RESEARCH ARTICLE

American Society of Plant Biologists | SEB SOCIETY FOR EXPERIMENTAL BIOLOGY | WILEY

Context effects on repair of 5′-overhang DNA double-strand breaks induced by Cas12a in Arabidopsis

Sébastien Lageix[1] | Miguel Hernandez[1] | Maria E. Gallego[1] | Jérémy Verbeke[1] | Yannick Bidet[2] | Sandrine Viala[2] | Charles I. White[1]

[1]Institut de Génétique, Reproduction et Développement, UMR CNRS 6293 - INSERM U1103 - Université Clermont Auvergne, Bât. CRBC, Faculté de Médecine, Clermont-Ferrand Cedex 1, France

[2]Laboratoire d'Oncologie Moléculaire, U1240 Imagerie Moléculaire et Stratégies Théranostiques, Université Clermont Auvergne, Centre Jean Perrin, INSERM, Clermont-Ferrand, France

Correspondence
Charles I. White, Institut de Génétique, Reproduction et Développement, UMR CNRS 6293 - INSERM U1103 - Université Clermont Auvergne, Bât. CRBC, Faculté de Médecine, 28 place Henri Dunant, 63001 Clermont-Ferrand Cedex 1, France.
Email: charles.white@cnrs.fr

Present address
Miguel Hernandez, Centro de Biología Molecular Severo Ochoa, Madrid, Spain.

Abstract

Sequence-specific endonucleases have been key to the study of the mechanisms and control of DNA double-strand break (DSB) repair and recombination, and the availability of CRISPR-Cas nucleases over the last decade has driven rapid progress in the understanding and application of targeted recombination in many organisms, including plants. We present here an analysis of recombination at targeted chromosomal 5′ overhang DSB generated by the FnCas12a endonuclease in the plant, *Arabidopsis thaliana*. The much-studied Cas9 nuclease cleaves DNA to generate blunt-ended DSBs, but relatively less is known about the repair of other types of breaks, such as those with 5′-overhanging ends. Sequencing the repaired breaks clearly shows that the majority of repaired DSB carry small deletions and are thus repaired locally by end-joining recombination, confirmed by Nanopore sequencing of larger amplicons. Paired DSBs generate deletions at one or both cut-sites, as well as deletions and reinsertions of the deleted segment between the two cuts, visible as inversions. While differences are seen in the details, overall the deletion patterns are similar between repair at single-cut and double-cut events, notwithstanding the fact that only the former involve cohesive DNA overhangs. A strikingly different repair pattern is however observed at breaks flanked by direct repeats. This change in sequence context results in the presence of a major alternative class of repair events, corresponding to highly efficient repair by single-strand annealing recombination.

KEYWORDS
Arabidopsis, Cas12a, DNA repair, DSB repair, recombination

1 | INTRODUCTION

DNA double-strand breaks (DSBs) are highly toxic lesions, causing both local damage and separating the chromatid arm distal to the break from its centromere. This acentric chromatid can be subsequently lost, leading to loss of many genes. Alternatively, either of the broken DNA ends can recombine with other sequences in the genome, to generate major structural genome rearrangements (dicentric/acentric chromosomes, deletions, translocations, inversions). A number of recombination pathways repair DSB very efficiently in living cells, and the specific outcome of a given repair event is determined by the recombination mechanism involved in its

repair. Thus, in somatic cells of animals and plants, chromosome breaks are predominantly repaired by DNA end-joining (EJ) recombination, which, while it frequently results in mutations at the repaired break-site, avoids structural genome damage by restricting repair to sequences immediately flanking the break. Readers are referred to excellent reviews on this subject (Chang et al., 2017; Kamoen et al., 2024; Knoll et al., 2014; Scully et al., 2019; Wright et al., 2018).

The possibility to specifically induce in vivo breakage of a chromosomal target sequence has been key to understanding of DSB repair and its consequences and applications. Starting with the use of natural and "designer" endonucleases to cleave specific chromosomal targets, this work has been tremendously accelerated in recent years by the availability of CRISPR/Cas endonucleases. Following on from earlier work using transgenic tester loci, zinc-finger and transcription activator-like effector (TALE) nucleases, the application of CRISPR/Cas tools to plant models over the last decade has enabled exciting new experimental approaches and resulted in major advances in the understanding and application of DNA repair and recombination in model and crop plants (Atia et al., 2024; Fauser et al., 2014; Gao, 2021; Jiang et al., 2013; Kamoen et al., 2024; Li et al., 2013; Nekrasov et al., 2013; Přibylová & Fischer, 2024; Schiml et al., 2016; Shan et al., 2013; Vu et al., 2017; Wolter et al., 2021; Xue & Greene, 2021).

The majority of studies of DSB-induced DNA repair/recombination in plants in recent years have involved the Cas9 endonuclease and its derivatives. Sequentially cleaving the two strands of the target DNA, Cas9 generates a blunt-ended or 1-base overhang DSB (Jinek et al., 2012; Kumar et al., 2023; Longo et al., 2024; Swarts & Jinek, 2018; Weiss et al., 2024). This means that most studies of Cas9-induced breaks involve the study of blunt-ended DSB, without single-stranded DNA (ssDNA) overhangs. The mutant Cas9-nickase protein, which cleaves only one strand, is an exception to this and has notably been used in studies of recombination at paired single-strand nicks (e.g., Bothmer et al., 2017; Fauser et al., 2014; Schiml et al., 2016, 2014). In contrast, the Cas12a (Cpf1) endonuclease produces a staggered break, with 4–5 base 5′ ssDNA overhangs at the ends (Swarts & Jinek, 2018; Zetsche et al., 2015). Furthermore, Cas12a has technical advantages, which facilitate its use in studies of DSB repair, notably that it depends upon an easily multiplexable, single guide RNA (gRNA) for targeting cleavage (Fonfara et al., 2016; Zetsche et al., 2017).

In this work, we present an analysis of the patterns of DSB repair recombination following cleavage at paired genomic target sites by FnCas12a in the plant, *Arabidopsis thaliana*. The majority of repaired DSB are repaired locally by EJ recombination. While differences are seen in the details, overall, the deletion patterns are not greatly affected by differing local sequence contexts flanking the DSB. The presence of flanking direct repeat sequences does however strongly impact the repair pattern—corresponding to highly efficient repair by single-strand annealing (SSA) recombination.

2 | MATERIALS AND METHODS

2.1 | Genomic target loci

To investigate endogenous repair profiles of CRISPR/Cas12a-induced DSBs, we screened the *A. thaliana* genome sequence and selected pairs of target sites at two distinct genomic loci. The first target locus, located in the middle of the left arm of chromosome 1, corresponds to the gene *LPP2* (At1G15080). gRNAs were designed to induce two DSBs in the first intron and second exon, respectively (see Figure 1a). The second target locus, Chr3_92, was identified by scanning the TAIR10 Arabidopsis genome sequence with the Tandem Repeat Finder (trf version 4.09; https://tandem.bu.edu/trf/trf.html) (Benson, 1999). Chr3_92 lies in the repeat-rich, pericentromeric region of Chromosome 3 and consists of two perfect 142 bp tandem direct repeats separated by 742 nt. The 142 bp repeats are themselves part of two degenerate 428 bp tandem repeats. Two unique Cas12a targets were chosen in the sequences between the two repeats. The Cas12a target sites in this locus are thus flanked by tandem direct repeat DNA sequences (see Figure 5a).

2.2 | T-DNA constructs and plant transformation

Recombinant CRISPR/Cas9 and sgRNA constructs (Table S1) were designed for unique genomic positions and cloned into the expression plasmids kindly provided by Dr Seiichi Toki (Endo et al., 2016), following the methods described in their publication. Floral-dip T-DNA transformation of *A. thaliana* plants of ecotype Columbia-0 by *Agrobacterium tumefaciens* strain C58C1 was carried out following standard methods (Bechtold et al., 1993; Clough & Bent, 1998).

2.3 | Plant selection

In order to obtain transgenic plant lines to address specifically each of the two selected targets, transformed T1 plants for each construct were grown and selected on Murashige and Skoog (MS) medium (4.9 g·L^{-1} MS medium, 10 g·L^{-1} sucrose, and 8 g·L^{-1}agar, pH 5.7) with kanamycin (50 μg·mL^{-1}) in a growth chamber with a 16/8 h light/dark cycle, temperature 23°C, and 60% relative humidity. The progeny of selected T1 individuals was grown for 10 days at 29°C on MS media, genomic DNA extracted using the Promega Wizard Genomic DNA Purification Kit and double-stranded DNA (dsDNA) quantified using the Qubit flourometer (High Sensitivity kit, Thermo Fisher). To assess the cutting efficiency of the transgenic lines, a PCR was used to amplify the region spanning the two cut-sites for each target loci. The T1 lines harboring the stronger activity for each construct was selected.
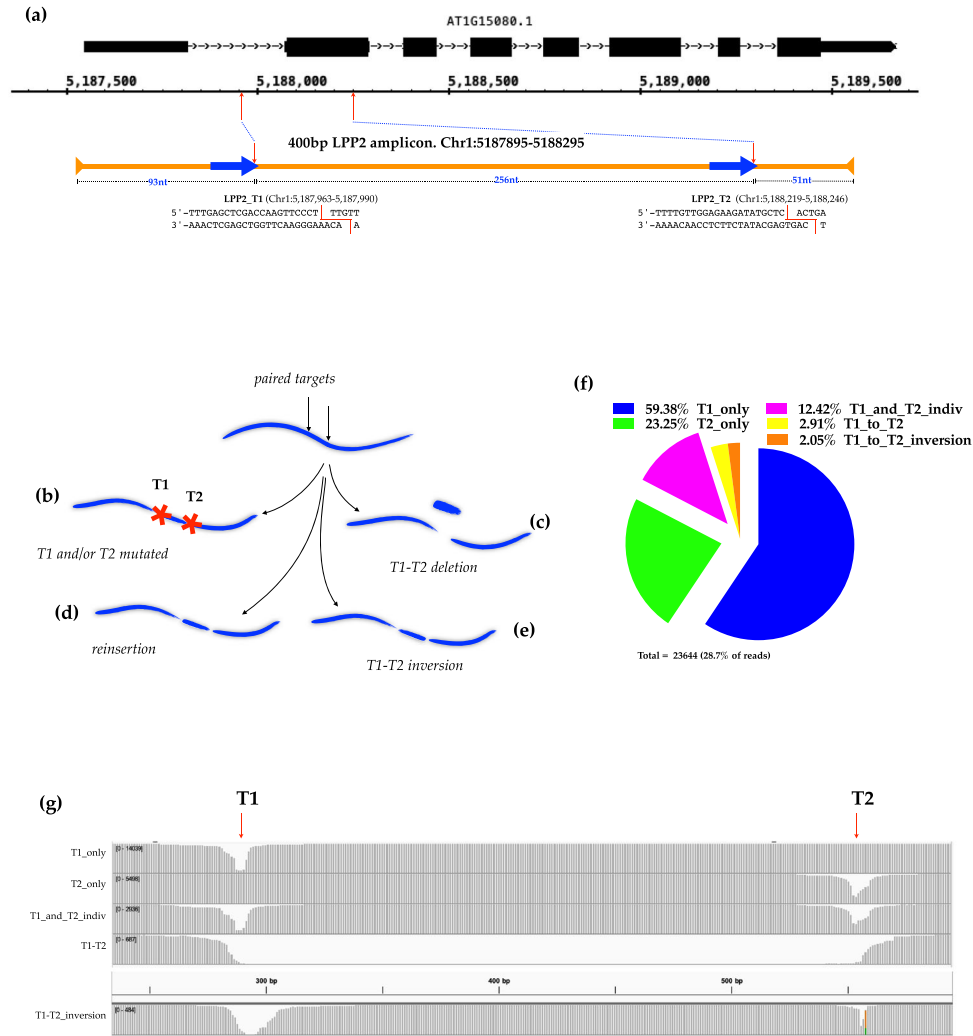
LAGEIX ET AL.

American Society
of Plant Biologists
Cultivating a better future through plant biology research

SEB
SOCIETY FOR EXPERIMENTAL BIOLOGY

WILEY

3 of 12

FIGURE 1 Patterns of repair of paired, 5′-overhang DNA breaks at the LPP2 locus. (a) A schematic showing the two Cas12a target sites of the target LPP2 locus and the 400 bp amplicon used for paired-end sequencing. The sequence, including the (TTT) protospacer adjacent motif (PAM) and the 23 nt protospacer, of the targets is shown along with staggered red lines indicating the expected cleavage of the target DNA strands. Cleavage of one or both targets is expected to result in mutation of one or both targets (b). Coincident cleavage at both sites can result in deletion of the DNA fragment between the two targets (c; T1-T2 deletion), possibly accompanied by its reinsertion in either the original (d; not distinguishable from T1 and T2 mutated class of Panel b), or inverted (e; T1-T2 inversion) orientation. (f) Sorting the paired-end sequences spanning the target locus confirmed the existence and relative proportions of these different classes of event induced in the genome of plants expressing Cas12a. (g) The validity of the sorting is confirmed by mapping the sequences of the different repair classes to the Arabidopsis genome, or to an artificial in silico inversion locus for the T1-T2_inversion class (bottom).

2.4 | Sequencing of DNA

The target regions from five pooled transgenic plants each were amplified by PCR (primers in Table S2) using Platinum SuperFi II DNA Polymerase (Invitrogen) and subjected to NGS. Two different methods were used for next-generation sequencing by Illumina MiSeq and Oxford Nanopore Technologies minION. The starting material for library preparation was 100 μL of PCR products at 10 ng·μL^{-1}. The Illumina MiSeq Nano V2 kit was used for paired-end sequencing of 400 bp amplicons. For the Nanopore Technology, library preparation was conducted using the SQK-LSK110 ligation sequencing kit following the manufacturer's protocol. Long amplicons were preferentially purified with magnetic beads using the Agencourt AMPure XP PCR purification protocol. Samples were sequenced using a minION Mk1c device using R9.4.1 (FLO-MIN106D) flow cells according to the manufacturer's protocols.

2.5 | Informatics analyses

In-house Python and R scripts using publicly available tools were written to analyze the DNA sequencing data produced in this study. The scripts and detailed instructions for their use can be found at https://github.com/chrlesw/dsbrepair.git.

Briefly, the scripts analyze NGS DNA sequence outputs from both Illumina paired-end sequencing and Oxford Nanopore

sequencing. The Nanopore sequences require no pre-processing, but prior to running the analyses, the paired-end Illumina sequences are merged into individual sequences using the BBMerge tool of the BBTools package (BBMap—Bushnell B.—sourceforge.net/projects/bbmap/Brian Bushnell [Bushnell et al., 2017]). These merged sequence files are the input for two main Python scripts, which carries out the analysis following two approaches:

1. Text-recognition approach: _dsb_sortAnal.py

Based on text searches and precise sequence lengths, this approach is best applied to the Illumina sequences due to the lower sequencing error rate compared to Nanopore sequences. The merged Illumina sequences are individually scanned for the presence/absence of the Cas12a target sequence(s) using Python text-search tools. The presence of inversions is identified by the presence of a reverse-complementary sequence taken from between the target sites. These, combined with their lengths, are used to sort and enumerate the merged sequences of different deletion classes: single site deletions, individual two-site deletions, deletions spanning the two cut-sites, and inversions of the sequences between the two cut-sites.

2. Mapping plus variant calling approach: _dsb_varAnal.py

Minimap2 (Li, 2018) is used to align the Nanopore or merged, paired-end Illumina sequences to the reference genome. Secondary or supplementary alignments are removed, and the BBMap callvariants.sh script of the BBTools package (BBMap—sourceforge.net/projects/bbmap/Brian Bushnell [Bushnell et al., 2017]) is used to produce lists of deletions, insertions, and substitutions. Shell commands then select variants affecting 10 nt windows centered on the cut-sites and sort them into lists of deletions, insertions, and substitutions. The potential implication of DNA sequence microhomologies is determined for each deletion junction, and plots showing each deletion as an arc linking its start- and end-points are generated.

3 | RESULTS

A FnCas12a construct expressing gRNAs targeting two sites, LPP2_T1 and LPP2_T2, respectively, located in the first intron and second exon of the lipid phosphate phosphatase 2 gene (LPP2, AT1G15080) was built. The key characteristic of this gene as a target for this work is that it is non-essential (Katagiri et al., 2005). The two target sites, T1 and T2, are 256 bp apart, oriented in the same direction and both are unique in the *A. thaliana* genome (Figure 1a). Stably transformed lines carrying this construct were produced by T-DNA transformation, and a line showing high levels of mutation at the target sites was chosen for further work. In parallel, a plant line carrying the same FnCas12a T-DNA construct but lacking the gRNA template was built as a −gRNA control.

Genomic DNA was extracted from 10 day old plantlets; a 400 bp region including the target sites was amplified (Figure 1a) and paired-end sequenced with an Illumina Miseq sequencer. The read pairs were merged with the BBMerge script of the BBMap package to give single sequences, and these were analyzed by an in-house Python script (see Section 2).

In principle, repair of the paired Cas12a-induced breaks can result in a number of outcomes: simple religation (not detectable) or mutation of either or both T1 or T2 (Figure 1b), deletion from T1 to T2 (Figure 1c), T1 + T2 cleavage followed by reinsertion of the cut-out fragment in either the original (Figure 1d), or inverted (Figure 1e) orientation.

A python script was written to quantify and sort the sequence reads into these classes. Based on the sequence length plus the presence/absence of the unmodified target sequences, this script gives a direct read-out of the events in each sequence. In the Cas12a + gRNA plants, 23,644 sequences with modified T1 and/or T2 were found in the 82,431 sequence reads (28.7%), while 11 were found in the 52,806 reads (.02%) from the Cas12a-gRNA controls. The proportions of the different classes of events found in the +gRNA sequences are shown in Figure 1f. Thus, of the 28.7% of sequences, which carry a deletion in T1 and/or T2, individual T1 deletions (59.38%) are 2.6-fold more frequent than individual T2 deletions (23.25%), presumably due to more efficient cleavage at the T1 target. Of the sequences with T1 and/T2 deletions, 12.42% have individual mutations at both T1 and T2, and 2.91% deletions extending from T1-T2, with a slightly lesser number (2.05%) having T1-T2 deletions associated with reinsertion of the intervening sequence in inverted orientation. It is important to keep in mind that the frequencies of the latter two classes (the T1-T2 events) will depend upon the likelihood of a given chromatid being cleaved at both sites co-incidentally and, thus, the relative and absolute frequencies of cleavage at T1 and T2. The relative frequencies of these paired-cut events would thus be expected to differ in different transformants, target pairs, or at different loci. Aligning the sorted sequence files of each class of events to the Arabidopsis genome (or an in silico T1-T2 inversion locus for the inversion sequences) confirms the efficacy of this approach and gives a first view of the nature of the deletions in each class (Figure 1g).

To take this analysis further, a second script was written to align the merged sequence pairs to the Arabidopsis genome and call variants with the callvariants.sh script of the BBMap package (see Section 2). Of 83,387 sequences mapped to the LPP2 locus from the Cas12a + gRNA line, deletions affecting one or both target sites represent 32.7%, insertions 1.0%, and substitutions 4.5% (Figure 2a). As expected, very few deletions (.02%) or insertions (.02%) were found in the 52,806 sequences from the Cas12a-gRNA control; however, numbers of substitutions affecting the target windows were similar in the sequences from the control and Cas12a + gRNA plants (3.3% vs. 4.5%). Expressing these numbers as percentages of detected variants (Figure 2b) gives 85.5% deletions, 2.7% insertions, and 11.8% substitutions in the +gRNA plantlets, and .58% deletions, .46% insertions, and 98.96% substitutions in the −gRNA controls. The majority of substitutions are thus not Cas12a targeted and presumably arise from sequencing errors, which occur at comparable frequencies in this material.

FIGURE 2   Deletions predominate in the repair of 5′-overhang DNA breaks at the LPP2 locus. (a, b) Analysis of variants affecting 10 bp windows centered on the cut-sites confirms the predominance of deletions in the repair products of Cas12a-induced breaks at the LPP2 locus. DNA from the −gRNA control plants contains almost exclusively base substitutions, presumably resulting from sequencing errors. (c) Plotting 1000 randomly chosen deletions as arcs (arcs from start to end of deletion) across the target locus (X-axis, horizontal orange bar) shows that the deletions are almost exclusively focussed locally at the cut-sites or span between them. This is confirmed by plotting the distributions of the numbers of deletions (d) and insertions (e) by their lengths (blue curves are +gRNA and red are −gRNA controls). Vertical orange lines in (c) mark the target sites.
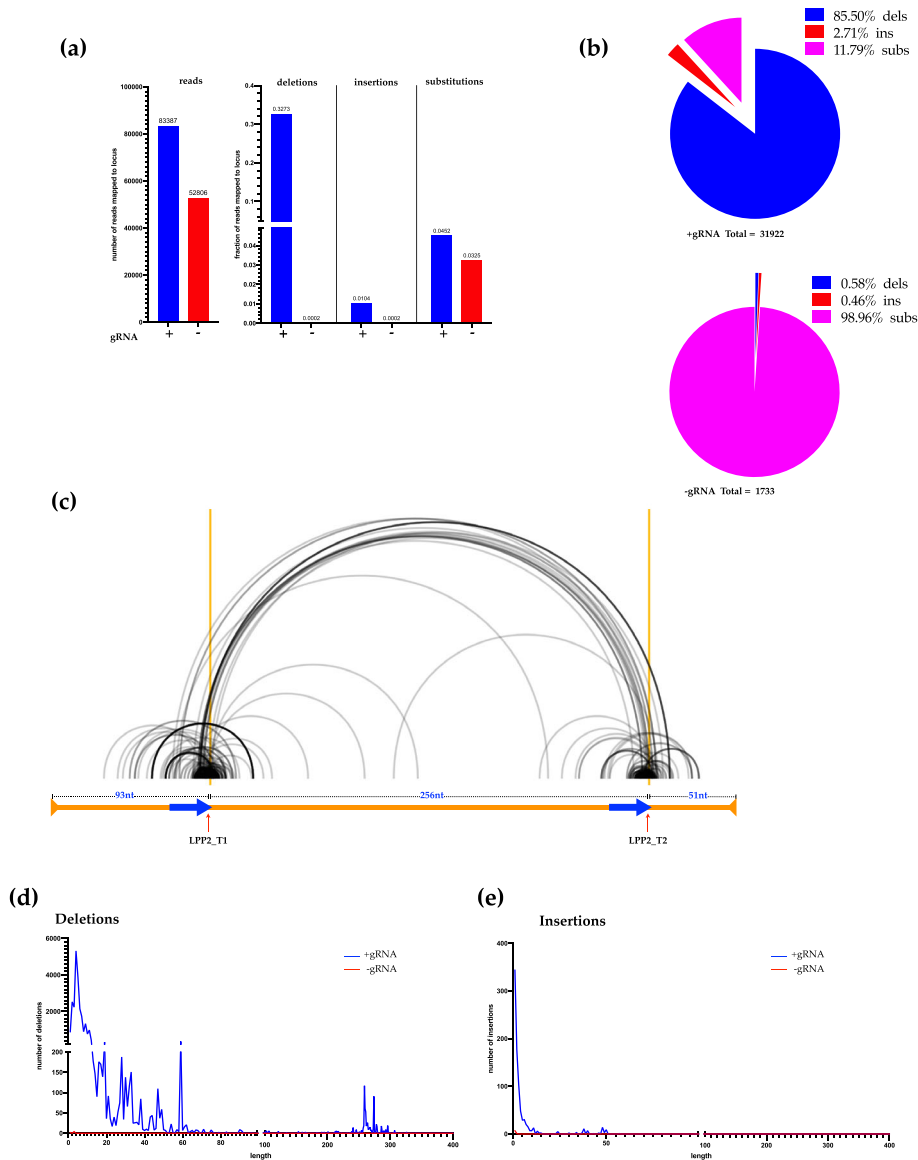
The great majority of mutations induced by Cas12a at this locus are deletions and are mostly short deletions. A clear, visual confirmation of this is shown in Figure 2c, which shows 1000 randomly chosen deletions, each plotted as an arc joining its start- and end-points. Both the clusters of small deletions at each cut-site and the larger deletions involving both cut-sites are clearly visible (black arcs). It is particularly striking to note the relatively small number of other deletions extending into the flanking chromosomal DNA (Figure 2c). Frequency distributions of deletion lengths confirm this observation, showing clearly that, with the exception of the T1-T2 deletions from 256 to 300 bp, the great majority are less than 20 bp, with only rare events being longer than 60 bp (Figure 2d). A similar analysis shows that practically, the totality of insertions is 5 bp or less, with a small minority extending to 10 bp (Figure 2e).

The FnCas12a endonuclease cleaves dsDNA to leave 4–5 nt, 5′ ssDNA overhangs (Zetsche et al., 2015), and we expected to find a major class of 4–5 nt deletions corresponding to loss of the ssDNA overhangs at the cut-sites, followed by religation. Frequency distributions of 4 and 5 nt deletion start- and end-sites confirm the presence of this class of "blunting/ligation" non-homologous end-joining (NHEJ) events (Figure 3a). This is particularly clearly visible at the T1 site, where practically the totality of 4- and 5-nt deletions is at the expected position. It is also seen at the T2 site; however, deletions starting 2 and 3 nt upstream of the expected peak are also clearly visible. Should these upstream deletion start-sites result from extended erosion of the upstream sequence prior to religation, this would result in deletions longer than the 4–5 bp and their exclusion from this analysis. We can thus confidently class these events as resulting from other cleavage sites upstream of the "primary" cut-site.

Each Cas12a cut produces complementary and potentially religatable overhanging 5′ ssDNA ends; however, this is not so for the double-cut events (deletions ≥256 bp), which involve joining of DNA ends from cuts in two different sequences. This distinction does not appear to significantly affect the loss of sequence in the deletions, as
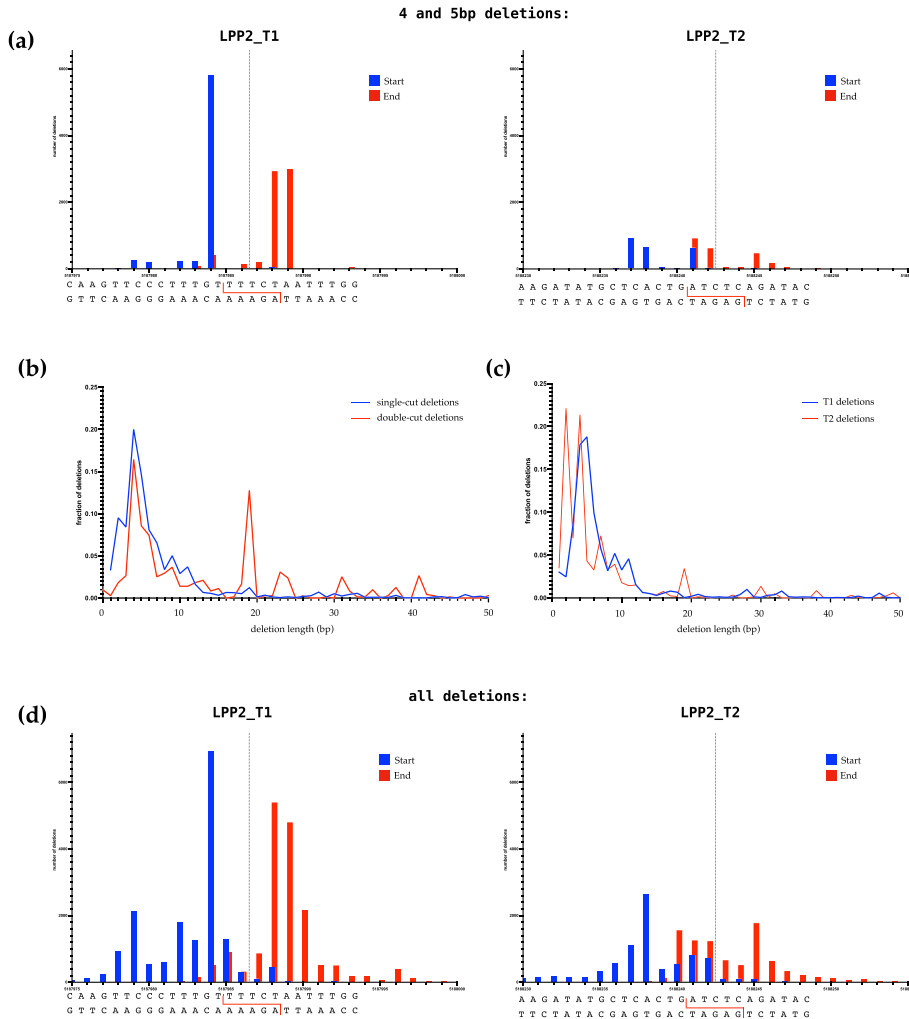
FIGURE 3 Patterns of deletions at the LPP2_T1 and LPP2_T2 targets. (a) Frequency distributions of start-sites (blue) and end-sites (red) of 4 and 5 nt deletion affecting the 10 bp windows centered on the cut-sites. The red bars have been offset slightly to the right for clarity. The target locus sequence with the cut-sites marked by staggered red lines is shown below the graph. Vertical dotted lines show the centers of the cut-sites to facilitate lecture of the graphs. The majority of these deletions begin at the base pair immediately upstream of the 5′ ssDNA overhang produced by Cas12a cleavage of the LPP2_T1 target, and the end-points are found four and five bases downstream. The LPP2_T2 target has three preferred start-points for 4 and 5 nt deletions, including deletions starting 2 and 3 nt upstream of the expected cut. (b) Frequency distributions of lengths of single cut and double cut (transposed 256 nt leftwards) confirm that the deletion patterns of the majority, small local deletions are not affected by the presence of cohesive (one-cut) versus non-cohesive (two-cut) 5′ ssDNA overhanging ends. Taken individually, frequency distributions of deletion lengths (c) and deletion start- and end-points (d; same presentation as in Panel a) at the two cut-sites are dominated by short local events, notwithstanding the presence of differences in the details of the distributions at the two sites.

can be seen by overlaying the distributions of the one-cut and the two-cut events (by subtracting 256 from the lengths of the two-cut deletions) (Figure 3b). To facilitate this comparison, the frequency distributions in Figure 3b are presented as fractions of total deletions. Plotting the distributions of T1 and T2 deletions separately confirms that, while overall very similar, unsurprisingly, there are differences in the details of the patterns of deletion lengths (Figure 3c) and the choice of start- and end-points (Figure 3d) at the two target sites. It appears reasonable to assume that this is at the origin of the minor differences between the patterns of local sequence loss in the single- and double-cut EJ events (Figure 3b). Furthermore, given that the targets both lie within the LPP2 gene, separated by only 256 bp, these specificities are presumably due to local DNA sequence context. Finally, the maximum number of nt of microhomology from 0 to 9 nt, potentially involved in each deletion, was calculated (see Figure S1a). Mean microhomology values for all deletions starting at each coordinate were calculated across the LPP2-T1 and LPP2_T2 target loci (Figure S1b,c). The number of deletions starting at each coordinate was also plotted to put these values into context, confirming the absence of correlation between microhomology and deletion frequency.

Due to restrictions on amplicon length for Illumina sequencing, the cut-sites analyzed above are 93 and 52 bp from the ends of the PCR-amplified fragment used for the sequencing (Figure 1a). Loci with deletions extending beyond these ends would not be amplified (nor sequenced), and this could clearly impact our analyses. We thus used Oxford Nanopore sequencing, which does not suffer from these constraints on amplicon sequence length, to test whether the Illumina approach had impacted our analyses and, if so, to build a clear picture of the DSB-induced recombination at this locus.

A 2.8 kb *LPP2* amplicon was chosen for Nanopore sequencing (Figure 4a). In this case, the T1 and T2 targets are, respectively, 1319 bp and 1226 nt from the amplicon ends. This segment was amplified from 10 day old plantlets expressing the Cas12a + gRNA (targeting LPP2 as above) and from plantlets expressing Cas12a + gRNA targeting another locus on Chromosome 3 as Cas12a-gRNA control (Chr3_92, see below). Although the Nanopore sequencing data includes a significant background of small deletions and insertions, these are almost all less than 6 bp long (Figure 4b,c), and when excluded from the analysis, the deletion patterns from the Illumina and Nanopore sequencing overlap very closely (Figure S2c).

LAGEIX ET AL.

American Society
of Plant Biologists **S·E·B** WILEY | 7 of 12
Cultivating a better future through plant biology research SOCIETY FOR EXPERIMENTAL BIOLOGY

FIGURE 4 Nanopore sequencing confirms patterns of Cas12a-induced deletions at the LPP2 targets. (a) Schematic showing the two Cas12a target sites of the target LPP2 locus and the 2.8 kb amplicon used for Nanopore sequencing. (b) Plot showing 1000 randomly selected deletions of >5 bp as arcs linking each deletion start- and end-point across the 400 bp amplicon spanning the target locus (X-axis). Vertical orange lines mark the target sites. The arcs are all of the same gray color, the darker lines reflect overlaying of multiple arcs. (c) Frequency distributions of numbers of deletions by their lengths in Cas12a + gRNA plants (blue curve) and Cas12a-gRNA controls (red curve). Deletions of less than 5 bp were excluded due to the high background of small indel errors in Nanopore sequences.
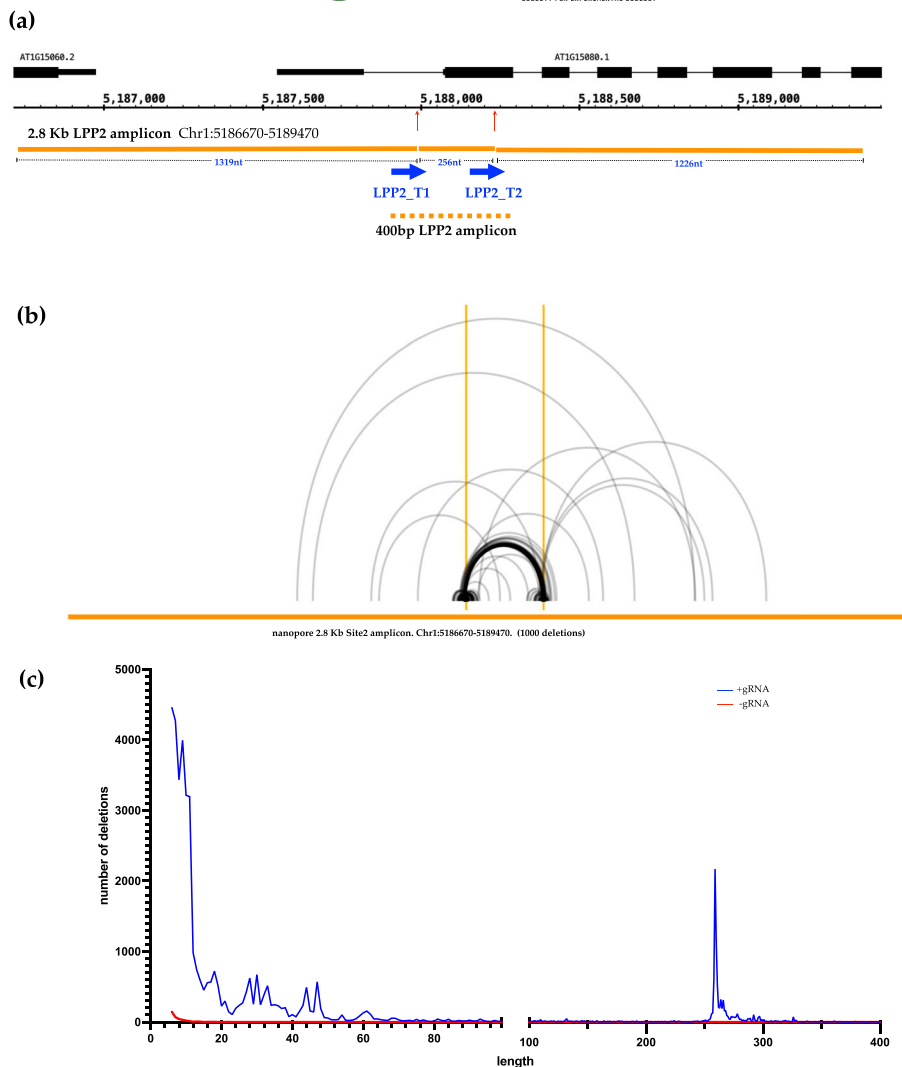
Of 85,439 sequences mapped to the LPP2 locus from the Cas12a + gRNA line, 48,587 indels of greater than 5 bp affecting one or both target sites were found, and of these, 98.7% were deletions and 1.3% insertions (Figure S2a,b). As expected, very few indels >5 bp (409: 379 deletions and 30 insertions) were found in the 87,331 sequences from the Cas12a-gRNA controls (Figure S2b). These numbers from the Nanopore sequencing of the 2.8 kb amplicon are thus comparable to those from the Illumina sequencing of the 400 bp amplicon.

Although care needs to be taken with the totals due to the exclusion of indels of 5 bp or less, Nanopore sequencing of the 2.8 kb amplicon does permit detection of longer indels that may have been artefactually excluded from the 400 bp amplicon/Illumina approach. Figure 4b presents an arc-diagram plot of 1000 random deletions of >5 bp from the +gRNA Nanopore sequencing, and Figure 4c shows the frequency distribution plot of all detected deletions of >5 bp. Comparing Figure 4b with the equivalent plot from the Illumina sequencing (Figure 2c) shows that, as expected, the longer amplicon Nanopore sequencing does permit detection of deletions extending further into the sequences flanking the cut-sites than those detected

with the 400 bp amplicon used for the Illumina sequencing. These are, however, clearly only a minority of the deletions and, furthermore, are restricted to the central part of the amplified sequence (Figure 4c).

Taken together these results give a clear picture of Cas12a-induced recombination at this locus in Arabidopsis plantlets, showing clearly that repair of the great majority of mutagenic repair of the breaks results in short, local deletions encompassing the breakpoint. To extend these conclusions and test the possible impact of different DNA sequence contexts on the repair mechanisms and outcomes, we applied this approach to a second target locus, in which the Cas12a target sites are flanked by tandem direct repeat DNA sequences. The chosen locus, Chr3_92, lies in the repeat-rich, pericentromeric region of Chromosome 3 (Figure 5a). This locus has two 142 nt direct repeats separated by 742 nt, and two unique Cas12a targets were chosen in the sequences between them. The perfect (no mismatches) 142 nt repeats are themselves embedded in two mildly degenerate (18 mismatches) 428 nt direct repeats (Figure 5a).

Following the same logic as for the analyses at the LPP2 locus, a 2.8 kb amplicon centered on the Chr3_92 locus was amplified from
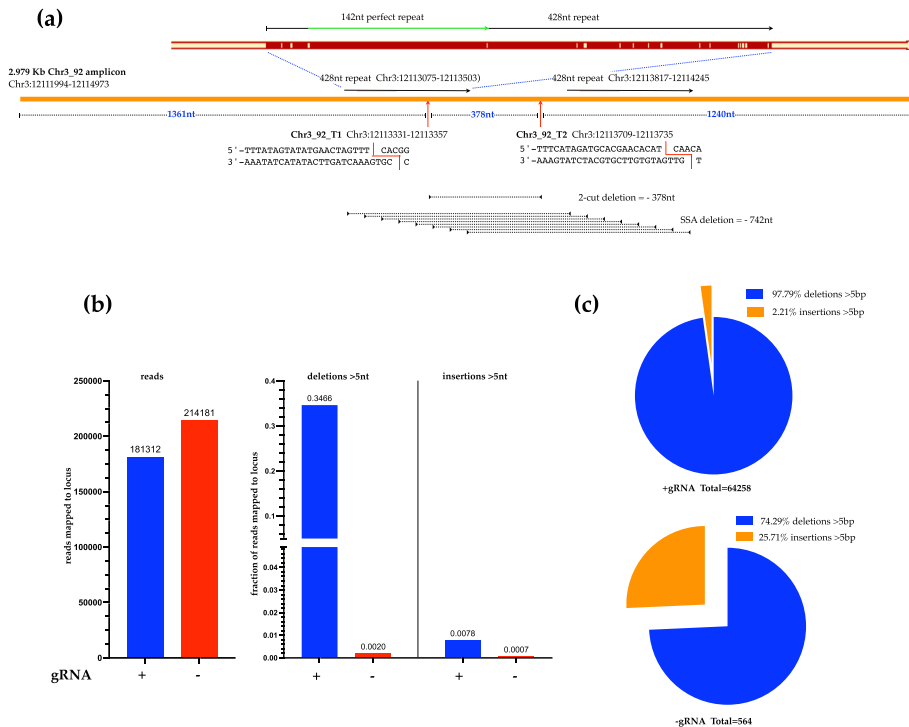
FIGURE 5 Patterns of Cas12a-induced deletions at the Chr3_92 target locus. (a) Schematic showing the two Cas12a target sites flanked by direct repeat sequences of the target Chr3_92 locus and the 2.979 kb amplicon used for Nanopore sequencing. The target sequences, including the (TTT) protospacer adjacent motif (PAM) and the 23 nt protospacer, are shown along with staggered red lines indicating the expected cleavage of the target DNA strands. (b) Numbers of sequences and the deletions and insertions expressed as fractions of mapped to the Chr3_92 locus from Cas12a + gRNA (blue) and Cas12a-gRNA plants (red). Indels of less than 5 bp were excluded due to the high background of small indel errors in Nanopore sequences. (c) As seen at the LPP2 locus, the majority of Cas12a-induced indels are deletions (b, c). SSA, single-strand annealing.

10 day old plantlets expressing the Cas12a + gRNA targeting Chr3_92_T1 and Chr3_92_T2 on Chromosome 3 and a Cas12a + gRNA targeting another locus as Cas12a-gRNA control (the LPP2 locus on Chromosome 1, see above). The Chr3_92_T1 and Chr3_92_T2 targets, separated by 378 nt, are 1361 bp and 1240 nt from the amplicon ends, respectively (Figure 5a). This 2979 bp segment (Chr3:12111994-12114973) was amplified by PCR and sequenced with the Oxford Nanopore Minion sequencer (as for the LPP2 locus, above).

Of 181,312 sequences mapped to the Chr3_92 locus from the Cas12a + gRNA line, 64,258 indels of greater than 5 bp affecting one or both target sites were found, and of these, 97.79% were deletions and 2.21% insertions (Figure 5b, c). As expected, very few indels >5 bp (564: 419 deletions and 145 insertions) were found in the 214,181 sequences from the Cas12a-gRNA controls (Figure 5b, c). These numbers from the Nanopore sequencing of the 2.8 kb amplicon are thus comparable to those from the Illumina sequencing of the 400 bp amplicon.

Analysis of the Cas12a-induced deletions at the Chr3_92 locus shows a pattern of local deletions, at and between the two cut-sites (Figure 6a), comparable to that observed at the LPP2 locus (Figure 4b). However, in contrast to the LPP2 deletion pattern, a new major deletion class at 742 bp represents 39.37% of the 61,689 deletions greater than 5 bp (Figure 6a,b). This length (742 bp) is precisely the spacing between the flanking repeat sequences, and these events thus presumably result from SSA recombination. The arc plot of the deletions confirms that this is the case: a major class of deletions with start/end points distributed across the repeat sequences, extend in-step from one repeat to the other (red arcs in Figure 6c). The presence

of these flanking direct repeat sequences thus generates a new class of recombinants representing almost 40% of the total that is not detected in their absence. The arc plot of the LPP2 deletions from Figure 3b is inset in Figure 6c for comparison.

The optimum temperature for activity of the FnCas12a endonuclease in plants is 29°C, (Bernabé-Orts et al., 2019; Huang et al., 2021; Lee et al., 2019; Malzahn et al., 2019; Schindele & Puchta, 2019), and the data presented in this article come from the analysis of plants grown at this temperature. While 29°C is a temperature well within the range that Arabidopsis plants will encounter in the environment and the plants grow well at this temperature, the standard laboratory growth temperature for *A. thaliana* is 23°. We thus verified that the higher growth temperature does not impact the choice of mechanism and thus outcomes of repair of Cas12a-induced DSB in the plants. Illumina sequencing of the 400 bp LPP2 locus amplicon from plantlets grown at 23 and 29°C shows similar proportions of the different deletion classes (Figure S3a), and the distributions of deletions (Figure S3b) and insertion (Figure S3c) lengths coincide at the two temperatures. The patterns of EJ repair of these breaks at 23 and 29° are thus equivalent. We also checked for a possible temperature-specific impact on the SSA recombination pathway by making an analogous comparison between results or repair of DSB targeted to the Chr3_92 locus. As for the results concerning EJ recombination at the LPP2 locus (above), similar proportions of the different deletion classes (Figure S4a) are observed, and the distributions of deletion (Figure S4b) and insertion (Figure S4c) lengths overlap at the two temperatures. The patterns of EJ and SSA repair of 5′-overhang DSB at these two loci are thus equivalent at these two temperatures.

FIGURE 6 The presence of flanking direct repeats generates a major new class of single-strand annealing (SSA) deletions. (a) Distributions of numbers of deletions by their lengths in Cas12a + gRNA plants (blue curve) and Cas12a-gRNA controls (red curve). Deletions of less than 5 bp were excluded due to the high background of small indel errors in Nanopore sequences. A major (39.4%), novel class of 742 bp deletions is created by the presence of the direct repeats flanking the Cas12a target sites. (c) Deletion start-end arc-plots confirm that these 742 bp deletions (red curves) result from in-step recombination between the flanking direct repeats, as expected from SSA recombination. This is strikingly different from the patterns observed at the LPP2 locus in the absence of the flanking repeats (inset in Panel c, repeated from Figure 4).
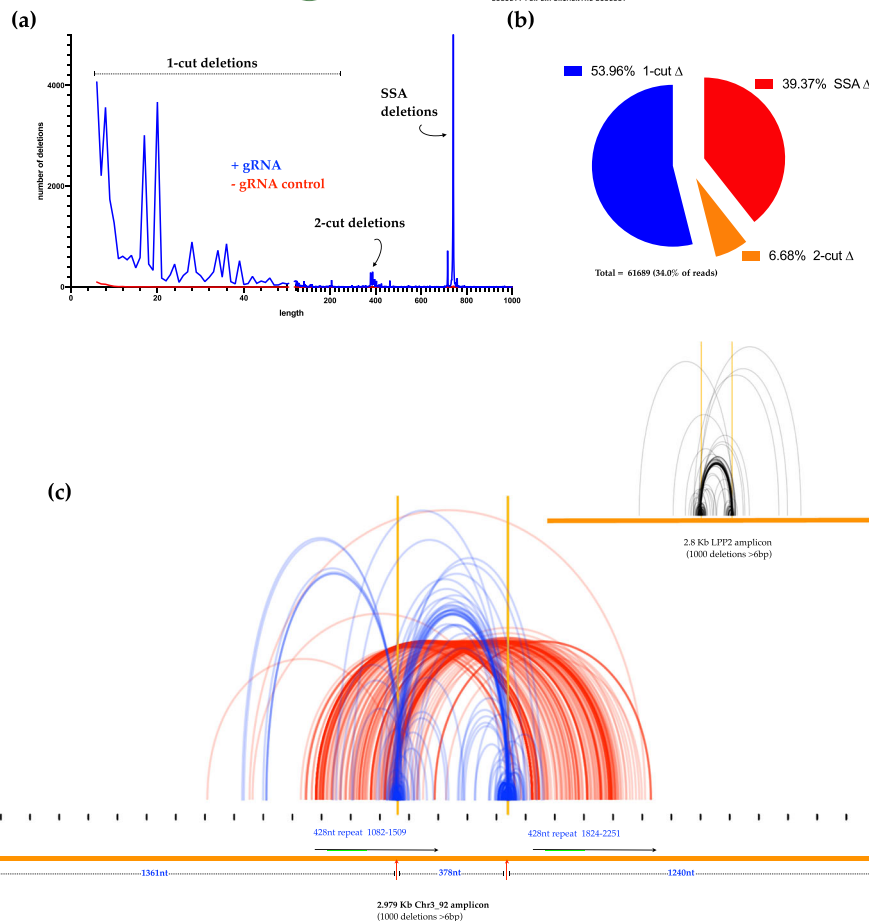
4 | DISCUSSION

This study concerns the repair of 5′-ssDNA overhang DSB in *A. thaliana*, using targeted *in planta* cleavage by FnCas12a (FnCpf1) (Endo et al., 2016). In addition to producing 5′-ssDNA overhang DSB, Cas12a has high specificity for its target and the very useful capability to process precursor RNAs into crRNAs (Fonfara et al., 2016; Makarova et al., 2015; Shmakov et al., 2015; Strohkendl et al., 2018; Swarts & Jinek, 2018; Zetsche et al., 2015). The presence of 5′ overhangs at Cas12a-induced DSB is of particular interest for study of DNA break repair and recombination. Like blunt-ended and 3′-ssDNA overhang DNA breaks, 5′-ssDNA overhang DSB can be repaired by simple religation. Such events are however very difficult to detect and analyze, as well as being potentially subject to further rounds of cleavage/repair. Degradation of the 4–5 nt 5′ overhangs or filling-in of the recessed 3′-ended strands left by Cas12a cleavage to generate blunt ends for EJ ligation would result in 4–5 bp deletions or insertions, respectively. Further resection of the 5′-ended strand would be needed to produce the 3′ overhangs needed for microhomology-dependent alternative EJ (aEJ or MMEJ) or homology-dependent recombination (HR) and SSA.

As well as being less studied, notably in plants, these breaks are thus of particular interest in terms of the study of recombination and DNA repair mechanisms. We chose to work with pairs of nearby

Cas12a cuts in order to characterize any specificities of repair of the (cohesive-ended) single Cas12a cuts versus the (non-cohesive-ended) twin-cuts and to check for more complex events involving reinsertion of a cut-out chromosome fragment. We note also that our PCR-amplicon-based analysis is necessarily restricted to "local" DSB repair outcomes. While we specifically include analysis of twin-break events and include long-read sequencing to take this into account, larger scale rearrangements (chromosome arm loss, translocations, etc.) are not directly analyzed in the work reported here. In this context, we direct readers to the recent study and discussion of Cas12a-induced mutations in the ascomycete fungus *Magnaporthe oryzae* (Huang et al., 2022). Finally, we note that while Cas12a enzymes are strongly impacted by protospacer/target mismatches, they cut the target DNA at the PAM-distal end of the target sequence (discussed by Strohkendl et al., 2018 and Swarts & Jinek, 2018). Thus, recognition and re-cleavage of repaired targets with small sequence changes (substitutions, small indels) will potentially be more frequent than would be the case with protospacer adjacent motif (PAM) proximal cleavage by enzymes such as Cas9 (Swarts & Jinek, 2018). While this effect could be expected to lead to longer deletions, our results however show a marked predominance of very short deletions. The impact of this in our plants is thus less important than might have been expected, perhaps due to the very efficient cutting.

Five principle conclusions emerge from the work presented here:

1. The great majority of mutations left by the repair of these 5′-overhang DSB are deletions, with a clear peak of deletion lengths at 4 bp. The frequencies then fall off rapidly with increasing deletion length. Strikingly, repair resulting in insertions (also very short) is >30-fold less frequent than that producing deletions. Examination of the patterns of deletion lengths shows them to be very similar at the different cut-sites, with however some differences seen in the details (see below). Plotting mean microhomology length against deletion start-point across the target sites shows no clear evidence for a relation between the probability of a given base pair serving as the start-point of a deletion and the mean number of nt of microhomology potentially involved in the deletions starting at that position in the sequence. While the existence of such microhomologies in the sequence in no way proves that they were actually involved in the recombination event under study, this result does show that the use of longer microhomologies is not a primary determinant in the choice of the start- and end-points of these very short deletions resulting from repair of 5′-overhang DSB in Arabidopsis.

2. The proximity of the FnCas12a cut-sites to the ends of the 400 bp LPP2 amplicon used for Illumina sequencing could clearly have impacted the patterns of deletions found in this work, excluding repair events involving longer deletions. We thus amplified a 2.8 kb fragment covering the same locus, for sequencing with the Oxford Nanopore Minion sequencer. The two FnCas12a cut-sites are 1319 and 1226 bp from the ends of this 2.8 kb amplicon, considerably more distant than the 93 and 51 bp in the 400 bp amplicon. Strikingly, while the results of sequencing the longer amplicon did permit identification of longer deletions, these were increasingly rare with increasing length, and the overall pattern of repair was not changed relative to that from the 400 bp amplicon sequencing.

3. The use of nearby pairs of Cas12a targets results in a spectrum of modifications at the repaired cuts: individual, local deletions at one or both targets, and deletion or inversion of the DNA sequences between the two targets (Figure 1). While the relative proportions of "one-cut" versus "two-cut" events are dependent upon the relative and absolute cutting efficiencies at the two sites, examination of the "excision" events involving coincident cutting of both target sites does yield some interesting conclusions. These include the 2.91% of sequences with deletions spanning the two cut-sites (T1-T2 deletions) and the events involving reinsertion of the excised T1-T2 fragment. Depending on the orientation of the reinsertion of the excised fragment, the "excision-reinsertion" events would include both the 2.05% of sequences with T1-T2 inversions and an equivalent number classed within the 12.42% of sequences with local deletions at both T1 and T2. Thus, the excision-reinsertion class would represent ∼4% of sequences. That this value is comparable in magnitude to the number of T1 to T2 deletions (excision without reinsertion) (∼3%) suggests that loss or reinsertion of the excised fragment in two-site cleavage events is equally likely.

4. The use of paired target sites permitted analysis of patterns of deletions at junctions resulting from joining DNA ends with complementary 5′-overhangs (1-cut deletions) and at those involving non-complementary 5′-overhangs (two cut deletions). The two cut-sites are separated by 256 bp, and subtracting this from the lengths of the T1-T2 deletions permits overlaying the pattern of the T1-T2 deletions with the one-cut events. While some minor differences in the patterns of deletions are made clearly visible by this approach, the predominance of 4 bp deletions, followed by a rapid fall-off with increasing deletion length, is strikingly similar. Thus, the presence/absence of cohesive 5′ overhangs does not affect the deletion pattern.

5. The pattern of EJ repair of the 5′-overhang DSB thus results overwhelmingly in the production of very short deletions, with the precise details of the deletion end-points being impacted by local sequence context. This is the case for the two target sites in the chromosome arm, expressed LPP2 gene and the two targets in the pericentromeric, heterochromatic Chr3_92 locus. The Chr3_92 cut-sites are flanked by tandem direct repeat sequences, and this leads to the presence of a new class of deletion events resulting from SSA recombination. It is particularly striking that these SSA events are a major repair class, greatly outnumbering the two-cut deletions and approaching in frequency the EJ repair events (39% vs. 61% of deletions, respectively). It is possible that these events result from microhomology-mediate EJ of a sequence close to the cut-site with a target in the flanking repeat. That this is not so is seen clearly in the fact that, rather than being concentrated close to the cut-sites, the start- and end-points of the SSA deletions are distributed throughout the (degenerate) repeat sequences. The presence of these flanking direct repeats thus leads to a striking change in the pattern of repair, channeling almost 40% of events into the SSA pathway.

AUTHOR CONTRIBUTIONS

Designed research and analysis: Sébastien Lageix, Maria E. Gallego, Yannick Bidet, Miguel Hernandez, and Charles I. White. *Carried out experiments*: Sébastien Lageix, Maria E. Gallego, Miguel Hernandez, Sandrine Viala, Jérémy Verbeke, and Charles I. White. *Resources*: Maria E. Gallego, Yannick Bidet, and Charles I. White. *Writing, with contributions from all of the authors*: Sébastien Lageix, Maria E. Gallego, and Charles I. White.

CONFLICT OF INTEREST STATEMENT

We declare no financial or other conflicts of interest in this work.

DATA AVAILABILITY STATEMENT

The sequence files are available under BioProject accession number PRJNA1113952 at https://www.ncbi.nlm.nih.gov/sra/PRJNA1113952 (Biosample Accessions: SAMN41471278, SAMN41471279, SAMN41471280, SAMN41471281, SAMN41471282, SAMN41471283, SAMN41488828, SAMN41488829, SAMN41488830). The Python and R scripts using publicly available tools written to analyze the DNA sequencing data produced in this study can be downloaded from https://github.com/chrlesw/dsbrepair.git.

ORCID

Miguel Hernandez https://orcid.org/0000-0002-7633-0810
Yannick Bidet https://orcid.org/0000-0002-9624-8195
Charles I. White https://orcid.org/0000-0001-8152-5797

REFERENCES

Atia, M., Jiang, W., Sedeek, K., Butt, H., & Mahfouz, M. (2024). Crop bioengineering via gene editing: Reshaping the future of agriculture. *Plant Cell Reports*, *43*, 98. https://doi.org/10.1007/s00299-024-03183-1

Bechtold, N., Ellis, J., & Pelletier, G. (1993). In planta *Agrobacterium*-mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *Comptes rendus de l'Académie des Sciences Paris, Life Science*, 316, 1194–1199.

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, *27*, 573–580. https://doi.org/10.1093/nar/27.2.573

Bernabé-Orts, J. M., Casas-Rodrigo, I., Minguet, E. G., Landolfi, V., Garcia-Carpintero, V., Gianoglio, S., Vázquez-Vilar, M., Granell, A., & Orzaez, D. (2019). Assessment of Cas12a-mediated gene editing efficiency in plants. *Plant Biotechnology Journal*, *17*, 1971–1984. https://doi.org/10.1111/pbi.13113

Bothmer, A., Phadke, T., Barrera, L. A., Margulies, C. M., Lee, C. S., Buquicchio, F., Moss, S., Abdulkerim, H. S., Selleck, W., Jayaram, H., Myer, V. E., & Cotta-Ramusino, C. (2017). Characterization of the interplay between DNA repair and CRISPR/Cas-induced DNA lesions at an endogenous locus. *Nature Communications*, *8*, 13905. https://doi.org/10.1038/ncomms13905

Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge—Accurate paired shotgun read merging via overlap. *PLoS ONE*, *12*, e0185056. https://doi.org/10.1371/journal.pone.0185056

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews. Molecular Cell Biology*, *18*, 495–506. https://doi.org/10.1038/nrm.2017.48

Clough, S. J., & Bent, A. F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal*, *16*, 735–743. https://doi.org/10.1046/j.1365-313x.1998.00343.x

Endo, A., Masafumi, M., Kaya, H., & Toki, S. (2016). Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from *Francisella novicida*. *Scientific Reports*, *6*, 38169. https://doi.org/10.1038/srep38169

Fauser, F., Schiml, S., & Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *The Plant Journal*, *79*, 348–359. https://doi.org/10.1111/tpj.12554

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., & Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, *532*, 517–521. https://doi.org/10.1038/nature17945

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. *Cell*, *184*, 1621–1635. https://doi.org/10.1016/j.cell.2021.01.005

Huang, J., Rowe, D., Subedi, P., Zhang, W., Suelter, T., Valent, B., & Cook, D. E. (2022). CRISPR-Cas12a induced DNA double-strand breaks are repaired by multiple pathways with different mutation profiles in Magnaporthe oryzae. *Nature Communications*, *13*, 7168. https://doi.org/10.1038/s41467-022-34736-1

Huang, T., Armstrong, B., Schindele, P., & Puchta, H. (2021). Efficient gene targeting in *Nicotiana tabacum* using CRISPR/SaCas9 and temperature tolerant LbCas12a. *Plant Biotechnology Journal*, *19*, 1314–1324. https://doi.org/10.1111/pbi.13546

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., & Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Research*, *41*, e188. https://doi.org/10.1093/nar/gkt780

Jinek, M. M., Chylinski, K. K., Fonfara, I. I., Hauer, M. M., Doudna, J. A. J., & Charpentier, E. E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, *337*, 816–821. https://doi.org/10.1126/science.1225829

Kamoen, L., Kralemann, L. E. M., Van Schendel, R., Van Tol, N., Hooykaas, P. J. J., De Pater, S., & Tijsterman, M. (2024). Genetic dissection of mutagenic repair and T-DNA capture at CRISPR-induced DNA breaks in *Arabidopsis thaliana*. *PNAS Nexus*, *3*, pgae094. https://doi.org/10.1093/pnasnexus/pgae094

Katagiri, T., Ishiyama, K., Kato, T., Tabata, S., Kobayashi, M., & Shinozaki, K. (2005). An important role of phosphatidic acid in ABA signaling during germination in *Arabidopsis thaliana*. *The Plant Journal*, *43*, 107–117. https://doi.org/10.1111/j.1365-313X.2005.02431.x

Knoll, A., Fauser, F., & Puchta, H. (2014). DNA recombination in somatic plant cells: Mechanisms and evolutionary consequences. *Chromosome Research.*, *22*, 191–201. https://doi.org/10.1007/s10577-014-9415-y

Kumar, J., Char, S. N., Weiss, T., Liu, H., Liu, B., Yang, B., & Zhang, F. (2023). Efficient protein tagging and cis-regulatory element engineering via precise and directional oligonucleotide-based targeted insertion in plants. *The Plant Cell*, *35*, 2722–2735. https://doi.org/10.1093/plcell/koad139

Lee, K., Zhang, Y., Kleinstiver, B. P., Guo, J. A., Aryee, M. J., Miller, J., Malzahn, A., Zarecor, S., Lawrence-Dill, C. J., Joung, J. K., Qi, Y., & Wang, K. (2019). Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnology Journal*, *17*, 362–372. https://doi.org/10.1111/pbi.12982

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, J.-F., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G. M., & Sheen, J. (2013). Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nature Biotechnology*, *31*, 688–691. https://doi.org/10.1038/nbt.2654

Longo, G. M. C., Sayols, S., Kotini, A. G., Heinen, S., Möckel, M. M., Beli, P., & Roukos, V. (2024). Linking CRISPR–Cas9 double-strand break profiles to gene editing precision with BreakTag. *Nature Biotechnology*. Advance online publication. https://doi.org/10.1038/s41587-024-02238-8

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., … Koonin, E. V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*, *13*, 722–736. https://doi.org/10.1038/nrmicro3569

Malzahn, A. A., Tang, X., Lee, K., Ren, Q., Sretenovic, S., Zhang, Y., Chen, H., Kang, M., Bao, Y., Zheng, X., Deng, K., Zhang, T.,

Salcedo, V., Wang, K., Zhang, Y., & Qi, Y. (2019). Application of CRISPR-Cas12a temperature sensitivity for improved genome editing in rice, maize, and *Arabidopsis*. *BMC Biology*, *17*, 9. https://doi.org/10.1186/s12915-019-0629-5

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G., & Kamoun, S. (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature Biotechnology*, *31*, 691–693. https://doi.org/10.1038/nbt.2655

Přibylová, A., & Fischer, L. (2024). How to use CRISPR/Cas9 in plants: From target site selection to DNA repair. *Journal of Experimental Botany*, *75*, erae147. https://doi.org/10.1093/jxb/erae147

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. *The Plant Journal*, *80*, 1139–1150. https://doi.org/10.1111/tpj.12704

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. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 7266–7271. https://doi.org/10.1073/pnas.1603823113

Schindele, P., & Puchta, H. (2019). Engineering CRISPR/LbCas12a for highly efficient, temperature-tolerant plant gene editing. *Plant Biotechnology Journal*, *17*, 1971–1973. https://doi.org/10.1111/pbi.13275

Scully, R., Panday, A., Elango, R., & Willis, N. A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews. Molecular Cell Biology*, *20*, 698–714. https://doi.org/10.1038/s41580-019-0152-0

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J. J., Qiu, J.-L., & Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology*, *31*, 686–688. https://doi.org/10.1038/nbt.2650

Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., Zhang, F., & Koonin, E. V. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Molecular Cell*, *60*, 385–397. https://doi.org/10.1016/j.molcel.2015.10.008

Strohkendl, I., Saifuddin, F. A., Rybarski, J. R., Finkelstein, I. J., & Russell, R. (2018). Kinetic basis for DNA target specificity of CRISPR-Cas12a. *Molecular Cell*, *71*, 816–824.e3. https://doi.org/10.1016/j.molcel.2018.06.043

Swarts, D. C., & Jinek, M. (2018). Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing. *WIREs RNA*, *9*, e1481. https://doi.org/10.1002/wrna.1481

Vu, G. T. H., Cao, H. X., Fauser, F., Reiss, B., Puchta, H., & Schubert, I. (2017). Endogenous sequence patterns predispose the repair modes of CRISPR/Cas9-induced DNA double-stranded breaks in *Arabidopsis thaliana*. *The Plant Journal*, *92*, 57–67. https://doi.org/10.1111/tpj.13634

Weiss, T., Kumar, J., Chen, C., Guo, S., Schlegel, O., Lutterman, J., Ling, K., & Zhang, F. (2024). Dual activities of an X-family DNA polymerase regulate CRISPR-induced insertional mutagenesis across species. *Nature Communications*, *15*, 6293. https://doi.org/10.1038/s41467-024-50676-4

Wolter, F., Schindele, P., Beying, N., Scheben, A., & Puchta, H. (2021). Different DNA repair pathways are involved in single-strand break-induced genomic changes in plants. *The Plant Cell*, *33*, 3454–3469. https://doi.org/10.1093/plcell/koab204

Wright, W. D., Shah, S. S., & Heyer, W.-D. (2018). Homologous recombination and the repair of DNA double-strand breaks. *The Journal of Biological Chemistry*, *293*, 10524–10535. https://doi.org/10.1074/jbc.TM118.000372

Xue, C., & Greene, E. C. (2021). DNA repair pathway choices in CRISPR-Cas9-mediated genome editing. *Trends in Genetics*, *37*, 639–656. https://doi.org/10.1016/j.tig.2021.02.008

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., & Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*, 759–771. https://doi.org/10.1016/j.cell.2015.09.038

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., Winblad, N., Choudhury, S. R., Abudayyeh, O. O., Gootenberg, J. S., Wu, W. Y., Scott, D. A., Severinov, K., van der Oost, J., & Zhang, F. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nature Biotechnology*, *35*, 31–34. https://doi.org/10.1038/nbt.3737

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.