Different DNA repair pathways are involved in single-strand break-induced genomic changes in plants

Felix Wolter [1], **Patrick Schindele** [1], **Natalja Beying** [1], **Armin Scheben** [2] and **Holger Puchta** [1,*,†]

1 Botanical Institute, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany
2 Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

*Author for correspondence: holger.puchta@kit.edu
†Senior author.

F.W., P.S., N.B., and H.P. designed the research; F.W., P.S., and N.B. performed the research; F.W., P.S., N.B., A.S., and H.P. analyzed the data; F.W., P.S., N.B., A.S., and H.P. wrote the paper.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the instructions for Authors (https://academic.oup.com/plcell) is: Holger Puchta (holger.puchta@kit.edu).

Abstract

In nature, single-strand breaks (SSBs) in DNA occur more frequently (by orders of magnitude) than double-strand breaks (DSBs). SSBs induced by the CRISPR/Cas9 nickase at a distance of 50–100 bp on opposite strands are highly mutagenic, leading to insertions/deletions (InDels), with insertions mainly occurring as direct tandem duplications. As short tandem repeats are overrepresented in plant genomes, this mechanism seems to be important for genome evolution. We investigated the distance at which paired 5′-overhanging SSBs are mutagenic and which DNA repair pathways are essential for insertion formation in *Arabidopsis thaliana*. We were able to detect InDel formation up to a distance of 250 bp, although with much reduced efficiency. Surprisingly, the loss of the classical nonhomologous end joining (NHEJ) pathway factors KU70 or DNA ligase 4 completely abolished tandem repeat formation. The microhomology-mediated NHEJ factor POLQ was required only for patch-like insertions, which are well-known from DSB repair as templated insertions from ectopic sites. As SSBs can also be repaired using homology, we furthermore asked whether the classical homologous recombination (HR) pathway is involved in this process in plants. The fact that RAD54 is not required for homology-mediated SSB repair demonstrates that the mechanisms for DSB- and SSB-induced HR differ in plants.

Introduction

Double-strand break (DSB) repair is crucial for the maintenance of genome stability. In the last millennium, site-specific nucleases like HO in yeast and I-SceI in mammals and plants were applied to study DSB repair in eukaryotes (Jasin and Haber, 2016; Puchta, 2016), which defined the two major pathways of DSB repair: homologous recombination (HR) and nonhomologous end joining (NHEJ; Puchta, 2005). Later research revealed that plants also have two different subpathways of NHEJ: the classical one (cNHEJ) characterized by the involvement of DNA ligase 4 (LIG4) and the Ku heterodimer that protects the double-stranded ends from degradation (Friesner and Britt, 2003; Huefner et al., 2011), and the

microhomology-mediated, alternative pathway (aNHEJ), in which polymerase Q (POLQ) plays a decisive role (van Schendel and Tijsterman, 2013; Mateos-Gomez et al., 2015). Following the development of artificial nucleases for gene editing, interest in DSB repair has grown and many studies have used synthetic enzymes to induce DSBs at different sites in various organisms (Schmidt et al., 2019a).

In contrast, only few studies have addressed the question of how single-strand breaks (SSB) can result in genetic change in eukaryotic genomes (Maizels and Davis, 2018). With the discovery of the CRISPR/Cas system, it became very easy to obtain a sequence-specific nickase from the Cas9 nuclease by simply introducing a point mutation in the active center of one of the two nuclease domains (Jinek et al., 2012; Mali et al., 2013). Thus, SSBs can now be induced at almost any position in the genome. Several studies have addressed how SSBs are repaired in mammalian cells. In the vast majority of cases, SSBs are simply religated by DNA Ligase 1 (Abbotts and Wilson, 2017). If replication occurs before ligation, SSBs are transformed into DSBs. If a single-ended DSB is formed, it might be repaired by HR. If a double-ended DSB is formed, repair occurs mainly by classical or alternative NHEJ, often leading to mutations at the break site (Wu, 2007; Shibata, 2017).

Interestingly, for mammalian cells, conflicting results have been published on whether the classical factors involved in homologous DSB repair are also involved if homology is used for SSB repair (Davis and Maizels, 2014; Vriend et al., 2016; Nakajima et al., 2018). This indicates that depending on cell-type, SSBs themselves, and not only SSBs transformed into DSBs, might induce HR but by a different mechanism. The induction of two SSBs at a certain distance on opposite strands efficiently induces mutations between these nicks. This strategy was originally developed to enhance specificity of DSB induction by Cas9 in mammalian cells: by the use of a nickase and two sgRNAs instead of a nuclease with one sgRNA, 46 instead of 23 bases are involved in defining specificity. In the case of the nickase, a DSB with overhanging ends is produced. The resulting repair led to a much broader spectrum of the allelic mutations, especially InDels (insertions/deletions) were larger in comparison to blunt-ended DSBs (Ran et al., 2013).

Using the Cas9 nickase, we previously showed that in the model plant *Arabidopsis thaliana*, the induction of a single SSB is hardly mutagenic but strongly enhances HR between intrachromosomal repeats (Fauser et al., 2014). Moreover, we were able to demonstrate that staggered breaks with 5′-overhangs of about 50 nucleotides were especially mutagenic, leading to a broad spectrum of insertions and deletions (Schiml et al., 2016). Similar results were later reported for mammalian cells (Bothmer et al., 2017). Of special interest was our finding that tandem sequence duplications arose at high frequency, even at a distance of 100 bp between nicks on opposite strands (Schiml et al., 2016).

SSBs are repair intermediates in base and nuclear excision repair (NER; Roldán-Arjona et al., 2019). Due to their sessile and autotrophic nature, plants are under constant threat due to harmful ultraviolet (UV) irradiation. Dimerization of pyrimidine bases is a major type of DNA damage induced by UV light. Besides photolyases, NER is the main pathway to repair UV-induced DNA damage in plants (Molinier, 2017). As it has been reported before that short tandem duplications are statistically overrepresented in genomes of plants such as rice (Vaughn and Bennetzen, 2014), the paired SSB-induced formation of tandem duplication seems to be an important mechanism for the evolution of plant genomes in general.

Therefore, we set out to define the precise mechanism of the process of duplication formation. A mechanism was proposed before (Vaughn and Bennetzen, 2014) and further refined by us (Schiml et al., 2016). According to this model, tandem duplications arise by fill-in synthesis from a staggered DSB with 5′-protrusions. It is tempting to speculate that the formation of such DSBs might be initiated by replicative DNA helicases. Both 5′-ends are set free and resected, while simultaneously new DNA synthesis begins from the 3′-ends. At some point, resection and synthesis will have completely consumed the single-stranded protrusion, leaving a blunt-ended DSB. The blunt-ended DSB can then be sealed by ligation by cNHEJ or aNHEJ using microhomologies. Depending on the extent of 5′-end resection, deletions or duplications of variable length are formed. Theoretically, microhomologies could also be used for hybridizing of the two 5′-overhangs early on. DNA synthesis fills in the single-stranded gaps, while the protruding 5′-flaps are removed. Finally, the remaining nicks are sealed by ligation (Figure 1A). In contrast, insertions that have been characterized before during DSB repair in plants often consist of various shorter, sometimes also longer patches of DNA sequences that have been copied from various sites mostly close to the break by a synthesis-dependent strand annealing like mechanisms (Figure 1B; Gorbunova and Levy, 1997; Salomon and Puchta, 1998). Thus, the nature of the insertion as well as their formation mechanism differs between paired, staggered SSBs, and DSBs in plant cells.

In the current study, we analyzed up to which distance the repair of paired 5′-overhanging SSBs can result in InDels. Furthermore, we tested by mutant analysis whether a series of key actors of different DNA repair pathways are required for tandem duplication formation. Finally, we tested whether the classical repair machinery of HR mediated, RAD54-dependent DSB repair is also involved in SSB repair by HR in plants.

Results

The influence of the 5′-overhang length on paired nickase efficiency

We were able to show before that the induction of paired SSBs on opposite strands at a distance of 52 bp resulting in 5′-overhangs is highly mutagenic (Schiml et al., 2014).

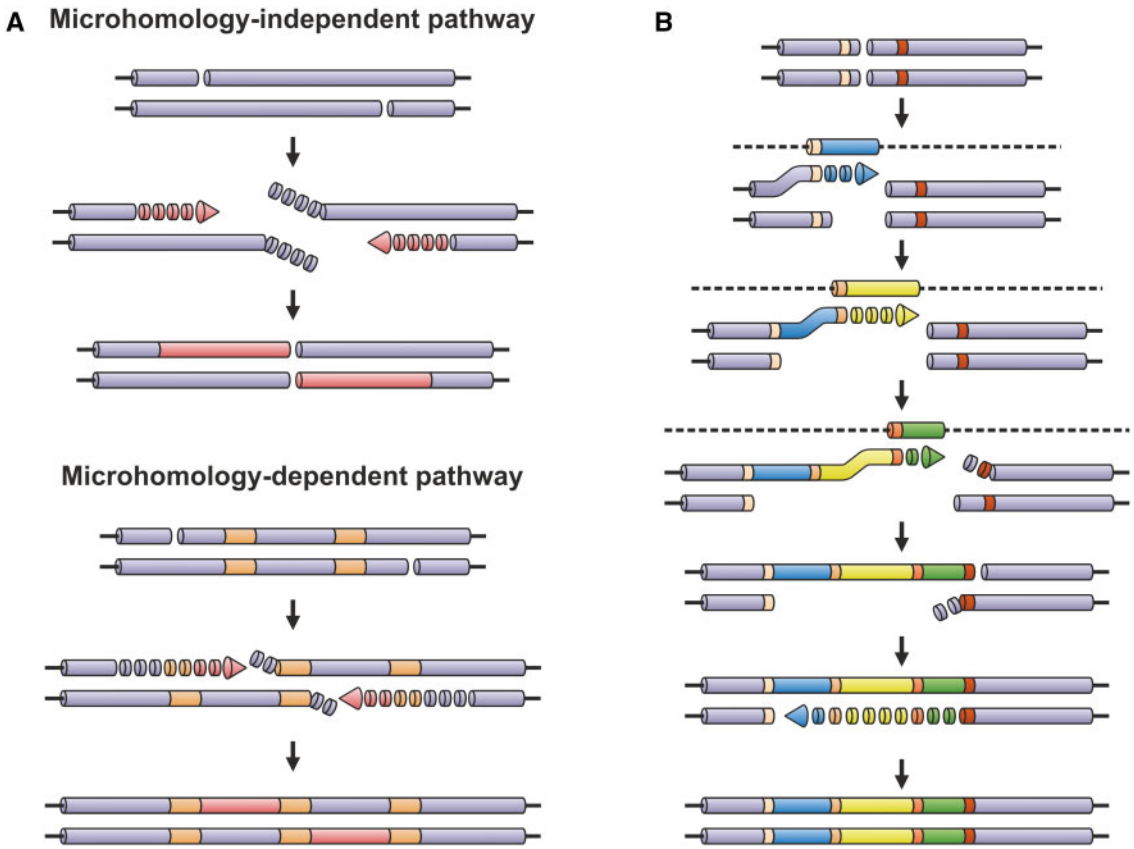

Figure 1 Models for the formation of tandem duplications and patch insertions. A, Models for the formation of tandem duplications from a staggered DSB with 5′-overhangs. Release and degradation of both 5′-overhangs with simultaneous synthesis from the 3′-ends results in sequence duplications of variable length. Microhomologies could theoretically be included for the hybridization of the 5′-overhangs. B, Model for the formation of patch insertions from a DSB. Microhomologies at the break site facilitate hybridization and copying of proximal sequence context through an SDSA-like mechanism. Inserted sequences may originate from several distinct sequence templates. Chopped fragments with arrows indicate DNA synthesis; chopped fragments without arrows indicate resection. Dotted bars with barrels indicate sequence context from remote chromosomal sites. Microhomologies are indicated in orange.

Furthermore, mutation efficiency did not decline when the distance between the nicks was extended to 100 bp (Schiml et al., 2016). Therefore, we were interested to study the repair of even larger spaced nicks (Figure 2A). To get comparable results, we used the same two loci in the genome as in our previous study, namely *ALCOHOL DEHYDROGENASE1* (*ADH1*) and a heterochromatic locus adjacent to the centromere of chromosome 3, which we refer to as heterochromatin locus (Schiml et al., 2016). Three nick distances of around 100 bp, 250 bp, and 600 bp were tested. Due to target sequence constraints, especially protospacer adjacent motif (PAM) requirements, distances could not always be matched to the nucleotide. All distances were tested in relation to the locus (Supplemental Table S1A), leading to six constructs that were generated and transformed into Arabidopsis Col-0 wild-type (WT) plants.

For all six approaches (three distances for each of the two loci), 30 primary transformants were selected, pooled and genomic DNA was purified. The respective target region was amplified using specific barcodes for each approach and subjected to next-generation sequencing (NGS) analysis. Due to the long distance between the nicks, we used the Single Molecule Real Time (SMRT) technology from Pacific Biosciences for this experiment, which can produce average read lengths of longer than 10 kb. In this way, we could ensure that the whole genomic region between the nicks is captured on a single read and even longer insertions can be properly identified. The long read length obtained from the SMRT sequencing technology came at the cost of reduced read number compared to Illumina technology, which was used for the analysis of 50-bp spaced paired nick repair in different mutant backgrounds below (read numbers for the different approaches are shown in Supplemental Table S2).

Background noise increases with increasing amplicon size due to a higher probability of a sequencing error within the amplicon. In this study, background InDel rates were 0.20% for the 1.1-kb amplicons (used for the amplification of the 600-bp nick distance approaches), and 0.15% for the 800-bp amplicons (used for amplification of 100- and 250-bp nick distances).

InDel for the 100-bp nick distance were found in every 3rd read at the *ADH1* locus and more than every 20th read at the heterochromatin locus (Supplemental Table S1B). InDel frequencies were reduced but clearly above

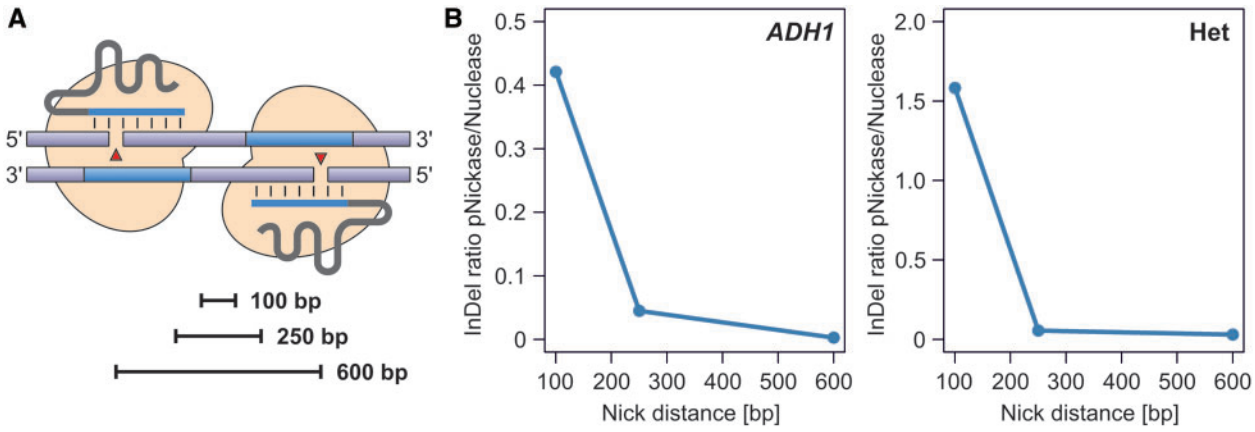

Figure 2 Influence of nick distance on the repair of paired nicks with 5′-overhangs. A, Paired nicks were induced in 5′-overhang configuration with nick distances of 100, 250, and 600 nt. B, Paired nick InDel frequencies as well as nuclease InDel frequencies at the respective target sites were determined at the ADH1 and heterochromatin (Het) loci. The ratio between the paired nick InDel frequencies and the nuclease InDel frequencies at the respective target sites are shown. Concerning the nuclease InDel frequencies, only the values of the respective downstream targets were used for calculation as, in case of the paired nicks, the upstream target is shared among the different approaches. Paired nickase is abbreviated as pNickase.

background level for a nick distance of 250 bp, with InDel rates in the 2% range for the *ADH1* and heterochromatin locus, respectively. In contrast, for a distance of 600 bp, rates were on the same level as in case of the untransformed control plants and thus not discernible from background noise (Supplemental Table S1B). Elevated InDel rates compared to the background were consistent even when using a range of filtering strategies based on InDel length and distance from the nick region (see "Materials and Methods" as well as Supplemental Tables S3 and S4).

These results indicate that the InDel frequency obtained with paired nicks is highly dependent on nick distances. However, as the drop in InDel frequency with increasing nick distances might also correlate with the efficiency of break induction at the individual target sites, we further analyzed the mutation rate as proxy for cleavage efficiency at the target sites using the Cas9 nuclease (Supplemental Table S1C). Figure 2B shows the ratio between the InDel frequencies of the paired nickase approaches and the cleavage efficiencies of the nuclease at the individual target sites. With increasing nick distances, a clear drop in relative InDel frequencies can be observed, with the most dramatic drop occurring between distances from 100 bp to 250 bp. This suggests that the occurrence of two SSBs in opposite strands with 5′-overhangs is only mutagenic at a maximum distance of hardly more than 100 nucleotides.

Defining DNA repair pathways involved in the formation of tandem duplications

In order to define which repair pathways are involved in tandem duplication formation, a larger number of Arabidopsis mutant lines deficient in important DNA repair factors were subjected to paired nick induction and the repair outcome was analyzed. As the tandem duplications were very efficiently induced by a paired nick distance of 50 bp generating a staggered break with 5′-overhangs, we used a previously applied experimental setup (Schiml et al., 2016) for testing the involvement of different DNA repair pathways. The enzyme machinery of three different DSB repair pathways has been characterized in Arabidopsis in detail and we tested two representative members of each pathway. KU70 (Tamura et al., 2002) and DNA LIG4 (West et al., 2000) are both involved in cNHEJ. X-ray repair cross complementing 1 (XRCC1; Charbonnel et al., 2010) and to a lesser extent poly(ADP-Ribose)-polymerase 1 (PARP1; Babiychuk et al., 1998; Metzger et al., 2013) are involved in aNHEJ, while Radiation Sensitive 51 (RAD51; Li et al., 2004) and Radiation Sensitive 54 (RAD54; Osakabe et al., 2006) are required for HR. As the *rad51* mutant is sterile in its homozygous form (Li et al., 2004), heterozygous mutants were transformed and the obtained primary transformants subsequently screened for homozygous mutants which were then used for the NGS analysis. We also tested nucleases that are involved in the processing of DNA repair intermediates, namely Meiotic Recombination 11 (MRE11; Hartung et al., 2006; Geuting et al., 2009), Excision Repair Cross Complementation 1 (ERCC1; Hefner et al., 2003; Dubest et al., 2004), and Exonuclease 1a/b (EXO1a/b; Kazda et al., 2012; Derboven et al., 2014). Moreover, the helicase RecQ4A (Hartung et al., 2007; Schröpfer et al., 2014) and polymerase epsilon Recovery Protein 3 (REV3; Kobbe et al., 2015), which are both involved in DNA crosslink repair were also included.

In a first set of experiments, paired nicks were introduced at the same sites in the heterochromatin locus as before (Schiml et al., 2016). Accordingly, the respective constructs were transformed into all different mutant backgrounds. If required, the transformation marker was changed to the phosphinothricin (PPT) resistance gene in some constructs, since some of the insertion mutants carried T-DNAs with a Kanamycin resistance marker. The seeds of the transformed plants were sown on PPT or kanamycin containing medium,

respectively. After 3 weeks of growth, 30 primary transformants for each of the mutant backgrounds were harvested and their pooled DNA was analyzed by polymerase chain reaction (PCR) amplification of the target site followed by Illumina sequencing of the amplicons.

In the WT, an InDel rate of around 24.6% was found, which is comparable to our previous findings for a 50-bp nick distance at the heterochromatic locus (27.7% mutation rate; Schiml et al., 2016). The occurrence of insertions and deletions was determined for all mutants (Supplemental Figure S1 and Supplemental Table S5). The two mutants representing the cNHEJ pathway *ku70* and *lig4* differed drastically from WT and all other tested mutants. While the number of deletions increased, the number of insertions is dramatically reduced, resulting in a substantial change in the ratio between both classes of events (Figure 3). The fact that both cNHEJ mutants show the same effect underlines that this outcome is neither dependent on T-DNA integration nor related to background mutations. Moreover, the obtained InDels differ from all other samples. Whereas WT and all other mutants show a bimodal deletion distribution with a peak at each of the two nick sites (Figure 4; Supplemental Figure S2), *lig4* and *ku70* mutants show a nonbimodal and unimodal distribution with a single peak in the center between the two nick sites, respectively. Interestingly, we also saw a deviation in the case of the *xrcc1* mutant. Here, the bimodal distribution of the deletion was even more pronounced (Figure 4).

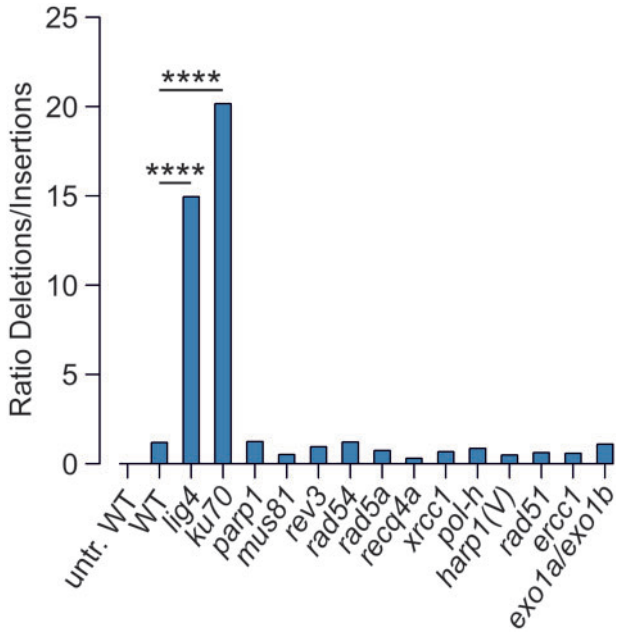

Figure 3 Ratio between deletion and insertion frequencies in different mutant backgrounds at the heterochromatin locus. Deletion and insertion frequencies as percentage of total reads were determined at the heterochromatin locus with nick distances of 50 nt in 5′-overhang configuration. Insertion and deletion frequencies were 11.25% and 13.33% in WT, 1.61% and 24.04% in *lig4*, and 1.21% and 24.33% in *ku70*, respectively. *P*-values were calculated using the Fisher's exact test, ****$P < 0.0001$.

Defining the nature of the insertions induced by paired nicks reveals that two distinct classes can be identified: tandem duplications and patch insertions. Patch insertions contain patches of DNA sequences that were copied from sites mostly close to the break. For all genetic backgrounds, in a first step insertions were ranked according to read number using the Cas-Analyzer software. Beginning with the most frequent insertions concerning read numbers, they were categorized either as tandem duplications or patch insertions, until the read number per read dropped below ten reads. This way, more than half of the reads could be evaluated for all mutants (with read counts being usually around 100,000 per mutant, see Supplemental Table S5 for the numbers per mutant). To make the classification as clear as possible, all insertions not meeting the requirements of a tandem duplication, as perfect repetition of the immediate upstream sequence, were classified as patch insertion, even though in few cases the origin of the insertion remained unclear.

In the WT, around 90% of the insertions could be classified as tandem duplications (Figure 5; Supplemental Table S6). In stark contrast, for both cNHEJ mutants, *lig4* and *ku70*, only 20% and 10% of the insertions could be categorized as tandem duplications, respectively. Interestingly, the *xrcc1* mutant showed almost exclusively tandem duplications. Since the total amount of insertions differs between the genetic backgrounds (Supplemental Figure S3), the absolute amounts of insertion types are best visualized as percentage of all reads (Supplemental Figure S4). Here, it can be seen that in the case of *lig4* and *ku70* genetic background, the absolute amount of patch insertions is not substantially different from the WT. However, tandem duplications are almost completely absent, which constitutes the only cause for the shift in relative amounts observed above. This strongly suggests a key role of the cNHEJ pathway in the formation of tandem duplications but no involvement of all other factors tested.

The classical NHEJ pathway is required for the formation of tandem duplication independently of the locus

In order to test whether cNHEJ factors are indeed essential for tandem duplication formation and the alteration not only arises from sequence specific peculiarities of a single locus, we also tested the same approach at the *ADH1* locus (Schiml et al., 2016). Analogous to the heterochromatin locus, we transformed the construct leading to a 5′-overhang configuration with a nick distance of 50 bp in the WT, as well as *lig4* and *ku70* mutants. In the WT, we obtained a general InDel frequency of 61% (Supplemental Figure S3), which is higher than what was obtained for the heterochromatic locus (24.6%) and close to the 63% obtained by Schiml et al. for the *ADH1* locus.

In *ku70* and *lig4* mutants, InDel frequencies were even higher with 68% and 80%, respectively. The increase in InDel frequency stems solely from an increase of deletion frequencies, whereas insertion frequencies were strongly reduced, just

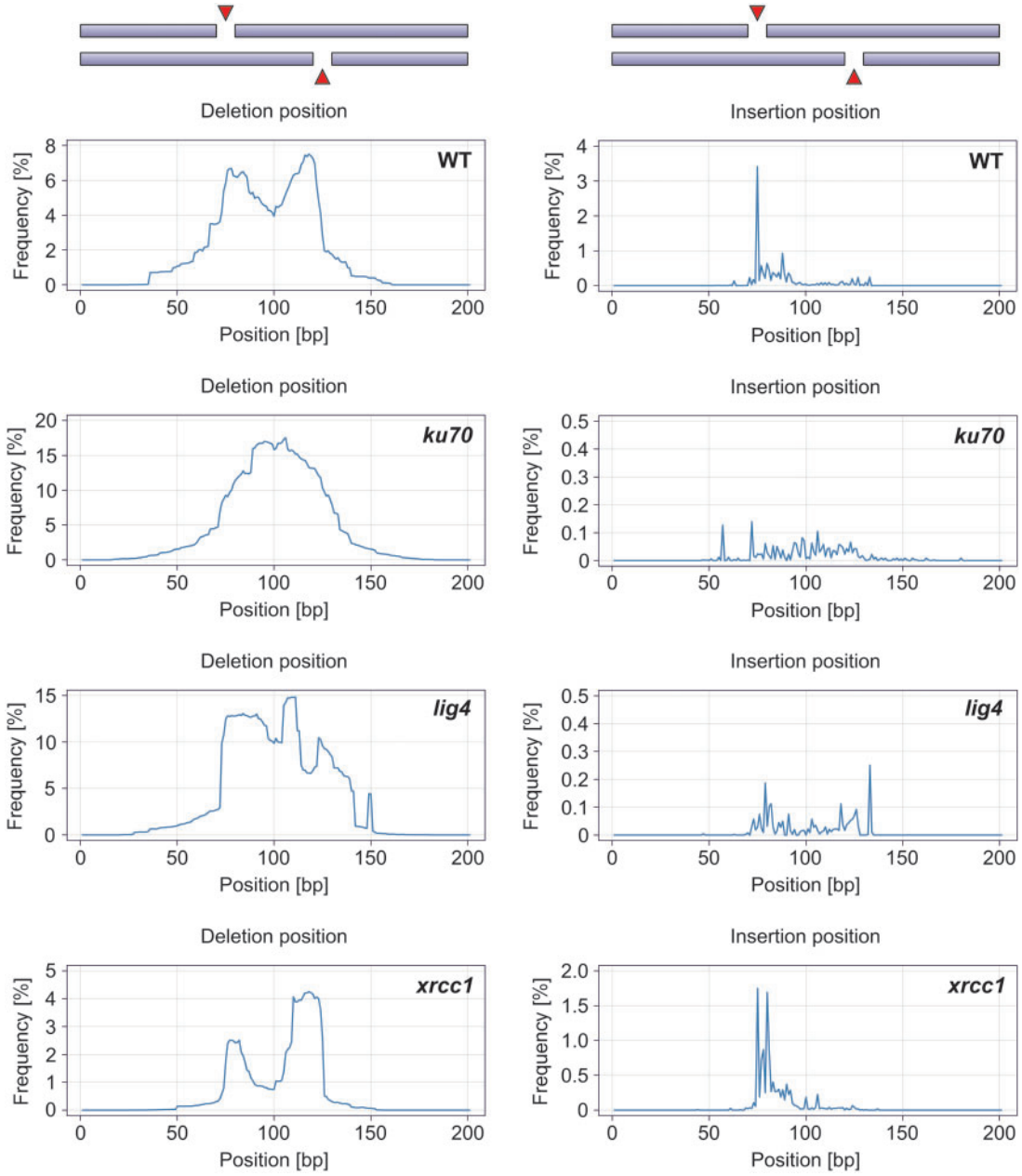

Figure 4 Distribution of deletions and insertions in different mutant backgrounds at the heterochromatin locus. The distribution was determined at the heterochromatin locus with nick distances of 50 nt in 5′-overhang configuration. Illumina sequencing data were evaluated using the Cas-Analyzer online tool. Nick induction of the upstream target occurs between positions 74 and 75, nick induction of the downstream target between positions 125 and 126.

as found for the heterochromatin locus. This leads to pronounced increases in the deletion/insertion ratio, which was 9.5 for the WT, but 35.2 for *ku70* and 44.9 for *lig4*. Even though the magnitude of this shift is smaller than at the heterochromatin locus, the direction of the effect is sustained.

Concerning the different insertion categories, we obtained in general lower tandem duplication frequencies at the *ADH1* locus, which might be an effect of local sequence context. In the WT, 54% of insertions could be classified as tandem duplications, compared to 90% at the heterochromatin locus. However, for *ku70* and *lig4* mutants, this was only 14% and 11%, again showing a tremendous reduction (Figure 6). This effect becomes even clearer when looking at the frequency of tandem duplications as percentage of all reads. In the WT 3.0% of all reads were tandem duplications, whereas in *lig4* and *ku70* only 0.27% and 0.19% of all reads, respectively, were tandem duplications (Supplemental Figure S5). These findings lend further support to the hypothesis

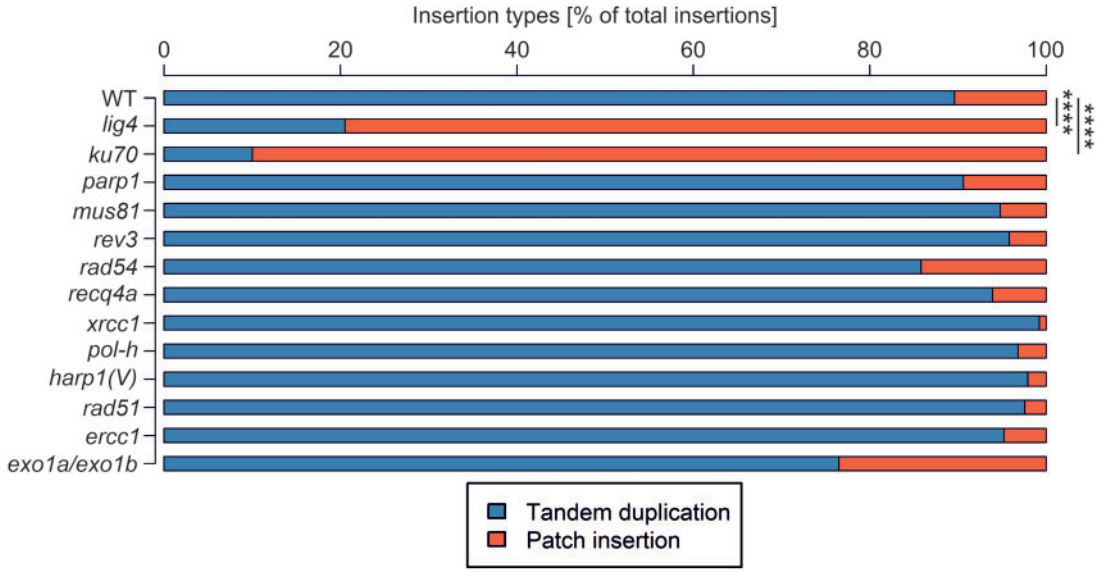

Figure 5 Distribution of insertion types in different mutant backgrounds at the heterochromatin locus. Insertion types were determined at the heterochromatin locus with nick distances of 50 nt in 5′-overhang configuration. Insertions were classified as tandem duplications and patch insertions. The respective amounts are presented as percentage of total insertions. *P*-values were calculated using the Fisher's exact test, ****$P<0.0001$.

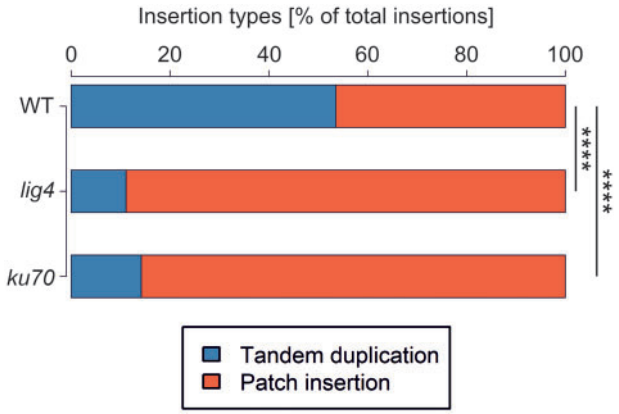

Figure 6 Distribution of insertion types in *lig4* and *ku70* at the ADH1 locus. Insertion types were determined at the ADH1 locus with nick distances of 50 nt in 5′-overhang configuration. Insertions were classified as tandem duplications and patch insertions. The respective amounts are presented as percentage of total insertions. *P*-values were calculated using the Fisher's exact test, ****$P<0.0001$.

that the cNHEJ pathway plays a key role in the formation of tandem duplications. Supplemental Tables S7 and S8 show a detailed overview of deletion and insertion frequencies as well as the amount of paired-end reads that met quality standards and were evaluated.

POLQ is required for patch insertions but not for tandem duplications

As Polymerase Q is a key actor of aNHEJ, we expected that the loss of this factor would have major consequences for paired nick repair. The obstacle to testing this was the fact that POLQ is indispensable for T-DNA integration (van Kregten et al., 2016). To obtain a *polq* mutant carrying a T-DNA expressing the Cas9 nickase, the following strategy was devised. In addition to the T-DNA carrying sgRNAs and Cas9 nickase for the heterochromatic locus, a second "helper" T-DNA was co-transformed by flower dipping that contained only a *POLQ* overexpression cassette with a *UBIQUITIN* promoter but no selection marker. The rationale behind this was that transient *POLQ* expression from the co-transformed T-DNA might lead, in some cases, to integration of the desired T-DNA without concomitant integration of the *POLQ*-expressing T-DNA, which would complement the mutant persistently.

Accordingly, primary transformants would have to be screened for plants carrying only the desired T-DNA but not the "helper" T-DNA. Eighty homozygous *polq* mutant plants (*teb-5*; Inagaki et al., 2006; Klemm et al., 2017) were transformed by flower dipping in this way. All obtained seeds were evaluated by sowing them on individual plates containing PPT. This way, PPT resistant plants were obtained and genotyped. Eventually, 16 individual primary transformants could be identified by PCR. However, screening for integration of the "helper" T-DNA revealed only four individual plants containing the desired T-DNA, but no *POLQ* expression cassette.

These plants were grown to maturity and from their progeny 30 T-DNA containing individuals each were analyzed by NGS. We were only able to detect the expected InDels at the heterochromatic locus in the progeny of one of the four lines, and only these individuals were evaluated. In the other cases, most probably due to aberrant T-DNA integration, only the selection marker but not the nickase was integrated in a functional form. Concerning this, it is essential to mention that the *polq* mutant has to be regarded separately

from all others as POLQ is indispensable for T-DNA integration. Thus, even by heterologous expression of POLQ, it cannot be expected that T-DNA integration is fully restored or that aberrant T-DNA integration occurs at a similar frequency (here being much more frequent in case of *polq*). In all other genetic backgrounds, primary transformants that harbor the T-DNA in a dysfunctional form are negligible when pools of 30 plants are analyzed. In *polq* mutants, dysfunctional T-DNA carriers or plants that gained resistance by other means are enriched as the rate of these abnormal resistance gainers is constant, whereas the rate of functional T-DNA carriers is strongly reduced in case of *polq* mutants, causing a shift in the ratio.

Fortunately, using the progeny of the remaining T1 line, we were able to characterize the repair of paired 5′-overhang SSBs in the absence of POLQ. Indeed, our analysis revealed a pattern of insertions and deletions differing from all other mutant backgrounds. A more pronounced bimodal deletion distribution was found in the WT (Figure 7A). The difference in deletion and insertion frequency becomes most apparent when the ratio between both is calculated. The *polq* mutants show a three-fold decrease in comparison to the WT (in contrast to the ten-fold increase of the cNHEJ mutants; Figure 7B; Supplemental Table S5). Moreover, *polq* mutants show almost exclusively tandem duplications, whereas patch insertions are almost lost completely (Figure 7C). Thus, the repair pattern observed in the absence of POLQ is strongly different than the WT and differs even more from the cNHEJ mutants. This is somewhat reminiscent but also more severe than that in *xrcc1* background (see Figures 3 and 4). The latter finding suggests the aNHEJ pathway with a key role of POLQ and to a lesser extent of XRCC1 in the formation of patch insertions.

Supplemental Figure S6 shows an overview of the five most frequent insertions and deletions in all different genetic backgrounds according to the read count.

The classical DSB-mediated HR pathway is not involved in homology-mediated SSB repair

In contrast to paired SSBs on opposing strands, single SSBs are extremely common lesions in Eukaryotic cells. Although those are mainly repaired by simple religation, homology might also be used for their repair. Various studies showed that SSBs induce HR in eukaryotes. Indeed, this might be due to the presence of this kind of lesion itself or due to its transformation into a DSB during replication. Therefore, we asked whether the classical HR pathway is required for this process in plants as conflicting results have been published for mammalian cells (Davis and Maizels, 2014; Vriend et al., 2016; Nakajima et al., 2018). In former studies, we were able to elucidate the role of various proteins involved in DSB-induced HR by crossing two different intra-chromosomal HR reporter constructs into a variety of DNA-repair mutants (Mannuss et al., 2010; Roth et al., 2012). DSB induction by

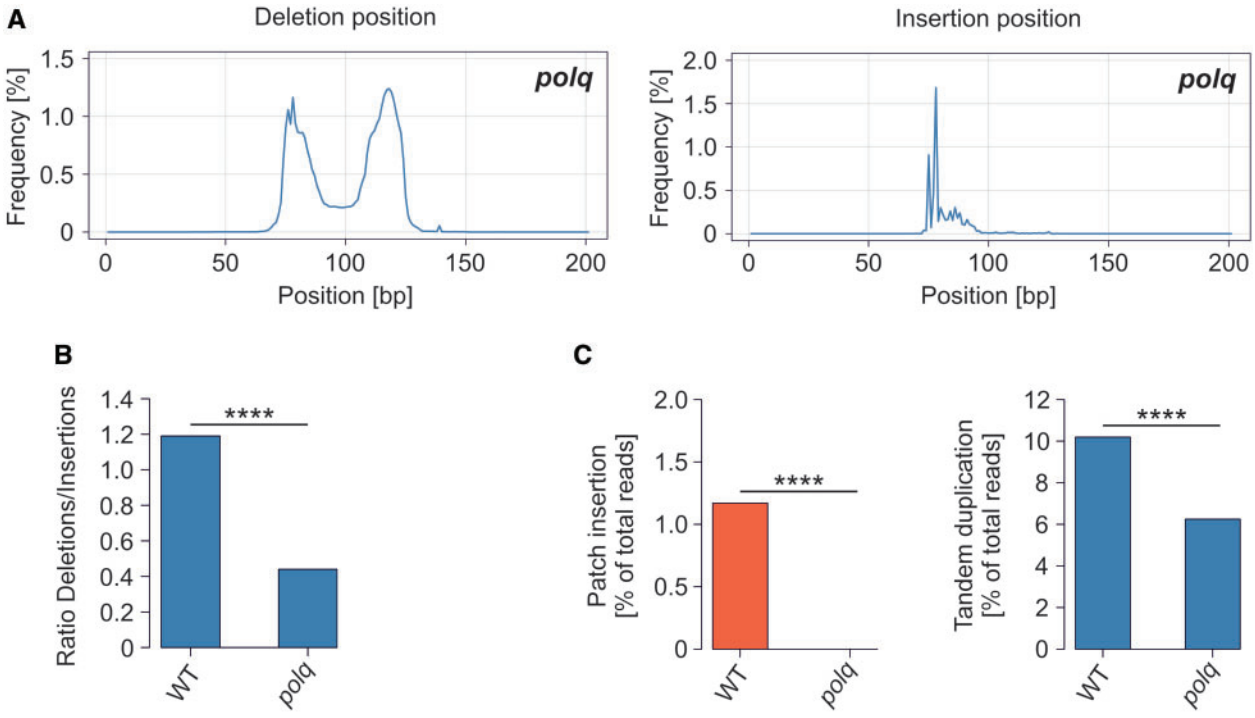

Figure 7 Evaluation of deletion and insertion analysis in *polq* mutant background. A, Distribution of deletions and insertions in *polq* T2. Illumina sequencing data were evaluated using the Cas-Analyzer online tool. Nick induction of the upstream target occurs between position 74 and 75, nick induction of the downstream target between position 125 and 126. B, Ratio of deletion/insertion in *polq* T2. C, Distribution of insertion types in *polq* T2. Insertions were classified as tandem duplications and patch insertions. The respective amounts are presented as percentage of total reads. All data were determined at the heterochromatin locus with nick distances of 50 nt in 5′-overhang configuration. *P*-values were calculated using the Fisher's exact test, ****$P<0.0001$.

sequence-specific nucleases in these reporter constructs results in the restoration of a marker by single-strand annealing (SSA; DGU.US line), or by a synthesis-dependent strand annealing (SDSA)-mediated gene conversion (IU.GUS line), respectively (Orel et al., 2003). Both lines carry as marker gene the bacterial *uidA* gene (GUS), which codes for the enzyme β-glucuronidase (Figure 8A). However, the sequence is interrupted by a short spacer sequence and is therefore not functional. The sequence can be restored by induction of a break in the spacer region and subsequent repair according to the respective mechanism. The efficiency of the reaction can be measured by the restored β-glucuronidase activity.

We were able to show that RAD54, which catalyzes the strand exchange reaction in HR (Tavares et al., 2019) is one of the major factors involved in gene conversion in somatic plant cells. In contrast, RAD54 plays no role in SSA (Mannuss et al., 2010; Roth et al., 2012). To investigate whether DSB-induced and SSB-induced HR share similar mechanisms in plants, we used the T-DNA insertion mutant line *rad54-1*. The setup was identical to our previous experiments (Mannuss et al., 2010; Roth et al., 2012) with the only difference being that single SSBs instead of DSB were induced in the same transgenic recombination lines as before.

For SSB induction, we used the previously published T-DNA constructs pDe-SpCas9-D10A+SSA and pDe-SpCas9-D10A+SDSA (Fauser et al., 2014). The sgRNAs of both constructs (SSA, SDSA) bind in the respective spacer of the corresponding reporter line. The expression vectors were transformed into the respective reporter lines via Agrobacterium-mediated transformation, the progeny sown on selection medium and 2-week-old primary transformants (approximately 20 seedlings) selected for further analysis. For measurement of the HR efficiency, we applied a 4-methylumbelliferyl-galactopyranoside (4-MUG) assay for β-glucuronidase activity in combination with a Lowry assay as previously described (Fauser et al., 2014; Steinert et al., 2015). Three biological replicates for each line were subsequently analyzed in the 4-MUG assay.

For better comparability between the reporter backgrounds, we normalized the data to the respective WT. No difference between the *rad54* mutant and the respective WT reporter lines could be identified (Figure 8B). Thus, involvement of RAD54 in the repair of these reporter substrates could be excluded. This demonstrates that SSB in contrast to DSB-induced gene conversion is RAD54-independent.

Discussion

A wealth of information has been accumulated over the years on how DSB are repaired in various eukaryotic genomes and what kind of genetic consequences this repair reaction can have. Due to the development of the field of gene editing, interest in the topic has recently massively increased. By now, DSB-induced genome engineering has

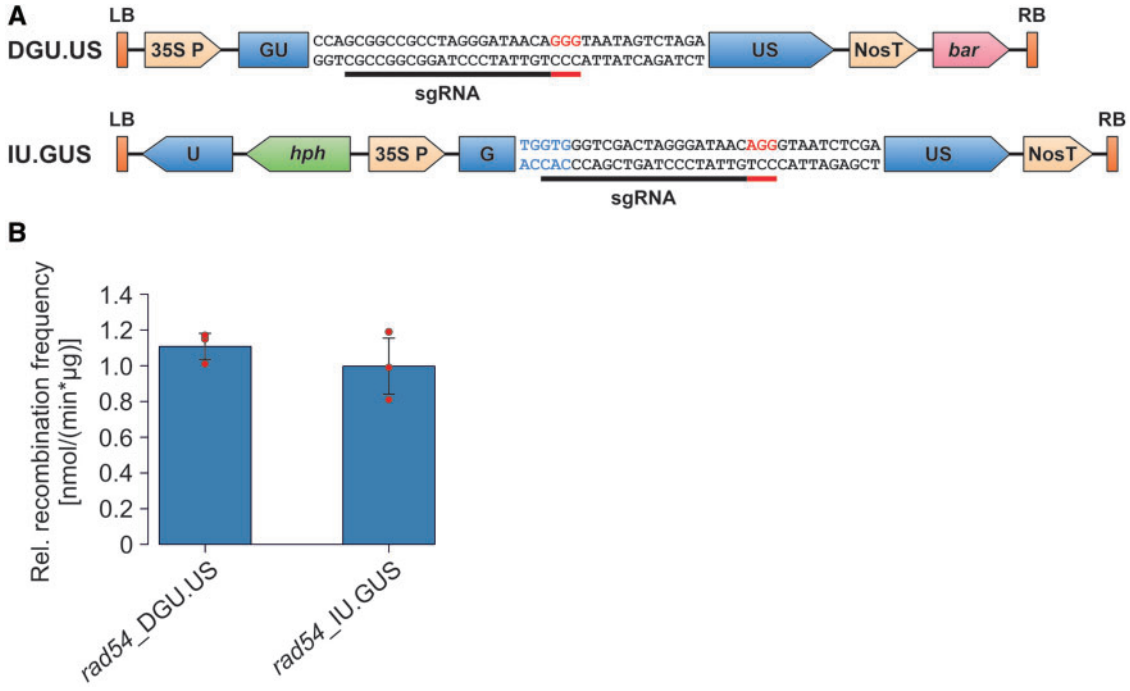

Figure 8 SSB-induced HR. A, Intrachromosomal HR reporter transgenes DGU.US and IU.GUS. In case of DGU.US, the ß-glucuronidase activity is restored by SSA; in case of IU.GUS via gene conversion. The guide sequence is indicated by a black bar, the PAM sequence is indicated by a red bar and highlighted in red. LB, left border, RB, right border, 35S P, 35S promoter, NosT, nopaline synthase terminator, bar, bar resistance, hph, hygromycin resistance, GUS, β-glucuronidase. B, Relative recombination frequency in *rad54*. The recombination frequencies of the *rad54-1* mutant lines in the respective reporter backgrounds DGU.US and IU.GUS relative to the WT WT = 1 are shown. Three biological replicates were performed for each line ($n = 3$), the error bars indicate the standard deviation.

become the standard technology for a large number of eukaryotes applied in many thousands of labs worldwide. Although SSBs occur in orders of magnitude more often than natural DSBs (Cao et al., 2019) in the lifetime of a cell, much less information has been obtained on SSB repair mechanisms. The availability of tools for SSB induction was limited for many years (Kim et al., 2012; Metzger et al., 2013; Katz et al., 2014; Wu et al., 2015). As by a simple mutation the Cas9 nuclease can be transferred into a nickase (Jinek et al., 2012), it became extremely easy to study SSB repair anywhere in the genome. Moreover, the induction of nicks by Cas9 has also been used for the development of different new genome engineering technologies. The first application of the Cas9 nickase was the induction of staggered SSBs on opposite strands to enhance Cas9 specificity of DSB induction (Mali et al., 2013; Ran et al., 2013). Later on, nickases were applied to mimic the replicating strand during base editing (Komor et al., 2016) and most recently for prime editing (Anzalone et al., 2019).

We started to use the Cas9 nickase for genome engineering in plants and could show that the induction of a single nick itself is not very mutagenic (Fauser et al., 2014). This was hardly surprising as most SSBs are repaired by simple religation. Only if a SSB is transformed into a DSB during replication, it might become mutagenic. In contrast, the paired nickase approach can be applied efficiently to induce heritable changes in Arabidopsis at a similar efficiency as DSB induction (Schiml et al., 2014), a fact that was later confirmed by others for rice (*Oryza sativa*; Mikami et al., 2016).

Over the course of these studies, we realized that besides deletions, insertions often arise during the repair of the paired nicks. Most of these insertions consist of simple tandem duplications of the sequence between the nicks (Schiml et al., 2014, 2016). The model proposed to explain their formation (Schiml et al., 2016) is simple and shown in Figure 1. By simple templated DNA synthesis from both ends, the stretch between the staggered nick is duplicated and then the resulting double stranded ends are ligated either with or without the use of microhomologies.

Interestingly, the nature of these insertions differs drastically from what we (Salomon and Puchta, 1998) and others (Gorbunova and Levy, 1997) found occurring during DSB repair in plant cells. Here, insertions are mostly a patchwork of different short and in some cases also longer sequences. These insertions are clearly templated, being copied from ectopic sites originating not from immediate context as the tandem duplications, but still mostly from close proximity. In addition, the sequence immediately upstream of the deletion and the sequence immediately upstream of the template for the insertion usually show a few nucleotides of microhomology. The formation of such insertions can be easily explained by a synthesis dependent strand annealing like copying mechanism (Salomon and Puchta, 1998) as shown in Figure 1.

As in plants, the germline is set aside late during somatic growth, insertions might well be transferred to the next generation. Thus, patterns of DNA repair are fixed genetically and play a role in genome evolution. This is indeed the case for short tandem duplications as they are overrepresented in different rice cultivars (Vaughn and Bennetzen, 2014). For a better understanding of plant genome evolution, it was important to learn more about their genesis.

Moreover, the controlled directed formation of tandem duplications might also be of great interest for application in genome engineering. Whereas by now chromosomal sequences can be deleted (Siebert and Puchta, 2002) or inverted (Schmidt et al., 2019b) or even chromosomal translocations (Pacher et al., 2007; Beying et al., 2020) can be achieved by DSB induction, no such technology has been developed for the formation of duplications. It has been shown in a seminal study that by Cas9 nuclease induced deletion formation within transcription factor binding sites of a promotor region, yield and fruit quality can be changed in tomato (Rodríguez-Leal et al., 2017). By the paired nickase approach, it should be possible to induce shorter sequence duplications in promoters to generate more binding sites for specific transcription factors (Wolter and Puchta, 2018).

To test up to which length the induction of such tandem duplications could be induced artificially, we tested paired nicks with distances of 100 bp to 600 bp. Inducing 5′-overhangs, we found that from the practical point of view only duplications of up to 100 bp can be achieved at a frequency feasible for genome engineering. Paired 5′-overhang nicks at a distance of 250 bp are still mutagenic but only to a very small extent, whereas by a longer distance the repair of paired 5′-overhang SSBs becomes totally independent from one another. Our finding also correlates with naturally occurring tandem duplications in rice. Here, few duplications of more than 100 bp could be detected (Vaughn and Bennetzen, 2014). It is tempting to speculate that structures carrying paired nicks are destablized by an incoming replicative DNA helicase. However, those structures might only fall apart and result in a DSB with overhanging ends, if both SSBs are close by. If the second nick is too far away, 250 bp or more, the structure might not immediately be transformed into a DSB.

We also sought to define which DNA repair factors are involved in the process of insertion formation during paired 5′-overhang SSB repair. Although we tested a dozen of different mutants of genes involved in DNA repair pathways, only a few seem to be relevant for insertion formation and most mutant backgrounds did not differ from WT. It turned out that the cNHEJ pathway is indispensable for the formation of tandem duplication. In contrast to all other mutants, both the *ku70* as well as the *lig4* mutant had a massive defect in the formation of this insertion class. We could confirm this finding by testing two independent loci for each cNHEJ mutant. This was surprising, as it indicates that the KU70/80 heterodimer that is protecting double stranded DNA ends from nucleolytic attacks either directly protects longer single stranded overlaps of up to 100 bases efficiently from degradation and/or is enhancing their fill-in reaction.

This might be related to the fact that cNHEJ is also involved in keeping the two ends of a DSB in close contact to avoid uncontrolled ligation with other broken chromosome ends (Schmidt et al., 2019b; Beying et al., 2020). Such a function is also supported by the fact that in the absence of cNHEJ more deletions are formed and beyond that, a unique unimodal deletion distribution can be observed with the most frequent deletion in the center of the two nicks. This distribution might be explained by a mechanism where, due to the lack of binding of the KU heterodimer, the 5′-overhangs are accessible for degradation. In case the 5′-overhangs are resected to a point where hybridization between the complementary 5′-overhangs is no longer possible, fill-in synthesis from the 3′-end and subsequent ligation would result in deletions predominantly in the center of the two nicks. Dependent on the extent of 5′-resection on both sites, deletions can also be further enlarged towards the nicks. This also becomes apparent, if the underlying sequences are taken into account (Supplemental Figure S6).

In *ku70*, deletions extend from the center of the two nicks either to both or one of the two nick sites, which can be explained by either an even or nearly even resection on both sites. Even though it shows much lower affinity to ssDNA, it was shown in vitro that KU70 can associate with ssDNA overhangs and protect them from exonuclease degradation (Krasner et al., 2015; Yuan et al., 2015). This is also in line with the deletion pattern observed in WT where deletions predominantly occur at the nick sites and also becomes apparent, if the underlying sequences are taken into account (Figure 4; Supplemental Figure S6). Next to a deletion similar to those of *ku70*, which is the most abundant concerning read number, deletions prone to one of the two nick sites are highly abundant that are completely absent in *ku70*. As KU70 has a much lower affinity to ssDNA, it is possible that in WT the induced staggered DSB is either repaired independently of KU70, generating deletions similar to those of *ku70*, or that KU70 preferentially binds to one of the two overhangs, generating deletions primarily at one of the two nick sites.

Previously, we found indications for a role of microhomologies in the formation of some tandem duplications (Schiml et al., 2016). As POLQ is the central player not only of aNHEJ (Mateos-Gomez et al., 2015) but also of T-DNA integration (van Kregten et al., 2016; Nishizawa-Yokoi et al., 2021), we were eager to test whether POLQ is also required for tandem duplication formation. Since the Arabidopsis *polq* mutant is not stably transformable by Agrobacterium via floral dip (van Kregten et al., 2016; Nishizawa-Yokoi et al., 2021), little information of its molecular role in DSB repair could previously be obtained. Here, we overcame this challenge using POLQ overexpression. Together with the T-DNA carrying the expression cassettes for Cas9 nickase and the sgRNAs, we transiently overexpressed POLQ on a second "helper" T-DNA to achieve integration. Although the procedure was far from efficient, we were able to analyze a larger number of repair events in the progeny from one T1 plant that carried the T-DNA with a functional nickase but no "helper" T-DNA. Hence, although the number of plants and the number of reads obtained is smaller than used for all other mutants, our results demonstrate that the formation of tandem duplications is not dependent on POLQ. In contrast, patch insertions were absent in the *polq* mutant background as well as in the *xrcc1* background indicating that aNHEJ is required for their formation. This is in line with various studies from different genetic systems, which show that templated insertions are a POLQ-mediated end joining specific signature (reviewed in Schimmel et al., 2019).

Thus, we could demonstrate that the two different classes of insertions arising during the repair of paired 5′-overhang SSBs depend on the action of two different pathways of NHEJ: Whereas the cNHEJ pathway is required for tandem insertions, the aNHEJ is essential for patch insertion formation. These results further suggest that tandem duplications are generally formed by a microhomology-independent pathway, as factors of cNHEJ but not aNHEJ, which is typically involved in microhomology-dependent repair, are required for tandem duplication formation.

Previously, we were also able to demonstrate that SSBs efficiently induce HR between intrachromosomal repeats (Fauser et al., 2014; Steinert et al., 2015). However, surprisingly, when applying the Cas9 nickase for gene targeting we found that SSBs are by far less efficient than DSBs in promoting HR reaction between a chromosomal DSB and extrachromosomal piece of DNA (Wolter et al., 2018). The mechanism of DSB-induced gene targeting in plant cells was elucidated long ago (Puchta, 1998). It is described best by the synthesis dependent strand annealing mechanism of HR (Huang and Puchta, 2019). We were able to demonstrate this before using the same experimental setup as in this study to show that DSB repair by HR induced gene conversion depends on the classical factors catalyzing strand exchange in the SDSA reaction: Knocking out RAD54 leads to a strong reduction of HR efficiency (Roth et al., 2012). The fact that we were not able to achieve this reduction using the same reporter line in an identical genetic background clearly demonstrates that SSB repair by HR is not accomplished via a mechanism involving strand exchange, at least in the context of local direct or inverted repeats. It is important to mention that SSB induction using those reporter lines in combination with 4-MUG in WT results in increased HR frequencies compared to DSB induction (Fauser et al., 2014). This shows that the different outcomes between DSB induction in the *rad54* mutant background from our previous experiments and SSB induction described here cannot be attributed to a reduced GUS restoration and thus more mutagenic NHEJ for a SSB (Roth et al., 2012).

Our findings are in agreement with studies in mammals. (Davis and Maizels, 2014) postulated an alternative pathway for repair of DNA nicks distinct from canonical double-strand break repair. This pathway is independent of RAD51 as well as BRCA2, and was dependent on the donor molecule. In fact, it was even inhibited by those two factors.

Therefore, we can conclude that SSBs are not processed into DSBs followed by HR repair but rather that the nick itself stimulates HR. Further research will help to identify the factors that are essential in SSB-induced gene conversion in eukaryotes.

Materials and methods

Plant growth and transformation

All *A. thaliana* lines used in this study were in the Columbia-0 (Col-0) background. All mutants were obtained as T-DNA insertion lines from the SALK or GABI collection, respectively (Supplemental Table S9). The *rad54-1* lines were crossed into the GUS reporter lines stably carrying the reporter constructs DGU.US-1 and IU.GUS-8 characterized in (Roth et al., 2012). Stable transformation of Arabidopsis was performed according to the floral dip method (Clough and Bent, 1998).

Seeds were sown on agar plates containing germination medium (GM: 4.9-g L1 Murashige and Skoog medium, 10-g L1 saccharose, pH 5.7, 7.6-g L1 plant-agar) or on substrate containing 1:1 Floraton 3 (Floragard Vertriebs GmbH, https://www.floragard.de/) and vermiculite (Deutsche Vermiculite Dämmstoff GmbH, http://www.vermiculite.de/), with 16-h light (Phillips, Master, TL-D 36W/840) and 8-h darkness at 22°C.

T-DNA constructs

The CRISPR/Cas9 nickase and nuclease constructs used in this study are based on the Gateway compatible pEn-Chimera and pDe-Cas9(-D10A) previously described (Fauser et al., 2014). The plant resistance cassette (*npt*II) of the destination vector was changed to PPT or gentamycin via *Hind*III, if the insertion lines carried T-DNAs with kanamycin resistance. CRISPR target sequences were inserted as annealed oligonucleotides into pEn-Chimera (see Supplemental Tables S10 and S11 for oligonucleotides used in this study). The programmed sgRNA expression cassettes were transferred into pDe-Cas9-D10A-Gent via Gateway LR reaction. The constructs for the paired nickase approach were generated as described before (Schiml et al., 2014).

DNA preparation from plant tissue

Two different protocols were used for DNA extraction from plant tissue. For routine applications a time optimized protocol (1) was used, for NGS analysis a quantity and purity optimized (2) protocol was used.

(a) Rosette leaves or whole plantlets were homogenized with a pestle. Five hundred microliters Shorty extraction buffer (200-mM Tris–HCl, 200-mM LiCl, 25-mM EDTA, 1% sodium dodecyl sulfate (SDS)) were added followed by centrifugation at 12,000*g* for 5 min. Three hundred microliters supernatant were transferred to 300 µL isopropanol and gently mixed for the precipitation of DNA. The precipitated DNA was pelleted by centrifugation at 13,500*g* for 10 min. The supernatant was discarded and the pellet washed once with 70% ethanol. The dried pellet was resuspended in 100–200-µL TE buffer (10-mM Tris, 1-mM EDTA) depending on the quantity of plant material used.

(b) About 15–35 plantlets 2–3 weeks post germination were ground below freezing temperature with mortar and pestle pre-cooled in liquid nitrogen. Subsequently, 5-mL isolation buffer (0.1-mg·mL ^{1} sodium disulfite, 4.2-mL DNA extraction buffer (63.76-g·L ^{1} sorbitol, 12.11-g·L ^{1} Tris, 1.86-g·L ^{1} EDTA, pH 7.5), 4.2-mL nuclei lysis buffer (24.22-g·L ^{1} Tris, 18.61-g·L ^{1} EDTA, 116.88-g·L ^{1} NaCl, 20-g·L ^{1} cetyl trimethylammonium bromide, pH 8.0), 800-mL 5% sarkosyl solution (50 g·L ^{1} N-lauroylsarcosin) was added. After thorough mixing, the sample was incubated at 65°C for 1 h. Protein separation was achieved by one round of chloroform extraction using 5 mL of chloroform and centrifugation at 5,000*g* and 4°C for 5 min. RNA was degraded by adding 100-µL RNase to the supernatant and incubating at room temperature for 20 min. The DNA was then precipitated by adding 5 mL of pre-cooled isopropanol (−80°C), gentle mixing by inversion, and centrifuged at 5,000*g* and 4°C for 10 min. The pellet was washed once with 70% ethanol, dried, and resuspended in 200–500-µL TE buffer depending on the amount of plant material used.

Sequencing of DNA

In order to determine guide RNA efficiencies, the respective target regions from 10 pooled transgenic plants each, transformed with the T-DNA harboring the Cas9 nuclease programmed to target the individual target sites, were amplified and subjected to TIDE analysis (Brinkman et al., 2014). In order to determine paired nick InDel frequencies, the respective target regions from untransformed plant DNA samples, to determine the background signal generated by sequencing errors, as well as plants transformed with the T-DNA were amplified and subjected to NGS. Two different methods were used for NGS applications, the Illumina HiSeq platform (Bentley et al., 2008) and the Pacific Biosciences SMRT platform (Eid et al., 2009). For both methods, sequencing as well as library preparation was performed by Eurofins GATC (Konstanz). The Illumina HiSeq platform was used for paired end sequencing of small amplicons up to a size of 300 bp. Up to 20 different amplicons were pooled and sequenced simultaneously. By including different barcodes on the primers, the reads could be demultiplexed after sequencing (see Supplemental Tables S12 and S13 for barcodes used). The starting material for library preparation was 100 µL of PCR products at 10 ng·µL ^{1}. Due to limitations of the HiSeq platform to amplicons below 400 bp, the SMRT sequencing platform was used for the larger amplicons up to 1,000 bp. Analogous to the HiSeq platform, multiple amplicons were sequenced simultaneously in the same sequencing run by using barcodes on the primers (see Supplemental Table S14 for barcodes used).

Analysis of InDels using Illumina short-reads

Analysis of paired nick repair in the different mutant backgrounds was based on reads generated using Illumina. HiSeq reads were demultiplexed and paired-end reads were

merged using CLC Genomics Workbench (Qiagen). The merged and sorted reads were aligned to the reference sequence and further analyzed using Cas-Analyzer (Park et al., 2017). Statistical comparison of InDel frequencies between WT and mutant backgrounds was conducted based on read counts using Fisher's exact test implemented in R.

Analysis of InDels using PacBio long reads

We based the distant nick repair analysis on reads generated using PacBio SMRT sequencing long reads. Demultiplexing was carried out with sabre v0.3 (https://github.com/najoshi/sabre), allowing one mismatch. A mean of 25% of reads could not be assigned to a barcode and were discarded. BBmap v38.38 (Bushnell, 2014) was used to attempt to rescue reads with unknown barcodes. However, a mean of $<$2% of reads with unknown barcodes could be recovered. To avoid any cross contamination, the rescued reads were not concatenated with the demultiplexed reads. All reads over 300-bp longer than the locus length were excluded. Reads were mapped to locus reference sequences using minimap2 v2.15 (Li, 2018) with default PacBio parameters (-x map-bp). InDel information for each read was extracted from the alignment CIGAR strings using pysam v0.15.2 (Heger et al., 2014). Since long InDels were the focus of the analysis, small InDels $<$3 bp were excluded from all samples including the WT background sample. To further validate this filtering approach, we also applied an alternative filter that excluded InDels $<$10 bp. For both InDel length filters, we further tested the effect of filtering InDels based on three possible distances to the nick region (overlapping, 200-bp upstream/downstream, and any distance).

Analysis of β-glucuronidase activity

To quantify SSB-induced HR, conventional GUS staining was performed as previously described (Orel et al., 2003) 14 days after sowing seeds on agar plates containing germination medium (GM) with the corresponding selection marker (gentamycin) and cefotaxime.

The detailed quantification of the Cas9-nickase activity was performed using a 4-MUG assay (Weigel and Glazebrook, 2002). For each line, T1 seeds were sown on selection media containing GM and gentamycin. After 14 days of growth, for each line 200 mg of plant material of independent primary transformants (approximately 20 seedlings) was collected thrice, constituting the biological replicates, and frozen in liquid nitrogen. Without thawing, the plant material was ground to fine powder and 150 µL of GUS extraction buffer (50-mM sodium phosphate, 10-mM EDTA, 0.1% SDS, and 0.1% Triton X-100) was added. After centrifugation at 4°C, 10 µL of the supernatant was transferred to 1-mL GUS extraction buffer containing 1-mM 4-MUG. To record the enzyme kinetics, after each incubation step of 10 min at RT, 100 µL were removed from the reaction and added to 900 µL of 1-M sodium carbonate stop buffer. The fluorescence measurement was carried out in an Infinite M200 PRO multimode plate reader (Tecan Group Ltd.) with excitation at 365 nm and emission at 455 nm. All replicates of the lines in the DGU.US background showed a β-glucuronidase conversion rate of approximately 1 nmol·(min*µg)$^{-1}$. For the samples in the IU.GUS background we measured rates of around 0.12 nmol·(min*µg)$^{-1}$. GUS data are shown in Supplemental Table S15.

Measurement of protein concentrations

For normalization of the obtained conversion rates to the protein concentration in the different samples, a Lowry assay was performed (Lowry et al., 1951). For this, proteins from 50 µL of the supernatant used for the 4-MUG assay diluted in 1 mL of H_2O were precipitated by adding 2% Na-Desoxychelate and 30% TCA. The precipitated proteins were dissolved in 2-ml Lowry D (2% Na_2CO_3 in 0.1 n NaOH + 2% $CuSO_4$*$5H_2O$ + 4% K/Na tartrate, 25:1:1) and 5% SDS as well as the Folin–Ciocalteu reagent (Folin/H_2O 1:2) were added under vortexing. After 60 min incubation in the dark, the protein complexes were measured at 650 nm. The protein concentrations were calculated using a BSA standard curve.

Accession numbers

Sequence data of the genes used in this article can be found at The Arabidopsis Information Resource with the accession numbers listed in Supplemental Table S9. Deep sequencing data that support the finding of this study have been deposited in SRA (accession: PRJNA727456). Custom scripts for data analysis of the PacBio reads are available on GitHub (https://github.com/ascheben/CutAnalyser/tree/master/pbindels). Statistical data are shown in Supplemental Table S16.

Supplemental data

The following materials are available in the online version of this article

Supplemental Figure S1. Deletion and insertion frequencies in different mutant backgrounds at the heterochromatin locus.

Supplemental Figure S2. Distribution of deletions and insertions in different mutant backgrounds at the heterochromatin locus.

Supplemental Figure S3. Deletion and insertion frequencies in *lig4* and *ku70* at the ADH1 locus.

Supplemental Figure S4. Distribution of insertion types in *lig4* and *ku70* at the heterochromatin locus.

Supplemental Figure S5. Distribution of insertion types in *lig4* and *ku70* at the *ADH1* locus.

Supplemental Figure S6. Representative illustration of the most frequent insertions and deletions in all different mutant backgrounds at the heterochromatin locus.

Supplemental Figure S7. Vector diagram.

Supplemental Table S1. Influence of nick distances on paired nickase efficiency.

Supplemental Table S2. Read numbers obtained by SMRT sequencing.

Supplemental Table S3. PacBio read counts and InDel rates using four different InDel-based filters.

Supplemental Table S4. Summary statistics of InDels based on PacBio long reads with InDels.

Supplemental Table S5. Data of deletion and insertion frequencies in different mutant backgrounds at the heterochromatin locus.

Supplemental Table S6. Insertion-type analysis of the heterochromatin locus.

Supplemental Table S7. Data of deletion and insertion frequencies in different mutant backgrounds at the *ADH1* locus.

Supplemental Table S8. Insertion-type analysis of the *ADH1* locus.

Supplemental Table S9. T-DNA insertion lines used in this study.

Supplemental Table S10. Oligonucleotides used in the paired nickase approach.

Supplemental Table S11. Oligonucleotides used for SSB induction in HR reporters.

Supplemental Table S12. Oligonucleotides used for *ADH1* locus NGS Illumina sequencing.

Supplemental Table S13. Oligonucleotides used for heterochromatin locus NGS Illumina sequencing.

Supplemental Table S14. Oligonucleotides used for NGS SMRT sequencing.

Supplemental Table S15. GUS data and *t* test results.

Supplemental Table S16. Results of Fisher's exact test for comparison of baseline and mutant read counts shown in Figures 3, 5, 6, and 7.

Supplemental Data Set S1. Illumina sequence reads evaluated by Cas-Analyzer.

Acknowledgments

The authors thank Luisa Telpl, Sakia Gabsi, Aruscha Baumanns and Waltraud Wehrle for excellent technical assistance and Amy Whitbread for critical reading of the article.

Funding

This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG; PU 137/19-1).

Conflict of interest statement. None declared.

References

Abbotts R, Wilson DM (2017) Coordination of DNA single strand break repair. Free Radical Biol Med **107:** 228 244

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature **576**: 149 157

Babiychuk E, Cottrill PB, Storozhenko S, Fuangthong M, Chen Y, O'Farrell MK, van Montagu M, Inzé D, Kushnir S (1998) Higher plants possess two structurally different poly(ADP-ribose) polymerases. Plant J **15**: 635 645

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature **456**: 53 59

Beying N, Schmidt C, Pacher M, Houben A, Puchta H (2020) CRISPR Cas9-mediated induction of heritable chromosomal translocations in Arabidopsis. Nat. Plants **19:** 778

Bothmer A, Phadke T, Barrera LA, Margulies CM, Lee CS, Buquicchio F, Moss S, Abdulkerim HS, Selleck W, Jayaram H, et al. (2017). Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. Nat Commun **8:** 13905

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

Bushnell B (2014) BBMap: A Fast, Accurate, Splice-Aware Aligner. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA

Cao H, Salazar-García L, Gao F, Wahlestedt T, Wu C-L, Han X, Cai Y, Xu D, Wang F, Tang L, et al. (2019) Novel approach reveals genomic landscapes of single-strand DNA breaks with nucleotide resolution in human cells. Nat Commun **10**: 5799

Charbonnel C, Gallego ME, White CI (2010) Xrcc1-dependent and Ku-dependent DNA double-strand break repair kinetics in Arabidopsis plants. Plant J **64**: 280 290

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

Davis L, Maizels N (2014) Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. Proc Natl Acad Sci USA **111**: E924 932

Derboven E, Ekker H, Kusenda B, Bulankova P, Riha K (2014) Role of STN1 and DNA polymerase α in telomere stability and genome-wide replication in Arabidopsis. PLoS Genet **10**: e1004682

Dubest S, Gallego ME, White CI (2004) Roles of the AtErcc1 protein in recombination. Plant J **39**: 334 342

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. (2009). Real-time DNA sequencing from single polymerase molecules. Science (New York, N.Y.) **323**: 133 138

Fauser F, Schiml S, Puchta H (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. Plant J **79**: 348 359

Friesner J, Britt AB (2003) Ku80- and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. Plant J **34**: 427 440

Geuting V, Kobbe D, Hartung F, Dürr J, Focke M, Puchta H (2009) Two distinct MUS81-EME1 complexes from Arabidopsis process Holliday junctions. Plant Physiol **150**: 1062 1071

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

Hartung F, Suer S, Bergmann T, Puchta H (2006) The role of AtMUS81 in DNA repair and its genetic interaction with the helicase AtRecQ4A. Nucleic Acids Res **34**: 4438 4448

Hartung F, Suer S, Puchta H (2007) Two closely related RecQ helicases have antagonistic roles in homologous recombination and DNA repair in Arabidopsis thaliana. Proc Natl Acad Sci USA **104**: 18836 18841

Hefner E, Preuss SB, Britt AB (2003) Arabidopsis mutants sensitive to gamma radiation include the homologue of the human repair gene ERCC1. J Exp Bot **54**: 669 680

Heger A, Belgrad TG, Goodson M, Jacobs K (2014) Pysam: Python interface for the SAM/BAM sequence alignment and mapping format. https://github.com/pysam-developers/pysam

Huang T-K, Puchta H (2019) CRISPR/Cas-mediated gene targeting in plants: finally a turn for the better for homologous recombination. Plant Cell Rep **38**: 443 453

Huefner ND, Mizuno Y, Weil CF, Korf I, Britt AB (2011) Breadth by depth: expanding our understanding of the repair of transposon-induced DNA double strand breaks via deep-sequencing. DNA Repair **10**: 1023 1033

Inagaki S, Suzuki T, Ohto M-A, Urawa H, Horiuchi T, Nakamura K, Morikami A (2006) Arabidopsis TEBICHI, with helicase and

DNA polymerase domains, is required for regulated cell division and differentiation in meristems. Plant Cell **18**: 879 892

Jasin M, Haber JE (2016) The democratization of gene editing: insights from site-specific cleavage and double-strand break repair. DNA Repair **44:** 6 16

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

Katz SS, Gimble FS, Storici F (2014) To nick or not to nick: comparison of I-SceI single- and double-strand break-induced recombination in yeast and human cells. PLoS One **9**: e88840

Kazda A, Zellinger B, Rössler M, Derboven E, Kusenda B, Riha K (2012) Chromosome end protection by blunt-ended telomeres. Genes Dev **26**: 1703 1713

Kim E, Kim S, Kim DH, Choi B-S, Choi I-Y, Kim J-S (2012) Precision genome engineering with programmable DNA-nicking enzymes. Genome Res **22**: 1327 1333

Klemm T, Mannuß A, Kobbe D, Knoll A, Trapp O, Dorn A, Puchta H (2017) The DNA translocase RAD5A acts independently of the other main DNA repair pathways, and requires both its ATPase and RING domain for activity in Arabidopsis thaliana. Plant J **91**: 725 740

Kobbe S, Trapp O, Knoll A, Manuss A, Puchta H (2015) The translesion polymerase ζ has roles dependent on and independent of the nuclease MUS81 and the helicase RECQ4A in DNA damage repair in Arabidopsis. Plant Physiol **169**: 2718 2729

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

Krasner DS, Daley JM, Sung P, Niu H (2015) Interplay between Ku and replication protein A in the restriction of Exo1-mediated DNA break end resection. J Biol Chem **290**: 18806 18816

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics (Oxford, England) **34**: 3094 3100

Li W, Chen C, Markmann-Mulisch U, Timofejeva L, Schmelzer E, Ma H, Reiss B (2004) The Arabidopsis AtRAD51 gene is dispensable for vegetative development but required for meiosis. Proc Natl Acad Sci USA **101**: 10596 10601

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the Folin phenol reagent. J Biol Chem **193**: 265 275

Maizels N, Davis L (2018) Initiation of homologous recombination at DNA nicks. Nucleic Acids Res **46**: 6962 6973

Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol **31**: 833 838

Mannuss A, Dukowic-Schulze S, Suer S, Hartung F, Pacher M, Puchta H (2010) RAD5A, RECQ4A, and MUS81 have specific functions in homologous recombination and define different pathways of DNA repair in Arabidopsis thaliana. Plant Cell **22**: 3318 3330

Mateos-Gomez PA, Gong F, Nair N, Miller KM, Lazzerini-Denchi E, Sfeir A (2015) Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. Nature **518**: 254 257

Metzger MJ, Stoddard BL, Monnat RJ (2013) PARP-mediated repair, homologous recombination, and back-up non-homologous end joining-like repair of single-strand nicks. DNA Repair **12**: 529 534

Mikami M, Toki S, Endo M (2016) Precision targeted mutagenesis via Cas9 paired nickases in rice. PlantCell Physiol **57**: 1058 1068

Molinier J (2017) Genome and epigenome surveillance processes underlying UV exposure in plants. Genes **8**:

Nakajima K, Zhou Y, Tomita A, Hirade Y, Gurumurthy CB, Nakada S (2018) Precise and efficient nucleotide substitution near genomic nick via noncanonical homology-directed repair. Genome Res **28**: 223 230

Nishizawa-Yokoi A, Saika H, Hara N, Lee L-Y, Toki S, Gelvin SB (2021) Agrobacterium T-DNA integration in somatic cells does not require the activity of DNA polymerase θ. New Phytol **229**: 2859 2872

Orel N, Kyryk A, Puchta H (2003) Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. Plant J **35**: 604 612

Osakabe K, Abe K, Yoshioka T, Osakabe Y, Todoriki S, Ichikawa H, Hohn B, Toki S (2006) Isolation and characterization of the RAD54 gene from Arabidopsis thaliana. Plant J **48**: 827 842

Pacher M, Schmidt-Puchta W, Puchta H (2007) Two unlinked double-strand breaks can induce reciprocal exchanges in plant genomes via homologous recombination and nonhomologous end joining. Genetics **175**: 21 29

Park J, Lim K, Kim J-S, Bae S (2017) Cas-analyzer: an online tool for assessing genome editing results using NGS data. Bioinformatics (Oxford, England) **33**: 286 288

Puchta H (1998) Repair of genomic double-strand breaks in somatic plant cells by one-sided invasion of homologous sequences. Plant J **13**: 331 339

Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot **56**: 1 14

Puchta H (2016) Breaking DNA in plants: how I almost missed my personal breakthrough. Plant Biotechnol J **14**: 437 440

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

Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB (2017) Engineering quantitative trait variation for crop improvement by genome editing. Cell **171**: 470 480.e8

Roldán-Arjona T, Ariza RR, Córdoba-Cañero D (2019) DNA base excision repair in plants: an unfolding story with familiar and novel characters. Front Plant Sci **10:** 1055

Roth N, Klimesch J, Dukowic-Schulze S, Pacher M, Mannuss A, Puchta H (2012) The requirement for recombination factors differs considerably between different pathways of homologous double-strand break repair in somatic plant cells. Plant J **72**: 781 790

Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO J **17**: 6086 6095

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

Schiml S, Fauser F, Puchta H (2016) Repair of adjacent single-strand breaks is often accompanied by the formation of tandem sequence duplications in plant genomes. Proc Natl Acad Sci USA **113**: 7266 7271

Schimmel J, van Schendel R, den Dunnen JT, Tijsterman M (2019) Templated insertions: a smoking gun for polymerase theta-mediated end joining. Trends Genet **35**: 632 644

Schmidt C, Pacher M, Puchta H (2019a) DNA break repair in plants and its application for genome engineering. Methods Mol Biol (Clifton, N.J.) **1864:** 237 266

Schmidt C, Pacher M, Puchta H (2019b) Efficient induction of heritable inversions in plant genomes using the CRISPR/Cas system. Plant J **98**: 577 589

Schröpfer S, Kobbe D, Hartung F, Knoll A, Puchta H (2014) Defining the roles of the N-terminal region and the helicase activity of RECQ4A in DNA repair and homologous recombination in Arabidopsis. Nucleic Acids Res **42**: 1684 1697

Shibata A (2017) Regulation of repair pathway choice at two-ended DNA double-strand breaks. Mutat Res **803–805:** 51 55

Siebert R, Puchta H (2002) Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. Plant Cell **14**: 1121 1131

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**: 1295 1305

Tamura K, Adachi Y, Chiba K, Oguchi K, Takahashi H (2002) Identification of Ku70 and Ku80 homologues in Arabidopsis thaliana: evidence for a role in the repair of DNA double-strand breaks. Plant J **29**: 771 781

Tavares EM, Wright WD, Heyer W-D, Le Cam E, Dupaigne P (2019) In vitro role of Rad54 in Rad51-ssDNA filament-dependent homology search and synaptic complexes formation. Nat Commun **10**: 4058

van Kregten M, de Pater S, Romeijn R, van Schendel R, Hooykaas PJJ, Tijsterman M (2016) T-DNA integration in plants results from polymerase-θ-mediated DNA repair. Nat Plants **2**: 16164

van Schendel R, Tijsterman M (2013) Microhomology-mediated intron loss during metazoan evolution. Genome Biol Evol **5**: 1212 1219

Vaughn JN, Bennetzen JL (2014) Natural insertions in rice commonly form tandem duplications indicative of patch-mediated double-strand break induction and repair. Proc Natl Acad Sci USA **111**: 6684 6689

Vriend LEM, Prakash R, Chen C-C, Vanoli F, Cavallo F, Zhang Y, Jasin M, Krawczyk PM (2016) Distinct genetic control of homologous recombination repair of Cas9-induced double-strand breaks, nicks and paired nicks. Nucleic Acids Res **44**: 5204 5217

Weigel D, Glazebrook J (2002) Arabidopsis: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

West CE, Waterworth WM, Jiang Q, Bray CM (2000) Arabidopsis DNA ligase IV is induced by gamma-irradiation and interacts with an Arabidopsis homologue of the double strand break repair protein XRCC4. Plant J **24**: 67 78

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**: 735 746

Wolter F, Puchta H (2018) Application of CRISPR/Cas to understand cis- and trans-regulatory elements in plants. Methods Mol Biol **1830:** 23 40

Wu H, Wang Y, Zhang Y, Yang M, Lv J, Liu J, Zhang Y (2015) TALE nickase-mediated SP110 knockin endows cattle with increased resistance to tuberculosis. Proc Natl Acad Sci USA **112**: E1530 1539

Wu L (2007) Role of the BLM helicase in replication fork management. DNA Repair **6**: 936 944

Yuan Y, Britton S, Delteil C, Coates J, Jackson SP, Barboule N, Frit P, Calsou P (2015) Single-stranded DNA oligomers stimulate error-prone alternative repair of DNA double-strand breaks through hijacking Ku protein. Nucleic Acids Res **43**: 10264 10276