Improving targeted gene editing in Arabidopsis thaliana

Kamoen, L.

Summary

In the realm of plant biotechnology, understanding the mechanisms of DNA repair and genetic modification is crucial for advancing crop improvement. Double-strand breaks (DSBs) in DNA are critical lesions that can be repaired through various pathways, including homologous recombination (HR), classical nonhomologous end joining (cNHEJ) and polymerase theta-mediated end joining (TMEJ). cNHEJ is a rapid repair process that directly ligates the broken DNA ends. This process typically results in error-free repair, although it can occasionally introduce small insertions or deletions. TMEJ, on the other hand, uses microhomologous sequences to align the DNA ends before joining, which can lead to larger deletions or complex rearrangements.

CRISPR/Cas technology has revolutionized plant biotechnology by providing a precise method for inducing DSBs at specific genomic locations. Several systems exist that utilize different Cas enzymes, such as Cas9 and Cas12a, each with its own specificities. The targeted breaks created by these enzymes can then be repaired by the plant's endogenous mechanisms, leading to targeted mutations or gene insertions.

Agrobacterium tumefaciens, a bacterium that naturally transfers DNA to plant cells, is widely used in plant biotechnology for genetic transformation. The bacterium carries a plasmid containing a region known as transfer DNA (T-DNA). This T-DNA can be transported into plant cells, where it integrates into the plant genome at a DSB site, facilitated by the plant's own repair machinery. T-DNA often integrates at random positions in the genome, frequently forming complex multimeric configurations, and is often accompanied by translocations. Achieving targeted T-DNA integration at a predefined locus would eliminate these positional variations, offering significant advantages for crop engineering. Early studies demonstrated that DSB induction could mediate targeted T-DNA integration. The introduction of the highly efficient CRISPR/Cas9 system alongside T-DNA transformation represents a promising step forward for the clean and precise integration of exogenous DNA.

In this thesis, we investigate the mechanisms of CRISPR/Cas-induced DSB repair in the model plant *Arabidopsis thaliana* (thale cress) and how these repair processes relate to T-DNA integration at the break sites. Additionally, we aim to develop new methods for targeted editing by exploring these mechanisms in detail. The combination of *Agrobacterium*-mediated transformation and CRISPR/Cas technology offers numerous opportunities for precise genetic modifications, but it remains poorly characterized. This study seeks to fill that gap by providing a comprehensive analysis of these processes.

Chapter 2 explores the intricate mechanisms of DNA repair and T-DNA capture at CRISPR/Cas9-induced DSBs. We discovered that two primary pathways, cNHEJ and TMEJ, are responsible for repairing these breaks, each leading to distinct mutational outcomes. Using advanced

sequencing techniques, we found that T-DNA capture at DSB sites is a frequent event. TMEJ plays a crucial role in attaching the 3' end of T-DNA to the break, while both TMEJ and cNHEJ can attach the 5' end. This dual involvement highlights the complexity and redundancy of the plant's DNA repair mechanisms. We also observed that T-DNA shards are often found within genomic DSB repair sites, indicating frequent temporary interactions during TMEJ. The findings suggest that CRISPR-induced DSBs can serve as efficient entry points for T-DNA integration, which has significant implications for genome engineering.

While Chapter 2 focuses on Cas9, **Chapter 3** provides a detailed comparison of the mutagenesis profiles induced by two CRISPR nucleases, Cas9 and Cas12a. Both nucleases are shown to induce DSBs that are predominantly repaired by cNHEJ and TMEJ, with cNHEJ accounting for 70% of the mutations and TMEJ for 30%. The study reveals that the configuration of DSB ends (blunt for Cas9 and staggered for Cas12a) does not significantly influence the choice of repair pathway. The major difference between the nucleases is the frequent occurrence of 1 bp insertions, which are shown to result from polymerase lambda (Polλ) action on Cas9-induced DSBs. Interestingly, both nucleases facilitate the integration of T-DNA at DSB sites. Long-read sequencing highlighted a higher degree of DNA loss during TMEJ repair. The findings suggest that both Cas9 and Cas12a are equally effective for genome engineering, offering flexibility in nuclease choice based on the availability of compatible PAM sequences. This research enhances our understanding of CRISPR-based genome editing in plants and supports the development of more precise genetic modification strategies for crop improvement.

CRISPR-based mutagenesis often leads to a diverse range of mutational outcomes. To investigate the potential for more predictable mutagenesis, **Chapter 4** focuses on the repair of DSBs flanked by short repeats. This process, referred to as extended microhomology-mediated end joining (eMMEJ), favors one specific outcome: the precise joining of the homologous flanks, resulting in deletion of one of the homologous regions and the sequence between them. We found that eMMEJ, prevalent in wild-type plants, was absent in polymerase theta (Polθ) deficient plants. To study the way Polθ is implicated in eMMEJ, we reintroduced the Polθ-encoding *TEB* gene into the *teb* mutant via *Agrobacterium* mediated transformation, fully restoring eMMEJ activity. Using this setup we investigated the two pivotal domains of Polθ: the N-terminal helicase domain and the C-terminal polymerase domain. While polymerase activity is required in standard sequence contexts, DSBs containing eMH rely solely on the helicase activity of Polθ, rendering the polymerase domain obsolete. This knowledge provides new possibilities for directing targeted mutagenesis towards a single predictable outcome.

Following a detailed analysis of CRISPR-mediated DSB repair and targeted T-DNA integration, **Chapter 5** aims to thoroughly understand this mechanism, considering all possible outcomes, including both the desired effects and any unintended consequences. We characterized CRISPR-induced DSBs at multiple genomic locations, as well as on the T-DNA itself, and found that the repair processes at the T-DNA are remarkably similar to those in the genome, showing consistent repair footprints. This facilitates the study of any desired DSB target at will, potentially paving the

way for multiplex strategies. Additionally, our research demonstrates that cleaved T-DNA can be efficiently incorporated into the genome, with a cleavage efficiency nearing 100%. This allows for the integration of T-DNA devoid of any unwanted sequences, such as border repeats or selective markers. Finally, we discovered that inducing multiple simultaneous DSBs often results in the fusion of break ends, leading to genomic aberrations. While this flexibility can be advantageous for targeted genome editing, it is crucial to monitor for potential side effects.

Lastly, **Chapter 6** combines the insights gained from the previous chapters to develop a novel method for targeted T-DNA integration and genome editing. Efficient techniques for precise and targeted gene modification in plant genomes hold immense potential for accelerating plant breeding efforts. Gene targeting (GT) is a widely used approach that achieves desired genetic modifications by introducing a donor DNA molecule to precisely alter a target locus. However, this method relies on homologous recombination (HR), resulting in a low frequency of successful gene targeting events. To overcome this challenge, we have developed an alternative method based on the more efficient eMMEJ process. By incorporating an extended microhomologous sequence (eMH) on the T-DNA, which matches the genomic target locus, we achieved high rates of targeted eMMEJ-mediated T-DNA capture. Similar to DSB repair, eMMEJ was the predominant mode of action, significantly outnumbering the targeted capture events involving random homology. Additionally, we observed an overall increase in targeted capture when eMH was present, as opposed to the absence of homology. Building on these results, we developed a technique for more subtle edits than complete T-DNA integration. By employing a dual-T-DNA approach, where the desired genetic changes were incorporated into one T-DNA, while another T-DNA contained the necessary reagents, including Cas9, sgRNA, and a selectable marker, we successfully obtained targeted mutations, ranging from single nucleotide polymorphisms (SNPs) to insertions and replacements. This method achieved an editing efficiency of approximately 2%, vastly higher than typical gene targeting efficiency.

This comprehensive exploration of DNA repair mechanisms, CRISPR/Cas technology, and targeted gene editing in plant biotechnology highlights the immense potential for advancing crop improvement and sustainable agriculture.