Strategies for the improvement of genome editing in Arabidopsis thaliana

Strunks, G.D.

Chapter 6

Summarizing discussion

Gary Strunks

DNA double-strand breaks (DSBs) are harmful lesions that can cause genomic instability and mutations if not repaired. Over the course of evolution, cells have evolved complex and highly conserved mechanisms to detect and repair DSBs in order to maintain genome integrity. DSBs can be repaired by two main pathways: nonhomologous end-joining (NHEJ) and homologous recombination (HR). NHEJ-mediated DSB repair can occasionally lead to mutations (insertions or deletions) at the break site, while HR ensures accurate DSB repair by utilizing homologous DNA sequences. In higher eukaryotes, the majority of DSBs are repaired through NHEJ in somatic cells, as this mechanism acts throughout the cell cycle. HR, on the other hand, is dependent on the availability of homologous DNA sequences at specific points during the cell cycle (1).

Understanding the NHEJ and HR pathways has been of great importance for the application of targeted mutagenesis and gene targeting in molecular biology and biotechnology. In **chapter 1**, the current understanding of DSB repair via NHEJ and HR is described in detail, as well as the exploitation of these pathways for plant genome editing. The advent of artificial sequence-specific nucleases (SSNs) such as zinc-finger nucleases (ZFNs), TALENs and CRISPR/Cas systems have substantially increased the feasibility of NHEJ-mediated targeted mutagenesis, which can now routinely be used as a plant biotechnology tool. However, the development and applicability of GT as a plant biotechnology tool has been much more challenging due to two roadblocks: the low frequency of HR in somatic cells and low transformation efficiencies in plants that impact GT frequency. Although several approaches led to increased GT efficiencies, including the use of SSNs and manipulation of the NHEJ and HR machinery, GT frequencies generally remained below the 1% range (2, 3). Therefore, new methods have to be developed in order to make GT an effective and feasible plant biotechnology tool.

To attempt to overcome the two roadblocks of low HR frequency in somatic cells and low transformation efficiencies in plants, we developed an *in planta* GT system aimed at performing GT in meiocytes, cells that by nature are prone to perform HR. The first step in developing this system was the construction and selection of efficient SSNs. In **chapter 2**, we tested the targeted mutagenesis efficiencies of two classes of SSNs: TALENs and CRISPR/Cas9, designed to target the Arabidopsis *CRU3* and *ADH1* genes. The CRISPR/Cas9 system performed noticeably better at both genes and performed best when a slightly different protospacer sequence was used ending with a GG sequence. These results were in accordance with other studies performed in *C. elegans* and human cells (4, 5), and may point towards a

more optimal protospacer design in general. The CRISPR/Cas9 constructs designed for the *CRU3* and *ADH1* genes were satisfactorily efficient for targeted mutagenesis. The Cas9 version used in all of our experiments was derived from *Streptococcus pyogenes* (SpCas9), but it has been reported that targeted mutagenesis efficiencies can get even higher when using Cas9 variants derived from different bacterial species. For example, average targeted mutagenesis frequencies at the Arabidopsis *ADH1* gene as high as 76% were obtained when using Cas9 derived from *Staphylococcus aureus* (SaCas9) (6), showing that SaCas9 can be an attractive alternative to SpCas9 for targeted mutagenesis and GT.

Elucidating the interplay between NHEJ and HR pathways in plants can give more insight in how these pathways can be further exploited for genome engineering. In **chapter 3**, we performed CRISPR/Cas9-mediated targeted mutagenesis at the Arabidopsis *CRU3* and *PPO* genes and analysed the repair outcomes in plants deficient in classical NHEJ (C-NHEJ) factor KU80 and alternative end-joining (A-EJ) factors PARP1 and PARP2. Absence of KU80 resulted in larger deletions, confirming its role in preventing DSB end resection. Absence of PARP1/2, however, did not seem to have a significant effect on repair outcomes, which suggested the activity of a KU- and PARP-independent a-EJ mechanism. We suggested that this activity could possibly involve Polymerase θ (POLQ), because POLQ is the factor essential for random T-DNA integration at DNA breaks in *Arabidopsis thaliana* (7). Indeed, POLQ was recently identified as the A-EJ key factor by Mara *et al*. 2019, in the moss *Physcomitrella patens* (8). In the absence of POLQ, CRISPR-induced mutation rate and A-EJ-based DSB repair was strongly reduced. Absence of POLQ also led to a dramatic increase in GT frequency by both reducing random integration of the repair template and increasing HR frequency (8). However, compared with flowering plants, GT rates in *P. patens* are high and it remains to be determined what effect disruption of POLQ has on GT efficiency in flowering plant such as Arabidopsis.

As GT frequency in higher plants is low, high transformation frequencies for the repair template are necessary for GT experiments. However, many crops cannot be transformed easily. To overcome this limitation, the group of Holger Puchta developed the *in planta* GT strategy (9). This strategy relies on the simultaneous induction of a DSB at a target gene and DSB induction at sequences flanking a stably integrated repair template, resulting in its release from the genome to be utilized for HR-mediated repair. GT events that occur in the germline will be transmitted to the next generation. With GT frequencies of about 0.14%, there was still room for further improvements (10). In **chapter 4** we describe an attempt to improve GT efficiency with the *in planta* strategy by directing GT to meiocytes, which should have an intrinsically

higher HR rate and could subsequently lead to more heritable GT events. We developed our in planta GT system for the seed-specific Arabidopsis *CRU3* gene based on the GT system developed by Shaked *et al*. (11), and used the efficient Cas9-CRU construct described in chapter 2 for DSB induction. A GFP sequence in the GT repair template was used to allow for detection of GT events as GFP-fluorescent seeds. To direct GT to meiocytes, Cas9 was expressed under the *SPO11-1*, *RPS5a* and *SPL* promoters, which are active during meiosis. Plants harbouring a CRU-GFP repair template were crossed with plants harbouring meiosis/germline-specific Cas9 expression cassettes, and GT events that arise during the development of F1 progeny should be visible in F2 seeds as inherited events.

Surprisingly, GFP-fluorescence was already detected in F1 seeds derived from all crossings. Molecular analysis did not reveal signatures of GT events at *CRU3* in plants derived form these seeds, and fluorescence was absent in F2 seeds derived from plants germinated from GFP-fluorescent F1 seeds. One possible explanation might be that the observed GFP signal was the result of a transiently expressed CRU-GFP fusion protein, created by one sided elongation of the CRU-GFP repair template, thereby obtaining a promoter sequence. Such an RT will be lost in the next generation when it is not integrated in the genome. It remains to be determined if this is indeed the case.

GFP-fluorescence was absent in F2 seeds, but true GT (TGT) events were detected in F2 plants with PCR analysis. All GT events were detected in plants in which the repair template could be excised from the genome, suggesting that repair template excision could benefit GT, which has also been shown in other studies (3, 9, 10). However, it is not a prerequisite (12). GT events that were detected with PCR analysis could not be confirmed with Southern blot analysis. This might indicate that the detected GT events were somatic events that occurred in a small number of somatic cells, and were therefore only detectable with PCR and not with Southern blot. This has been observed by others (13), and is also in line with the observation that most GT events occurred in F2 plants that constitutively expressed Cas9 and were not transmitted to the germline, which is supported by the lack of heritable mutations using pUbi-Cas9 and absence of GFP fluorescence in F3 seeds.

Heritable mutations could not be detected when using the *SPL* promoter for targeted mutagenesis, which is in contrast with a study by Mao *et al*. 2016 (14). However, heritable mutations were detected using the egg cell 1.2 (*EC1.2*) promoter, although at lower levels than described by others (14, 15). It thus seems that this promoter is a suitable candidate for germline-specific Cas9 expression for *in planta* GT, but we were not yet able to test this

promoter for *in planta* GT. Recently, however, the *EC1.1* promoter fused with the *EC1.2* enhancer was used to drive SaCas9 expression for *in planta* GT at the Arabidopsis *ALS* gene, which resulted in a substantial increase in GT frequency up to 6% (3). This promoter combined with an efficient Cas9 thus seems to be very suitable for gene targeting, and it may be worthwhile to test this approach in our *CRU3 in planta* GT system. That being said, the seed-specificity of *CRU3* might hamper the occurrence of GT in other plant tissue due to a condensed chromatin state.

Besides *in planta* GT, other ways to enhance GT efficiency were explored. **Chapter 5** describes attempts to identify plants with higher GT frequency using the genome interrogation method. Previously, genome interrogation was successfully applied in *Arabidopsis* to uncover novel phenotypes, including increased somatic HR, high salinity tolerance and enhanced growth characteristics (16–18). We used the 3F-EAR genome interrogation library instead of the library based on 3F-VP16 fusions, as plants transformed with 3F-EAR fusions previously showed larger phenotypic variation compared to plants transformed with 3F-VP16 fusions (18). Initial screening of plants transformed with the 3F-EAR library, using the *PPO* GT system, yielded true GT events, and 3F-EAR sequences were retrieved from these plants. Although four out of seven retrieved 3F sequences contained mutations, we decided to continue with the analysis of GT frequencies in reconstituted 3F-EAR lines, as these 3F-EAR constructs might still be able to have an effect on GT.

When screening for GT at *PPO* in reconstituted 3F-EAR plants we detected an increase in GT frequency in plants harbouring the 3F-EAR4 construct. However, we could not detect a significant GT enhancement in reconstituted 3F-EAR4 plants for the *CRU3* and *ADH1* genes. Based on these results, we cannot rule out that 3F-EAR4 could indeed trigger GT enhancement, but in order to find statistically significant numbers of GT events, this experiment should be performed on a larger scale. Although genome interrogation has proven to be a powerful method to uncover novel phenotypes, using this method in combination with GT experiments on a larger scale would be time consuming and logistically challenging. In addition, the ZF-VP16 collection could be screened for GT mutants.

In conclusion, this thesis describes the development of novel tools for the improvement of plant genome editing, in particular gene targeting. Although we could not conclusively show improvement in GT efficiency, the tools developed could be further improved in order to be able to induce higher GT frequencies in future experiments. For example, the *S. aureus* Cas9 ortholog (SaCas9) has been shown to be more efficient than the widely used SpCas9 for both

targeted mutagenesis and GT (2, 3). Germline-specific expression of Cas9 with an efficient promoter turns out to be a crucial step in generating both heritable targeted mutations and GT events. Germline-specific SaCas9 expression by the *EC1.1* promoter fused to the *EC1.2* enhancer resulted in feasible GT frequencies at the Arabidopsis *ALS* gene (3). SaCas9 expression with the *EC1.1* promoter or other efficient germline-specific promoters may also help to increase *in planta* GT efficiencies at other genes, including *CRU3*. Beside optimizing tools for DSB induction, inactivation or repression of POLQ in higher plants might help to enhance GT frequencies even further, preferably via transient methods (7, 8). Combining such a method with efficient germline-specific Cas9 expression and DSB induction could eventually lead to GT frequencies that facilitate to apply routinely gene modification and gene replacement in crop plants.

References

1. Mao Z, et al. (2008) DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle* 7(18):2902–2906.
2. Steinert J, Schiml S, Puchta H (2016) Homology-based double-strand break-induced genome engineering in plants. *Plant Cell Rep* 35(7):1–10.
3. 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. *Plant J* 94(4):735–746.
4. Farboud B, Meyer BJ (2015) Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. *Genetics* 199:959–971.
5. Malina A, et al. (2015) PAM multiplicity marks genomic target sites as inhibitory to CRISPR-Cas9 editing. *Nat Commun* 6:10124.
6. Steinert J, Schiml S, Fauser F, Puchta H (2015) Highly efficient heritable plant genome engineering using Cas9 orthologues from Streptococcus thermophilus and Staphylococcus aureus. *Plant J* 84(6):1295–1305.
7. van Kregten M, et al. (2016) T-DNA integration in plants results from Polymerase Theta-mediated DNA repair. *Nat Plants* 2:1–6.
8. Mara K, et al. (2019) POLQ plays a key role in the repair of CRISPR / Cas9-induced double-stranded breaks in the moss Physcomitrella patens. *New Phytol* 222:1380–1391.
9. Fauser F, et al. (2012) In planta gene targeting. *Proc Natl Acad Sci U S A* 109(19):7535–40.
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* 80:1139–1150.

11. Shaked H, Melamed-Bessudo C, Levy AA (2005) High-frequency gene targeting in Arabidopsis plants expressing the yeast RAD54 gene. *Proc Natl Acad Sci U S A* 102(34):12265–12269.
12. de Pater S, Klemann BJPM, Hooykaas PJJ (2018) True gene-targeting events by CRISPR/Cas-induced DSB repair of the PPO locus with an ectopically integrated repair template. *Sci Rep* 8(1):3338.
13. Feng Z, et al. (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc Natl Acad Sci U S A* 111(12):4632–7.
14. Mao Y, et al. (2016) Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in Arabidopsis. *Plant Biotechnol J*:1–14.
15. Wang ZP, et al. (2015) Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biol* 16(1):1–12.
16. Lindhout BI, Pinas JE, Hooykaas PJJ, Van Der Zaal BJ (2006) Employing libraries of zinc finger artificial transcription factors to screen for homologous recombination mutants in Arabidopsis. *Plant J* 48(3):475–483.
17. van Tol N, Pinas J, Schat H, Hooykaas PJJ, van der Zaal BJ (2016) Genome interrogation for novel salinity tolerant Arabidopsis mutants. *Plant Cell Environ*. doi:10.1111/pce.12805.
18. Tol N Van, et al. (2017) Enhancement of Arabidopsis growth characteristics using genome interrogation with artificial transcription factors. *PLoS One* 12(3):e0174236.