Chapter 1

CRISPR/Cas-Mediated *In Planta* Gene Targeting

Simon Schiml, Friedrich Fauser, and Holger Puchta

Abstract

The recent emergence of the CRISPR/Cas system has boosted the possibilities for precise genome engineering approaches throughout all kingdoms of life. The most common application for plants is targeted mutagenesis, whereby a Cas9-mediated DNA double-strand break (DSB) is repaired by mutagenic nonhomologous end joining (NHEJ). However, the site-specific alteration of a genomic sequence or integration of a transgene relies on the precise repair by homologous recombination (HR) using a suitable donor sequence: this poses a particular challenge in plants, as NHEJ is the preferred repair mechanism for DSBs in somatic tissue. Here, we describe our recently developed *in planta* gene targeting (ipGT) system, which works via the induction of DSBs by Cas9 to activate the target and the targeting vector at the same time, making it independent of high transformation efficiencies.

Key words Gene technology, Genome engineering, Double-strand break repair, Engineered nucleases, Cas9

1 Introduction

Modern methods for genome engineering in plants, but also in other eukaryotes, rely on the targeted induction of a DSB into the DNA. Thus, natural DSB repair mechanisms can be stimulated and exploited to achieve a desired outcome. Basically, DSBs in somatic plant tissues can be repaired via two distinctive pathways [1]. The major pathway is marked by NHEJ, involving processing of the DSB ends followed by a ligation reaction. Due to this processing, NHEJ generally incorporates small insertions or deletions into the genomic sequence, thus potentially generating a frameshift in an open reading frame. This approach is therefore referred to as targeted mutagenesis. The second pathway is homologous recombination, where a homologous donor sequence can be utilized as repair template for an error-free repair [2]. If an ectopic sequence is offered, usually termed donor sequence, that is, homologous or a transgene flanked by homologies, the respective sequence can be inserted into the repaired site, hence changing its information in a predefined manner.

Wolfgang Busch (ed.), *Plant Genomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1610,
DOI 10.1007/978-1-4939-7003-2_1,

This experimental approach is termed gene targeting (GT) and can be used to perform amino acid exchanges or to guide a transgene to a desired position within the genome [3].

Widespread use of both targeted mutagenesis and gene targeting has become possible through the development of programmable nucleases which enable the induction of a precise DSB at a desired position in the genome [4]. The most recent yet most versatile class of programmable nucleases is formed by the bacterial clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system. Originating from being a bacterial adaptive immune system, CRISPR/Cas was adapted as a programmable nuclease, which is composed of two components [5]. The small so-called sgRNA binds directly to its complementary sequence on the target DNA, next to the protospacer adjacent motif (PAM), usually "NGG." The endonuclease Cas9 then cleaves the DNA within the bound segment, 3 bp away from the PAM. Owing to this simple, yet highly efficient architecture together with its applicability to a vast range of organisms, RNA-guided Cas9 has rapidly become the most important tool for targeted genome engineering [6–9].

Achieving GT in plants is to date still challenging, as it relies on the rarely occurring repair via HR and therefore depends on highly efficient DSB induction together with the constant availability of a donor. Nevertheless, gene targeting efficiencies in plants are only in the percent range. As this requires a bigger number of transformation events, GT is hardly achievable for plant species with low transformation efficiencies. Within recent years, we were able to establish the efficient ipGT system in the model plant *Arabidopsis thaliana* that is independent of transformation efficiencies or the use of a mutant background to enhance GT [10, 11]. In this approach, a T-DNA is stably integrated that contains an expression cassette for a nuclease as well as the donor sequence, consisting of the desired transgene flanked by homologies to the desired target locus. Upon expression of the nuclease inside the plant cells, the donor sequence is excised via two DSBs and a third DSB is induced in the target locus, thus activating it for HR (Fig. 1). By using the flanking homologies, the donor sequence can then be integrated into the repaired locus. Since in plants the germ line is developed out of somatic tissue, the GT event can become heritable, thus creating offspring that stably carries the GT event.

Here, we describe the procedure to perform the *Streptococcus pyogenes* Cas9-mediated ipGT approach in *A. thaliana*, enabling stably heritable targeting of a transgene to a desired locus, for example, to specifically tag an endogenous gene, or the predefined alteration of a genomic sequence, to achieve amino acid substitutions.

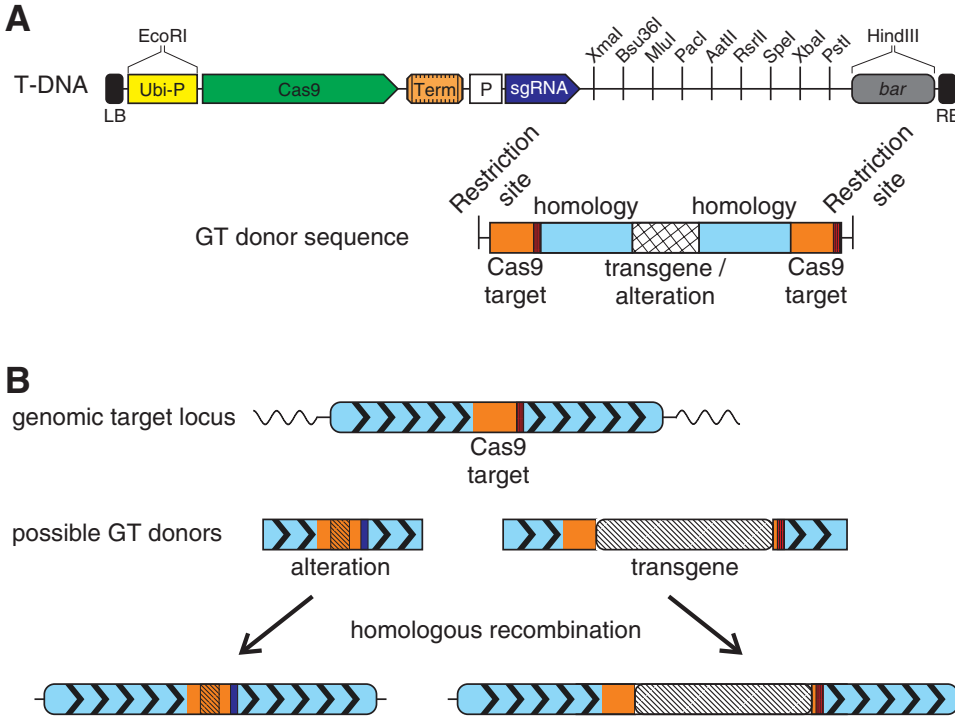

Fig. 1 Outline of the ipGT approach. (**a**) T-DNA construct. The Cas9 sequence is controlled by a constitutive ubiquitin promoter, which is exchangeable with EcoRI. The GT donor sequence can be cloned into the depicted restriction sites in the MCS. The general structure of the donor consists of the desired sequence flanked by homologies to the target locus, the Cas9 target sites, and the required restriction sites. The PPT resistance (*bar*) can be exchanged using HindIII. (**b**) Upon expression of the nuclease, a DSB is induced in the target locus and two DSBs in the T-DNA release the donor sequence. The latter should contain specific sequence alterations – silent mutation in case single amino acid changes will be targeted in an ORF to avoid Cas9 cutting. Alternatively, the cutting site may be replaced by foreign sequence flanked by the required homologies to ensure integration into the genomic site

2 Materials

2.1 Plasmids

All plasmids are available directly from the authors or through the Arabidopsis Biological Resource Center (ABRC). Full sequence information is deposited at www.botanik.kit.edu/crispr.

1. pDe-CAS9 (ABRC CD3-1928)

   Binary vector for stable transformation into plants via *A. tumefaciens*. Contains a constitutive Cas9 expression system, along with a Gateway destination sequence with a *ccdB* gene to take up the sgRNA expression sequence. Confers plant resistance to phosphinothricin (PPT). This plasmid also serves as the vector for the desired GT donor sequence.

2. pEn-Chimera (ABRC CD3-1931)

   Ampicillin-resistant Gateway entry vector, containing the sgRNA expression cassette, flanked by attL1/2 sites.

2.2 Organisms

1. *Escherichia coli*, standard cloning strain for all cloning steps; *ccdB*-resistant strain for propagation of pDe-CAS9 (e.g., DB3.1 [12], *see* **Note 1**).
2. *A. thaliana* plants, either wild-type (e.g., Col-0) or any transformable mutant.
3. *Agrobacterium tumefaciens*, any conventional transformation strain, e.g., GV3101 [13].

2.3 Reagents

1. Restriction enzyme BbsI, additional restriction enzymes as required (*see* Subheading 3, Fig. 1 and **Note 2**).
2. T4 Ligase for conventional cloning steps.
3. Proofreading DNA polymerase for the generation of the donor sequence.
4. A robust Taq polymerase for *E. coli* colony PCRs and for screening of putative GT plants (*see* **Note 3**).
5. Gateway LR Clonase II (ThermoFisher Scientific, supplied with proteinase K).
6. LB medium (for *E. coli)*: 10 g/L peptone, 5 g/L yeast extract, 10 g/L NaCl. Solid media: 7.5 g/L agar.
7. YEB medium (for *A. tumefaciens)*: 5 g/L beef extract and 5 g/L peptone. 1 g/L yeast extract, 5 g/L sucrose, 439 mg/L $MgSO_4$, and 7.5 g/L agar for solid media.
8. Germination medium (GM): 4.9 g/L Murashige & Skoog, 10 g/L sucrose, pH 5.7, and 8 g/L agar. For selection media, add ampicillin (100 mg/L), spectinomycin (100 mg/L), or PPT (6 mg/L).
9. TE buffer: 10 mM Tris–HCl and 1 mM EDTA at pH 8.

3 Methods

3.1 Experimental Design: Site-Specific Integration of a Foreign Sequence

1. After selecting the desired target locus for your transgene sequence, identify potential Cas9 target sites (*see* **Note 4**).
2. Add at least 400 bp upstream and downstream from the Cas9 cutting site to the desired transgene, therefore defining the exact position and orientation (*see* **Notes 5** and **6**).
3. For correct excision of the donor sequence, add the Cas9 target site including the PAM to the proximal and distal end of the donor sequence (*see* **Note 7**).

3.2 Specific Sequence Alteration

1. Select a Cas9 target site close to the site to be changed.
2. Choose the flanking homologies as described above; a total size of 0.8–1.5 kb is recommended (0.4–0.8 kb on either side).

3. Additionally to the desired base exchange(s), it is crucial to introduce silent mutations to abolish any Cas9 activity in your donor sequence (*see* **Note 8**). This should also enable the detection of the GT event by PCR, if a primer is used that can bind to the altered sequence but not to the genomic site.
4. As described above (*see* Subheading 3.1, **step 3**), flank the donor sequence by the correct Cas9 target sites (including the PAM) to enable its excision from the T-DNA (*see* **Note 7**).
5. Finally, add restriction sites to the donor sequence to enable cloning into the T-DNA construct. Fig. 1 depicts the T-DNA construct and the available restriction sites in the MCS (*see* **Note 2**).
6. The construct itself can be assembled by overlap extension PCR, Gibson assembly (generate fragments with a potent proofreading polymerase), or via gene synthesis (*see* **Note 9**).

3.3 Cloning of the T-DNA Construct for ipGT

1. Order oligonucleotides for your Cas9 target sequence. For an NGG PAM, the fw oligo should contain the 20 nt upstream of the PAM with ATTG added to the 5′ end. The second oligo should contain the reverse complement of the target sequence with AAAC added 5′.
2. Dilute and mix your oligos in ddH_2O to a final concentration of 2 pmol/μL for each oligo in a total volume of 50 μL. Incubate for 5 min at 95 °C and put at room temperature for an additional 10 min for annealing.
3. Digest pEn-Chimera with BbsI as recommended by the supplier for at least 2 h. Purify the reaction and dilute the final concentration to 5 ng/μL.
4. Perform a ligation reaction with 2 μL digested vector, 3 μL prepared oligos, 1 μL T4 ligase buffer, 1 μL T4 ligase, and 3 μL ddH_2O, and incubate as recommended by the vendor. Transform into *E. coli* and select for colonies on ampicillin-containing LB plates.
5. Set up a colony PCR as recommended by the vendor of the Taq polymerase to identify positive colonies, using your fw oligo and M13 rev as primers, which generate a band at approx. 370 bp.
6. Purify plasmids from a small number (1–4) of correct clones. Validate by sequencing with primer SS42.
7. Transfer the correct sgRNA expression sequence to pDe-CAS9 by Gateway cloning. Set up a reaction with 100 ng entry vector, 300 ng destination vector (pDe-CAS9), 4 μL TE buffer, and 1 μL LR Clonase II in a total volume of 10 μL, and incubate for 2–3 h at room temperature.
8. Stop the reaction by adding 1 μL Proteinase K for 10 min at 37 °C (crucial step).

9. Transform the whole reaction into *E. coli* (*see* **Note 1**), and select on spectinomycin-containing LB plates.
10. Check for correct clones by colony PCR with primers SS42/SS43, producing a 1 kb band.
11. Isolate correct plasmids. A control restriction digestion is possible with AflII and NheI, producing bands at approx. 5.9, 5, and 3.8 kb.
12. Add your GT donor sequence by conventional cloning. If one restriction enzyme is used, make sure to dephosphorylate the vector backbone prior to ligation (*see* **Note 10**).
13. Transform into *E. coli* and grow on spectinomycin-containing LB plates. Identify and verify correct clones by a suitable colony PCR, restriction digestion, and sequencing.
14. Transform your final plasmid into *A. tumefaciens* and further into *A. thaliana* (e.g., by floral dipping [14]).

3.4 Identification of GT Plants

1. Select primary transformant plants by sowing seeds from transformed plants on germination medium containing PPT.
2. Pick at least 40 plants for further cultivation in soil (*see* **Note 11**).
3. Qualitative control of the nuclease activity at this stage is optional. This can be done by T7 endonuclease assay, restriction digestion assay, or high-resolution melting analysis.
4. Isolate DNA from single leaves of the primary transformants with a fast extraction protocol [15]. Set up a PCR as depicted in Fig. 2, using one primer outside of the homologous region and a corresponding primer within the transgene or specific for the defined sequence alteration. That way, only the correct GT event should produce a band (*see* **Note 12**).
5. Cultivate plants in soil for progeny seeds.
6. Test the T2 lines for single-locus integration of the T-DNA by sowing a small amount (~50–100) of progeny seeds from each T1 plant on PPT selection medium, and verify a correct Mendelian segregation pattern after 10–14 d (75% germination, representing homozygous and heterozygous plants with respect to the T-DNA).
7. For ten or more correctly segregating T2 lines, sow at least 100 seeds without applying a selection marker (*see* **Note 13**). Isolate DNA from each plant and check again for the GT event with PCR. Cultivate candidate plants in soil individually to obtain progeny seeds.
8. In T3, check for absence of the T-DNA by sowing a small amount of seeds on selection medium. Confirm the presence of the GT event to assure stable inheritance (*see* **Note 14**).

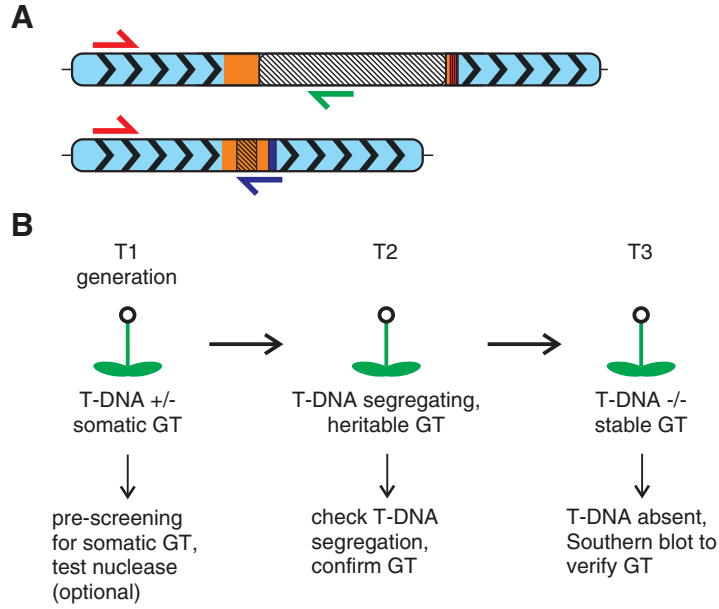

Fig. 2 Identification of GT plants. (**a**) PCR-based identification. Primers should be placed outside the homologous donor flanks and inside the transgene or the altered sequence segment (*see* **Note 12**). (**b**) Simplified procedure for stable GT plants. In T1, confirm functionality of the GT approach by PCR-based prescreening as depicted above. Correct segregation of the T-DNA has to be confirmed in T2, along with further PCRs for the GT event. In absence of the T-DNA in T3 (plants not germinating on selection marker), any positive GT PCR indicates stable inheritance. The procedures in T2/T3 may be repeated in later generations until the stable event is obtained

9. To assure the correct, two-sided GT (i.e., both flanks were correctly integrated without the occurrence of NHEJ) and physical linkage, verification with a Southern blot is highly recommended, exploiting the introduction (or destruction) of novel restriction sites along with the GT event.

4 Notes

1. All cloning steps require a conventional *E. coli* strain, e.g., DH5α. The *ccdB*-resistant strain is only used to amplify pDe-CAS9 in advance to any cloning steps described here.
2. Choose restriction enzymes for the cloning according to Fig. 1, avoiding restriction sites that are already present in the donor construct design.
3. We recommend DreamTaq polymerase (ThermoFisher Scientific) for colony and plant screening PCRs.

4. Target site selection can be done manually simply by looking for an "NGG" as PAM; take the 20 bp upstream by avoiding to have five or more consecutive T in it. Software-assisted selection is possible with CRISPR-P or CCTop [16, 17].
5. The actual size can vary greatly depending on your experiment. Generally, longer homologies (0.8–1 kb) should improve HR frequencies. However, if the homology contains parts of a promoter region, one has to be aware of potential expression from the T-DNA itself.
6. Since both sides of the cutting site are included in the homologies, assure that the sgRNA cannot bind there, as it would lead to the degradation of your donor sequence.
7. Using a vector set which is capable of expressing more than one sgRNA [10], it is also possible to have different Cas9 target sites within the target locus and to release the donor sequence.
8. The most effective way is to alter the PAM and the seed region of the target site [18]. Note that for *S. pyogenes* Cas9, NAG is also reported to be a functional PAM sequence [19]. Furthermore, if you plan to confirm your GT experiment with a Southern blot, consider also degrading a restriction site in the donor sequence that is present in the genomic sequence or vice versa. CRISPR-P illustrates restriction sites in a chosen target site [17].
9. Having the donor sequence synthesized is the easiest, yet most expensive method. However, overlap PCR or Gibson assembly may be challenging due to reoccurring sequence elements flanking the construct.
10. Since the orientation of the donor construct within the T-DNA is arbitrary, cloning with one restriction enzyme is generally sufficient. With two enzymes, however, no dephosphorylation is required.
11. All numbers given in this paragraph are based on experience and may be scaled up to actually detect the GT event.
12. Assure the functionality of the PCR: primers may be tested individually with genomic DNA and the T-DNA construct with suitable corresponding primers. Also, consider performing a similar PCR for the downstream flank of your GT event, to exclude false positives such as one-sided GT events. Note that the GT may be a rare event, so a high number of PCR cycles (>40) is required.
13. Preferably choose lines that were already tested positive for the GT in T1, as these indicate a functional integration site of the T-DNA.
14. The described processes can be repeated in following generations if necessary.

Acknowledgment

We thank Amy Whitbread for the critical reading of the manuscript. Our work on Cas9-mediated genome engineering and GT was funded by the European Research Council (Advanced Grant "COMREC") as well as the Federal Ministry of Education and Research (PLANT 2030, Pflanzenbiotechnologie fur die Zukunft – TAMOCRO, Grant 0315948).

References

1. Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot 56(409):1–14. doi:10.1093/jxb/eri025
2. Puchta H, Dujon B, Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc Natl Acad Sci U S A 93(10):5055–5060
3. Steinert J, Schiml S, Puchta H (2016) Homology-based double-strand break-induced genome engineering in plants. Plant Cell Rep 35(7):1429–1438. doi:10.1007/s00299–016–1981-3
4. Puchta H, Fauser F (2014) Synthetic nucleases for genome engineering in plants: prospects for a bright future. Plant J 78(5):727–741. doi:10.1111/tpj.12338
5. Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096):816–821. doi:10.1126/science.1225829
6. Schiml S, Puchta H (2016) Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas. Plant Methods 12:8. doi:10.1186/s13007–016–0103-0
7. Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157(6):1262–1278. doi:10.1016/j.cell.2014.05.010
8. Mali P, Yang L, Esvelt KM et al (2013) RNA-guided human genome engineering via Cas9. Science 339(6121):823–826. doi:10.1126/science.1232033
9. Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823. doi:10.1126/science.1231143
10. Schiml S, Fauser F, Puchta H (2014) The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. Plant J 80(6):1139–1150. doi:10.1111/tpj.12704
11. Fauser F, Roth N, Pacher M et al (2012) In planta gene targeting. Proc Natl Acad Sci U S A 109(19):7535–7540. doi:10.1073/pnas.1202191109
12. Bernard P, Couturier M (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. J Mol Biol 226(3):735–745
13. Koncz C, Kreuzaler F, Kalman Z et al (1984) A simple method to transfer, integrate and study expression of foreign genes, such as chicken ovalbumin and alpha-actin in plant tumors. EMBO J 3(5):1029–1037
14. Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J 16(6):735–743
15. Zhang J, Stewart JM (2000) Economical and rapid method for extracting cotton genomic DNA. J Cotton Sci 4:193–201
16. Stemmer M, Thumberger T, Del Sol Keyer M et al (2015) CCTop: an intuitive, flexible and reliable crispr/cas9 target prediction tool. PLoS One 10(4):e0124633. doi:10.1371/journal.pone.0124633
17. Lei Y, Lu L, Liu H et al (2014) CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. Mol Plant 7(9):1494–1496. doi:10.1093/mp/ssu044
18. Jiang W, Bikard D, Cox D et al (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31(3):233–239. doi:10.1038/nbt.2508
19. Hsu PD, Scott DA, Weinstein JA et al (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31(9):827–832. doi:10.1038/nbt.2647