CRISPR-Cas9 integrates exogeneous sorting new recombinant DNA in the model tree *Populus trichocarpa*

Ali Movahedi[1], Hui Wei[1], Zhong-Hua Chen[2], Weibo Sun[1], Jiaxin Zhang[3], Dawei Li[1], Honghua Ruan[1], and Qiang Zhuge[1*]

[1] Co-Innovation Center for Sustainable Forestry in Southern China, Key Laboratory of Forest Genetics & Biotechnology, Ministry of Education, Nanjing Forestry University, Nanjing 210037, China

[2] School of Science, Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW 2751, Australia

[3] School of Food Science and Pharmaceutical Engineering, Nanjing Normal University, Nanjing 210046, China

[*]Correspondence should be addressed to Qiang Zhuge: Co-Innovation Center for Sustainable Forestry in Southern China, Key Laboratory of Forest Genetics and Biotechnology, Ministry of Education, Nanjing Forestry University, Nanjing 210037, China. E-mail: qzhuge@njfu.edu.cn; Fax: +86 25 85428701

Running title: XRCC4 deficient vividly enhanced HDR efficiency

Abstract

One of the major challenges facing researchers is using an efficient homology-directed DNA repair (HDR) to replace the targeted fragment on the genome of tree species with the desired DNA fragment. In the past, researches have been conducted on genetic modifications in mammals using HDR effector proteins to guide double-stranded breaks (DSBs) more precisely. Here, we targeted XRCC4, a cofactor for DNA ligase IV that is a key role in nonhomologous end-joining (NHEJ), to enhance HDR. XRCC4 deficient incorporated with efficient homologous recombination factors CtIP and MRE11 promoted HDR efficiency in poplar tree lines up to 10% and dramatically decremented polymorphisms. We also could introduce exogenous bleomycin to the poplar genome and generate stable lines resistant to the zeocin antibiotic.

Keywords: CRISPR, XRCC4, CtIP, MRE11, HDR, *Populus trichocarpa*

Introduction

Precise, targeted genetic modification in trees has been challenging because of the efficient NHEJ factors and also the difficulty of delivery of template DNA for HDR into the cell nucleus. HDR has always been used to precisely repair DSBs, while NHEJs do the random and irregular repair in the DSB area. NHEJ has shown a predominant pathway and occurs among the cell cycle widely (Panier and Boulton, 2014), and HDR rarely occurs (Puchta, 2005) because of the low efficiency of transferring DNA or RNA fragments into the cell nucleus resulting in difficulties to accurately target and modify plant genomes. Up to date, many studies have been carried out to improve the genetic modification of crops by HDR. For instance, one study has been carried out to increase *ARGOS8* expression by replacing the GOS2 promoter via HDR to drive *ARGOS8* expression (Shi et al., 2017). Another report showed ten-fold enhancement in the efficiency of the insertion of the 35S promoter in the upstream of the *ANT1* gene in tomato (Cermak et al., 2015). Some researchers also exhibited the promotion in gene-targeting efficiency in potato (Butler et al., 2016), tomato (Dahan-Meir et al., 2018), rice (Sun et al., 2016; Wang et al., 2017), wheat (Gil-Humanes et al., 2017), cassava (Hummel et al., 2018), soybean (Li et al., 2015), and maize (Svitashev et al., 2015).

Although it has been previously reported the use of protoplasts for this purpose (Svitashev et al., 2016), the methods have been very inefficient. Due to the difficulty of moving the donor DNA template (DDT) to the required location within the cell nucleus, researches of the use of particle bombardment and virus-based replicons have been reported to increase this translocation, but the target relocation or replacement is still ongoing (Gil-Humanes et al., 2017; Wang et al., 2017). Therefore, this is one of the major problems in the genetic modification of trees. Previous data indicate that to increase HDR efficiency, it is necessary to increase the number of cells containing DDTs at S/G2 cell division phases (Yang et al., 2016). *Agrobacterium* has been widely used to transduce genes into plant cell nuclei (Movahedi et al., 2015). This method was improved to increase transformation efficiency (Movahedi et al., 2014), but there was no report of improving the Agrobacteria method delivery to increase the efficiency of transferring DDT and consequently the recovery of DSBs as HDR. Furthermore, the positive effects of recombinant homologous factors and their effects on enhancing HDR efficiency in mammalians have been already reported (Tran et al., 2019), with at least a 2-fold increase of HDR and a 6-fold increase of HDR / NHEJ ratio. Cas9 integrates with MRE11, CtIP, and Rad51,

Rad52, and promotes significant HDR efficiency in human cells and decreases significantly NHEJ (Tran et al., 2019). On the other hand, inhibition of DNA ligase IV (LIG4), Ku 70, Ku 80, and XRCC4, which are known as the most important NHEJ factors (Friesner and Britt, 2003; Maruyama et al., 2015; Pierce et al., 2001; Tran et al., 2019), increase the HDR efficiency up to 19-fold (Tran et al., 2019). XRCC4 is one cofactor of LIG4 to interact with KU 70 and KU 80 and ligate the DSB (Grawunder et al., 1998b). Ku70 and Ku80 protect DSB from discrediting with forming one heterodimeric complex to bind tightly and load additional repair proteins such as DNA ligase IV, which is outwardly involved only in NHEJ (Grawunder et al., 1998a). On the other hand, the effect of off-targets on efficient precise genome editing and ways to reduce their effects has been already studied (Hajiahmadi et al., 2019; Li et al., 2019; Wu and Yin, 2019).

In this study, we hypothesized that inhibition of nonhomologous recombination cofactors XRCC4 promotes the HDR efficiency in poplar trees. We could overcome to make one recombinant chromosomal DNA in *P. trichocarpa* with a haploid chromosome of 19 by inhibiting activities of the NHEJ pathway (Maruyama et al., 2015) and enhancing the activities of the HDR pathway (Tran et al., 2019). Using this approach, we used the optimized length of homologous arms 400 bp (Song and Stieger, 2017).

Results and discussion

Poplars genome exhibited no exogenous integration without impacting HDR effector proteins

We started to recognize CRISPR sites located on the *MKK2* gene near the 3’ UTR and scored all detected targets (Doench et al., 2014; Hsu et al., 2013) concerning higher scores denoting higher activity, specificity, and less off-target activity. We discovered one target that is located on exon 8 (Non off-target on CDS throughout the whole *Populus trichocarpa* genome) (Figure 1A and Supplementary Table 1). To investigate whether XRCC4 deficiency improves the integration efficiency of DDT at the desired locus, we first tested the insertion of DDT without targeting *XRCC4* and also fusing CtIP and MRE11. For this, we designed one plasmid for harboring the gRNA scaffold (pgRNA) and one plasmid harboring DDT (pDDT) (Figure 1A; Supplementary 1). We designed a special unique target (with no on- and -off-targets through whole poplar genome) besides DDT to cleave and release DDT and increase HDR efficiency (Zhang et al., 2017) (Figure 1A, pDDT). We used the optimized of *Agrobacterium tumefaciens*

strain EHA105 stimulant and pathogenic suspension OD_{600} = 2.5 (~2×10^9 cell ml^{-1}) with the optimized ratio of 4:1 pDDT/pgRNA to increase DDT fragments during S/G2 cell division (Tran et al., 2019) and to avoid off-target editing caused by the extra accumulation of pgRNA (Hajiahmadi et al., 2019) (Figure 1B). We carried out five different sets of HDR experiments (Figure 1C) in *Populus* calli to test whether HDR happened with various HDR factors to integrate exogenous *BleoR* CDS precisely and its expression to recover poplars on zeocin, contained DDT transferring alone, DDT and CtIP fusion with Cas9, DDT and MRE11 fusion with Cas9, DDT and complex of CtIP and MRE11 fusion with Cas9, and finally DDT and complex of CtIP and MRE11 accompanied with XRCC4 deficiency. We then used TaqMan real-time PCR to detect HDR possibilities and investigate XRCC4 deficiency (Figure 1D). To detect HDR efficiency, it was necessary to have two probes FAM1 including exon8 and exon9 from *MKK2* (114bp) attached by 36 bp from 5' *BleoR* CDS and FAM2 including 57 bp from 3' homology arm, 48 bp from attached poly-A and 6xHis, and 45 bp from attached 3' *BleoR* CDS (Figure 1D). Because we didn't achieve recovered poplar on zeocin in Experiment I (ExI) (Figure 1C), it was clear that DDT couldn't integrate *BleoR* into the poplar genome. We carried out real-time PCR of nine regenerated shoots from calli (with quadruplicate), and the 2D kernel density plot exhibited more signal accumulation ~0.022 in wild-type (WT) or mutated events, and lesser ~0.0073 in partial FAM2 (Figure 1E; Supplementary 2). Then, we carried out Sanger sequencing to align events and reveal fluorescent FAM signals (Figure 1F; Supplementary 3).

BleoR exogenous integrated scantily to the poplar genome by enhancing both expressions of CtIP and MRE11

We decided to design plasmid included a fused CtIP and Cas9 (pgCtIP) instead of pgRNA with the ratio of 4:1 pDDT/pgCtIP, to test whether it is possible to promote HDR efficiency in poplars via Experiment II (ExII) (Figure 2A). We obtained more fluorescent signal accumulation in FAM1&2 (Edited events) at ~0.022 and lesser in partial FAM1, and -FAM2, mutated, and WT events at ~0.018 (Figure 2B; Supplementary 2). DNA sequencing alignment of 17 regenerated shoots from calli on zeocin exhibited HDR efficiency with one recovered poplar (Appendant *BleoR* into poplar genome) (Figure 2C; Supplementary 4). Furthermore, to investigate whether MRE11 is as effective as CtIP in increase HDR efficiency in poplar, we designed plasmid harboring a combined MRE11 and Cas9 (pgMR) instead of pgRNA with the same ratio 4:1 via Experiment III (ExIII) (Figure 2D). MRE11 enhanced signal accumulation in

partial FAM1 ~0.022 more than partial FAM1 and edited events ~0.015 (Figure 2E; Supplementary 2). In this experiment, we discovered three events included signals in both FAM1&2 and one recovered poplar (Figure 2C, Supplementary 5). DNA sequencing alignment of 15 regenerated shoots revealed that recovered event included *BleoR* attached by only 3' homology arm (partial FAM2) (Figure 2C, Supplementary 5). Because we couldn't achieve the desired result in overcoming NHEJ to integrate *BleoR* CDS, we decided to design Experiment IV (ExIV), and one plasmid harboring fused both MRE11 and CtIP with Cas9 (pgCtMR) (Figure 2G). This complex enhanced HDR efficiency and exhibited more edited events. The 2D kernel density plot revealed FAM1&2 signal accumulation ~0.036, partial FAM1 ~0.024, partial FAM2 ~0.018, and mutated and WT events ≤0.012 (Figure 2H; Supplementary 2). The complex of CtIP and MRE11 with Cas9 also promoted the integration of *BleoR* sequences and increased recovered events (Figure 2I; Supplementary 6).

XRCC4 deficient dramatically enhanced HDR efficiency and decreased polymorphisms

We concluded that for improving the integration of *BleoR* into the poplar genome, it might be better to target *XRCC4*. We used pgCtMR plasmid and re-designed that by adding one CRISPR site (Located on 5' region of target CDS) to mutate *XRCC4* (Non-off-target site on whole poplar genome; Activity score: 0.415; Specificity score: 100%) (Doench et al., 2014; Hsu et al., 2013) (pggCtMR) (Figure 3A) and carried out the Experiment V (ExV) (Figure 1C) with the ratio of 4:1 pDDT/pggCtMR. XRCC4 deficiency meaningfully enhanced HDR efficiency and increased edited events. The 2D kernel density plot exhibited FAM1&2 signal accumulation ~0.052, partial FAM1 ≤0.026, partial FAM2 ≤0.017, and mutated and WT events ≤0.0087 (Figure 3B; Supplementary 2). We then carried out sequencing of 31 regenerated seedlings and discovered 12 recovered events, and it means we could overcome NHEJ and integrate *BleoR* in the precise quarry in the poplar genome (Figure 3C; Supplementary 7). We performed real-time PCR to evaluate *BleoR* expression in all experiments (ExI to ExV) (Figure 1C) to calculate HDR efficiency (Figure 3D; Supplementary 8). We discovered that XRCC4 deficiency (ExV) promoted HDR efficiency meaningfully 10 times in comparing with the fusion of Cas9 with HDR factors CtIP (ExII), and MRE11 (ExIII) and also 2.8 times in comparing with the fusion of Cas9 and complex of CtIP and MRE11 (ExIV) (Figure 3D). Based on the promising results, we decided to analyze the total achieved FAM fluorescent signals. The results exhibited that we significantly accomplished the most partial FAM1, -FAM2, and edited events (FAM1&2) from

the ExV in comparison with the other experiments (Figure 3E). To test whether the HDR efficiency depends on the happened polymorphisms we analyzed the polymorphisms of homology arms and also different types of variants, protein effects, and genotyping (Supplementary Table 2). The box and whisker plot exhibited that XRCC4 deficiency in the ExV improved HDR in homology arms and statistically decremented polymorphisms in comparison with the ExI and also more than the ExII, -III and -IV (Figure 3F). Our analysis detected eight types of polymorphisms (Supplementary table 2) including deletions, deletion tandem repeats, insertions, insertion tandem repeats, SNPs, SNP transitions (A to C or G to T and reversely), SNP transversions (Purines to pyrimidines or reversely), and substitutions (Figure 3G). We detected the highest frequency of deletion polymorphisms in the ExI, and the least of that in the ExV (Figure 3G). The HDR factors CtIP decremented deletion tandem repeats (ExII) and MRE11 decremented SNPs, SNP transitions, and SNP transversions, and the complex of HDR factors with Cas9 (ExIV) decremented substitution polymorphisms (Figure 3G). Moreover, the box (Data overlap) and whisker plot (With normal data distribution curve) of total polymorphisms revealed that using of HDR recombination factors and also suppression of NHEJ factors decrement total polymorphisms and lead to the occurrence of HDR in poplar (Figure 3G). Moreover, we evaluated the ratio of *BleoR* and *MKK2* expressions and discovered the direct relative between increase HDR efficiency to the expression of HDR and NHEJ recombination factors (Figure 3H) and their expression. The heat-map exhibited that XRCC4 deficiency enhanced the expression on *BleoR* (HDR efficiency %) and *MKK2* in the ExV more than the overloading of HDR recombination factors (ExII, ExIII, and ExIV) (Figure 3H). We also performed real-time PCR to evaluate the interaction proteins involved in NHEJ and HDR pathways in poplar in our experiments. The heat-map plot analysis exhibited that the overloading of proteins CtIP, MRE11, and the complex of them upregulated BRCA1, Rad50, and Rad51, and downregulated Lig4, and XRCC4 (Figure 3I). In the ExI, NHEJ revealed the dominant pathway with the more average expression 145% of Lig4 and XRCC4 than the HDR pathway with the average expression 115% of CtIP, MRE11, BRCA1, Rad50, and Rad51 (Figure 3I). The role of these pathways was changed a little in ExII, ExIII, and ExV with the average expression ~124% of BRCA1, Rad50, and Rad51, and the decreased average expression 104% of Lig4 and XRCC4, leading to increase HDR and discover recovered events (Figure 3I). XRCC4 deficiency (ExV) caused to significantly change the dominant pathway to HDR and revealed the average

expression 146% of BRCA1, Rad50, and Rad51 in comparison with the decreased expression 93% of Lig4 (Figure 3I) leading to enhance HDR efficiency and discover significant recovered events. In summary, we have proved that XRCC4 deficiency can apply for promoting HDR in poplar, therefore greatly expanding our capacity to improve hereditary developments in poplar. This breakthrough technology is likely to promote the biotechnological research, breeding program, and forest conservation of tree species.

Methods

Targets and protein detection

We decided to target the *MKK2* gene from *Populus trichocarpa* (POPTR_0018s05420g; Chromosome 18). We used Uniprot (https://www.uniprot.org/) to download MKK2 protein and then used the BLAST database of the National Center for Biotechnology Information (NCBI) (https://blast.ncbi.nlm.nih.gov/) to download full DNA sequences and CDS, which was then ligated into the pEASY-TA cloning vector and verified by Sanger sequencing (GenScript, Nanjing). We used Geneious Prime® 2020.1.1 to analyze locus and detect targets (Supplementary Table 1)(Doench et al., 2014; Hsu et al., 2013). The target sequences heading the PAM motif was concerned from the exon8 area of the *MKK2* and exon1 area of the *XRCC4* (POPTR_0010s08650g, Chromosome 10) genes. Furthermore, we used Uniprot to detect CtIP (POPTR_001G269700v3), MRE11 (POPTR_0001s41800g), BRCA1 (POPTR_0005s26150g), Rad50 (POPTR_0001s32760g), Rad51 (POPTR_0014s06360g), Lig4 (POPTR_0018s13870g).

Target oligo synthesis

For targeting indicated DNA sequences, we designed a pair of oligos *MKK2* Oligo-F and -R (Supplementary Table 3) flanked by *Bsa*I adaptors. Synthesized oligos were then ligated into digested pRGEB31 vectors by *Bsa*I restriction enzyme(Xie and Yang, 2013) to construct pgRNA (Supplementary 1, pgRNA). Afterward, we transferred all vectors into *E. coli* (DH5α) and propagated under the normal conditions. Vectors were then extracted using the plasmid midi kit (Qiagen, USA) and confirmed by sanger sequencing (GenScript, Nanjing).

Synthesis of DDT and pDDT

To design DDT, we divided DDT into five fragments (Supplementary 9a). To construct fragment one (Supplementary 9a) we used primers OS1-F and -R (Supplementary Table 3) to

isolate OsU3 promoter and gRNA scaffold from pRGEB31 and add *Hind*III and *BamH*I endonucleases to the 5' and 3' respectively. Then we used designed special gRNA oligo1-F and -R (Sgo1-F and -R)(Supplementary Table 3) as the described details 30 to form special gRNA target1 duplex (Sgt1) and ligated to the fragment one. Then, we carried out normal PCR and purified products to clone into a pEASY T3 cloning vector (TransGen Biotech Co.), screen, and sequences. To construct fragment two (Supplementary 9), we isolated 5' homology arm (400 bp) sequences from *Populus trichocarpa* genome applying primers 5' Ho-F-1 and -R-1 (Supplementary Table 3), cloned into the pEASY vector and sequenced. Then, we applied primers *BamH*I-special target1 (St1)-5' Ho-F-2 and 5' Ho-R-2-39 bp from complemented 5' of fragment 3 (Supplementary 9a). Then we carried out normal PCR and ligated that into the pEASY vector, screened and sequenced. We designed primers BleoR-1092F and -2276R (Supplementary 9a; Supplementary Table 3) for running PCR and isolate *BleoR* CDS from PCR®-XL-Topo® vector, purified and sequenced to confirm. Then we utilized primers BleoR-overlap1-3 (BP1-3) forward and reverse (Supplementary 9a; Supplementary Table 3) to add the remained sequences from exon8 and also exon9 to the 5' region of *BleoR* CDS and also 6XHis and PolyA tail to the 3' region (Supplementary 9a). Therefore, we designed overlap PCR primers (Supplementary 9a; Supplementary Table 3) and using isolated *BleoR* CDS as the template to form fragment three. Afterward, we isolated 3' homology arm (400 bp) using primers 3' Ho-F-1 and -R-1 (Supplementary Table 3) and carried out normal PCR to clone into the pEASY vector and sequence. Then, we applied the primers 30 bp Poly-T-3' Ho-F-2 and 3' Ho-R-2-St2-*Nco*I (Supplementary 9a; Supplementary Table 3) and constructed the fragment four. Finally, we used primers Os2-F-*Nco*I and *Sda*I-Os2-R to perform normal PCR and isolate the OsU3 promoter and gRNA scaffold (Supplementary Table 3), to purify and clone into the pEASY vector and sequence. Then we formed sgt2 using oligos sgo2-F and -R and ligated into the purified to construct fragment five, perform PCR and sequence. To construct pDDT, we ligated fragments three and two using PCR (Supplementary 9b). For this, we designed 39 bp overhang on fragment 2 that was complementary to the end of fragment three to form preliminary DDT (Supplementary 9b). In this PCR, we prepared a PCR reaction with 500 ng of each fragment. So basically, we used everything in PCR reaction except primers. We denatured fragments at 95 degrees for 5 minutes. Then we performed two cycles annealing and extension. For this, we allowed PCR products to anneal at 68 degrees to avoid nonspecific hybridization amongst the long PCR

products for 30 seconds and then allowed to extend for one minute at 74 degrees to have a double-stranded product. Then we added the primers to the distal ends of the fragments two and three and performed one normal PCR. We purified PCR products and ligated into the pEASY vector to sequence and confirm. Then we ligated preliminary DDT product to fragment four as described before and formed secondary DDT products (Supplementary 9b). After sequencing and confirmation, we used the restriction cloning technique (We used digested fragments each of them 50 ng, 10x T4 DNA ligase buffer 0.5 ul, T4 DNA ligase (NEB) 1 ul and H_2O to 5 ul and then we incubated that reaction at 25 degrees for 4 hours and transferred into *E. Coli* DH5α competent cells to sequence and confirm) to ligate secondary DDT product to the fragments one and four (Supplementary 9b) to achieve DDT products. Subsequently, we used again the restriction cloning technique to merge the DDT product and pRGEB31 vector (Supplementary 9b) and form the pDDT vector (Supplementary 9b).

Synthesis of pgCtIP, pgMR, pgCtMR, and pggCtMR

To design a fused CtIP and Cas9 cassette, we isolated the CaMV35S promoter, 3xFLAG, and Cas9 CDS from pRGEB31 (Supplementary 10a; Supplementary Table 3), and then we performed PCR and ligated them into the pEASY vector to sequence and confirm. To achieve CtIP CDS, we extracted total RNA (450 ng/ L) from young leaves of *Populus trichocarpa* applying TRIzol and carried out reverse transcription with 2 ug RNA and oligo-dT primers to synthesize the first cDNA strand (PrimeScript One-Step RT-PCR Kit Ver.2, Takara Biotechnology, Dalian, China). Then we used primers *BamH*I-CtIP-F and *StuI*-CtIP-R (Supplementary 10a; Supplementary Table 3) to isolate CtIP CDS. We used primers *Stu*I-PolyA-F and *Pme*I-PolyA-R (Supplementary Table 3) to isolated 3'UTR and PolyA fragments from the pCAG-T3-hCAS-pA plasmid and then ligated into the pEASY vector and sequenced. We ligated CaMV35S and 3xFLAG fragments using restriction cloning and formed backbone 1 (Supplementary 10a). Then we used restriction cloning to ligate Cas9 and CtIP and form the backbone 2 (Supplementary 10a). We ligated backbone 1 and -2 using *Hind*III restriction cloning to form backbone 3 (Supplementary 10a). Finally, we ligated backbone 3 and 3'UTR-PolyA fragment using *Stu*I restriction cloning to construct the CtIP cassette. We digested and then ligated CtIP cassette and pRGEB31 by *Sda*I and *Pme*I endonucleases respectively to form pgCtIP (Figure 2A; Supplementary 10a). To construct a fusion of MRE11 and Cas9, we isolated CaMV35 promoter, 3xFLAG, Cas9, 3'UTR, and PolyA like the previous steps (Supplementary

10b; Supplementary Table 3). To isolate MRE11 CDS we recruited extracted total RNA and RT-PCR again and used the primers *Not*I-MRE-F and *Nde*I-MRE-R (Supplementary 10b; Supplementary Table 3). The isolated CaMV35S and 3xFLAG fragments ligated together according to the *Xho*I endonuclease homology site to form backbone 1 (Supplementary 10b). We built backbone 2 using the isolated Cas9 and 3'UTR-PolyA fragments (Supplementary 10b; Supplementary Table 3). Then, we ligated backbone 1, backbone 2 and also isolated MRE11 CDS from RT-PCR products applying *Not*I and *Nde*I endonucleases restriction cloning to form MR cassette (Supplementary 10b) and afterward ligated the digested MR cassette and pRGEB31 by *Sda*I and *Pme*I endonuclease homology sites to construct pgMR vector (Figure 2D; Supplementary 10b). To construct the CtMR cassette, we isolated CaMV35S and 3xFLAG fragments and ligated them using *Xho*I restriction cloning to form backbone 1 (Supplementary 11a, Supplementary Table 3). We isolated MRE11 CDS by *Not*I-MRE-F and *Nde*I-MRE-R primers and then ligated to the backbone 1 to release backbone 2 (Supplementary 11a, Supplementary Table 3). Afterward, we isolated Cas9 and RT-PCR product CtIP CDS applying primers (Supplementary Table 3) and then ligated them using *BamH*I restriction cloning to form backbone 3 (Supplementary 11a). We then used backbone 3 and isolated 3'UTR-PolyA fragment to form backbone 4 (Supplementary 11a). Finally, we used backbone 2 and -4 to form CtMR cassette and then used *Sda*I and *Pme*I to digest and then ligate CtMR and pRGEB31 to form pgCtMR vector (Supplementary 11a). We then decided to construct the *XRCC4* cassette to target *MKK2* and *XRCC4* genes. We recruited primers *Hind*III-XR-Cass1-F and *Kas*I-XR-Cass1-R (Supplementary 11b; Supplementary Table 3) to isolate the OsU3 promoter and gRNA scaffold from pRGEB31 vector and then recruited MKK2 Oligo-F and -R (Supplementary 10d; Supplementary Table 3) to form MKK2 target duplex and ligated. We isolated again the OsU3 promoter and gRNA scaffold using primers *Ksa*I-XR-Cass2-F and *Sda*I-XR-Cass2-R (Supplementary 11b; Supplementary Table 3) and also recruited XRCC4-Oligo1 and -2 to form XRCC4 target duplex and ligated. We ligated two completed fragments to form XRCC4-Cassette (Backbone 1) (Supplementary 11b). Afterward, we used backbone 1 and pRGEB31 vector to form backbone 2 (Supplementary 11b). Finally, we digested backbone 2 and constructed the CtMR vector by *Sda*I and *Pme*I and then ligated them to form the pggCtMR vector (Supplementary 11b).

TaqMan real-time PCR

To test whether the designed experiments were able to increase the exact merging of exogenous *BleoR* on both sides to the genome binding regions, we decided to use Applied Biosystem real-time PCR (Applied Biosystems, USA) and run TaqMan assay using dye labels such as FAM and VIC that attaches to the 5' region of the probe and non-fluorescent quencher (NFQ) that attaches to the 3' region of the probe. That's why we designed primers (Supplementary Table 3) to probe two 150 bp fragments FAM1 and -2 (Figure 1D; Supplementary 12) on the *MKK2* gene as the target assay and also designed primers to probe one 106 bp fragment VIC on the *actin* gene as the reference (Stable copy number) (Supplementary 12).

Evaluation of HDR efficiency

Regarding the purpose of this study to integrate the exogeneous *BleoR* CDS into the poplar genome, the ΔΔCt mean of *BleoR* was calculated in our experiments. We picked randomly twelve of the most grown events (Regenerated calli on selection media) from each experiment (Supplementary 8) to extract genomic DNA using the DNeasy Plant Mini Kit (Qiagen, USA) following the manufacturer's instructions. We used genomic DNA as the template and applied Fast Start Universal SYBR Green Master (Rox; No. 04913914001: Roche, USA) and primers BleoR-52F and BleoR-151R (Supplementary Table 3) with three technical repeats for each event.

Transformation

We cultivated *Populus trichocarpa* seedlings in a phytotron at 23±2°C under a 16/8 light/dark time(Movahedi et al., 2015). To generate transgenic lines, we removed leaves from four weeks old grown young poplars and dipped them in the optimized (Movahedi et al., 2014)of *Agrobacterium tumefaciens* stimulant and pathogenic suspension (OD_{600}: 2.5, 120 min, pH ~5, Acetosyringone (As): 200 µM) for 5 min with gentle shaking and then transferred in semi-solid woody plant medium (WPM) enriched with 0.5 mg/L 6-benzylaminopurine (6-BA), 0.004 mg/L thidiazuron (TDZ), 200 µM As and 0.5% (w/v) agar. We then incubated stimulated explants in a dark area at 23°C for 2 days. Afterward, assumed transformants were co-cultivated in selection media enriched with 0.5 mg/L 6-BA, 0.004 mg/L TDZ, 400 mg/L cefotaxime, 8 mg/L hygromycin, 100 mg/L zeocin and 0.8% (w/v) agar. After 4 weeks, regenerated shoots were sub-cultured independently in media with 0.5 mg/L 6-BA, 0.001 mg/L TDZ, 400 mg/L cefotaxime, 8

mg/L hygromycin, 100 mg/L zeocin and 0.8% (w/v) agar to lengthen. After 6-8 weeks, elongated shoots were introduced in MS media with 100 mg/L cefotaxime, 100 mg/L zeocin, and 0.8% (w/v) agar to root. Three lines were used for each experiment independently and each line included 18 individuals.

DNA sequencing and Statistical analysis

Poplar seedlings were cultivated to generate T_0 poplars. Young leaves then were cut off and introduced to the Agrobacterium for transferring DDT and regenerating T_1 calli events. Genomic DNAs were extracted from three weeks regenerated resistant shoots using the DNeasy Plant Mini Kit (Qiagen, USA). We performed PCR using Easy Taq polymerase (TransGene Biotech), 100 ng genomic DNA as a template, and primers MKK2-S-7F and MKK2-S-1139R (Supplementary Table 3) to sequence and analysis (Figure 1F, 2C, 2F, 2I, and 3C; Supplementary 3,4,5,6, and 7). All data were analyzed using ANOVA One-Way with Turkey means comparison calculated by OriginPro 2018 and Excel 2019 software (Microsoft, Redmond, WA, USA). Differences were analyzed statistically meaningful when the confidence intervals presented no overlap of the mean values with an error value of 0.05.

Supplementary information

Supplemental Information is available for this paper.

Funding

This project was funded by the National Key Program on Transgenic Research (2018ZX08020002), the National Science Foundation of China (No. 31570650), the Priority Academic Program Development of Jiangsu Higher Education Institutions and the Nanjing forestry university fund (No. 163108059).

Author contribution

A.M. conceived the project. A.M and H.W. performed the experiments. Z.H.C., W.S., J.Z., D.L., H.R., and Q.Z. reviewed the manuscript and confirmed it. A.M. and H.W. wrote the manuscript. Q.Z. supervised this project.

Acknowledgment

We thank all researchers especially professor Zhong-Hua Chen that helped us to improve this research with their directions.

Reference

Butler, N.M., Baltes, N.J., Voytas, D.F., and Douches, D.S. (2016). Geminivirus-Mediated Genome Editing in Potato (Solanum tuberosum L.) Using Sequence-Specific Nucleases. Front Plant Sci 7:1045.

Cermak, T., Baltes, N.J., Cegan, R., Zhang, Y., and Voytas, D.F. (2015). High-frequency, precise modification of the tomato genome. Genome Biol 16:232.

Dahan-Meir, T., Filler-Hayut, S., Melamed-Bessudo, C., Bocobza, S., Czosnek, H., Aharoni, A., and Levy, A.A. (2018). Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. Plant J 95:5-16.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol 32:1262-1267.

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

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C.V., Sanchez-Leon, S., Baltes, N.J., Starker, C., Barro, F., Gao, C., et al. (2017). High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. Plant J 89:1251-1262.

Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., and Lieber, M.R. (1998a). DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. Mol Cell 2:477-484.

Grawunder, U., Zimmer, D., Kulesza, P., and Lieber, M.R. (1998b). Requirement for an interaction of XRCC4 with DNA ligase IV for wild-type V(D)J recombination and DNA double-strand break repair in vivo. J Biol Chem 273:24708-24714.

Hajiahmadi, Z., Movahedi, A., Wei, H., Li, D.W., Orooji, Y., Ruan, H.H., and Zhuge, Q. (2019). Strategies to Increase On-Target and Reduce Off-Target Effects of the CRISPR/Cas9 System in Plants. Int J Mol Sci 20.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31:827-832.

Hummel, A.W., Chauhan, R.D., Cermak, T., Mutka, A.M., Vijayaraghavan, A., Boyher, A., Starker, C.G., Bart, R., Voytas, D.F., and Taylor, N.J. (2018). Allele exchange at the EPSPS locus confers glyphosate tolerance in cassava. Plant Biotechnol J 16:1275-1282.

Li, J., Hong, S., Chen, W., Zuo, E., and Yang, H. (2019). Advances in detecting and reducing off-target effects generated by CRISPR-mediated genome editing. J Genet Genomics.

Li, Z., Liu, Z.B., Xing, A., Moon, B.P., Koellhoffer, J.P., Huang, L., Ward, R.T., Clifton, E., Falco, S.C., and Cigan, A.M. (2015). Cas9-Guide RNA Directed Genome Editing in Soybean. Plant Physiol 169:960-970.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat Biotechnol 33:538-542.

Movahedi, A., Zhang, J.X., Amirian, R., and Zhuge, Q. (2014). An Efficient Agrobacterium-Mediated Transformation System for Poplar. Int J Mol Sci 15:10780-10793.

Movahedi, A., Zhang, J.X., Gao, P.H., Yang, Y., Wang, L.K., Yin, T.M., Kadkhodaei, S., Ebrahimi, M., and Qiang, Z.G. (2015). Expression of the chickpea CarNAC3 gene enhances salinity and drought tolerance in transgenic poplars. Plant Cell Tiss Org 120:141-154.

Panier, S., and Boulton, S.J. (2014). Double-strand break repair: 53BP1 comes into focus. Nat Rev Mol Cell Biol 15:7-18.

Pierce, A.J., Hu, P., Han, M., Ellis, N., and Jasin, M. (2001). Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. Genes Dev 15:3237-3242.

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

Shi, J., Gao, H., Wang, H., Lafitte, H.R., Archibald, R.L., Yang, M., Hakimi, S.M., Mo, H., and Habben, J.E. (2017). ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. Plant Biotechnol J 15:207-216.

Song, F., and Stieger, K. (2017). Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. Mol Ther Nucleic Acids 7:53-60.

Sun, Y., Zhang, X., Wu, C., He, Y., Ma, Y., Hou, H., Guo, X., Du, W., Zhao, Y., and Xia, L. (2016). Engineering Herbicide-Resistant Rice Plants through CRISPR/Cas9-Mediated Homologous Recombination of Acetolactate Synthase. Mol Plant 9:628-631.

Svitashev, S., Schwartz, C., Lenderts, B., Young, J.K., and Cigan, A.M. (2016). Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. Nat Commun 7.

Svitashev, S., Young, J.K., Schwartz, C., Gao, H., Falco, S.C., and Cigan, A.M. (2015). Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA. Plant Physiol 169:931-945.

Tran, N.T., Bashir, S., Li, X., Rossius, J., Chu, V.T., Rajewsky, K., and Kuhn, R. (2019). Enhancement of Precise Gene Editing by the Association of Cas9 With Homologous Recombination Factors. Front Genet 10.

Wang, M.G., Lu, Y.M., Botella, J.R., Mao, Y.F., Hua, K., and Zhu, J.K. (2017). Gene Targeting by Homology-Directed Repair in Rice Using a Geminivirus-Based CRISPR/Cas9 System. Mol Plant 10:1007-1010.

Wu, J., and Yin, H. (2019). Engineering guide RNA to reduce the off-target effects of CRISPR. J Genet Genomics.

Xie, K.B., and Yang, Y.N. (2013). RNA-Guided Genome Editing in Plants Using a CRISPR-Cas System. Mol Plant 6:1975-1983.

Yang, D., Scavuzzo, M.A., Chmielowiec, J., Sharp, R., Bajic, A., and Borowiak, M. (2016). Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. Sci Rep 6:21264.

Zhang, J.P., Li, X.L., Li, G.H., Chen, W., Arakaki, C., Botimer, G.D., Baylink, D., Zhang, L., Wen, W., Fu, Y.W., et al. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. Genome Biol 18:35.

Figure legends:

Figure 1. Exogenous *BleoR* integration into the poplar genome generated by Cas9. (A) Protospacer Adjacent Motif (PAM) was detected at the end of exon8 to lead Cas9. 400 bp sequences from both sides of the CRISPR target were selected for HDR in this study. The 5' homology arm included part sequences of the intron between exon6 and -7, exon7, intron sequences between exon7 and -8, and a part of exon8. The 3' homology arm included intron sequences between exon8 and -9 and 3' UTR of the *MKK2* locus up to 400 bp. We designed DDT included remained sequences of exon8, exon9, BleoR CDS, 6xHis, and PolyA sequences flanked by the 3'- and 5' homology arms. We added two special targets besides of DDT. The DDT then was ligated into the pRGEB31 vector to form pDDT. **(B)** pDDT and pgRNA were mixed 4:1 and introduced to the *Agrobacterium tumefaciens* to form inoculator suspension. we

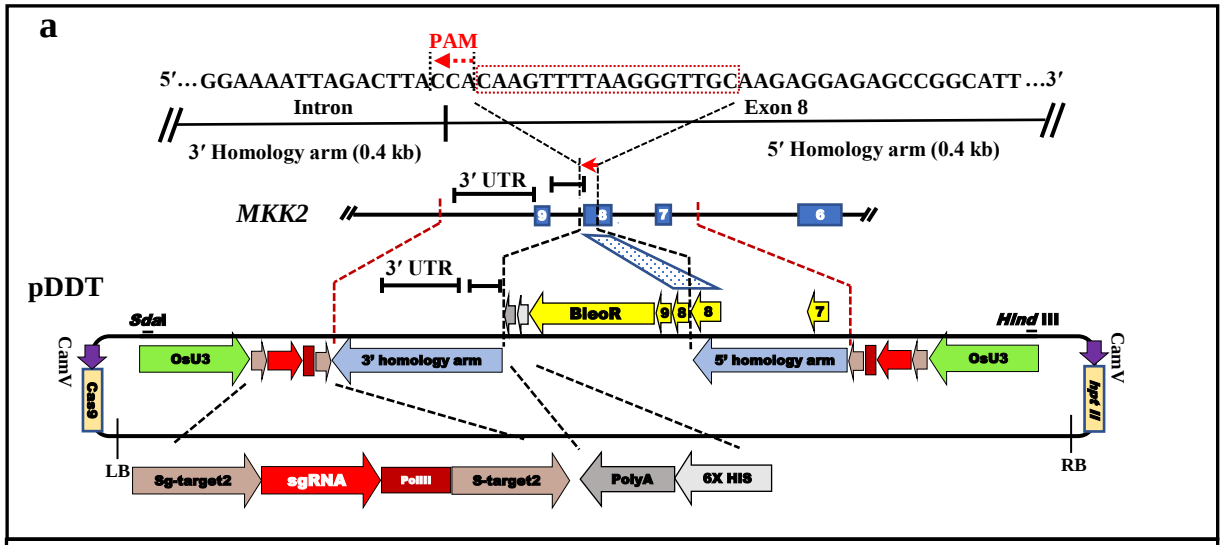
condensed inoculator up to OD600=2.5 and then dipped all cut off leaves. The putatively transformed calli were regenerated on zeocin. We allowed transformed calli to elongate and shoot. The properly transformed shoots were then transferred on rooting media and allowed to be recovered events. **(C)** Designed experiments for this study involved a normal transformation of pDDT and pgRNA (I), Overloaded CtIP (II), MRE11 (III), CtIP+MRE11 (IV), and CtIP+MRE11 and XRCC4 target (V). **(D)** The design of the TaqMan real-time PCR assay for the detection of HDR sequences and evaluation included FAM1 and -2 DNA binding probes. **(E)** the 2D kernel density plot of TaqMan real-time PCR for FAM1 and FAM 2 fluorescent intensities related to the experiment I. There is no density for the edited event. Samples were analyzed in quadruplicate. **(F)** Schematic alignment of events involved in experiment I. Regarding the alignment, five events revealed no FAM signals and one event revealed sequences similar to the wild-type.

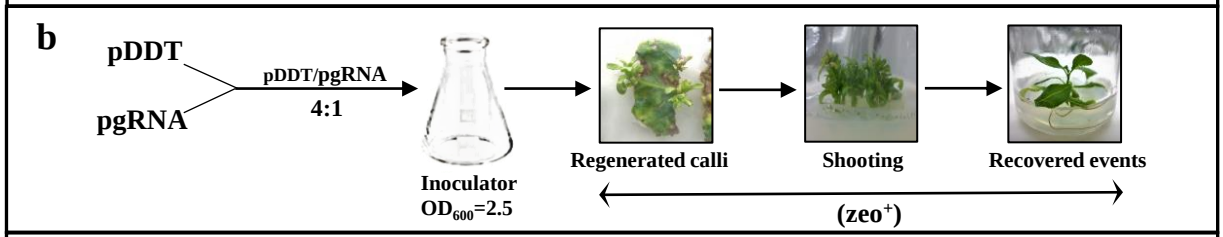
Figure 2. Comparison of improvement of the edited events with the other transformants partial and mutated through experiments II, III, and IV. (A) The fusion of CtIP to the C-terminal of Cas9 caused to form pgCtIP vector. **(B)** Density plot of FAM1 and -2 intensities revealed an expansion in edited events against partial, mutant, and wild-types. **(C)** Event alignments involved in experiment II. Three events expressed *MKK2* (Partial FAM1 #21, FAM1 and -2 #29, and partial FAM1 #35). **(D)** The fusion of MRE11 to the N-terminal of Cas9 caused to form pgMR vector. **(E)** The density plot of FAM1 and -2 ΔΔCt revealed an increased intensity of partial FAM1 events. **(F)** Event alignments involved in experiment III. Four events expressed *MKK2* (#10, #23, #36, and #45) and one partial FAM2 event (#6) expressed *BleoR* and caused to be recovered. Only one edited event (#45 FAM1 and -2) expressed *MKK2*. **(G)** The fusion of both MRE11 and CtIP to Cas9 began to form pgCtMR vector. **(H)** Density plot revealed a remarkable increase of edited events signals in confronting with three earlier experiments. **(I)** event alignments involved in experiment IV. Nine events expressed *MKK2* (#9, #17, #39, #45, #54, #60, #68, #79, and #83) and four events expressed *BleoR* (Recovered) (#17, #54, #68, and #90). Three recovered events (#54, #68, and #90) exhibited both FAM1 and -2, and one recovered event exposed partial FAM1 (#17).

Figure 3. *XRCC4* targeting, HDR efficiency, and polymorphism evaluation. (A) XRCC4 target sequences were added to the pgCtMR to form the pggCtMR vector. **(B)** The density plot revealed a significant increase of FAM1 and -2 intensities in edited events in comparison with

the earlier experiments and significant decrease intensities in WT and mutated events. **(C)** Event alignments involved in experiment V. Surprisingly, twenty events expressed *MKK2* and twelve events exhibited *BleoR* and rooted (recovered). Three of recovered events (#37, #53, and #59) showed partial HDR (partial FAM). Fifteen events revealed FAM1 and -2 signals (edited) and only one event was visualized as mutated. **(D)** Bar plot of promotion of HDR efficiency in experiments. HDR efficiency plot revealed that XRCC4 deficient (ExV) let to form HDR significantly more than the fusion of CtIP (ExII), MRE11 (ExIII), and CtIP+MRE11 (Ex.V) with the Cas9. Also, ExIV meaningfully revealed more HDR than ExII and -III. Twelve biological replicates of regenerated resistant calli were selected for this assay with three technical repeats. Bars represent the mean of date. Curves describe the normal distribution of data. Error bars interpret standard error (SE). **(E)** The identification of total FAM signals (FAM1 and -2) visualized in the experiments showed more signs remarkably measured in ExV than ExI, II, and-III. **(F)** Identification of the polymorphisms that happened in homology arms through experiments. Box and Whisker plot revealed that most polymorphisms happened in homology arms by ExI and it was significantly more than those in ExV and -IV. Error bars represent SE. **(G)** Stacked column plot of total polymorphisms happened in DDT integration into the poplar genome. Deletion and insertion happened much more than the other types. SNP and substitution happened less than the other types. Whisker and normal distribution curves exposed that the total polymorphisms caused by *XRCC4* deficient were less than the other experiments. **(H)** Evaluation of the expression of *MKK2* and *BleoR* through the experiments. Heat map plot revealed that the highest expression of *MKK2* and *BleoR* occurred in ExV. Nine biological replicates with three technical repeats were used in this assay. WT and pDDT were used as the control for *MKK2* and *BleoR* expression respectively. **(I)** The Heat map of the interaction of proteins included in HDR and NHEJ. Overexpression *CtIP* and/or *MRE11* caused to enhance the expression of *BRCA1*, *Rad50*, and *Rad51* and to demote the expression of *Lig4* and *XRCC4*. The highest expression of the HDR factors visualized in ExV means that XRCC4 deficient caused to decrease the expression of NHEJ factor Lig4 and led to intensifying HDR efficiency. Three biological replicates with three technical repeats were used in this assay.

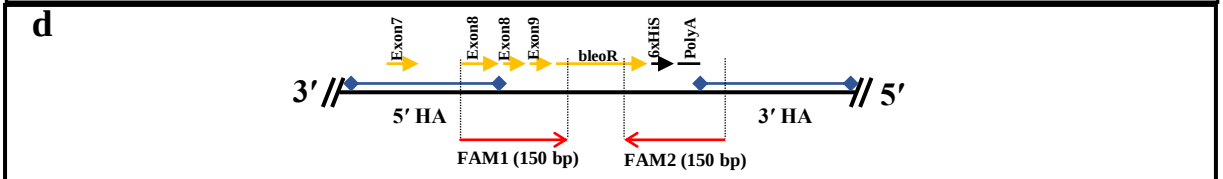
Figure 1

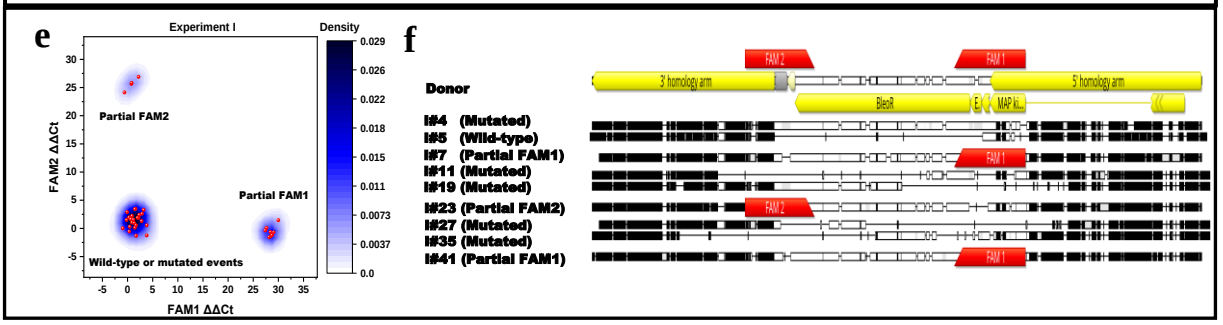




c

Experiments	Plasmid			Number of leaves for transformation	Regenerated resistant calli	Regenerated resistant shoots from calli	Non recovered mutant events	WT events	Recovered edited events (Rooting)
	XRCC4 target	MRE11	CtIP						
I	-	-	-	66	13	9	8	1	0
II	-	-	+	67	23	17	14	1	1
III	-	+	-	68	25	15	12	2	1
IV	-	+	+	70	33	22	17	1	4
V	+	+	+	65	42	31	19	0	12





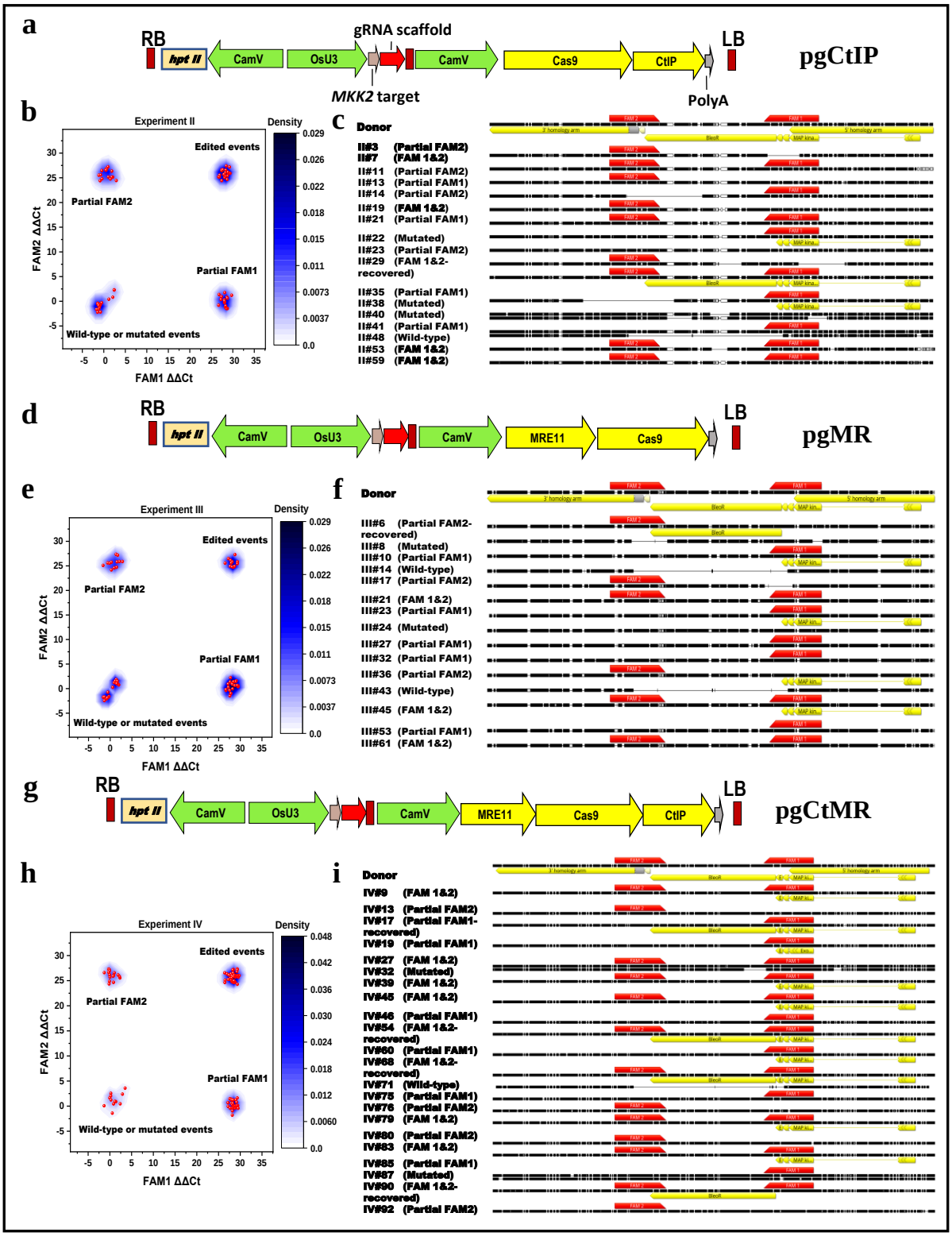
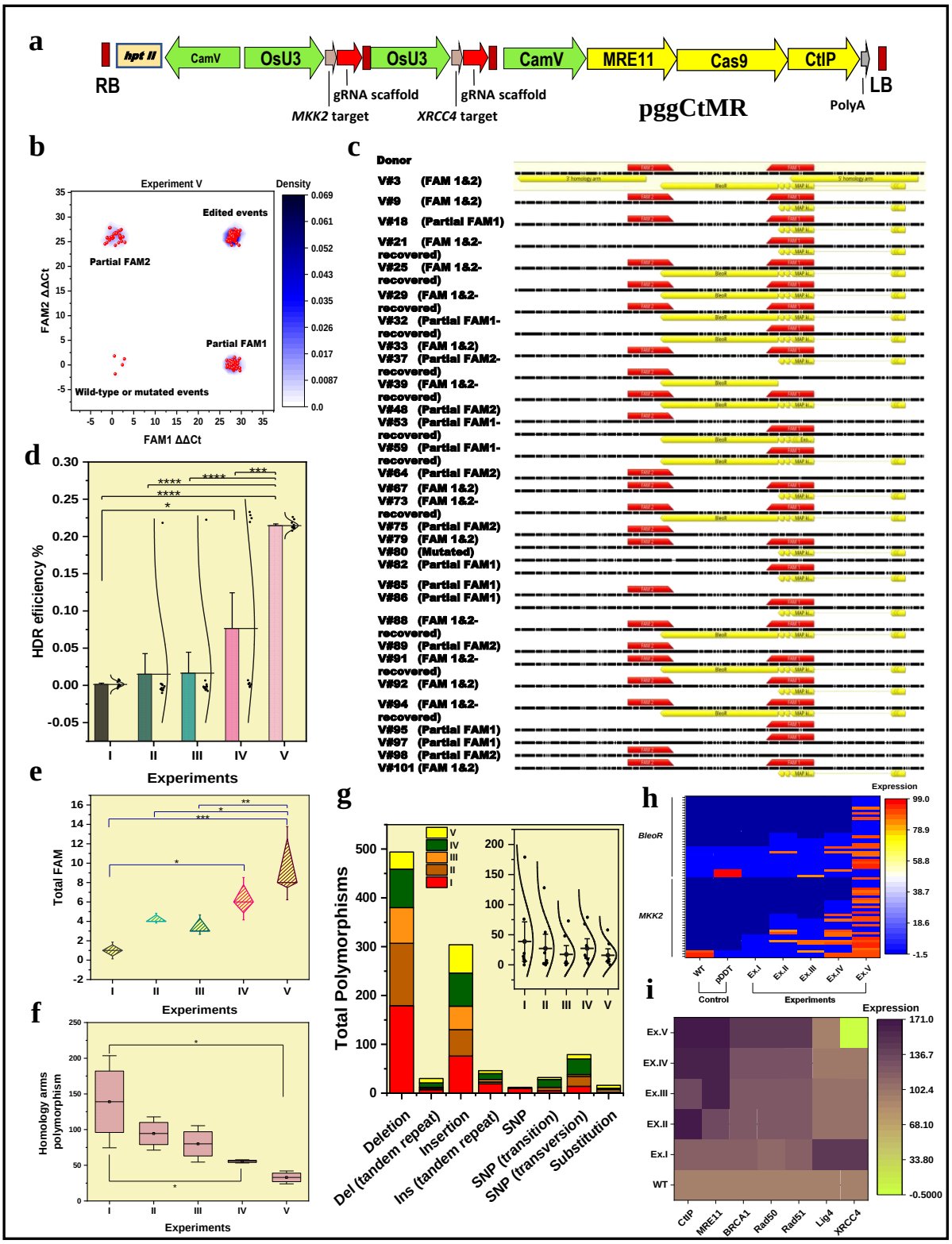

a
RB
hpt II
CamV
OsU3
gRNA scaffold
CamV
Cas9
CtIP
LB
pgCtIP
MKK2 target
PolyA
b
Experiment II
Density
Edited events
Partial FAM2
Partial FAM1
Wild-type or mutated events
FAM2 ΔΔCt
FAM1 ΔΔCt
c
Donor
II#3 (Partial FAM2)
II#7 (FAM 1&2)
II#11 (Partial FAM2)
II#13 (Partial FAM1)
II#14 (Partial FAM2)
II#19 (FAM 1&2)
II#21 (Partial FAM1)
II#22 (Mutated)
II#23 (Partial FAM2)
II#29 (FAM 1&2-recovered)
II#35 (Partial FAM1)
II#38 (Mutated)
II#40 (Mutated)
II#41 (Partial FAM1)
II#48 (Wild-type)
II#53 (FAM 1&2)
II#59 (FAM 1&2)
d
RB
hpt II
CamV
OsU3
CamV
MRE11
Cas9
LB
pgMR
e
Experiment III
Density
Edited events
Partial FAM2
Partial FAM1
Wild-type or mutated events
FAM2 ΔΔCt
FAM1 ΔΔCt
f
Donor
III#6 (Partial FAM2-recovered)
III#8 (Mutated)
III#10 (Partial FAM1)
III#14 (Wild-type)
III#17 (Partial FAM2)
III#21 (FAM 1&2)
III#23 (Partial FAM1)
III#24 (Mutated)
III#27 (Partial FAM1)
III#32 (Partial FAM1)
III#36 (Partial FAM2)
III#43 (Wild-type)
III#45 (FAM 1&2)
III#53 (Partial FAM1)
III#61 (FAM 1&2)
g
RB
hpt II
CamV
OsU3
CamV
MRE11
Cas9
CtIP
LB
pgCtMR
h
Experiment IV
Density
Edited events
Partial FAM2
Partial FAM1
Wild-type or mutated events
FAM2 ΔΔCt
FAM1 ΔΔCt
i
Donor
IV#9 (FAM 1&2)
IV#13 (Partial FAM2)
IV#17 (Partial FAM1-recovered)
IV#19 (Partial FAM1)
IV#27 (FAM 1&2)
IV#32 (Mutated)
IV#39 (FAM 1&2)
IV#45 (FAM 1&2)
IV#46 (Partial FAM1)
IV#54 (FAM 1&2-recovered)
IV#60 (Partial FAM1)
IV#68 (FAM 1&2-recovered)
IV#71 (Wild-type)
IV#75 (Partial FAM1)
IV#76 (Partial FAM2)
IV#79 (FAM 1&2)
IV#80 (Partial FAM2)
IV#83 (FAM 1&2)
IV#85 (Partial FAM1)
IV#87 (Mutated)
IV#90 (FAM 1&2-recovered)
IV#92 (Partial FAM2)

Figure 2

Figure 3


a
RB
hpt II
CamV
OsU3
OsU3
CamV
MRE11
Cas9
CtIP
LB
gRNA scaffold
gRNA scaffold
MKK2 target
XRCC4 target
pggCtMR
PolyA
b
Experiment V
Density
Edited events
Partial FAM2
Partial FAM1
Wild-type or mutated events
FAM2 ΔΔCt
FAM1 ΔΔCt
c
Donor
V#3 (FAM 1&2)
V#9 (FAM 1&2)
V#18 (Partial FAM1)
V#21 (FAM 1&2-recovered)
V#25 (FAM 1&2-recovered)
V#29 (FAM 1&2-recovered)
V#32 (Partial FAM1-recovered)
V#33 (FAM 1&2)
V#37 (Partial FAM2-recovered)
V#39 (FAM 1&2-recovered)
V#48 (Partial FAM2)
V#53 (Partial FAM1-recovered)
V#59 (Partial FAM1-recovered)
V#64 (Partial FAM2)
V#67 (FAM 1&2)
V#73 (FAM 1&2-recovered)
V#75 (Partial FAM2)
V#79 (FAM 1&2)
V#80 (Mutated)
V#82 (Partial FAM1)
V#85 (Partial FAM1)
V#86 (Partial FAM1)
V#88 (FAM 1&2-recovered)
V#89 (Partial FAM2)
V#91 (FAM 1&2-recovered)
V#92 (FAM 1&2)
V#94 (FAM 1&2-recovered)
V#95 (Partial FAM1)
V#97 (Partial FAM1)
V#98 (Partial FAM2)
V#101 (FAM 1&2)
d
HDR efficiency %
Experiments
e
Total FAM
Experiments
f
Homology arms polymorphism
Experiments
g
Total Polymorphisms
Deletion
Del (tandem repeat)
Insertion
Ins (tandem repeat)
SNP
SNP (transition)
SNP (transversion)
Substitution
h
Expression
BleoR
MKK2
WT
pDDT
Ex.I
Ex.II
Ex.III
Ex.IV
Ex.V
Control
Experiments
i
Expression
Ex.V
EX.IV
Ex.III
EX.II
Ex.I
WT
CtIP
MRE11
BRCA1
Rad50
Rad51
Lig4
XRCC4