and CRISPR/Cas9 systems: challenges and perspectives. *Plant Biotechnology Journal*, 12(8), 1006–1014.
- Malzahn, A., Lowder, L., & Qi, Y. (2017). Plant genome editing with TALEN and CRISPR. *Cell & Bioscience*, 7, 21.
- Marks, D., & Feldmann, K. (1989). Trichome Development in *Arabidopsis thaliana*. *The Plant Cell*, 1, 1043–1050.
- Mashiko, D., Fujihara, Y., Satouh, Y., Miyata, H., Isotani, A., & Ikawa, M. (2013). Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Scientific Reports*, 3, 3355.
- Miki, D., Zhang, W., Zeng, W., Feng, Z., & Zhu, J.-K. (2018). CRISPR/Cas9-mediated gene targeting in *Arabidopsis* using sequential transformation. *Nature Communications*, 9(1), 1967.
- Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y.-L., Rupniewski, I., ... Rebar, E. J. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnology*, 25(7), 778–785.

- 
- Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., ... Rebar, E. J. (2011). A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology*, 29(2), 143–148.
- Minkenberg, B., Xie, K., & Yang, Y. (2017). Discovery of rice essential genes by characterizing a CRISPR-edited mutation of closely related rice MAP kinase genes. *The Plant Journal: For Cell and Molecular Biology*, 89(3), 636–648.
- Moscou, M. J., & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*, 326(5959), 1501.
- Murashige, T., & Skoog, F. (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum*, 15(3), 473–497.
- Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G., & Kamoun, S. (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature Biotechnology*, 31(8), 691–693.
- Osakabe, Y., Watanabe, T., Sugano, S. S., Ueta, R., Ishihara, R., Shinozaki, K., & Osakabe, K. (2016). Optimization of CRISPR/Cas9 genome editing to modify abiotic stress responses in plants. *Scientific Reports*, 6, 26685.
- Peng, C., Wang, H., Xu, X., Wang, X., Chen, X., Wei, W., ... Xu, J. (2018). High-throughput detection and screening of plants modified by gene editing using quantitative real-time polymerase chain reaction. *The Plant Journal: For Cell and Molecular Biology*, 95(3), 557–567.
- Peterson, B. A., Haak, D. C., Nishimura, M. T., Teixeira, P. J. P. L., James, S. R., Dangel, J. L., & Nimchuk, Z. L. (2016). Genome-Wide Assessment of Efficiency and Specificity in CRISPR/Cas9 Mediated Multiple Site Targeting in Arabidopsis. *Plos One*, 11(9), e0162169.
- Ramirez, C. L., Foley, J. E., Wright, D. A., Müller-Lerch, F., Rahman, S. H., Cornu, T. I., ... Joung, J. K. (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature Methods*, 5(5), 374–375.
- Schimpl, S., & Puchta, H. (2016). Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas. *Plant Methods*, 12, 8.
- Tálas, A., Kulcsár, P. I., Weinhardt, N., Borsy, A., Tóth, E., Szebényi, K., ... Welker, E. (2017). A convenient method to pre-screen candidate guide RNAs for CRISPR/Cas9 gene editing by NHEJ-mediated integration of a “self-cleaving” GFP-expression plasmid. *DNA Research*, 24(6), 609–621.
- Thomas, H. R., Percival, S. M., Yoder, B. K., & Parant, J. M. (2014). High-throughput genome editing and phenotyping facilitated by high resolution melting curve analysis. *Plos One*, 9(12), e114632.
- Van der Oost, J. (2013). Molecular biology. New tool for genome surgery. *Science*, 339(6121), 768–770.
- Vision, T. J., Brown, D. G., & Tanksley, S. D. (2000). The origins of genomic duplications in Arabidopsis. *Science*, 290(5499), 2114–2117.
- Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., & Qiu, J.-L. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology*, 32(9), 947–951.
- Wang, Z.-P., Xing, H.-L., Dong, L., Zhang, H.-Y., Han, C.-Y., Wang, X.-C., & Chen, Q.-J. (2015). Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biology*, 16, 144.
- Wolter, F., Klemm, J., & Puchta, H. (2018). Efficient in planta gene targeting in Arabidopsis using egg cell-specific expression of the Cas9 nuclease of

- 
- Staphylococcus aureus*. *The Plant Journal: For Cell and Molecular Biology*, 94(4), 735–746.
- Yan, W., Chen, D., & Kaufmann, K. (2016). Efficient multiplex mutagenesis by RNA-guided Cas9 and its use in the characterization of regulatory elements in the AGAMOUS gene. *Plant Methods*, 12, 23.
- Zhang, Z., Mao, Y., Ha, S., Liu, W., Botella, J. R., & Zhu, J.-K. (2016). A multiplex CRISPR/Cas9 platform for fast and efficient editing of multiple genes in *Arabidopsis*. *Plant Cell Reports*, 35(7), 1519–1533.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 3406–3415.