Context effects on repair of 5'-overhang DSB induced by Cas12a in Arabidopsis.

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

[1] 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

[2] Laboratoire d'Oncologie Moléculaire
U1240 Imagerie Moléculaire et Stratégies Théranostiques
Université Clermont Auvergne, Centre Jean Perrin, INSERM
58 rue Montalembert, 63000 Clermont-Ferrand, France

[&] Current address: Centro de Biología Molecular Severo Ochoa,
Calle Nicolás Cabrera 1, 28049 Madrid

*Corresponding author : charles.white@cnrs.fr
ORCID: 0000-0001-8152-5797

All authors have seen and approved the manuscript. It hasn't been accepted or published elsewhere.
The authors are aware of no competing interests concerning this work.

Keywords: Recombination, DNA Repair, DSB repair, Cas12a, Arabidopsis

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 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, double-strand breaks (DSB), 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 DSB 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 an alternative class of repair events, corresponding to highly efficient repair by Single-strand Annealing recombination.

Introduction

DNA double-strand breaks (DSB) 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 recombination (EJ) 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 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 and Fischer, 2024; Schiml et al., 2016; Shan et al., 2013; Vu et al., 2017; Wolter et al., 2021; Xue and Greene, 2021).

The majority of studies 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 and 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 (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 (eg. (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 and 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

87 majority of repaired DSB are repaired locally by End-Joining recombination. While
88 differences are seen in the details, overall the deletion patterns are not greatly affected by
89 differing local sequence contexts flanking the DSB. The presence of flanking direct repeats
90 sequences does however strongly impact the repair pattern - corresponding to highly
91 efficient repair by Single-strand Annealing recombination.
92

93 ## Materials and Methods

94

95 **Genomic target loci**

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

108

109 **T-DNA constructs and plant transformation**

110 Recombinant CRISPR/Cas9 and sgRNA constructs (Table S1) were designed for unique
111 genomic positions and cloned into the expression plasmids kindly provided by Dr Seiichi Toki
112 (Endo et al., 2016). T-DNA transformation of *Arabidopsis thaliana* plants of ecotype Columbia-
113 0 by *Agrobacterium tumefaciens* strain C58C1 was by the floral-dip method (Bechtold, N. et al.,
114 1993; Clough and Bent, 1998).

115

116 **Plant selection**

117 In order to obtain transgenic plant lines to address specifically each of the two selected
118 targets, transformed T1 plants for each construct were grown and selected on Murashige and
119 Skoog (MS) medium (4.9 $g.l^{-1}$ Murashige and Skoog medium, 10 $g.l^{-1}$ sucrose, and 8 $g.l^{-1}$agar,
120 pH 5.7) with kanamycin (50 $\mu g.mL^{-1}$) in a growth chamber with a 16/8-hour light/dark

121 cycle, temperature 23°C and 60% relative humidity. The progeny of individuals T1 were
122 grown for 10 days at 29°C on MS media and gDNA was extracted using Promega's Wizard
123 Genomic DNA Purification Kit. To assess the cutting efficiency of the transgenic lines, a PCR
124 amplifying the deletion between the two cuts sites for each target loci was performed. The T1
125 harboring the stronger activity for each construct was selected.

126

127 **Sequencing of DNA**

128 The respective target regions from 5 pooled transgenic plants each were amplified by PCR
129 (primer in Table S2) using Platinum SuperFi II DNA Polymerase (Invitrogen) and subjected
130 to NGS. Two different methods were used for NGS applications, the Illumina MiSeq and the
131 Nanopore Technology. The starting material for library preparation was 100 µL of PCR
132 products at 10 ng·μL^{-1}. The Illumina MiSeq Nano V2 was used for paired end sequencing of
133 amplicons of a size of 400 bp. For the Nanopore Technology, library preparation was
134 conducted using the SQK-LSK110 ligation sequencing kit following the manufacturer's
135 protocol. Samples were sequenced using a minION Mk1c device using R9.4.1 (FLO-
136 MIN106D) flow cells.
137
138 **Informatics analyses.**

139 In-house Python and R scripts using publicly available tools were written to analyse the
140 DNA sequencing data produced in this study (https://github.com/chrlesw/dsbrepair.git).
141 Briefly, the scripts analyse NGS DNA sequence outputs from both Illumina paired-end
142 sequencing or Oxford Nanopore sequencing. The Nanopore sequences require no pre-
143 processing, but prior to running the analyses, the paired-end Illumina sequences are merged
144 into individual sequences using the BBMerge tool of the BBTools package (BBMap –
145 Bushnell B. – sourceforge.net/projects/bbmap/ Brian Bushnell (Bushnell et al., 2017)). These
146 merged sequence files are the input for two main Python scripts, which carries out the
147 analysis following two approaches:
148 **1) Text-recognition approach: _dsb_sortAnal.py**
149 Based on text searches and precise sequence lengths, this approach is best applied to the
150 Illumina sequences due to the lower sequencing error rate compared to Nanopore sequences.
151 The merged Illumina sequences are individually scanned for the presence/absence of the
152 Cas12a target sequence(s) using Python text-search tools. The presence of inversions is
153 identified by the presence of a reverse-complementary sequence taken from between the
154 target sites. These, combined with their lengths, are used to sort and enumerate the merged
155 sequences of different deletion classes: single site deletions, individual two-site deletions,
156 deletions spanning the two cut-sites and inversions of the sequences between the two cut-
157 sites.
158 2) **Mapping plus variant calling approach: _dsb_varAnal.py**

159 Minimap2 (Li, 2018) is used to align the Nanopore or merged, paired-end Illumina sequences
160 to the reference genome. Secondary or supplementary alignments are removed and the
161 BBMap callvariants.sh script of the BBTools package (BBMap
162 – sourceforge.net/projects/bbmap/ Brian Bushnell (Bushnell et al., 2017)) used to produce
163 lists of deletions, insertions and substitutions. Shell commands then select variants affecting
164 10 nt windows centered on the cut-sites and sort them into lists of deletions, insertions and
165 substitutions. The potential implication of DNA sequence microhomologies is determined
166 for each deletion junction and plots showing each deletion as an arc linking its start and end
167 points are generated.
168

169 ## Results

170 A FnCas12a construct expressing guide RNAs targeting two sites, LPP2_T1 and LPP2_T2,
171 respectively located in the first intron and second exon of the Lipid Phosphate Phosphatase 2
172 gene (LPP2, AT1G15080). The key characteristic of this gene as a target for this work is that it
173 is non-essential (Katagiri et al., 2005). The two target sites, T1 and T2, are 256 bp apart,
174 oriented in the same direction and both are unique in the *Arabidopsis thaliana* genome (Figure
175 1a). Stably transformed lines carrying this construct were produced by T-DNA
176 transformation and a line showing high levels of mutation at the target sites was chosen for
177 further work. In parallel a plant line carrying the same FnCas12a T-DNA construct, but
178 lacking the gRNA template was built as a -gRNA control.
179 Genomic DNA was extracted from 10-day old plantlets, a 400bp region including the target
180 sites was amplified (Figure 1a) and paired-end sequenced with an Illumina Miseq sequencer.
181 The read-pairs were merged with the BBMerge script of the BBMap package to give single
182 sequences and these were analysed by an in-house Python script (see Materials and
183 Methods).
184
185 In principle, repair of the paired Cas12a-induced breaks can result in a number of outcomes:
186 simple religation (not detectable) or mutation of either or both T1 or T2 (Fig. 1b), deletion
187 from T1 to T2 (Fig. 1c), T1+T2 cleavage followed by re-insertion of the cut-out fragment in
188 either the original (Fig. 1d), or inverted (Fig. 1e) orientation.
189 A python script was written to quantify and sort the sequence reads into these classes. Based
190 on the sequence length plus the presence/absence of the unmodified target sequences, this
191 script gives a direct read-out of the events in each sequence. In the Cas12a+gRNA plants,
192 23644 sequences with modified T1 and/or T2 were found in the 82431 sequence reads
193 (28.7%), while 11 were found in the 52806 reads (0.02%) from the Cas12a-gRNA controls. The
194 proportions of the different classes of events found in the +gRNA sequences are shown in
195 Figure 1f. Thus, of the 28.7% of sequences which carry a deletion in T1 and/or T2, individual
196 T1 deletions (59.38%) are 2.6-fold more frequent than individual T2 deletions (23.25%),
197 presumably due to more efficient cleavage at the T1 target. 12.42% of the sequences with T1
198 and/T2 deletions have individual mutations at both T1 and T2, and 2.91% deletions

199 extending from T1-T2, with slightly lesser number (2.05%) having T1-T2 deletions associated
200 with reinsertion of the intervening sequence in inverted orientation. It is important to keep in
201 mind that the frequencies of the latter two classes (the T1-T2 events) will depend upon the
202 likelihood of a given chromatid being cleaved at both sites co-incidentally, and thus the
203 relative and absolute frequencies of cleavage at T1 and T2. The relative frequencies of these
204 paired-cut events would thus be expected to differ in different transformants, target pairs or
205 at different loci. Aligning the sorted sequence files of each class of events to the Arabidopsis
206 genome (or an in-silico T1-T2 inversion locus for the inversion sequences) confirms the
207 efficacy of this approach and gives a first view of the nature of the deletions in each class
208 (Fig. 1g).
209
210 To take this analysis further, a second script was written to align the merged sequence-pairs
211 to the Arabidopsis genome and call variants with the callvariants.sh script of the BBMap
212 package (see Materials and Methods). Of 83387 sequences mapped to the LPP2 locus from
213 the Cas12a+gRNA line, deletions affecting one or both target sites represent 32.7%, insertions
214 1.0% and substitutions 4.5% (Figure 2a). As expected, very few deletions (0.02%) or
215 insertions (0.02%) were found in the 52806 sequences from the Cas12a-gRNA control,
216 however numbers of substitutions affecting the target windows were similar in the
217 sequences from the control and Cas12a+gRNA plants (3.3% versus 4.5%). Expressing these
218 numbers as percentages of detected variants (Figure 2b) gives 85.5% deletions, 2.7%
219 insertions and 11.8% substitutions in the +gRNA plantlets, and 0.58% deletions, 0.46%
220 insertions and 98.96% substitutions in the -gRNA controls. The majority of substitutions are
221 thus not Cas12a targeted and presumably arise from sequencing errors, which occur at
222 comparable frequencies in this material .
223 The great majority of mutations induced by Cas12a at this locus are deletions and are mostly
224 short deletions. A clear, visual confirmation of this is shown in Figure 2c, which shows 1000
225 randomly-chosen deletions, each plotted as an arc joining its start- and end-points. Both the
226 clusters of small deletions at each cut-site and the larger deletions involving both cut-sites
227 are clearly visible (black arcs). It is particularly striking to note the relatively small number of
228 other deletions extending into the flanking chromosomal DNA (Figure 2c). Frequency
229 distributions of deletion lengths confirm this observation, showing clearly that, with the
230 exception of the T1-T2 deletions from 256-300bp, the great majority are less than 20bp, with
231 only rare events being longer than 60bp (Figure 2d). A similar analysis shows that practically
232 the totality of insertions are 5bp or less, with a small minority extending to 10bp (Figure 2e).
233 The FnCas12a endonuclease cleaves double-stranded DNA (dsDNA) to leave 4-5nt, 5' single-
234 strand (ssDNA) overhangs (Zetsche et al., 2015) and we expected to find a major class of 4-
235 5nt deletions corresponding to loss of the ssDNA overhangs at the cut sites, followed by
236 religation. Frequency distributions of 4 and 5nt deletion start- and end-sites confirm the
237 presence of this class of "blunting/ligation" NHEJ events (Figure 3a). This is particularly
238 clearly visible at the T1 site, where practically the totality of 4- and 5-nt deletions are at the
239 expected position. It is also seen at the T2 site, however deletions starting 2 and 3 nt

240 upstream of the expected peak are also clearly visible. Should these upstream deletion start
241 sites result from extended erosion of the upstream sequence prior to religation, this would
242 result in deletions longer than the 4-5bp and their exclusion from this analysis. We can thus
243 confidently class these events as resulting from other cleavage sites upstream of the
244 "primary" cut site.
245 Each Cas12a cut produces complementary and potentially religatable overhanging 5' ssDNA
246 ends, however this is not so for the double-cut events (deletions ≥256bp), which involve
247 joining of DNA ends from cuts in two different sequences. This distinction does not appear
248 to significantly affect the loss of sequence in the deletions, as can be seen by overlaying the
249 distributions of the 1-cut and the 2-cut events (by subtracting 256 from the lengths of the 2-
250 cut deletions) (Figure 3b). To facilitate this comparison, the frequency distributions in Figure
251 3b are presented as fractions of total deletions. Plotting the distributions of T1 and T2
252 deletions separately confirms that, while overall very similar, unsurprisingly there are
253 differences in the details of the patterns of deletion lengths (Figure 3c) and the choice of
254 start and end-points (Figure 3d) at the two target sites. It appears reasonable to assume that
255 this is at the origin of the minor differences between the patterns of local sequence loss in the
256 single- and double-cut EJ events (Figure 3b). Furthermore, given that the targets both lie
257 within the LPP2 gene, separated by only 256bp, these specificities are presumably due to
258 local DNA sequence context. Finally, the maximum number of nt of microhomology from 0
259 to 9 nt, potentially involved in each deletion was calculated (see Figure S1a). Mean
260 microhomology values for all deletions starting at each coordinate were calculated across the
261 LPP2-T1 and LPP2_T2 target loci (Figure S1b and c). The number of deletions starting at
262 each coodinate was also plotted to put these values into context, confirming the absence of
263 correlation between microhomology and deletion frequency.
264
265 Due to restrictions on amplicon length for Illumina sequencing, the cut-sites analysed above
266 are 93 and 52bp from the ends of the PCR-amplified fragment used for the sequencing
267 (Figure 1a). Loci with deletions extending beyond these ends would not be amplified (nor
268 sequenced) and this could clearly impact our analyses. We thus used Oxford Nanopore
269 sequencing, which doesn't suffer from these constraints on amplicon sequence length, to test
270 whether the Illumina approach had impacted our analyses and, if so, to build a clear picture
271 of the DSB-induced recombination at this locus.
272 A 2.8kb *LPP2* amplicon was chosen for Nanopore sequencing (Figure 4a). In this case, the T1
273 and T2 targets are respectively 1319bp and 1226nt from the amplicon ends. This segment was
274 amplified from 10-day old plantlets expressing the Cas12a+gRNA (targeting LPP2 as above)
275 and from plantlets expressing Cas12a+gRNA targeting another locus on chromosome 3 as
276 Cas12a-gRNA control (Chr3_92, see below). Although the Nanopore sequencing data
277 includes a significant background of small deletions and insertions, these are almost all less
278 than 6bp long (Figure 4b, c), and when excluded from the analysis, the deletion patterns
279 from the Illumina and Nanopore sequencing overlap very closely (Figure S2c).

280 Of 85439 sequences mapped to the LPP2 locus from the Cas12a+gRNA line, 48587 indels of
281 greater than 5bp affecting one or both target sites were found and of these, 98.7% were
282 deletions and 1.3% insertions (Figure S2a, S2b). As expected, very few indels >5bp (409: 379
283 deletions and 30 insertions) were found in the 87331 sequences from the Cas12a-gRNA
284 controls (Figure S2b). These numbers from the Nanopore sequencing of the 2.8kb amplicon
285 are thus comparable to those from the Illumina sequencing of the 400bp amplicon.
286 Although care needs to be taken with the totals due to the exclusion of indels of 5bp or less,
287 Nanopore sequencing of the 2.8kb amplicon does permit detection of longer indels that may
288 have been artefactually excluded from the 400bp amplicon/Illumina approach. Figure 4b
289 presents an Arc-diagram plot of 1000 random deletions of >5bp from the +gRNA Nanopore
290 sequencing and Figure 4c the frequency distribution plot of all detected deletions of >5bp.
291 Comparing Figure 4b with the equivalent plot from the Illumina sequencing (Figure 2c)
292 shows that, as expected, the longer amplicon Nanopore sequencing does permit detection of
293 deletions extending further into the sequences flanking the cut-sites than those detected with
294 the 400bp amplicon used for the Illumina sequencing. These are, however, clearly only a
295 minority of the deletions and furthermore, are restricted to the central part of the amplified
296 sequence (Figure 4c).
297 Taken together these results give a clear picture of Cas12a-induced recombination at this
298 locus in Arabidopsis plantlets, showing clearly that repair of the great majority of mutagenic
299 repair of the breaks results in short, local deletions encompassing the breakpoint. To extend
300 these conclusions and test the possible impact of different DNA sequence contexts on the
301 repair mechanisms and outcomes, we applied this approach to a second target locus, in
302 which the Cas12a target sites are flanked by tandem direct repeat DNA sequences. The
303 chosen locus, Chr3_92, lies in the repeat-rich, pericentromeric region of Chromosome 3
304 (Figure 5a). This locus has two 142nt direct repeats separated by 742nt and two unique
305 Cas12a targets were chosen in the sequences between them. The perfect (no mismatches) 142
306 nt repeats are themselves embedded in two mildly degenerate (18 mismatches) 428nt direct
307 repeats (Figure 5a).
308 Following the same logic as for the analyses at the LPP2 locus, a 2.8Kb amplicon centered on
309 the Chr3_92 locus was amplified from 10 day-old plantlets expressing the Cas12a+gRNA
310 targeting Chr3_92_T1 and Chr3_92_T2 on chromosome 3 and a Cas12a+gRNA targeting
311 another locus as Cas12a-gRNA control (the LPP2 locus on Chromosome 1, see above). The
312 Chr3_92_T1 and Chr3_92_T2 targets, separated by 378nt, are 1361bp and 1240nt from the
313 amplicon ends respectively (Figure 5a). This 2979bp segment (Chr3:12111994-12114973) was
314 amplified by PCR and sequenced with the Oxford Nanopore Minion sequencer (as for the
315 LPP2 locus, above).
316 Of 181312 sequences mapped to the Chr3_92 locus from the Cas12a+gRNA line, 64258 indels
317 of greater than 5bp affecting one or both target sites were found and of these, 97.79% were
318 deletions and 2.21% insertions (Figure 5b, 5c). As expected, very few indels >5bp (564: 419
319 deletions and 145 insertions) were found in the 214181 sequences from the Cas12a-gRNA

controls (Figure 5b, 5c). These numbers from the Nanopore sequencing of the 2.8kb amplicon are thus comparable to those from the Illumina sequencing of the 400bp 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 742bp represents 39.37% of the 61689 deletions greater than 5bp (Figure 6a, 6b). 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 and Puchta, 2019) and the data presented in this article comes from 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 *Arabidopsis thaliana* is 23°. We thus verified that the higher growth temperature doesn't impact the choice of mechanism and thus outcomes of repair of Cas12a-induced DSB in the plants. Illumina sequencing of the 400bp LPP2 locus amplicon from plantlets grown at 23°C 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 End-Joining 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 End-Joining and SSA repair of 5'-overhang DSB at these two loci are thus equivalent at these two temperatures.

Discussion

This study concerns the repair of 5'-ssDNA overhang DSB in *Arabidopsis thaliana*, using targeted *in planta* cleavage by FnCas12a (FnCpf1) (Endo et al., 2016). In addition to

358 producing 5'-ssDNA overhang DSB, Cas12a has high specificity for its target and the very
359 useful capability to process precursor RNAs into crRNAs (Fonfara et al., 2016; Makarova et
360 al., 2015; Shmakov et al., 2015; Strohkendl et al., 2018; Swarts and Jinek, 2018; Zetsche et al.,
361 2015). The presence of 5' overhangs at Cas12a-induced DSB is of particular interest for study
362 of DNA break repair and recombination. Like blunt-ended and 3'-ssDNA overhang DNA
363 breaks, 5'-ssDNA overhang DSB can be repaired by simple religation. Such events are
364 however very difficult to detect and analyse, as well as being subject to further
365 cleavage/repair. Degradation of the 4-5nt 5' overhangs or filling-in of the recessed 3'-ended
366 strands left by Cas12a cleavage to generate blunt ends for EJ ligation would result in 4-5 bp
367 deletions or insertions respectively. Further resection of the 5'-ended strand would be
368 needed to produce the 3' overhangs needed for microhomology-dependent alternative end-
369 joining (aEJ or MMEJ) or homology-dependent recombination (HR) and single strand
370 annealing (SSA).
371 As well as being less studied, notably in plants, these breaks are thus of particular interest in
372 terms of recombination. We chose to work with pairs of nearby Cas12a cuts in order to
373 characterise any specificities of repair of the (cohesive-ended) single Cas12a cuts, versus the
374 (non-cohesive-ended) twin-cuts and to check for more complex events involving re-insertion
375 of a cut-out chromosome fragment. We note also that our PCR-amplicon based analysis is
376 necessarily restricted to "local" DSB repair outcomes. While we specifically include analysis
377 of twin-break events and include long-read sequencing to take this into account, larger scale
378 rearrangements (chromosome arm loss, translocations...) are not directly analysed in the
379 work reported here. In this context, we direct readers to the recent study and discussion of
380 Cas12a-induced mutations in the ascomycete fungus *Magnaporthe oryzae* (Huang et al., 2022).
381 Finally, we note that while Cas12a enzymes are strongly impacted by protospacer/target
382 mismatches, they cut the target DNA at the PAM-distal end of the target sequence (see
383 (Strohkendl et al., 2018; Swarts and Jinek, 2018)). Thus, recognition and re-cleavage of
384 repaired targets with small sequence changes (substitutions, small indels) will potentially be
385 more frequent than would be the case with PAM-proximal cleavage by enzymes such as
386 Cas9 (Swarts and Jinek, 2018). While this effet could be expected to lead to longer deletions,
387 our results however show a marked predominance of very short deletions, suggesting that
388 the impact of this in our plants is less than might have been expected, perhaps due to the
389 very efficient cutting.
390
391 Five principle conclusions emerge from the work presented here:
392 1) The great majority of mutations left by the repair of these 5'-overhang DSB are deletions,
393 with a clear peak of deletion lengths at 4bp. The frequencies then fall off rapidly with
394 increasing deletion length. Strikingly, repair resulting in insertions (also very short) is > 30-
395 fold less frequent than that producing deletions. Examination of the patterns of deletion
396 lengths shows them to be very similar at the different cut-sites, with however some
397 differences seen in the details (see below). Plotting mean microhomology length against
398 deletion start point across the target sites shows no clear evidence for a relation between the

399 probability of a given base pair serving as the start-point of a deletion and the mean number
400 of nt of microhomology potentially involved in the deletions starting at that position in the
401 sequence. While the existence of such microhomologies in the sequence in no way proves
402 that they were actually involved in the recombination event under study, this result does
403 indicate that the use of longer microhomolgies is not a primary determinant in the choice of
404 the start and end-points of these very short deletions resulting from repair of 5'-overhang
405 DSB in Arabidopsis.
406
407 2) The proximity of the FnCas12a cut-sites to the ends of the 400bp LPP2 amplicon used for
408 Illumina sequencing could clearly impact the patterns of deletions found in this work,
409 excluding repair events involving longer deletions. We thus amplified a 2.8 Kb fragment
410 covering the same locus, for sequencing with the Oxford Nanopore Minion sequencer. The
411 two FnCas12a cut-sites are 1319 and 1226 bp from the ends of this 2.8 Kb amplicon,
412 considerably more distant than the 93 and 51bp in the 400 bp amplicon. Strikingly, while the
413 results of sequencing the longer amplicon did permit identification of longer deletions, these
414 were increasingly rare with increasing length and the overall pattern of repair wasn't
415 changed relative to that from the 400 bp amplicon sequencing.
416
417 3) The use of nearby pairs of Cas12a targets results in a spectrum of modifications at the
418 repaired cuts: individual, local deletions at one or both targets, and deletion and inversions
419 of the DNA sequences between the two targets (Figure 1). While the relative proportions of
420 "one-cut" versus "two-cut" events will be dependent upon the relative and absolute cutting
421 efficiencies at the two sites, examination of the "excision" events involving coincident cutting
422 of both target sites does yield some interesting conclusions. These include the 2.91% of
423 sequences with deletions spanning the two cut sites (T1-T2 deletions) and the events
424 involving reinsertion of the excised T1-T2 fragment. Depending on the orientation of the
425 reinsertion of the excised fragment, the "excision-reinsertion" events would include both the
426 2.05% of sequences with T1-T2 inversions and an equivalent number classed within the
427 12.42% of sequences with local deletions at both T1 and T2. Thus the excision-reinsertion
428 class would represent ~4% of sequences. That this value is comparable in magnitude to the
429 number of T1 to T2 deletions (excision without reinsertion (~3%) suggests that loss or
430 reinsertion of the excised fragmant in two-site cleavage events are equally likely.
431
432 4) The use of paired target sites permitted analysis of patterns of deletions at junctions
433 resulting from joining DNA ends with complementary 5'-overhangs (1-cut deletions) and at
434 those involving non-complementary 5'-overhangs (two-cut deletions). The two cut-sites are
435 separated by 256 bp and subtracting this from the lengths of the T1-T2 deletions permits
436 overlaying the pattern of the T1-T2 deletions with the 1-cut events. While some differences
437 are made clearly visible by this approach, the predominance of 4 bp deletions, followed by a
438 rapid fall-off with increasing deletion-length is strikingly similar. Thus, the
439 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 Single-Strand Annealing (SSA) recombination. It is particularly striking that these SSA events are a major repair class, greatly outnumbering the 2-cut deletions and approaching in frequency the EJ repair events (39% versus 61% of deletions respectively). It is possible that these events result from microhomology-mediate end-joining 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.

Acknowledgements

We thank Olivier Mathieu, Yoan Renaud and the members of the recombination group for their help and suggestions.
This work was supported by the Agence National de Recherche (ANR-16-CE91-0010-01 RecInChromatin), the European Marie Sklodowska Curie Actions (EU H2020-MSCA-ITN-2017-765212), the Centre National de Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (Inserm) and the Université Clermont Auvergne.

Data 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 analyse the DNA sequencing data produced in this study can be downloaded from:
https://github.com/chrlesw/dsbrepair.git

Supporting Information

Supporting Figures and Tables:

476 **Figure S1. Potential implication of microhomologies in deletion formation.**

477 **Figure S2. Cas12a-induced Indel patterns are confirmed by 2.8 Kb amplicon Nanopore**

478 **Figure S3. Cas12a-induced Indel patterns at the LPP2 locus from plants grown at 23°C and**
479 **29°C.**

480 **Figure S4. Cas12a-induced Indel patterns at the Chr3_92 locus from plants grown at 23°C and**
481 **29°C.**

482 **Table S1. DNA sequences used for sgRNA constructs.**

483 **Table S2. List of primers used in this study.**

484

485 ## Author Contributions

486 Designed Research and Analysis: SL, MG, YB, MH, CIW; Carried out experiments: SL, MG,
487 MH, SV, JV, CIW; Resources: MG, YB, CIW;
488 Writing, with contributions from all of the authors: SL, MG, CIW.

489
490 ## Conflict of Interest
491 We declare no financial or other conflicts of interest in this work.
492
493

494 ## Figure Legends

495 **Figure 1. Patterns of repair of paired, 5'-overhang DNA breaks at the LPP2 locus.**

496 (a) A schematic showing the two Cas12a target sites of the target LPP2 locus and the 400 bp
497 amplicon used for paired-end sequencing. The sequence, including the (TTT) PAM and the
498 23nt protospacer, of the targets is shown along with staggered red lines indicating the
499 expected cleavage of the target DNA strands. Cleavage of one or both targets is expected to
500 result in mutation of one or both targets (b). Coincident cleavage at both sites can result in
501 deletion of the DNA fragment between the two targets (c ; T1-T2 deletion), possibly
502 accompanied by its reinsertion in either the original (d ; not distinguishable from T1 and T2
503 mutated class (b)), or inverted (e ; T1-T2 inversion) orientation. (f) Sorting the paired-end
504 sequences spanning the target locus confirmed the existence and relative proportions of
505 these different classes of event induced in the genome of plants expressing Cas12a. (g) The
506 validity of the sorting is confimed by mapping the sequences of the different repair classes to
507 the Arabidopsis genome, or to an artifical *in silico* inversion locus for the T1-T2_inversion
508 class (bottom).
509

510 **Figure 2. Deletions predominate in the repair of 5'-overhang DNA breaks at the LPP2 locus.**

511 (a, b) Analysis of variants affecting 10 bp windows centered on the cut-sites confirms the
512 predominance of deletions in the repair products of Cas12a-induced breaks at the LPP2
513 locus. DNA from the -gRNA control plants contains almost exclusively base-substitutions,
514 presumably resulting from sequencing errors. (c) Plotting 1000 randomly-chosen deletions as
515 arcs (arcs from start to end of deletion) across the target locus (X-axis, horizontal orange bar),
516 shows that the deletions are almost exclusively focussed locally at the cut-sites or span
517 between them. This is confirmed by plotting the distributions of the numbers of deletions (d)
518 and insertions (e) by their lengths (blue curves are +gRNA and red are -gRNA controls).
519 Vertical orange lines in (c) mark the target-sites.
520
521 **Figure 3. Patterns of deletions at the LPP2_T1 and LPP2_T2 targets.**

522 (a) Frequency distributions of start-sites (blue) and end-sites (red) of 4 and 5nt deletion
523 affecting the 10 bp windows centered on the cut-sites. The red bars have been offset slightly
524 to the right for clarity. The target locus sequence with the cut-sites marked by staggered red
525 lines, is shown below the graph. Vertical dotted lines show the centers of the cut-sites to
526 facilitate lecture of the graphs. The majority of these deletions begin at the base pair
527 immediately upstream of the 5' ssDNA overhang produced by Cas12a cleavage of the
528 LPP2_T1 target, and the end points are found 4 and 5 bases downstream. The LPP2_T2 target
529 has 3 preferred start-points for 4 and 5nt deletions, including deletions starting 2 and 3 nt
530 upstream of the expected cut. (b) Frequency distributions of lengths of single-cut and
531 double-cut (transposed 256nt leftwards) confirm that the deletion patterns of the majority,
532 small local deletions aren't affected by the presence of cohesive (1-cut) versus non-cohesive
533 (2-cut) 5' ssDNA overhanging ends. Taken individually, frequency distributions of deletion
534 lengths (c) and deletion start- and end-points (d ; same presentation as in (a)) at the two cut-
535 sites are dominated by short local events, notwithstanding the presence of differences in the
536 details of the distributions at the two sites.
537
538 **Figure 4. Nanopore sequencing confirms patterns of Cas12a-induced deletions at the LPP2**
539 **targets.**

540 (a) Schematic showing the two Cas12a target sites of the target LPP2 locus and the 2.8 Kb
541 amplicon used for Nanopore sequencing. (b) Plot showing arcs linking each deletion start-
542 and end-point across the target locus (X-axis, horizontal orange bar). Vertical orange lines
543 mark the target-sites. The arcs are all of the same grey colour, the darker lines reflect
544 overlaying of multiple arcs. (c) Distributions of numbers of deletions by their lengths in
545 Cas12a+gRNA plants (blue curve) and Cas12a-gRNA controls (red curve). Deletions of less
546 than 5 bp were excluded due to the high background of small indel errors in Nanopore
547 sequences.
548
549 **Figure 5. Patterns of Cas12a-induced deletions at the Chr3_92 target locus.**

550 (a) Schematic showing the two Cas12a target sites flanked by direct-repeat sequences of the
551 target Chr3_92 locus and the 2.979 Kb amplicon used for Nanopore sequencing. The target
552 sequences, including the (TTT) PAM and the 23nt protospacer, are shown along with
553 staggered red lines indicating the expected cleavage of the target DNA strands. (b) Numbers
554 of sequences and the deletions and insertions expressed as fractions of mapped to the
555 Chr3_92 locus from Cas12a+gRNA (blue) and Cas12a-gRNA plants (red). Indels of less than
556 5 bp were excluded due to the high background of small indel errors in Nanopore sequences.
557 (c) As seen at the LPP2 locus, the majority of Cas12a-induced indels are deletions (b, c).
558
559 **Figure 6. The presence of flanking direct-repeats generates a major new class of Single-Strand**
560 **Annealing deletions.**

561 (a) Distributions of numbers of deletions by their lengths in Cas12a+gRNA plants (blue
562 curve) and Cas12a-gRNA controls (red curve). Deletions of less than 5 bp were excluded due
563 to the high background of small indel errors in Nanopore sequences. A major (39.4%), novel
564 class of 742 bp deletions is created by the presence of the direct repeats flanking the Cas12a
565 target sites. (c) Deletion start-end arc-plots confirm that these 742 bp deletions (red curves)
566 result from in-step recombination between the flanking direct-repeats, as expected from SSA
567 recombination. This is strikingly different from the patters observed at the LPP2 locus in the
568 absence of the flanking repeats (inset in (c), repeated from Figure 4).
569

570 ## Supporting Figure Legends

571 **Supporting Figure S1. Potential implication of microhomologies in deletion formation.**

572 (a) The maximum number of nt of microhomology, from 0 to 9 nt, by comparing the
573 sequences of the upstream DSB-end with the downstream end of the deletion, and the
574 downstream DSB-end with the upstream end of the deletion (black bars connected by dotted
575 lines). The lengths of such microhomologies, potentially involved in the deletion, were
576 determined for each deletion and the mean microhomology values calculated for each
577 deletion start-point across the target sequence. These are plotted as vertical blue bars for the
578 sequences upstream of the LPP2_T1 (b) and LPP2_T2 (c) targets. The number of deletions
579 starting at each start-point are overlaid (red curve) and the cut-site (vertical dotted lines) are
580 included for reference.
581
582 **Supporting Figure S2. Cas12a-induced Indel patterns are confirmed by 2.8 Kb amplicon**
583 **Nanopore sequencing.**

584 (a) Numbers of sequences and the deletions and insertions expressed as fractions of mapped
585 to the LPP2 locus from Cas12a+gRNA (blue) and Cas12a-gRNA plants (red). Indels of less
586 than 5 bp were excluded due to the high background of small indel errors in Nanopore
587 sequences. As seen from the Illumina sequencing of the 400 bp amplicon (Fig. 2), the

588 majority of Cas12a-induced indels are deletions (a, b). Deletion-length frequency
589 distributions from the 400 bp paired-end Illumina (blue curve) and 2.8 Kb Nanopore (red
590 curve) sequencing data (c).
591
592 **Supporting Figure S3. Cas12a-induced Indel patterns at the LPP2 locus from plants grown at**
593 **23°C and 29°C.**

594 Illumina sequencing of the 400bp LPP2 locus amplicon from plantlets grown at 23°C and
595 29°C shows similar proportions of the different deletion classes (a). This is also seen in the
596 distributions of deletion-lengths (b) and insertion-lengths (c) from plantlets grown at 29°
597 (blue curves) and 23° (red curves).
598
599 **Supporting Figure S4. Cas12a-induced Indel patterns at the Chr3_92 locus from plants grown**
600 **at 23°C and 29°C.**

601 Nanopore sequencing of the 2.8 Kb Chr3_92 locus amplicon from plantlets grown at 23°C
602 and 29°C shows similar proportions of the different deletion classes (a). This is also seen in
603 the distributions of deletion-lengths (b) and insertion-lengths (c) from plantlets grown at 29°
604 (blue curves) and 23° (red curves).
605
606 **Table S1. DNA sequences used for sgRNA constructs.**

607

608 **Table S2. List of primers used in this study.**

609

610 ## Bibliography

611

612 Atia, M., Jiang, W., Sedeek, K., Butt, H., Mahfouz, M., 2024. Crop bioengineering via gene editing:
613 reshaping the future of agriculture. Plant Cell Rep 43, 98. https://doi.org/10.1007/s00299-024-03183-1
614 Bechtold, N., Ellis, J., Pelletier, G., 1993. In planta Agrobacterium-mediated gene transfer by
615 infiltration of adult Arabidopsis thaliana plants. C. R. Acad. Sci. Paris, Life Sci. 1194–1199.
616 Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids
617 Research 27, 573–580. https://doi.org/10.1093/nar/27.2.573
618 Bernabé-Orts, J.M., Casas-Rodrigo, I., Minguet, E.G., Landolfi, V., Garcia-Carpintero, V., Gianoglio,
619 S., Vázquez-Vilar, M., Granell, A., Orzaez, D., 2019. Assessment of Cas12a-mediated gene editing
620 efficiency in plants. Plant Biotechnol J 17, 1971–1984. https://doi.org/10.1111/pbi.13113
621 Bothmer, A., Phadke, T., Barrera, L.A., Margulies, C.M., Lee, C.S., Buquicchio, F., Moss, S.,
622 Abdulkerim, H.S., Selleck, W., Jayaram, H., Myer, V.E., Cotta-Ramusino, C., 2017. Characterization
623 of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus.
624 Nat Commun 8, 13905. https://doi.org/10.1038/ncomms13905
625 Bushnell, B., Rood, J., Singer, E., 2017. BBMerge – Accurate paired shotgun read merging via
626 overlap. PLOS ONE 12, e0185056. https://doi.org/10.1371/journal.pone.0185056
627 Chang, H.H.Y., Pannunzio, N.R., Adachi, N., Lieber, M.R., 2017. Non-homologous DNA end joining
628 and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol 18, 495–506.
629 https://doi.org/10.1038/nrm.2017.48
630 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. Sci Rep 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 : for cell and molecular biology 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. Nat Commun 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 Biotechnol J 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, NY) 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 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. 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 Biotechnol J 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. Nat Biotechnol 1–15. 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., M, T.R., P, T.M., White, M.F., Yakunin, A.F., Garrett, R.A., Oost, J. van der, Backofen, R., 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, Yingxiao, Chen, H., Kang, M., Bao, Y., Zheng, X., Deng, K., Zhang, T., Salcedo, V., Wang, K., Zhang, Yong, Qi, Y., 2019. Application of CRISPR-Cas12a temperature sensitivity for improved genome editing in rice, maize, and Arabidopsis. - PubMed - NCBI. 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

690 the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nature Biotechnology
691 31, 691–693. https://doi.org/10.1038/nbt.2655
692 Přibylová, A., Fischer, L., 2024. How to use CRISPR/Cas9 in plants: from target site selection to
693 DNA repair. Journal of Experimental Botany erae147. https://doi.org/10.1093/jxb/erae147
694 Schiml, S., Fauser, F., Puchta, H., 2016. Repair of adjacent single-strand breaks is often accompanied
695 by the formation of tandem sequence duplications in plant genomes. Proceedings of the National
696 Academy of Sciences of the United States of America 113, 7266–7271.
697 https://doi.org/10.1073/pnas.1603823113
698 Schiml, S., Fauser, F., Puchta, H., 2014. The CRISPR/Cas system can be used as nuclease for in planta
699 gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable
700 progeny. The Plant journal : for cell and molecular biology 80, 1139–1150.
701 https://doi.org/10.1111/tpj.12704
702 Schindele, P., Puchta, H., 2019. Engineering CRISPR/ LbCas12a for highly efficient, temperature-
703 tolerant plant gene editing. Plant Biotechnology Journal 17, 1971–3. https://doi.org/10.1111/pbi.13275
704 Scully, R., Panday, A., Elango, R., Willis, N.A., 2019. DNA double-strand break repair-pathway
705 choice in somatic mammalian cells. Nat Rev Mol Cell Biol 20, 698–714.
706 https://doi.org/10.1038/s41580-019-0152-0
707 Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.-L., Gao,
708 C., 2013. Targeted genome modification of crop plants using a CRISPR-Cas system. Nature
709 Biotechnology 31, 686–688. https://doi.org/10.1038/nbt.2650
710 Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E.,
711 Minakhin, L., Joung, J., Konermann, S., Severinov, K., Zhang, F., Koonin, E.V., 2015. Discovery and
712 Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Molecular Cell 60, 385–397.
713 https://doi.org/10.1016/j.molcel.2015.10.008
714 Strohkendl, I., Saifuddin, F.A., Rybarski, J.R., Finkelstein, I.J., Russell, R., 2018. Kinetic Basis for
715 DNA Target Specificity of CRISPR-Cas12a. Molecular Cell 71, 816-824.e3.
716 https://doi.org/10.1016/j.molcel.2018.06.043
717 Swarts, D.C., Jinek, M., 2018. Cas9 versus Cas12a/Cpf1: Structure–function comparisons and
718 implications for genome editing. WIREs RNA 9, e1481. https://doi.org/10.1002/wrna.1481
719 Vu, G.T.H., Cao, H.X., Fauser, F., Reiss, B., Puchta, H., Schubert, I., 2017. Endogenous sequence
720 patterns predispose the repair modes of CRISPR/Cas9-induced DNA double-stranded breaks in
721 Arabidopsis thaliana. Plant J 92, 57–67. https://doi.org/10.1111/tpj.13634
722 Weiss, T., Kumar, J., Chen, C., Guo, S., Schlegel, O., Lutterman, J., Ling, K., Zhang, F., 2024. Dual
723 activities of an X-family DNA polymerase regulate CRISPR-induced insertional mutagenesis across
724 species. Nat Commun 15, 6293. https://doi.org/10.1038/s41467-024-50676-4
725 Wolter, F., Schindele, P., Beying, N., Scheben, A., Puchta, H., 2021. Different DNA repair pathways
726 are involved in single-strand break-induced genomic changes in plants. The Plant Cell 33, 3454–3469.
727 https://doi.org/10.1093/plcell/koab204
728 Wright, W.D., Shah, S.S., Heyer, W.-D., 2018. Homologous recombination and the repair of DNA
729 double-strand breaks. J Biol Chem 293, 10524–10535. https://doi.org/10.1074/jbc.TM118.000372
730 Xue, C., Greene, E.C., 2021. DNA Repair Pathway Choices in CRISPR-Cas9-Mediated Genome
731 Editing. Trends in Genetics 37, 639–656. https://doi.org/10.1016/j.tig.2021.02.008
732 Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P.,
733 Volz, S.E., Joung, J., Oost, J. van der, Regev, A., Koonin, E.V., Zhang, F., 2015. Cpf1 is a single
734 RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771.
735 https://doi.org/10.1016/j.cell.2015.09.038
736 Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad,
737 N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S., Wu, W.Y., Scott, D.A., Severinov, K., Oost,
738 J. van der, Zhang, F., 2017. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array.
739 Nature Biotechnology 35, 31–34. https://doi.org/10.1038/nbt.3737
740

Figure 1

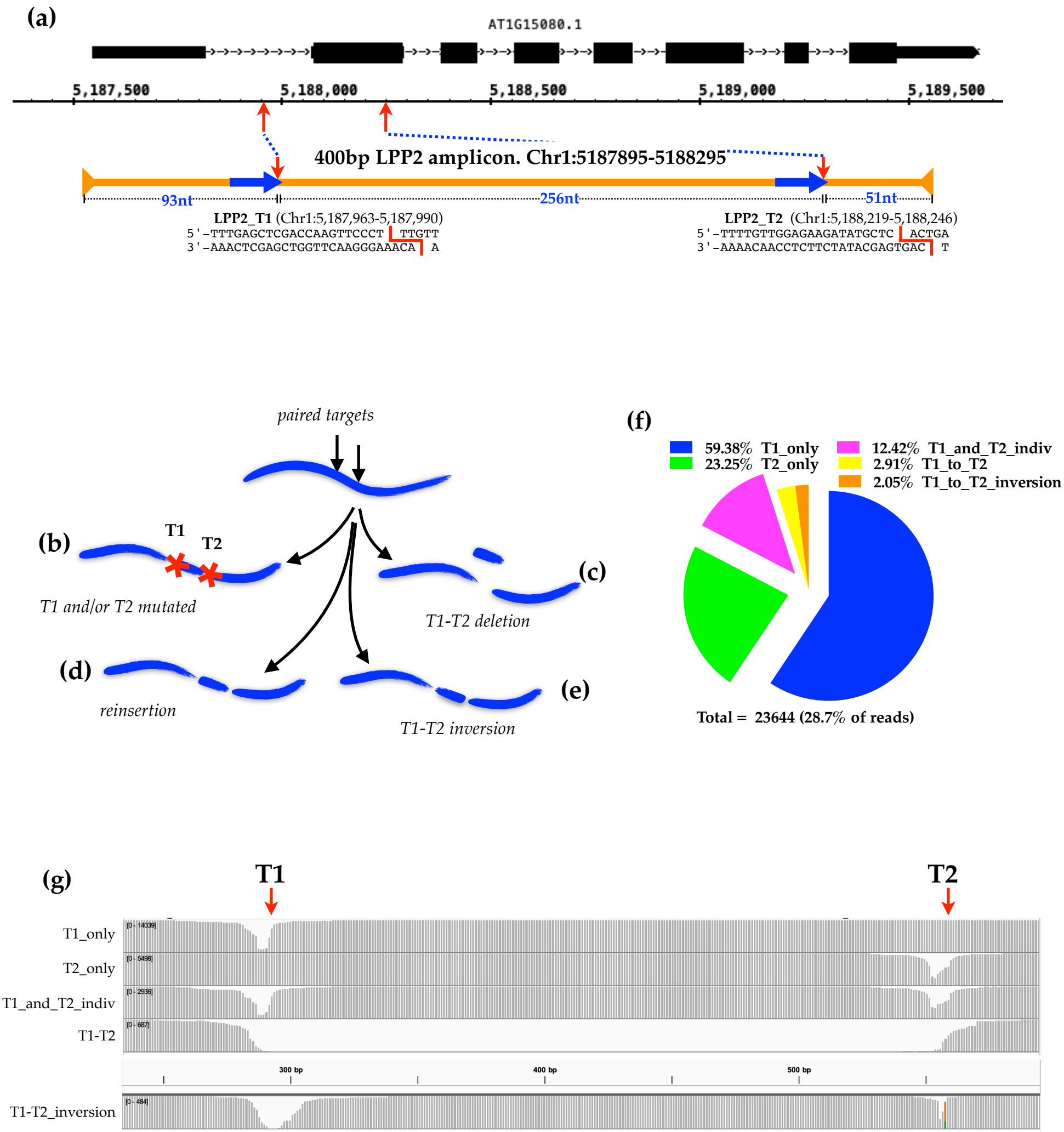
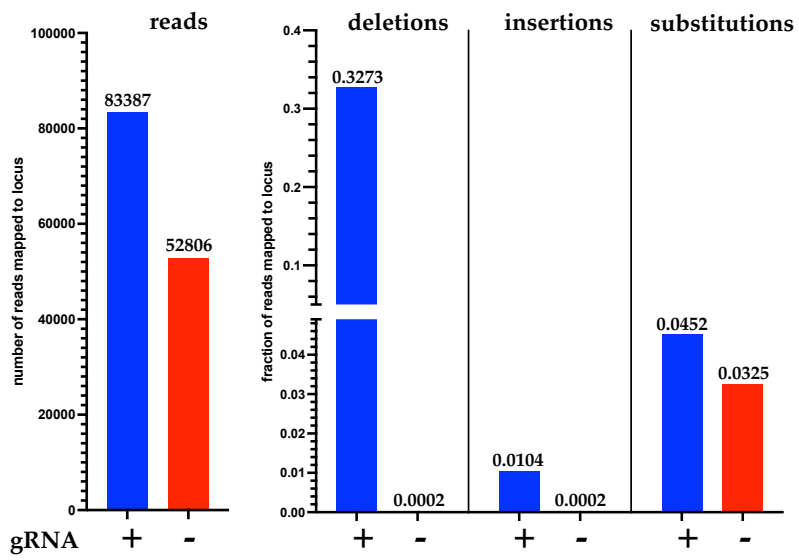
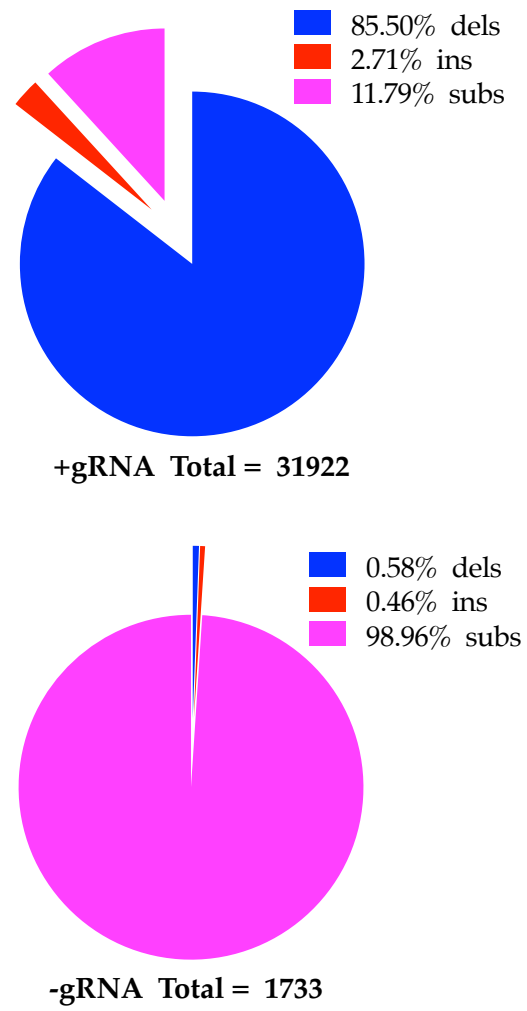

(a)
AT1G15080.1
5,187,500
5,188,000
5,188,500
5,189,000
5,189,500
400bp LPP2 amplicon. Chr1:5187895-5188295
93nt
256nt
51nt
LPP2_T1 (Chr1:5,187,963-5,187,990)
5'-TTTGAGCTCGACCAAGTTCCCT TTGTT
3'-AAACTCGAGCTGGTTCAAGGGAAACA A
LPP2_T2 (Chr1:5,188,219-5,188,246)
5'-TTTTGTTGGAGAAGATATGCTC ACTGA
3'-AAAACAACCTCTTCTATACGAGTGAC T
paired targets
(b)
T1
T2
T1 and/or T2 mutated
(c)
T1-T2 deletion
(d)
reinsertion
(e)
T1-T2 inversion
(f)
59.38% T1_only
23.25% T2_only
12.42% T1_and_T2_indiv
2.91% T1_to_T2
2.05% T1_to_T2_inversion
Total = 23644 (28.7% of reads)
(g)
T1
T2
T1_only
T2_only
T1_and_T2_indiv
T1-T2
300 bp
400 bp
500 bp
T1-T2_inversion

Figure 2

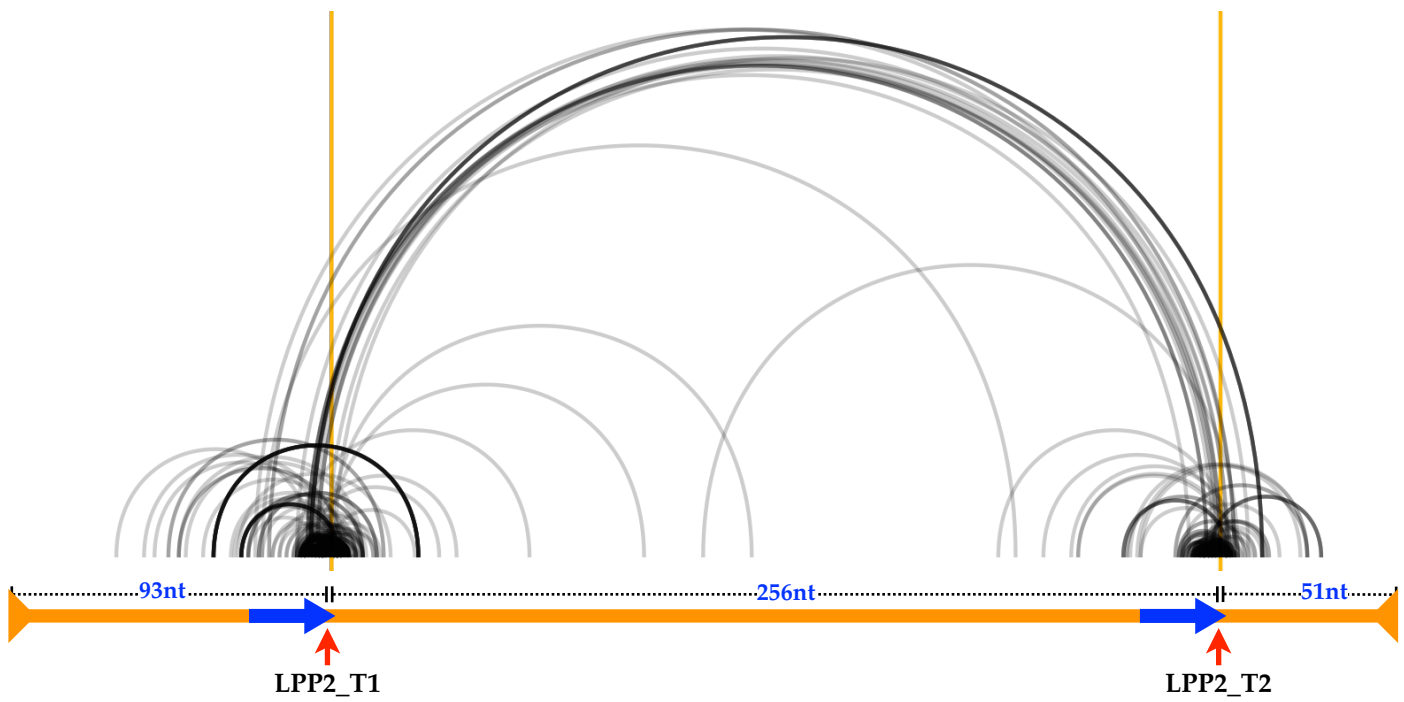
(a)



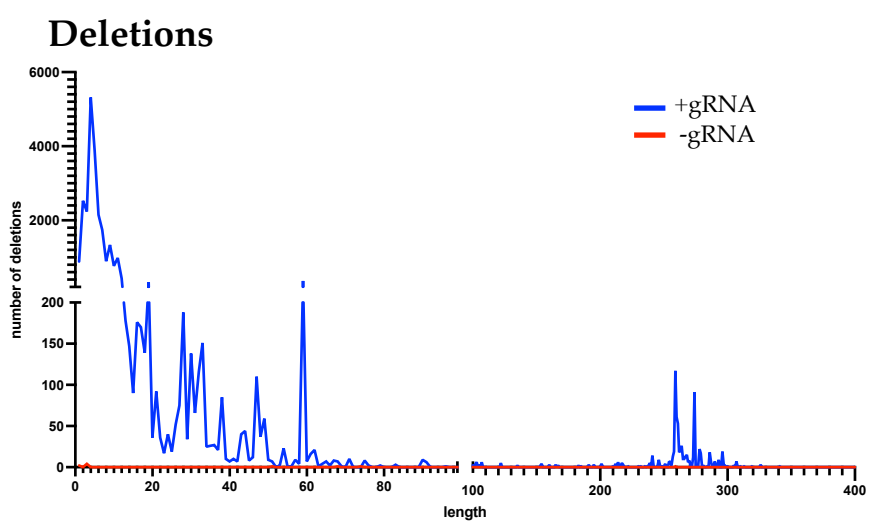
(b)



(c)



(d)



(e)

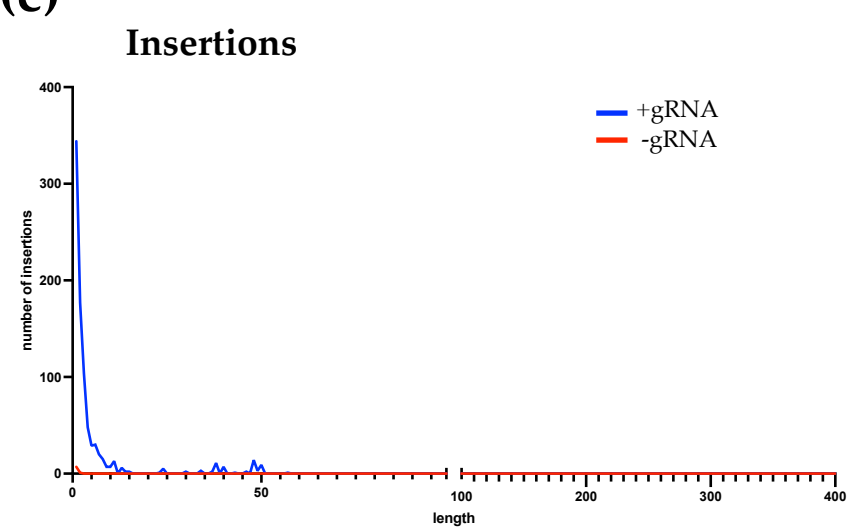


Figure 3

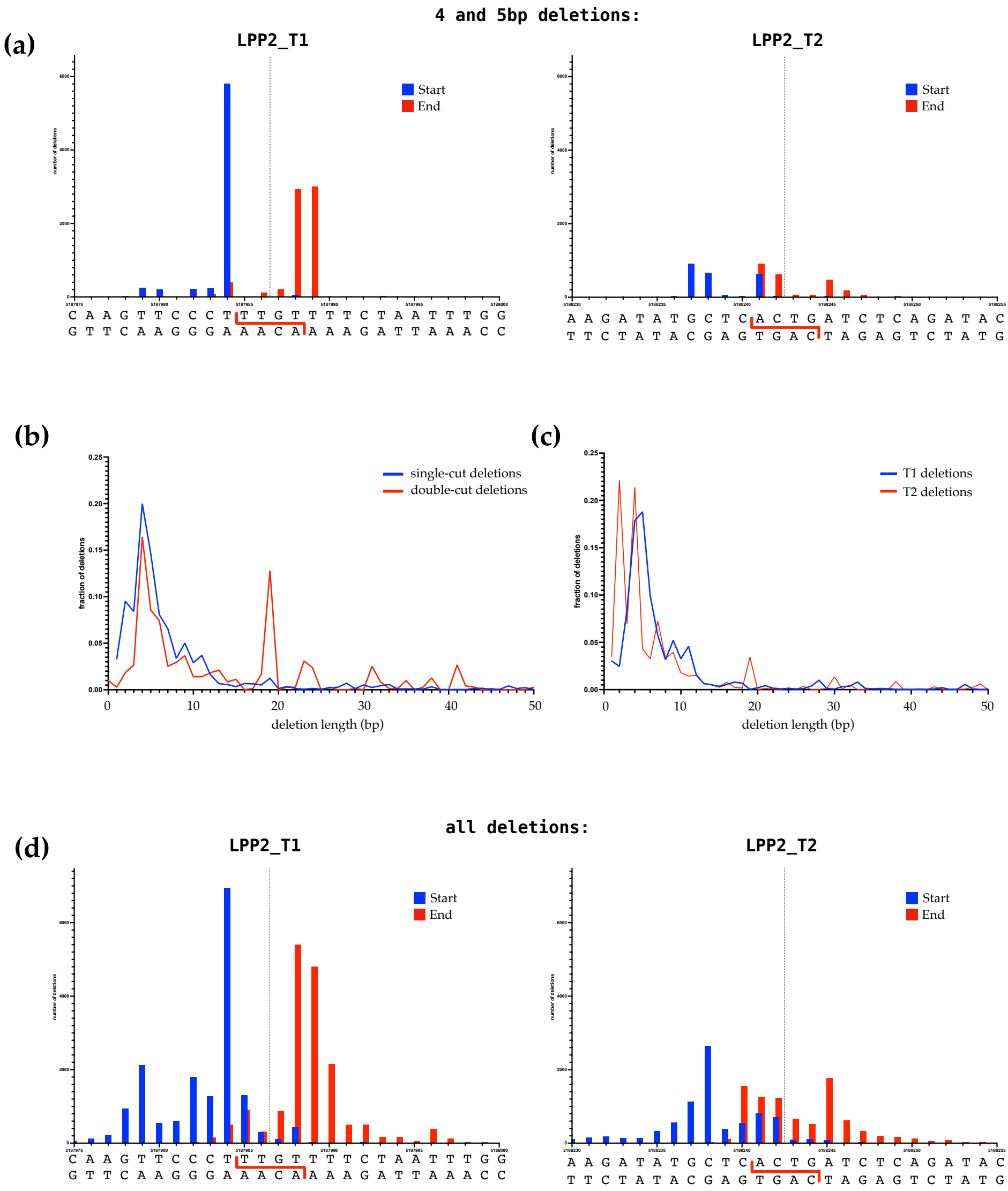
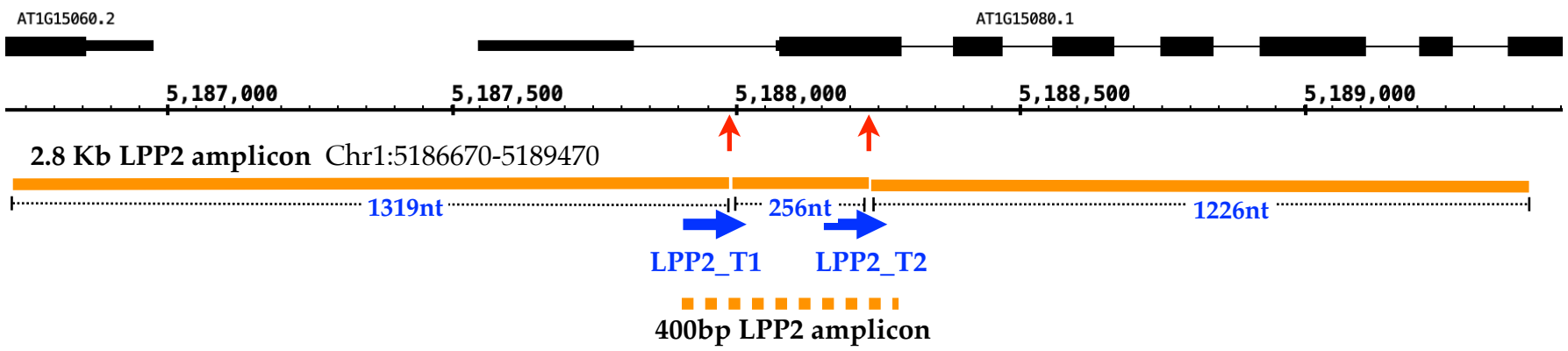
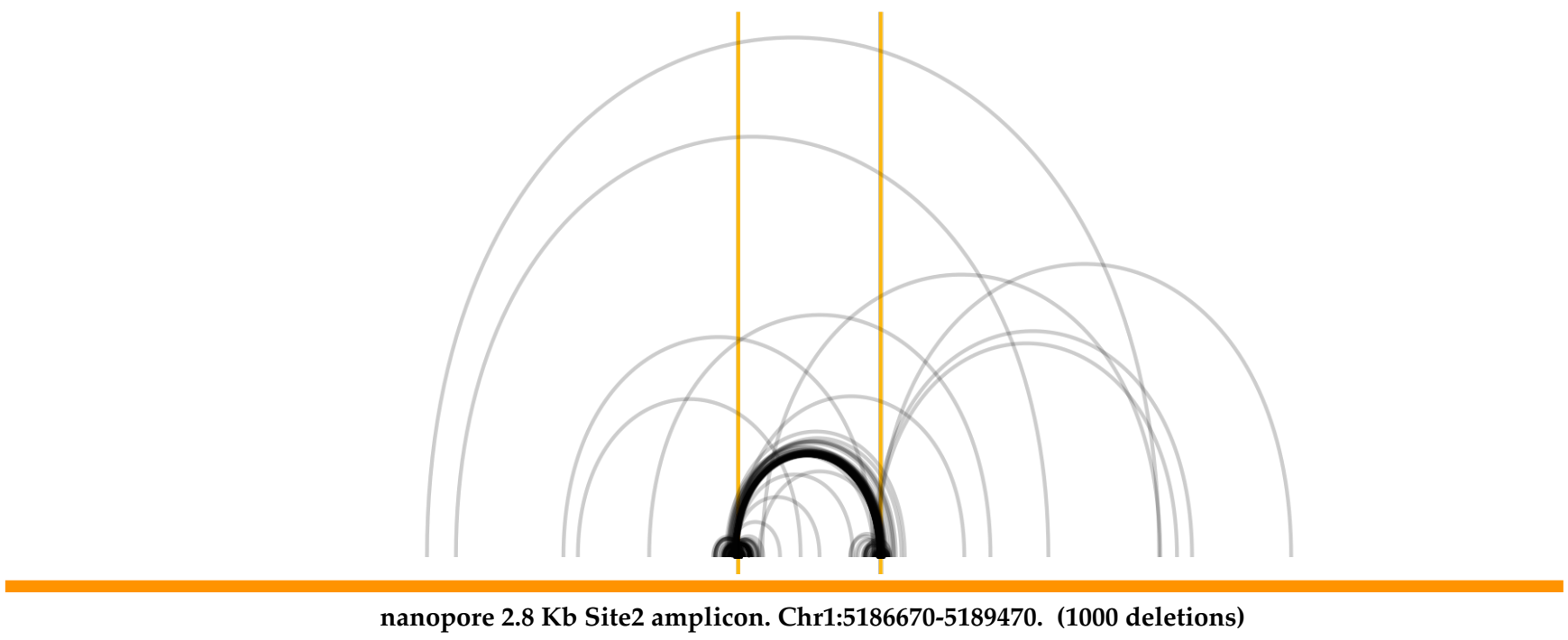

4 and 5bp deletions:
(a)
LPP2_T1
LPP2_T2
Start
End
number of deletions
C A A G T T C C C T T T G T T T T C T A A T T T G G
G T T C A A G G G A A A C A A A A G A T T A A A C C
A A G A T A T G C T C A C T G A T C T C A G A T A C
T T C T A T A C G A G T G A C T A G A G T C T A T G
(b)
single-cut deletions
double-cut deletions
fraction of deletions
deletion length (bp)
(c)
T1 deletions
T2 deletions
fraction of deletions
deletion length (bp)
all deletions:
(d)
LPP2_T1
LPP2_T2
Start
End
number of deletions
C A A G T T C C C T T T G T T T T C T A A T T T G G
G T T C A A G G G A A A C A A A A G A T T A A A C C
A A G A T A T G C T C A C T G A T C T C A G A T A C
T T C T A T A C G A G T G A C T A G A G T C T A T G

Figure 4

(a)


AT1G15060.2
AT1G15080.1
5,187,000
5,187,500
5,188,000
5,188,500
5,189,000
2.8 Kb LPP2 amplicon Chr1:5186670-5189470
1319nt
256nt
1226nt
LPP2_T1
LPP2_T2
400bp LPP2 amplicon

(b)


nanopore 2.8 Kb Site2 amplicon. Chr1:5186670-5189470. (1000 deletions)

(c)

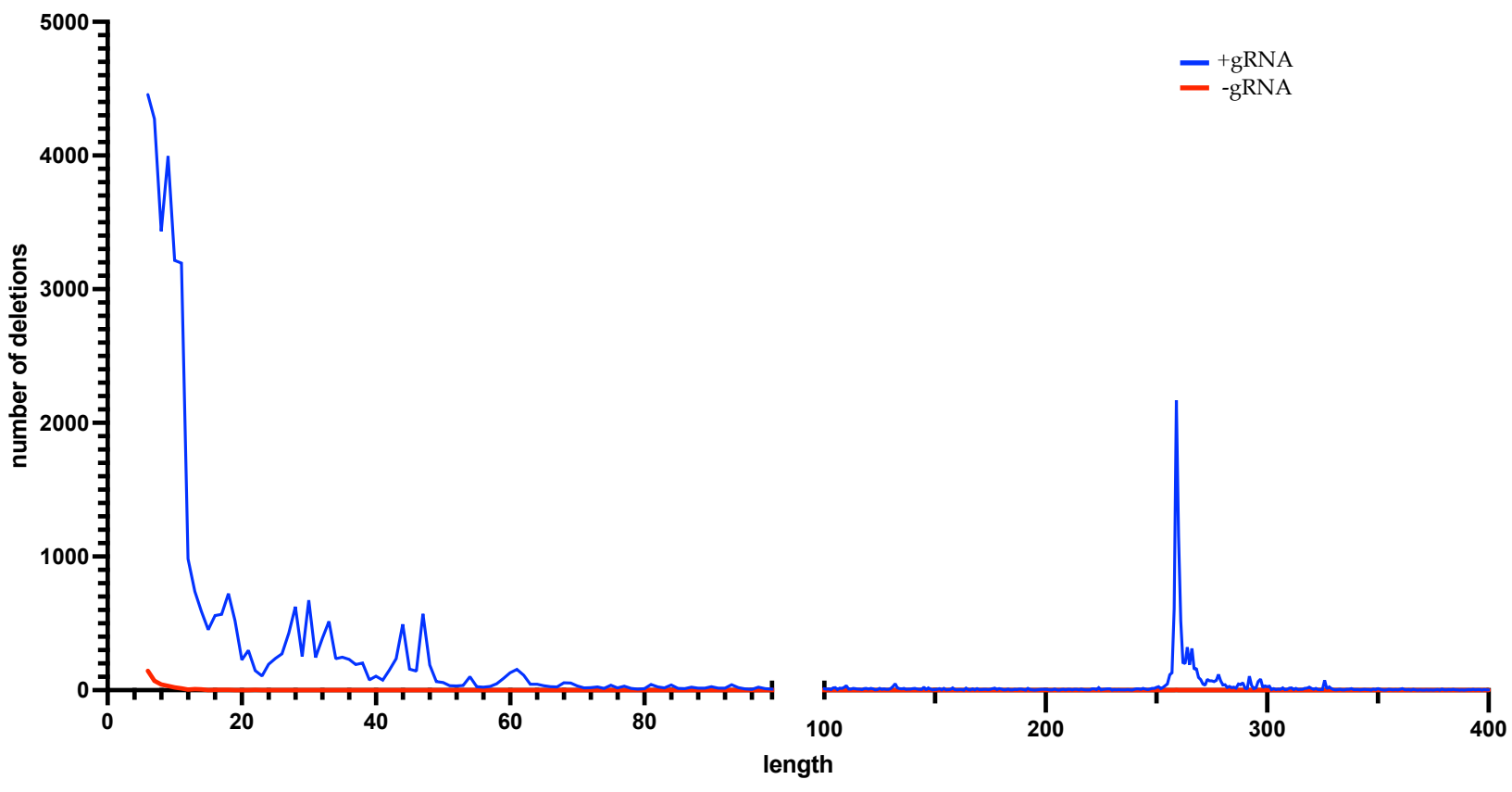
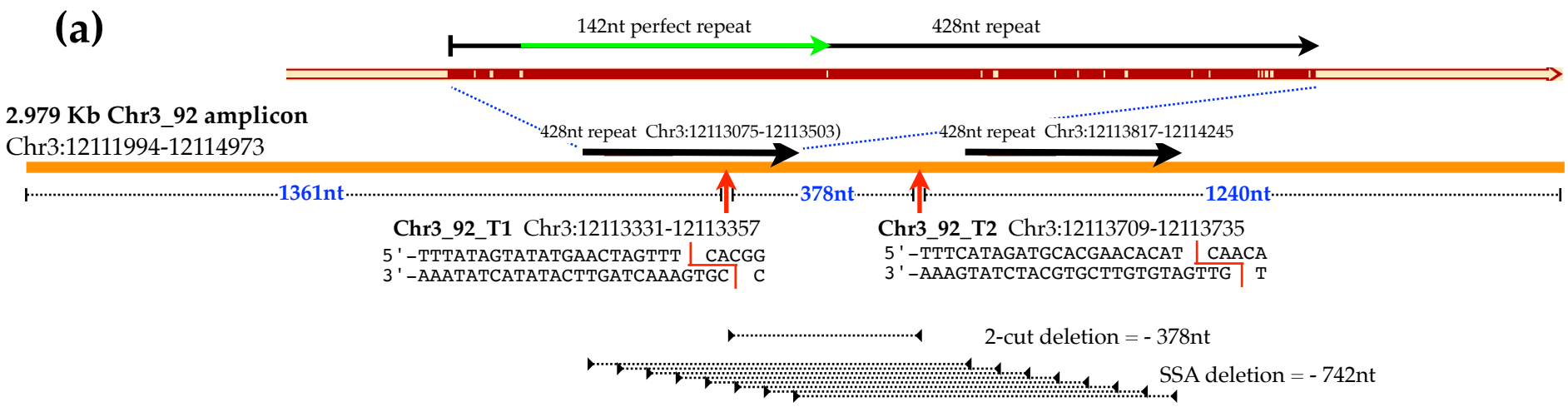
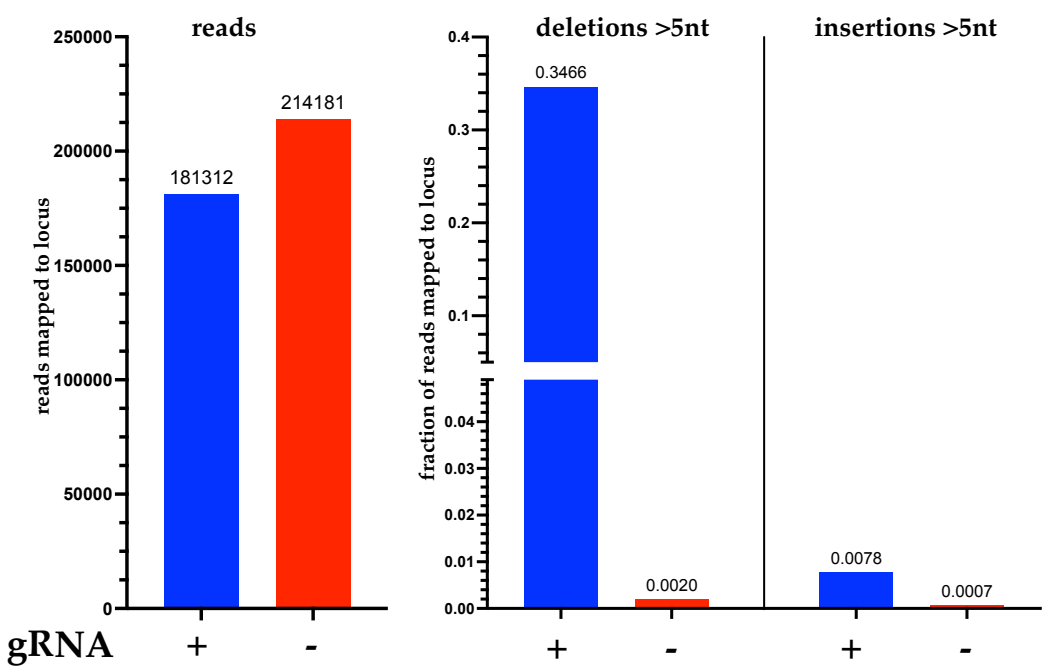

+gRNA
-gRNA
number of deletions
5000
4000
3000
2000
1000
0
0
20
40
60
80
100
200
300
400
length

Figure 5


(a)
142nt perfect repeat
428nt repeat
2.979 Kb Chr3_92 amplicon
Chr3:12111994-12114973
428nt repeat Chr3:12113075-12113503)
428nt repeat Chr3:12113817-12114245
1361nt
378nt
1240nt
Chr3_92_T1 Chr3:12113331-12113357
5'-TTTATAGTATATGAACTAGTTT CACGG
3'-AAATATCATATACTTGATCAAAGTGC C
Chr3_92_T2 Chr3:12113709-12113735
5'-TTTCATAGATGCACGAACACAT CAACA
3'-AAAGTATCTACGTGCTTGTGTAGTTG T
2-cut deletion = - 378nt
SSA deletion = - 742nt


(b)
reads
deletions >5nt
insertions >5nt
reads mapped to locus
fraction of reads mapped to locus
181312
214181
0.3466
0.0020
0.0078
0.0007
gRNA
+
-

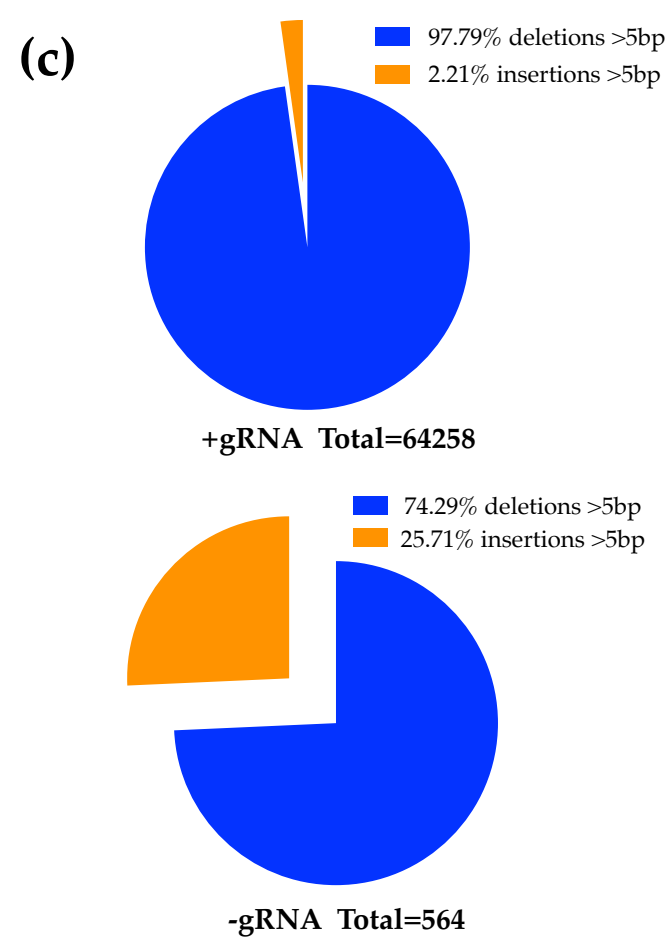
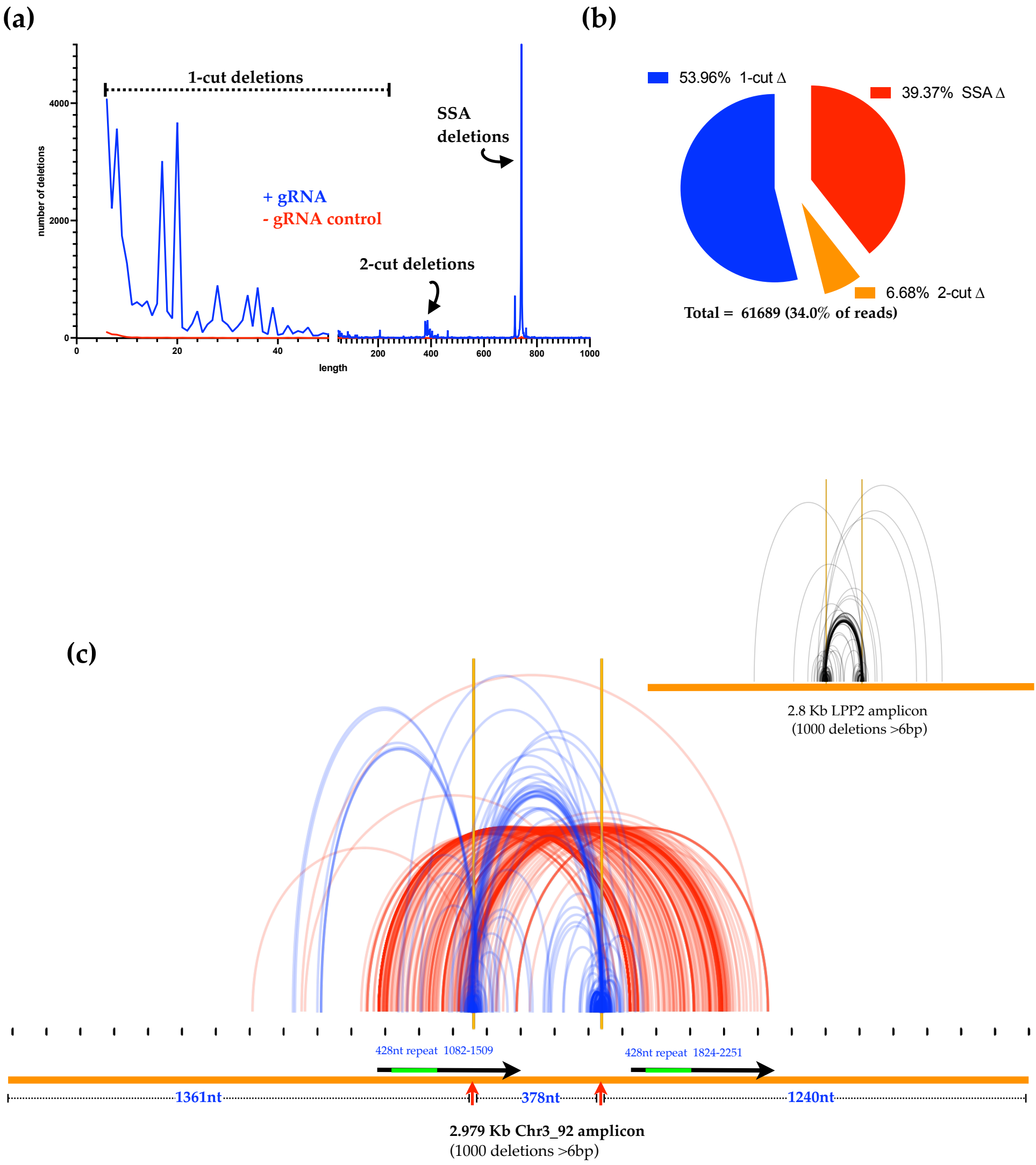

(c)
97.79% deletions >5bp
2.21% insertions >5bp
+gRNA Total=64258
74.29% deletions >5bp
25.71% insertions >5bp
-gRNA Total=564

Figure 6


(a)
1-cut deletions
SSA deletions
+ gRNA
- gRNA control
2-cut deletions
number of deletions
length
(b)
53.96% 1-cut Δ
39.37% SSA Δ
6.68% 2-cut Δ
Total = 61689 (34.0% of reads)
(c)
2.8 Kb LPP2 amplicon
(1000 deletions >6bp)
428nt repeat 1082-1509
428nt repeat 1824-2251
1361nt
378nt
1240nt
2.979 Kb Chr3_92 amplicon
(1000 deletions >6bp)