Repair of adjacent single-strand breaks is often accompanied by the formation of tandem sequence duplications in plant genomes

Simon Schiml[a], Friedrich Fauser[a,1], and Holger Puchta[a,2]

[a]Botanical Institute II, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

Edited by James A. Birchler, University of Missouri–Columbia, Columbia, MO, and approved May 16, 2016 (received for review March 8, 2016)

Duplication of existing sequences is a major mechanism of genome evolution. It has been previously shown that duplications can occur by replication slippage, unequal sister chromatid exchange, homologous recombination, and aberrant double-strand break-induced synthesis-dependent strand annealing reactions. In a recent study, the abundant presence of short direct repeats was documented by comparative bioinformatics analysis of different rice genomes, and the hypothesis was put forward that such duplications might arise due to the concerted repair of adjacent single-strand breaks (SSBs). Applying the CRISPR/Cas9 technology, we were able to test this hypothesis experimentally in the model plant *Arabidopsis thaliana*. Using a Cas9 nickase to induce adjacent genomic SSBs in different regions of the genome (genic, intergenic, and heterochromatic) and at different distances (~20, 50, and 100 bps), we analyzed the repair outcomes by deep sequencing. In addition to deletions, we regularly detected the formation of direct repeats close to the break sites, independent of the genomic context. The formation of these duplications as well as deletions may be associated with the presence of microhomologies. Most interestingly, we found that even the induction of two SSBs on the same DNA strand can cause genome alterations, albeit at a much lower level. Because such a scenario reflects a natural step during nucleotide excision repair, and given that the germline is set aside only late during development in plants, the repair of adjacent SSBs indeed seems to have an important influence on the shaping of plant genomes during evolution.

double-strand break repair | CRISPR/Cas | genome editing | homologous recombination

The opportunity to experimentally investigate the outcome of double-strand break (DSB) repair in eukaryotic organisms was enabled by the use of site-specific endonucleases targeting unique genomic sequences. Early work using the meganuclease I-SceI more than 20 y ago demonstrated that the induction of a DSB leads to enhanced homologous recombination (HR) in multicellular eukaryotes (1). This finding led to the application of I-SceI as a tool for detailed analysis of genomic changes that might occur due to DSB repair via nonhomologous end-joining (NHEJ) in plants. Thus, it was possible to demonstrate that along with inducing various kinds of deletions (2, 3), DSB repair also can be associated with the integration of T-DNA, as well as with the copying of genomic sequences into the break (4, 5).

Recently, a novel type of programmable nuclease was introduced by the discovery of the molecular mechanism of the bacterial clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system (6, 7). In this system, the endonuclease Cas9 is guided by two specialized RNAs (CRISPR RNA and transactivating CRISPR RNA) via direct base-pairing to bind and cleave invading harmful DNA. It has been demonstrated in vitro and in vivo that by fusing the two RNAs to a so-called single-guide RNA (sgRNA), the system can serve as a highly efficient and precisely programmable nuclease for genome engineering (6, 8, 9). Furthermore, Cas9 also can be used as a single-strand break (SSB)-inducing nickase by introducing a specific point mutation in one of the nuclease domains (D10A or H840A).

Owing to its simple principle and applicability to a vast range of organisms, RNA-guided Cas9 was successfully established as a valuable tool for DSB-mediated genome engineering. We, among other groups, were able to demonstrate both the efficient induction of heritable targeted mutagenesis events by Cas9 in plants (10, 11) and HR-mediated in planta gene targeting (12). In contrast to the induction of a DSB by the Cas9 nuclease, the induction of a single SSB by the Cas9 nickase is not mutagenic (10); however, when used in a "paired nickase" approach, in which a DSB is created by the combined action of two sgRNAs and the Cas9 nickase in close proximity on both DNA strands (13), mutations can be induced that also can be transferred to the germline of plants (12).

It is commonly assumed that natural duplications in eukaryotic genomes originate from unequal sister chromatid exchange; from intrachromosomal, allelic, or ectopic recombination during DSB repair (14, 15); or from the action of transposable elements (16–18). However, the most common form of insertions consists of tandem duplications, which can vary greatly in size and repeat number and contribute to genome expansion and the development of novel gene functions (19). Classically, the generation of tandem duplications is explained by replication slippage (20, 21), although the initial microhomology required for priming a slipping replication cannot always be found. More recent models involve the participation of DSBs and a subsequent NHEJ-mediated formation of tandem duplications (22). A recent bioinformatics study of the genomes of different rice species proposed a model in which adjacent SSBs in opposing strands contribute to the formation of tandem duplications (23), based on the interference of long patch-mediated repair of two SSBs (24).

Significance

A major principle in genome evolution is the duplication of existing sequences. Various mechanisms linked to DNA replication, homologous recombination, or double-strand break repair have been elucidated over the years. Using the single-strand break (SSB)-inducing nickase variant of the CRISPR/Cas system, we examined genomic alterations following the occurrence of two adjacent SSBs in the model plant *Arabidopsis thaliana*. In addition to deletions, tandem duplications were found regularly, with both outcomes indicating partial involvement of microhomologies. These results demonstrate that the repair of adjacent SSBs plays a surprising and important role in shaping plant genomes during evolution.

Author contributions: S.S., F.F., and H.P. designed research; S.S. performed research; S.S. and H.P. analyzed data; and S.S., F.F., and H.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data have been deposited in the Sequence Read Archive (SRA) database (accession ID SRR3614304).

[1]Present address: Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305.

[2]To whom correspondence should be addressed. Email: holger.puchta@kit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1603823113/-/DCSupplemental.

Here we used the Cas9 nickase to experimentally test this theoretically proposed model for the formation of tandem duplications. Paired SSBs were introduced in three different genomic locations and at different distances in *Arabidopsis thaliana*, and the results were examined by next-generation sequencing (NGS). We found clear evidence that supports the model for SSB-mediated duplications, as well as the partial involvement of microhomologies in the formation of insertions and deletions. Intriguingly, even the induction of paired SSBs on the same DNA strand led to similar alterations, although at much lower frequencies.

Results

To assess genomic changes caused by the induction of adjacent SSBs, we chose three different genomic loci on three different chromosomes of *A. thaliana* (Fig. 1). We selected the *ADH1* gene (alcohol dehydrogenase 1; AT1G77120) on chromosome 1 to represent a constitutively expressed coding sequence, a long intergenic region of almost 8 kb between the genes AT2G22620 (Rhamnogalacturonate lyase family protein) and AT2G22630 (Agamous-like 17) on chromosome 2 and a locus closely adjacent to the centromere of chromosome 3. The latter can be considered a heterochromatic region owing to its high degree of C methylation and its enrichment of H3.1, H3K9me2, and H3K27me1 as indicated by respective genome-wide mapping data (25, 26). Paired Cas9-D10A nickase constructs were designed and cloned for all three loci as described previously (12). These constructs enabled the paired induction of SSBs at distances of ∼20, 50, and 100 bps on each strand of the DNA, all creating 5′ overhangs. In addition, for every distance, constructs of paired SSBs on the same DNA strand were cloned as well (Fig. 1).

The constructs were transformed stably into wild type *A. thaliana* plants using *Agrobacterium*-mediated transformation. Primary transformed plants were grown on selection media, and for each construct, DNA was extracted from a batch of 2-wk-old seedlings. The respective sgRNA target sites were amplified by PCR, and the pooled amplicons were subjected to NGS to detect the outcomes of the repair of induced breaks. The datasets were mapped to their respective reference sequence, and the differences from the reference, indicated by deletions, insertions, or substitutions, were calculated either by their position in the reference or by the length of the alteration.

DSB Induction by Paired SSBs on Opposite DNA Strands. Using the foregoing experimental setup, we first looked at overall mutation frequencies. To our surprise, we found only very low frequencies for all three genomic sites at the distance of ∼20 bps, with 2.2% of reads exhibiting mutations for the genic locus, 6.1% exhibiting mutations for the intergenic locus, and 1.3% exhibiting mutations for the heterochromatic locus. This finding stands in contrast to the results of previous studies in mammalian cells, in which even shorter distances of paired SSBs produced considerable amounts of mutagenesis (13). At a distance of ∼50 bp, we found frequencies of 63.1% for the genic locus, 30.4% for the intergenic locus, and 27.7% for the heterochromatic locus, values consistent with those reported in our earlier study on the use of paired Cas9 nickases in plants (12). When SSBs were introduced at a distance of ∼100 bps, the frequencies of detected mutations remained somewhat constant for the intergenic locus (26.6%) and the heterochromatic locus (30.5%). For the genic locus, we found a surprisingly high frequency of 77%. This result was due mainly to the presence of a certain kind of deletion of 105 bps, which alone represented 17.6% of all reads in this respective dataset. The details of this deletion will be discussed below.

Generally, for all three loci and independent of the distance between SSBs, deletions were more frequent than insertions. A general overview of the size of the deletions and insertions is presented in Fig. 2. The sizes of deletions correlate with the SSB distance; that is, deletions of >50 bps are scarce for the 50-bps SSB distance and become more frequent at the 100-bps SSB distance. The largest numbers of insertions are found for the 50-bps distance, with length typically ranging from ∼20 to 50 bps. The distribution of mutations based on their position along the reference sequence is shown in Figs. S2–S4. The distribution of deletions shows two maximums at the sgRNA target sites, but insertions are roughly equally distributed around the target sites.

In-Depth Analysis of Individual Insertions. To define the origin of insertions, we took a closer look at respective individual events. Reads with insertions were chosen and sorted by the length of the inserted sequence. For the longest insertions, the origin of the inserted sequence was investigated. A corresponding representation for the genic locus is shown in Fig. 3. Data for the intergenic and heterochromatic loci are provided in Figs. S5 and S6. We found that along with the very low overall frequency of mutagenesis, there were no long insertions (≥10 bps) when SSBs were induced at a distance of 20 bps; however, at the longer distances, we found insertions of up to 100 bps. When the inserted sequences were aligned, they always matched the sequence that was directly adjacent to the position of the insertion, thus forming a tandem duplication.

Interestingly, for the 100-bps SSB distance, the duplications are formed by a single insertion of a long sequence segment. For the 50-bps SSB distance, however, although the overall insertion length is comparable, the insertions are mostly formed by three similar repeats of shorter individual length that align to the same position adjacent to the insertion site, thus forming a total of four repeats that may differ slightly from one another. This likely indicates repeated rounds of duplication formation by several consecutive rounds of paired SSB induction, given that the sgRNA recognition sites are not altered by such repair events.

Analysis of Mutagenesis Events After Paired Induction of SSBs on the Same DNA Strand. The induction of paired SSBs on one strand of DNA was conducted to resemble a situation during nucleotide excision repair (NER), exhibiting a segment of single-stranded

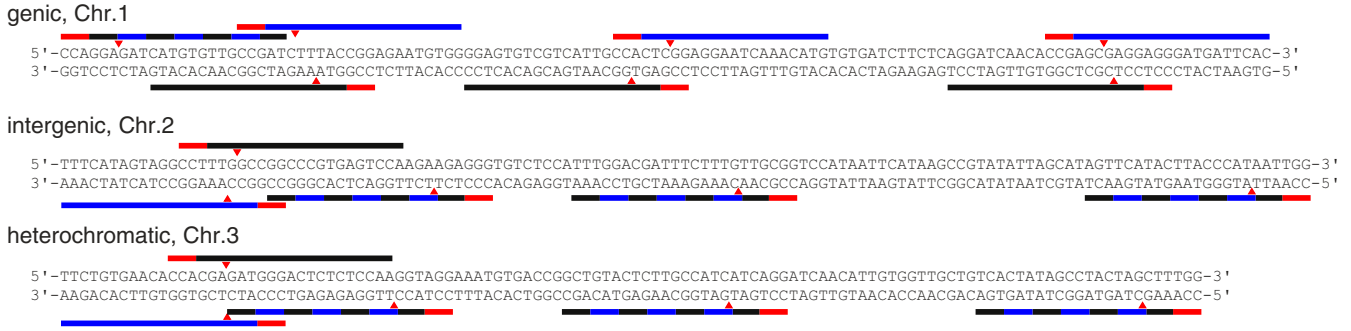

Fig. 1. Experimental setup for the investigation of genomic changes due to paired SSBs. Induction of paired SSBs was conducted at three different distances for three different loci each. The respective sgRNAs were positioned as indicated, thus generating paired SSBs either on opposing DNA strands (black bars) or on the same DNA strand (blue bars). In some cases, sgRNAs were used for both arrangements (black/blue bars). The position of the bars depicts the DNA strand, to which the sgRNA is complementary. Red triangles mark the exact position and strand where the SSB is induced.

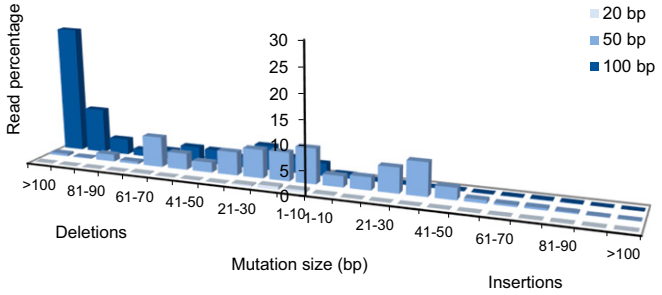

Fig. 2. Size distributions of insertions and deletions following paired SSBs on opposing DNA strands for the genic locus. For all distances and loci, deletions are more frequent than insertions; insertions are most frequent at a distance of ~50 bps. Corresponding representations for the intergenic and heterochromatic locus are given in Fig. S1.

DNA between SSBs on the same strand (27). Again, the amplicon NGS data were mapped to the respective reference sequence, and the frequencies of mutated reads were assessed (Table 1). Because this approach does not directly involve the formation of a DSB, the mutagenic potential is expectedly low; in most cases, the detected frequencies of both insertions and deletions are below 1%. Nonetheless, the investigation of individual reads can clearly reveal the presence of deletions that correlate with the position of SSB induction and also affect longer sequence segments. In addition, we detected insertions resembling the situation found for SSBs on opposing DNA strands. Although the frequencies for long insertions (>10 bps) were extremely low for all of the different experimental situations investigated, we found the origin of longer insertions again to be in the vicinity of the site of insertions. However, unlike for SSBs on both DNA strands, the observed insertions did not always originate from the directly adjacent sequence, but did always originate in close proximity, within 20–50 bps (Fig. 4).

Discussion

Implications for Using the Cas9 Paired Nickase Approach for Genome Engineering in Plants. Using the data obtained in our experiments, we were able to provide refined recommendations for the use of Cas9 paired nickases for the production of mutant plant lines. The paired nickase approach is especially desirable if off-target effects are expected as, for example, when targeting duplicated or otherwise highly homologous genes, because individual off-target binding of the sgRNAs will lead only to SSBs and thus not to the formation of mutations (10). We have previously shown that paired Cas9 nickases can efficiently generate heritable mutations in a desired gene in plants when the SSBs are placed at a distance of ~50 bps and in an orientation that leads to the formation of 5′ overhangs (12). The results presented in the present study indicate that a distance of 20 bps between the individual SSBs is insufficient to introduce efficient gene disruption in plants. This observation is in line with a previous study applying paired Cas9 nickases in human cell cultures (13).

In principle, there are two possible reasons why we see such an effect: either two SSBs at such a distance are not induced efficiently owing to binding constraints of neighboring sgRNA-Cas9-D10A complexes, or two SSBs at a closer distance are not as mutagenic.

PLANT BIOLOGY

```
genic reference CAGGAGATCATGTGTTGCCGATCTTTACCGGAGAATGTGGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTTCTCAGGATCAACACCGAGCGAGGAGGGATG
~50-bps distance
Read ID
JN35DQI01D0G3O  CAGGAGATCATGTGTTGCCGATCTTTACCGGAGAATGT*GGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTTCTCAGGATCAACACCGAGCGAGGAGGGATG
                            GTTGCCGATCTTTACCGGAGAATGT-GGGGAGTGTCGTCATT
                                    TTACCGGAGAATGT-GGGGAGTGT
                            GTTGCCGATCTTTACCGGAGAATGT-GGGGAGTGT

JN35DQI01AZAIW  CAGGAGATCATGTGTT*GCCGATCTTTACCGGAGAATGTGGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTTCTCAGGATCAACACCGAGCGAGGAGGGATG
                                  CCGATCTTTACCGGAGAATGTGGGGAGT
                                   CGATCTTTACCGGAGAATGTGGGGAGTGTCG
                                  CCGATCTTTACCGGAGAATGTGGGGAGTGTCG

JN35DQI01CFVET  CAGGAGATCATGTGTTGCCGATCTTTACCGGAGAATGTGGGGAGTGTCGTCATTG*CCACTCGGAGGAATCAAACATGTGTGATCTTCTCAGGATCAACACCGAGCGAGGAGGGATG
                                      TTACCGGAGAATGTG
                                      TTGCCG
                                   ATCTTTACCGGAGAATGTGGGGAGTGTCGTCATTG
                                      TTACCGGAGAATGTGGGGAGTGTCGTCATTG

JN35DQI01BRKZS  CAGGAGATCATGTGTTGCCGATCTTACCGGAGAATGTGGGGAGTGTCGTCATT*GCCACTCGGAGGAATCAAACATGTGTGATCTTCTCAGGATCAACACCGAGCGAGGAGGGATG
                                       TACCGGAGAATGTGGGGAGTGTCGT
                                       TACCGGAGAATGTGGGGAGTGTCGTCATT
                                       TACCGGAGAATGTGGGGAGTGTCGTCATT

JN35DQI01DQXSR  CAGGAGATCATGTGTTGCCGATCTTTACCGGAGAATGTGGGGAGTGTCGT*CATTGCCACTCGGAGGAATCAAACATGTGTGATCTTCTCAGGATCAACACCGAGCGAGGAGGGATG
                                      TTACCGGAGAATGTGGGGAGTGTCG
                                  GCAATTTACCGGAGAATGTGGGGAGTGTCGT
                                      TTACCGGAGAATGTGGGGAGTGTCGT

~100-bps distance

JN35DQI01CU3D3  CAGGAGATCATGTGTTGCCGATCTTTACCGGAGAATGTGGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTTCTCAGGATC*AACACCGAGCGAGGAGGGATG
                                     TTTACCGGAGAATGTGGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTTCTCAGGATC

JN35DQI01E1UPD  CAGGAGATCATGTGTTGCCGATCTTACCGGAGAATGTGGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTT*CTCAGGATCAACACCGAGCGAGGAGGGATG
                                       TACCGGAGAATGTGGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTT

JN35DQI01A6T0N  CAGGAGATCATGTGTTGCCGATCTTTACCGGAGAATGTGGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTT*CTCAGGATCAACACCGAGCGAGGAGGGATG
                                      TTACCGGAGAATGTGGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTT

JN35DQI01BRQ81  CAGGAGATCATGTGTTGCCGATCTTTACCGGAGAATGTGGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTTC*TCAGGATCAACACCGAGCGAGGAGGGATG
                                    CGATCTTTACCGGAGAATGTGGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTTC

JN35DQI01CDVK8  CAGGAGATCATGTGTTGCCGATCTTTACCGGAGAATGTGGGG-AGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTTCTC*AGGATCAACACCGAGCGAGGAGGGATG
                            GTTGCCGATCTTTACCGGAGAATGTGGGGGAGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTTCTC
                                          TGTGGGG-AGTGTCGTCATTGCCACTCGGAGGAATCAAACATGTGTGATCTTCTC-AGGATCAA
```

Fig. 3. Alignment of long insertion in the genic locus for SSBs in 50 bps and 100 bps distance. The site of the insertion is marked with an asterisk. The inserted sequence is given below each read and is aligned with the respective origin. Line breaks indicate continuing insertion of the same sequence, thus creating multiple repeats. Microhomologies of at least 2 bp that may be involved in formation of the duplications are indicated in orange. Intergenic and heterochromatic locus are shown in Figs. S5 and S6.

Table 1. Mutagenesis frequencies after paired SSB induction on the same DNA strand as determined by the deep-sequencing experiment

Experiment	% Deletions	% Insertions
Genic 20 bps	<0.01	0.25
Genic 50 bps	0.62	0.23
Genic 100 bps	0.65	0.28
Intergenic 20 bps	4.36	0.48
Intergenic 50 bps	0.12	0.27
Intergenic 100 bps	0.44	0.4
Heterogenic 20 bps	0.41	0.19
Heterogenic 50 bps	0.32	0.23
Heterogenic 100 bps	<0.01	0.41

We favor the former hypothesis, considering the lack of indications for a specific distance-dependent error-free repair mechanism. Our analysis shows that this effect is not locus-specific. Surprisingly, the distance of 100 bps between the SSBs did not detectably reduce the mutation frequency compared with 50 bps. Based on our results, we conclude that any spacing of 50–100 bps between SSBs should induce efficient mutagenesis in plants in a paired nickase approach. In addition, we demonstrate that cytosine methylation seems not to influence the cleavage by Cas9 in plants, with no reduction in mutagenesis detected for the methylated, heterochromatic locus. This finding is in accordance with previous experiments on the target specificity and binding efficiency of Cas9 (28, 29).

The Role of Paired SSBs on Opposite Strands in the Formation of Genomic Alterations. Over the last 3 y, multiple studies have been published describing the use of the CRISPR/Cas system for precise genome engineering approaches in various eukaryotes. Here we used Cas9 as a tool to elucidate the molecular mechanisms of genomic changes that might play an important role in plant genome evolution. In contrast to other synthetic nucleases, such as ZFNs and TALENs, the CRISPR/Cas system offers the unique possibility to induce site-specific SSBs by converting the Cas9 nuclease into a nickase by exchanging a single amino acid (6). In a previous study using deep sequencing, we found no indication that the induction of a unique SSB is mutagenic in *Arabidopsis* (12). In the present study, we applied Cas9-D10A nickase to determine the mutagenic potential of two SSBs occurring in close proximity to each other, and found that the occurrence of two SSBs at distances of 50–100 bps is highly mutagenic if they are induced in opposing strands. Interestingly, we found no noticeable differences in the mutation patterns among the three genomic loci investigated. Although we cannot draw any general conclusions based on this limited number of loci, our experiments can be taken as a hint that the same kind of repair mechanisms might operate in genic, intergenic, and heterochromatic regions of *Arabidopsis*.

The induction of paired SSBs on opposing strands of DNA leads to deletions as the most abundant type of mutation. We assume that their generation is based on the transformation of a double-stranded DNA with two SSBs in opposing strands to two pieces of broken DNA with free single-stranded overhangs (Fig. 5). Because the positioning of the SSBs was chosen to create 5′ overhangs, the free, single-stranded DNA is prone to processing by endonucleases (30). If resection of both 5′ ends exceeds the limit in which there is still complementarity between the two strands, the result is ultimately a sequence deletion. The synthesis of new DNA starting from the 3′ ends occurs simultaneously to the processing of the 5′ ends, finally forming a blunt-ended DSB, once synthesis reaches the ends of the matrix strands on both sides. Sealing of the gap can then be accomplished by a simple ligation mechanism, resembling the situation for classical NHEJ. We also find strong indications for the involvement of microhomology-based alternative end-joining. We observed a specific deletion for the genic locus at an SSB distance of 100 bps, which accounts for >17% of all reads in this dataset. This is due to microhomologies of 4 bps that are exposed directly adjacent to the cutting sites of the Cas9 nickase. Here a direct hybridization of the two 3′ ends can be envisaged, followed by cleavage of the overhanging 5′ single strands (Fig. S7).

Our analysis of insertions revealed an important source for tandem duplications. Based on our results, we conclude that these sequence duplications are indeed formed as suggested by the model of Vaughn and Bennetzen (23), which can be further refined and

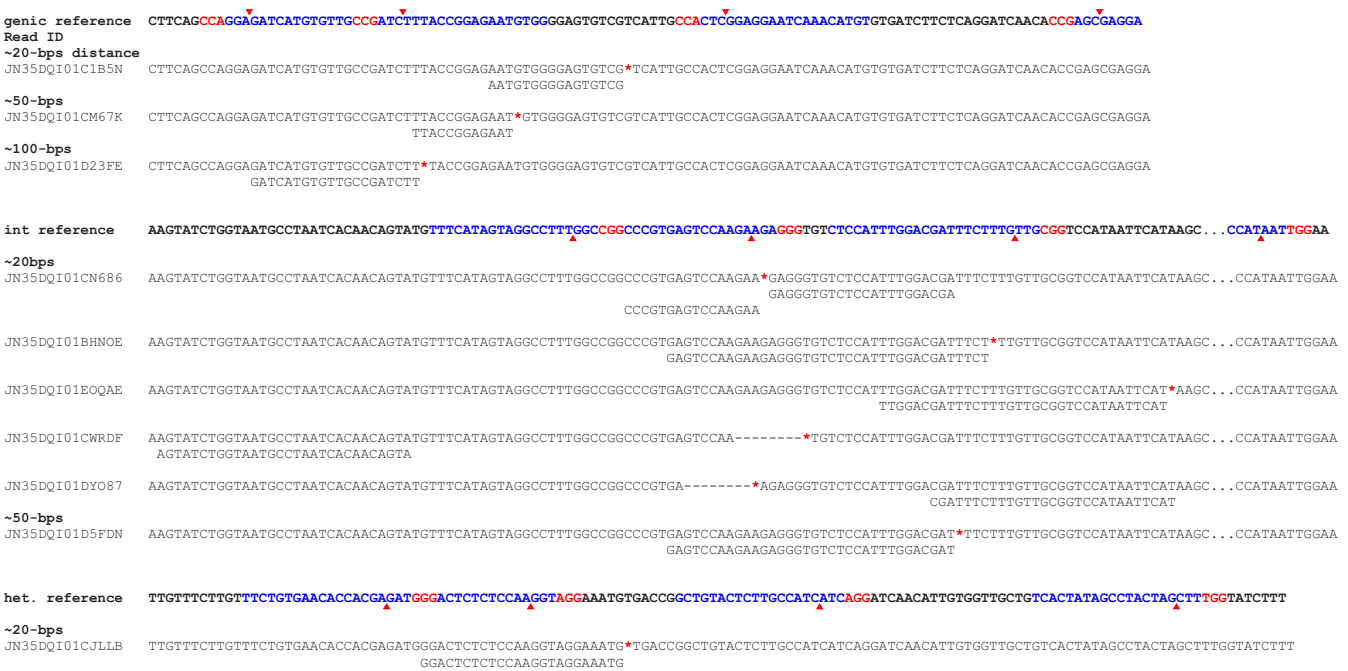

Fig. 4. Formation of long insertions (>10 bps) after induction of paired SSBs on the same DNA-strand. An asterisk marks the position of the insertion; the inserted sequence is aligned with its respective origin. Again, the insertions originate from the direct vicinity of the site of the SSB induction.

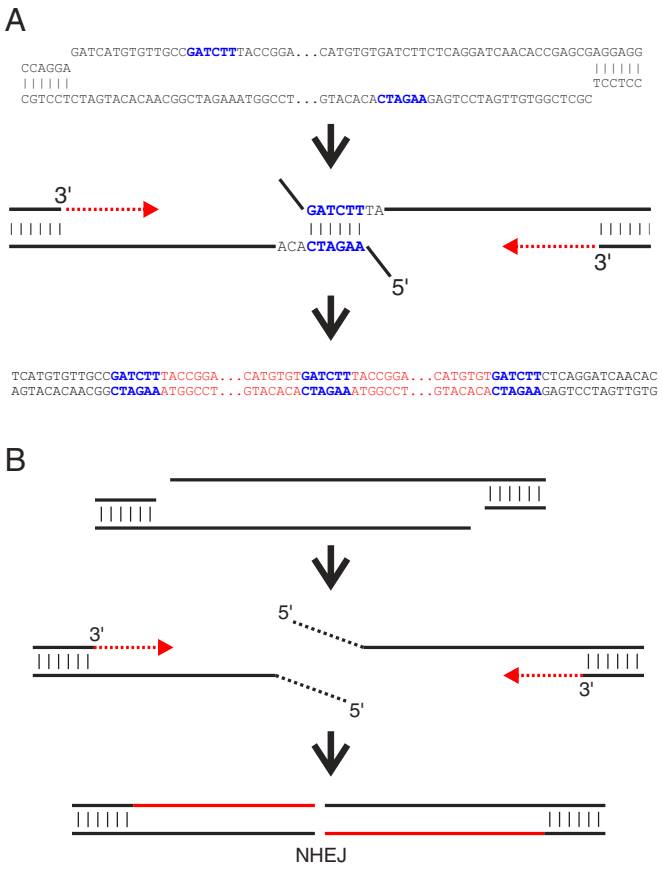

Fig. 5. Model for the formation of tandem sequence duplications as a result of paired SSBs on opposite DNA strands. (*A*) Microhomologies (blue) can be identified upstream of the duplicated sequence and at the 3′ end of the duplication, and an intermediate structure can be stabilized by hybridization. (*B*) Without microhomologies, parallel degradation of the 5′ ends and synthesis from the 3′ ends with final ligation of the blunt ends can also yield duplicated sequences.

extended (Fig. 5). Analyzing our data, we come to the conclusion that two related pathways exist that differ by whether or not microhomologies are involved in the reaction. In the microhomology-independent pathway, both 5′ ends are set free to be resected. At the same time, starting from the 3′ ends, new DNA synthesis can commence until the protruding strands' ends are reached. Thus, a blunt-ended DSB is formed, which can be sealed by classical NHEJ-mediated ligation. The concerted DNA synthesis on both sides can lead to the formation of sequence duplications of variable length, depending on the extent of the 5′-end resection. Furthermore, we found evidence of the utilization of microhomologies in the formation of sequence duplications. This is indicated by the presence of short homologous sequence motifs on either the 5′ or 3′ side of the insertion as well as in the parental sequence directly downstream or upstream of the insertion site, respectively. In these cases, we assume that the single-stranded protruding 5′ ends can hybridize to each other via regions of microhomology, forming a more or less stable intermediate structure. The nonpairing 5′ flaps are removed at the same time as internal gaps are repaired by fill-in synthesis. Finally, the two arising staggered nicks in the filled up double-stranded intermediate are ligated.

Our data indicate that paired nicks in opposing strands are highly mutagenic, so they might be relevant for plant genome evolution even if they occur at a lower frequency in natural situations. Naturally, SSBs in opposing strands at short distances might occur during base excision repair, especially if there is a high incidence of incorporation of uracils into DNA (24). Alternatively, nicking errors on the complementary strand by NER might lead to paired SSBs in opposite strands (23). However, it must be noted that nicks induced from DNA damage are obviously induced only once in the natural environment, whereas the Cas9 protein can cleave multiple times until the target is destroyed. This is also documented by the fact that we found both the presence of single repeats, but also the formation of multiple repeats. Nevertheless, our experimental data strongly suggest that SSBs in opposing strands are responsible for tandem duplications also under natural conditions.

Adjacent SSBs in the Same Strand Can Also Result in Genomic Alterations. Interestingly, also the induction of adjacent SSBs on the same DNA strand led to the formation of mutations. We assume that formation of both deletions and insertions is based on the initial release of the single-stranded DNA patch in between the two SSBs (Fig. 6), either spontaneously or driven by DNA helicases. The resulting single-stranded gap might be prone to breakage, resulting in a DSB with protruding ends. After processing of the ends by resection and fill-in, the resulting DSB is repaired by NHEJ, leaving behind a larger deletion. For the formation of insertions, the synthesis of new DNA has to begin, starting from the 3′ end. The possibility for an insertion from the vicinity of the break sites may be given by replication slippage, either back to an already synthesized segment or back to a more distant site, yet still in close proximity.

Because the initial situation of neighboring SSBs resembles a step within NER, we assume that although the mutagenic potential of SSBs in the same strand in contrast to opposing strands is, according to our data, less severe by at least two to three orders of magnitude, it nevertheless might have important consequences for genome evolution, especially in plants.

Owing to their sessile and autotrophic nature, plants, and especially their genetic information, are under constant threat from harmful UV irradiation. Dimerization of pyrimidine bases is a major DNA damage induced by UV light. Along with photolyases (31) and translesion synthesis (32), excision repair (33, 34) is the main pathway for the repair of such UV-induced DNA damage. Thus, by the sheer number of events, a process that itself is hardly mutagenic might become a significant cause for genome alterations. Interestingly, a recent study performed with mammalian cells indicates that, most likely due to delayed damage recognition, NER is slow in heterochromatic regions and leads to the accumulation of a significant number of mutations (35). In this sense, paired SSBs, whether they occur at high frequency in the same strand or at low frequency in opposite strands in close proximity, might be a major threat to genome stability. This is especially relevant for genome evolution in

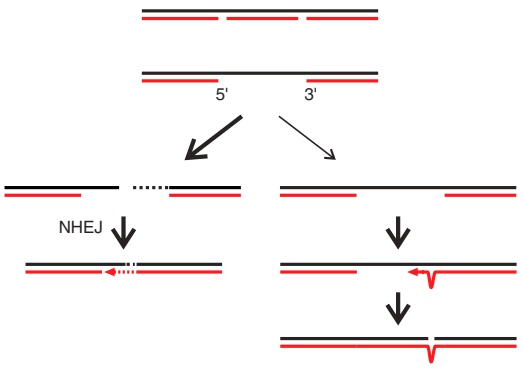

Fig. 6. Formation of mutations after occurrence of adjacent SSBs on the same DNA strand. After release of the single-stranded segment, the remaining gap can be prone to further breakage, leading to end-joining accompanied by deletions. Insertions may be a result of template switching to an adjacent sequence area. Subsequent solution of the intermediate structure can result in a full duplication.

plants, given that their germline is set aside only late in development. The presence of multiple short direct repeats in plant genomes (23) is a strong indication that this is indeed the case.

Materials and Methods

T-DNA Construction. Constructs for paired nickases were cloned as described previously (10, 12). Expression cassettes for sgRNAs were assembled by cloning respective oligonucleotides into linearized pEn-C1.1, with subsequent transformation into *Escherichia coli* strain NEB5α, and were then transferred into pDe-CAS9-D10A by conventional and Gateway cloning. This procedure resulted in the final T-DNA constructs, each harboring a constitutive expression system for Cas9-D10A and two sgRNA sequences for respective induction of paired SSBs. The primers used in this study are listed in Table S1.

Plant Transformation and Selection. Plant lines of *A. thaliana* with a Columbia-0 background were used. *Arabidopsis* plants were transformed by *Agrobacterium*-mediated transformation as described previously (36), using *A. tumefaciens* strain GV3101. Selection of the primary transformant plants was done on agar plates with germination medium (4.9 g/L Murashige and Skoog medium, 10 g/L sucrose, and 8 g/L agar, pH 5.7) containing 30 mg/L kanamycin and 0.5 g/L cefotaxime.

Amplicon Deep Sequencing. Batches of 30 primary transformants for each construct were used for DNA extraction, which was performed as described previously (4). Using a proofreading polymerase, MID-labeled amplicons for deep-sequencing analysis were generated by PCR and purified using the peqGOLD Cycle-Pure Kit (Peqlab Biotechnologie). NGS was performed on a Roche 454 FLX+ System by Eurofins Genomics. Data analysis was performed with the module lastz of the Galaxy web server (37–39) to obtain an overall sequence mapping and for individual variant detection. Calculations of differences by position were done with the CRISPResso program. Reads taken into the calculations covered at least 70% of the reference sequence. Total read numbers used for analysis are given in Table S2.

ACKNOWLEDGMENTS. We thank Maren Scheidle, Simon Stowasser, and Waltraud Wehrle for their excellent technical assistance. This work was funded by the European Research Council (Advanced Grant COMREC 26852).

1. Puchta H, Dujon B, Hohn B (1993) Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. *Nucleic Acids Res* 21(22):5034–5040.
2. Kirik A, Salomon S, Puchta H (2000) Species-specific double-strand break repair and genome evolution in plants. *EMBO J* 19(20):5562–5566.
3. Siebert R, Puchta H (2002) Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. *Plant Cell* 14(5):1121–1131.
4. Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J* 17(20):6086–6095.
5. Knoll A, Fauser F, Puchta H (2014) DNA recombination in somatic plant cells: Mechanisms and evolutionary consequences. *Chromosome Res* 22(2):191–201.
6. Jinek M, et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096):816–821.
7. Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 109(39):E2579–E2586.
8. Cong L, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121):819–823.
9. Mali P, et al. (2013) RNA-guided human genome engineering via Cas9. *Science* 339(6121):823–826.
10. Fauser F, Schiml S, Puchta H (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J* 79(2):348–359.
11. Schiml S, Puchta H (2016) Revolutionizing plant biology: Multiple ways of genome engineering by CRISPR/Cas. *Plant Methods* 12:8.
12. Schiml S, Fauser F, Puchta H (2014) The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny. *Plant J* 80(6):1139–1150.
13. Ran FA, et al. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154(6):1380–1389.
14. Puchta H (2005) The repair of double-strand breaks in plants: Mechanisms and consequences for genome evolution. *J Exp Bot* 56(409):1–14.
15. Pace JK, 2nd, Sen SK, Batzer MA, Feschotte C (2009) Repair-mediated duplication by capture of proximal chromosomal DNA has shaped vertebrate genome evolution. *PLoS Genet* 5(5):e1000469.
16. Woodhouse MR, Pedersen B, Freeling M (2010) Transposed genes in *Arabidopsis* are often associated with flanking repeats. *PLoS Genet* 6(5):e1000949.
17. Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci USA* 101(34):12404–12410.
18. Wicker T, et al. (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8(12):973–982.
19. Thomas EE (2005) Short, local duplications in eukaryotic genomes. *Curr Opin Genet Dev* 15(6):640–644.
20. Levinson G, Gutman GA (1987) Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol Biol Evol* 4(3):203–221.
21. Chen JM, Chuzhanova N, Stenson PD, Férec C, Cooper DN (2005) Meta-analysis of gross insertions causing human genetic disease: Novel mutational mechanisms and the role of replication slippage. *Hum Mutat* 25(2):207–221.
22. Messer PW, Arndt PF (2007) The majority of recent short DNA insertions in the human genome are tandem duplications. *Mol Biol Evol* 24(5):1190–1197.
23. Vaughn JN, Bennetzen JL (2014) Natural insertions in rice commonly form tandem duplications indicative of patch-mediated double-strand break induction and repair. *Proc Natl Acad Sci USA* 111(18):6684–6689.
24. Vispé S, Satoh MS (2000) DNA repair patch-mediated double strand DNA break formation in human cells. *J Biol Chem* 275(35):27386–27392.
25. Zhang X, et al. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 126(6):1189–1201.
26. Sequeira-Mendes J, et al. (2014) The functional topography of the *Arabidopsis* genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* 26(6):2351–2366.
27. de Laat WL, Jaspers NG, Hoeijmakers JH (1999) Molecular mechanism of nucleotide excision repair. *Genes Dev* 13(7):768–785.
28. Hsu PD, et al. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31(9):827–832.
29. Yaung SJ, Esvelt KM, Church GM (2014) CRISPR/Cas9-mediated phage resistance is not impeded by the DNA modifications of phage T4. *PLoS One* 9(6):e98811.
30. Symington LS, Gautier J (2011) Double-strand break end resection and repair pathway choice. *Annu Rev Genet* 45:247–271.
31. Britt AB, Chen JJ, Wykoff D, Mitchell D (1993) A UV-sensitive mutant of *Arabidopsis* defective in the repair of pyrimidine-pyrimidinone(6-4) dimers. *Science* 261(5128):1571–1574.
32. Kobbe S, Trapp O, Knoll A, Manuss A, Puchta H (2015) The translesion polymerase ζ has roles dependent on and independent of the nuclease MUS81 and the helicase RECQ4A in DNA damage repair in *Arabidopsis*. *Plant Physiol* 169(4):2718–2729.
33. Gallego F, Fleck O, Li A, Wyrzykowska J, Tinland B (2000) AtRAD1, a plant homologue of human and yeast nucleotide excision repair endonucleases, is involved in dark repair of UV damages and recombination. *Plant J* 21(6):507–518.
34. Fidantsef AL, Mitchell DL, Britt AB (2000) The *Arabidopsis* UVH1 gene is a homolog of the yeast repair endonuclease RAD1. *Plant Physiol* 124(2):579–586.
35. Adar S, Hu J, Lieb JD, Sancar A (2016) Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. *Proc Natl Acad Sci USA* 113(15):E2124–E2133.
36. Clough SJ, Bent AF (1998) Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16(6):735–743.
37. Giardine B, et al. (2005) Galaxy: A platform for interactive large-scale genome analysis. *Genome Res* 15(10):1451–1455.
38. Blankenberg D, et al. (2010) Galaxy: A web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* Chapter 19:Unit 19.10.1–21.
39. Goecks J, Nekrutenko A, Taylor J; Galaxy Team (2010) Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11(8):R86.