NHEJ deficiency develops homologous recombination in poplar meaningfully further than the overexpression of HDR factors

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

[1] College of Biology and the Environment, 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, Ali Movahedi, and Liming Yang: College of Biology and the Environment, Co-Innovation Center for Sustainable Forestry in Southern China, Key Laboratory of Forest Genetics & Biotechnology, Ministry of Education, Nanjing Forestry University, Nanjing 210037, China. E-mail: qzhuge@njfu.edu.cn; ali_movahedi@njfu.edu.cn; yangliming@njfu.edu.cn; Fax: +86 25 85428701

§ These authors are equally as the first author

Running title: *XRCC4* deficiency improves HDR efficiency about 50 times in poplar

Abstract

- Efficient homology-directed DNA repair (HDR) is a vital difficulty confronting researchers to replace the target genome's desired fragment. In plants, scientists have performed meticulous investigations on herbal, crops, and citrus trees using HDR effector proteins, CtIP and MRE11, to obtain double-stranded breaks (DSBs) more precisely. Although HDR efficiency in plants previously has been reported, no record has been declared about HDR efficiency in poplars.
- Here, we hypothesized that inhibition of nonhomologous recombination cofactors XRCC4, together with enhancing the HDR pathway activities, enables us to generate the HDR efficiency in poplar trees. In this study, the *BleoR* gene was used to integrate into the interested site and develop resistant poplars against Zeocin antibiotics. We designed plasmids, including different fusions of HDR proteins and, together with the *XRCC4* target. Furthermore, real-time PCR, western blotting, RT-PCR, RT-qPCR, southern blotting, and DNA sequencing were applied to exhibit and evaluate HDR efficiency.
- While both applying HDR proteins and *XRCC4* deficiency simultaneously could improve HDR efficiency, which showed about 50 times more than usual editing by CRISPR-Cas9, the only using HDR proteins without *XRCC4* deficiency showed about 16 times more. We developed a new recombinant poplar genome to generate stable lines resistant to the Zeocin antibiotic.

Keywords: CRISPR; XRCC4; Homologous recombination; BleoR; *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 unexpected and irregular repair in the DSB area. NHEJ has shown a predominant pathway and occurs among the cell cycle widely (Panier & Boulton, 2014), while HDR rarely occurs (Puchta, 2005) because of the low efficiency of transferring DNA or RNA fragments into the cell nucleus resulting in difficulties to target and modify plant genomes accurately. 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 a ten-fold enhancement in the efficiency of the insertion of the 35S promoter in 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 using protoplasts for this purpose (Svitashev *et al.*, 2016), the methods have been very inefficient. Concerning the difficulty of donor DNA template (DDT) delivery to the cell nucleus, the 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 significant problems in the genetic modification of trees. Previous data indicated it is necessary to increase the number of cells containing DDTs at S/G2 cell division phases to increase HDR efficiency (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). Still, there was no report on enhancing 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 impact on enhancing HDR efficiency in mammalians have already been reported (Tran *et al.*, 2019), with at least a 2-fold increase HDR and a 6-fold increase in HDR / NHEJ ratio. Cas9 integrates with MRE11, CtIP, and Rad51, Rad52, and promotes significant HDR efficiency in human cells

and decreases NHEJ (Tran *et al.*, 2019) significantly. On the other hand, inhibition of DNA ligase IV (LIG4), Ku 70, Ku 80, and XRCC4, which are known as the most critical NHEJ factors (Pierce *et al.*, 2001; Friesner & Britt, 2003, 2003; Maruyama *et al.*, 2015; 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 by 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). Programmable endonucleases affect DSBs at target positions in genomic DNA but can also create undesired breaks outside of on-target positions and create off-target mutations. Cleavage at off-target sites direct to chromosomal rearrangements, including translocations, insertions, and deletions, which happen in the interruption of regular gene expression and the activation of oncogenes (Li *et al.*, 2019). Today, scientists have realized that reducing off-target may allow efficient and accurate genome editing (Wu & Yin, 2019). For this why, the effect of off-targets on efficient, precise genome editing and ways to reduce their impacts has already been studied (Hajiahmadi *et al.*, 2019; Li *et al.*, 2019; Wu & Yin, 2019).

In this study, we hypothesized that inhibition of nonhomologous recombination cofactors XRCC4 promotes the HDR efficiency in poplar trees. We used the optimized length of homologous arms 400 bp (Song & Stieger, 2017) and 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).

Materials and Methods

Plant 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 used stems from four weeks old young poplars and dipped them in the optimized of *Agrobacterium tumefaciens* stimulant and pathogenic suspension (OD_{600}: 2.5, 120 min, pH ~5, Acetosyringone (As): 200 µM) (Movahedi *et al.*, 2014) for 5 min with gentle shaking and then transferred in semi-solid woody plant medium (WPM) enriched with 0.05 mg/L Indole-3-butyric acid (IBA), 0.006 mg/L thidiazuron (TDZ), 200 µM As and 0.5% (w/v) agar. Afterward, the stimulated

stems were incubated in a dark area at 23°C for two days. The assumed transformants were then co-cultivated in selection media enriched with 0.1 mg/L IBA, 0.006 mg/L TDZ, 100 mg/L cefotaxime, 8 mg/L hygromycin, 50 mg/L Zeocin and 0.8% (w/v) agar. Two weeks regenerated buds were then sub-cultured independently in media including 0.1 mg/L IBA, 0.001 mg/L TDZ, 100 mg/L cefotaxime, 8 mg/L hygromycin, 50 mg/L Zeocin and 0.8% (w/v) agar to grow. After six weeks, grown buds (Each bud included four to six small leaves) were introduced in MS media with 0.1 mg/L IBA, 200 mg/L cefotaxime, 8 mg/L hygromycin, 100 mg/L Zeocin, and 0.8% (w/v) agar to root. Five lines were used for each experiment independently, and each line included about 30 individuals.

Targets and protein detection

We decided to target the *MKK2* gene from *Populus trichocarpa* (POPTR_0018s05420g; Chromosome 18). Therefore, 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. To detect targets, we used Geneious Prime® 2020.1.1 to analyze *MKK2* locus and to detect targets in comparison with the whole genome of *P. trichocarpa*, which has been already downloaded from NCBI (Supplementary Table 1) (Hsu *et al.*, 2013; Doench *et al.*, 2014). Geneious Prime® also has been used to analyze the *XRCC4* (POPTR_0010s08650g, Chromosome 10) gene for knocking out. The PAM motif target sequences were concerned with the exon 8 area from the *MKK2* and exon 1 area from the *XRCC4* genes. Furthermore, and to evaluate the effect of HDR proteins and also proper function of edited *MKK2* gene in achieved poplar trees, we used Uniprot to use CtIP (POPTR_001G269700v3), MRE11 (POPTR_0001s41800g), BRCA1 (POPTR_0005s26150g), Rad50 (POPTR_0001s32760g), Rad51 (POPTR_0014s06360g), Lig4 (POPTR_0018s13870g).

***MKK2* locus target oligo synthesis**

For targeting indicated DNA sequences, we designed a pair of oligos (Supplementary Table 2; *MKK2* Oligo-F and -R) flanked by *Bsa*I adaptors. Synthesized oligos were then ligated into digested pRGEB31 vectors by *Bsa*I restriction enzyme (Xie & Yang, 2013) to construct pgRNA (Supplementary 1a). Afterward, we transferred all vectors into *E. coli* (DH5α) and propagated under normal conditions. Vectors were then extracted using the plasmid midi kit (Qiagen, USA) and confirmed by sanger sequencing (GenScript, Nanjing).

Construction of DDT and pDDT

To produce DDT, we designed five fragments, constructed and ligated them, respectively (Supplementary2a). To construct fragment one, the OsU3 promoter and gRNA scaffold were isolated from pRGEB31 (Supplementary Table 2, OS1-F and -R) flanked by HindIII and BamHI endonucleases. To increase the amount of DDT in the cell nucleus and improve HDR efficiency, we decided to use the cleavage property of Cas9 with designing two unique gRNA targets 1 and -2 (No on- and -off-targets through whole poplar genome) besides DDT (Zhang *et al.*, 2017) (Supplementary 1b). Thus, we designed special gRNA oligos (Sgo1-F and -R) (Supplementary 2a; Supplementary Table 2, special gRNA oligo1-F and -R) as the described details (Xie & Yang, 2013) to form special gRNA target1 (Sgt1) and to ligate into the fragment one. To construct fragment two, we isolated 5' homology arm (400 bp) sequences from *P. trichocarpa* genomic DNA (Supplementary Table 2, 5' Ho-F-1 and -R-1). Afterward, regular PCR was carried out using primers with the extensions of BamHI-special target1 (St1) and 39 bp from complemented 5' of fragment 3 (Supplementary Table 2, 5' Ho-F-2 and -R-2) (Supplementary 2a) to achieve component two.

To construct fragment three, we isolated the BleoR CDS from the PCR®-XL-Topo® vector (Supplementary Table 2, BleoR-1092F and -2276R). Then, the overlap-PCR was performed (Supplementary Table 2, BP1,2,3-F and -R) using isolated BleoR CDS as the template to add the remained sequences from exon 8 and exon 9 to the 5' region of BleoR CDS and also 6XHis and PolyA tail to the 3' area of BleoR CDS (Supplementary 2a). To assemble fragment four, we isolated a 3' homology arm (400 bp) (Supplementary Table 2, 3' Ho-F-1 and -R-1) from *P. trichocarpa* genomic DNA. Then, we performed PCR to extend 3' homology arm with 30 bp Poly-T and *Nco*I-special target2 (St2) sequences (Supplementary Table 2, Ho-F-2 and -R-2) (Supplementary 2a). Finally, we performed standard PCR to isolate the OsU3 promoter and gRNA scaffold again from pRGEB31 (Supplementary Table 2, Os2-F and Os2-R). Moreover, we designed special gRNA oligos (Sgo2-F and -R) (Supplementary 2a; Supplementary Table 2, special gRNA oligo2-F and -R) again as the described details (Xie & Yang, 2013) to form special gRNA target2 (Sgt2) and to ligate into the fragment five.

To construct pDDT, we ligated fragments three and two using PCR (Supplementary 2b). For this, we designed a 39 bp overhang on fragment two that was complementary to the end of fragment three to form preliminary DDT (Supplementary 2b). In this PCR, we prepared a PCR

reaction with 500 ng of each component. We used everything in PCR reaction except primers and then denatured fragments at 95 degrees for 5 minutes and allowed two annealing and extension cycles. We allowed PCR products to anneal at 68 degrees to avoid nonspecific hybridization amongst the long PCR products for 30 seconds and then extend for one minute at 74 degrees to have a double-stranded outcome. Then we added the primers to the distal ends of fragments two and three and performed one standard PCR. We purified PCR products and ligated into the pEASY vector to sequence and confirm. Then we ligated the preliminary DDT product to fragment four as described before and formed secondary DDT products (Supplementary 2b). After sequencing and confirmation, we used the restriction cloning technique to ligate secondary DDT product to the fragments one and four (Supplementary 2b) to achieve DDT products. Briefly, we incubated a reaction including 50 ng of each digested fragments, 10x T4 DNA ligase buffer 0.5 ul, T4 DNA ligase (NEB) 1 ul, and H2O to 5 ul at 25 degrees for 4 hours and transferred into *E. Coli* DH5α competent cells for sequencing and confirmation. Subsequently, we used the restriction cloning technique to merge the DDT product and pRGEB31 vector and form the pDDT vector (Supplementary 2b).

Synthesis of pgCtIP and pgMR

To design a fused CtIP and Cas9 cassette, we isolated the CaMV35S promoter, 3xFLAG, and Cas9 CDS from pRGEB31 (Supplementary 3a) using designed primers (Supplementary Table 2). In the next step, we obtained CtIP CDS using RT-PCR from the *Populus trichocarpa* genome (Supplementary 3a; Supplementary Table 2, CtIP-F and -R). The 3'UTR and PolyA fragments were isolated from the pCAG-T3-hCAS-pA plasmid (Supplementary 3a; Supplementary Table 2, PolyA-F and -R). To complete pgCtIP, we ligated CaMV35S and 3xFLAG fragments using restriction cloning and formed backbone 1 (Supplementary 4a). The isolated Cas9 and the obtained CtIP CDS were also ligated, applying restriction cloning to form the backbone 2 (Supplementary 4a). The backbones 1 and 2 were then ligated using *Hind*III restriction cloning to form backbone 3 (Supplementary 4a). In the next step, the resulted backbone 3 was ligated to the assembled 3'UTR-PolyA using *Stu*I restriction cloning to form the CtIP cassette (Supplementary 4a; Supplementary 5a). We used *Sda*I and *Pme*I restriction enzymes to restrict the cloning of the CtIP cassette and pRGEB31 and achieve the pgCtIP plasmid (Supplementary 4a; Supplementary 5a).

To construct a fusion of MRE11 and Cas9, we isolated CaMV35 promoter, 3xFLAG, Cas9, 3'UTR, and PolyA as same the previous steps (Supplementary 3b; Supplementary Table 2). The MRE11 CDS was obtained recruiting extracted total RNA from *Populus trichocarpa* genome and RT-PCR as mentioned above (Supplementary 3b; Supplementary Table 2, MRE-F and R). To complete pgMR, we ligated the isolated CaMV35S and 3xFLAG fragments concerning *Xho*I endonuclease to form backbone 1 (Supplementary 4b). On the other hand, we constructed backbone 2 using the isolated Cas9 and 3'UTR-PolyA fragments (Supplementary 4b). The backbone 1, backbone 2, and MRE11 CDS product were then merged concerning *Not*I and *Nde*I restriction cloning to form MR cassette (Supplementary 4b; Supplementary 5b). Afterward, we used restriction cloning with *Sda*I and *Pme*I to construct pgMR plasmid (Supplementary 4b; Supplementary 5b).

Synthesis of pgCtMR and pggCtMR

To construct the CtMR cassette, we prepared all the required fragments, as mentioned above (Supplementary 3c). Afterward, we merged the CaMV35S and 3xFLAG components using *Xho*I restriction cloning to form backbone 1 (Supplementary 6a). We then ligated backbone 1 and the already obtained MRE11 CDS product (Supplementary Table 2, MRE-F and -R) using *Not*I restriction cloning to form backbone 2 (Supplementary 6a). On the other hand, the isolated Cas9 and the obtained RT-PCR product CtIP CDS were ligated using *BamH*I restriction cloning to form backbone 3 (Supplementary 6a). We then used backbone 3 and isolated 3'UTR-PolyA fragment to form backbone 4 (Supplementary 6a). Eventually, we cloned backbones 2 and 4 to construct the CtMR cassette (Supplementary 6a; Supplementary 5c) and thereupon implemented *Sda*I and *Pme*I restriction cloning to ligate CtMR cassettes into pRGEB31, forming pgCtMR plasmid (Supplementary 6a; Supplementary 5c). To target the *XRCC4* gene and *MKK2* simultaneously, we designed one cassette, including both *XRCC4*, 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%) (Hsu *et al.*, 2013; Doench *et al.*, 2014), and *MKK2* gRNAs. For this purpose, we used primers (Supplementary Table 2, XR-Cass1-F and -R) to isolate the OsU3 promoter and gRNA scaffold from the pRGEB31 vector and then used *MKK2* designed oligos (Supplementary Table 2, *MKK2* Oligo-F and -R) to ligate *MKK2* target duplex (Supplementary 3d). Besides, we used primers (Supplementary Table 2; XR-Cass2-F and -R) to isolate the OsU3 promoter and gRNA scaffold again. In this process,

we applied *XRCC4* designed oligos (Supplementary Table 2; XRCC4-Oligo1 and -2) to ligate *XRCC4* target duplex (Supplementary 3d). The achieved fragments were then cloned using *Kas*I restriction cloning to form *XRCC4*-Cassette (Backbone 1) (Supplementary 6b; Supplementary 5d). Afterward, the *XRCC4*-Cassette was cloned into pRGEB31 using *Hind*III and *Sda*I restriction cloning to form backbone 2 (Supplementary 6b). Finally, we used *Sda*I and *Pme*I restriction cloning to clone the CtMR cassette into the backbone 2 forming pggCtMR plasmid (Supplementary 6b; Supplementary 5d). We performed PCR, cloning into pEASY T3 vector, and DNA sequencing in all the above processes for confirming the right ligation.

RT-PCR, DNA sequencing, Southern blotting, and Western blotting

Total RNA (100 ng/ml) was extracted from young leaves of five weeks grown buds on Zeocin applying TRIzol. We then carried out reverse transcription using total RNA and oligo-dT primers to synthesize the first cDNA strand (PrimeScript One-Step RT-PCR Kit Ver.2, Takara Biotechnology, Dalian, China) according to the manufacturer's instructions. Afterward, we designed two RT-PCR for both investigations of right *MKK2* transcription and right happening HDR. The first RT-PCR was intended to isolate a 920 bp of MKK2 CDS (Supplementary Table 2, RT-F and R) while the forward primer was designed from 5' region of exon 9 (15 bp) and 3' region of exon 8 (15 bp). This RT-PCR purpose was to show the precise attaching of exon 8 and 9 to direct the transcription of *MKK2* correctly. The second RT-PCR was performed to isolate a 413 bp of recombinant CDS (Supplementary Table 2, RT-F-107 and RT-R-519). The forward primer was designed from *BleoR*, and the reverse primer was designed from exon 7 of *MKK2*. The purpose of this RT-PCR was to show the explicit HDR happening through our experiments via transcription of single mRNA from *MKK2* and BleoR.

Genomic DNA was extracted from young leaves of five weeks grown buds on Zeocin, applying the DNeasy Plant Mini Kit (Qiagen, USA). The quality of the extracted genomic DNA (250–350 ng/μl) was determined by a BioDrop spectrophotometer (UK). To DNA sequencing, we carried out PCR using designed primers (Supplementary Table 2, MKK2-S-7F and MKK2-S-1139R), Easy Taq polymerase (TransGene Biotech), and 50 ng of extracted genomic DNA as a template. All desired bonds were then cut off from gels, purified, and sent to the company for sequencing (GeneScript, Nanjing), alignment, and analysis (Supplementary 7-11). Southern blotting was performed to verify the integration of *BleoR* into the poplar genome. 500 ng of genomic DNA was cleaved with *BamH*I and *Hind*III at 37 °C for4 h. The digested DNA was

then used as a template for PCR to label a 160 bp probe from integrated *BleoR* CDS into the genomic DNA (Supplementary Table 2; S-F and -R). In this step, we used the DIG (digoxigenin) reagent, according to the manufacturer's instruction (catalog number 11745832910; Roche, Basel, Switzerland). PCR product was then segregated on a 0.8% agarose gel. The separated fragments were shifted on a Hybond N+ nylon membrane (Amersham Biosciences BV, Eindhoven, The Netherlands).

For extraction of proteins, 150 mg fresh leaves of five weeks grown buds were milled in 500 µl extraction bu□er (125 mM Tris, pH 6.8, 4 M Urea, 5% β -mercaptoethanol, 4% w/v SDS). The centrifuge was then performed at 13,000 rpm for 10 min, and the supernatant was obtained for gel analysis. The extracted protein was then boiled in loading bu□er (24% w/v glycerol, 100 mM Tris, a drop amount of Bromophenol Blue, 4% v/v β -mercaptoethanol, 8% w/v SDS) for 10 min. The extracted protein was analyzed by SDS-PAGE and conceived using coomassie brilliant blue R-250 staining. After that, we carried out western blotting according to Sambrook *et al.* (1989) using a rabbit anti-His polyclonal antibody developed in our laboratory as the primary and peroxidase-conjugated goat antirabbit IgG (Zhongshan Biotechnique, Beijing, China) as the secondary antibody.

TaqMan real-time PCR

To test the effect of designed parameters in all experiments on the proper integration of exogenous *BleoR* with both homology arms, we decided to run the TaqMan assay applying dye labels such as FAM and VIC adopting Applied Biosystem real-time PCR (Applied Biosystems, USA). We used high quality extracted genomic DNA (Refer to the southern blotting) as the template for running TaqMan real-time PCR. In this assay, two fluorescent FAM and VIC will attach to the 5' region of the probe, while a non-fluorescent quencher (NFQ) binds to the 3' region. Thus, we designed primers to probe two 150 bp fragments FAM1 (Supplementary 2, FAM1-F and -R) and FAM2 (Supplementary 2, FAM2-F and -R). These primers were designed in such a way that FAM1 was able to probe 114 bp nucleotides from the 5' homology arm and also 36 bp from BleoR. Besides, FAM2 was able to probe 105 bp nucleotides from the 3' homology arm and 45 bp from the BleoR (Supplementary 12). We also designed primers (Supplementary 2, VIC-F, and -R) to probe one 106 bp fragment VIC on the *actin* gene as the reference with stable copy number (Supplementary 12). All samples were analyzed in quadruplicate.

Evaluation of HDR efficiency

To evaluate the HDR efficiency, we decided to calculate and compare the ΔΔCt mean of BleoR expression integrated into the poplar genome from all grown buds in five designed experiments separately. In this step, we used the synthesized cDNA (Point to the RT-PCR section) and designed primers (Supplementary Table 2, BleoR-52F and -151R) to carry out real-time PCR. We used the Fast Start Universal SYBR Green Master (Rox; No. 04913914001: Roche, USA) and performed three technical repeats for each event. Then, we used ANOVA-One way to analyze the achieved mean data and compared.

RT-qPCR

We applied RT-qPCR using synthesized cDNA from grown buds on Zeocin (as mentioned above) as the template and designed primers (Supplementary Table2, RT-qPCR part) to investigate the expression of *BleoR* and *MKK2* genes and their impact on each other. We also explored our method's impact to develop HDR efficiency on HDR (CtIP, MRE11, BRCA1, Rad50, and Rad51) and NHEJ (Lig4, XRCC4) influential factors.

Statistical analysis

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 when the confidence intervals presented no overlap of the mean values with an error value of 0.05.

Results

HDR strategy and target detection

Our purpose in this study was to increase HDR efficiency to integrate one exogenous into the poplar genome. For this purpose, we decided to integrate *BleoR* exogenous with *the MKK2* gene (Figure 1a) to generate resistant poplar against the Zeocin antibiotic. We started to recognize CRISPR sites located on the *MKK2* gene near the 3' UTR and scored all detected targets (Hsu *et al.*, 2013; Doench *et al.*, 2014) concerning higher scores denoting higher activity specificity and less off-target activity. We discovered one target located on exon 8 with the highest activity score and no detected off-target on CDS throughout the whole *Populus trichocarpa* genome (Figure 1a and Supplementary Table 1). According to Figure 1b, the strategy of integrating *BleoR* was to target exon 8 in *MKK2*. Protospacer Adjacent Motif (PAM), including two nucleotides from exon 8 and one nucleotide from intron 8, was detected close to

the end of exon 8 to lead Cas9. 400 bp from upstream sequences of the desired target, including a few sequences from intron 6, exon 7, intron 7, and a few sequences from exon 8), was selected to be 5' homology arm (Figure 1b). Regarding the desired CRISPR target and to avoid *MKK2* damaging, we decided to add particular sequences instead of remained nucleotides from exon 8 (Leu-Ala-Thr-Leu-Lys-Thr-Cys) and also particular sequences instead of exon 9 (Val-Leu-Val-Lys-Met) to the end of 5' homology arm (Supplementary 1b). Then, 375 bp *BleoR* CDS, 18 bp 6xHis tag, and 30 bp Poly A were designed to attach the achieved DDT sequences, and finally, 400 bp from downstream sequences of the desired target was selected to be 3' homology in the designed DDT.

Transformant poplars with no integrated *BleoR* revealed no recovery on selection media

To investigate whether *XRCC4* deficiency enables to improve the HDR efficiency at the desired locus, we first tested the integration of DDT without both *XRCC4* deficiency and fusing CtIP and MRE11. Therefore, to transform pgRNA and pDDT into plant cells, we utilized the optimized of *Agrobacterium tumefaciens* strain EHA105 stimulant and pathogenic suspension OD_{600} = 2.5 (~2×10^9 cell ml^{-1}) with the 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 2a).

For transformation, cut stems were carried on regeneration media (Figure 2b-a). Regenerated calli were then transported on selection media, including Zeocin, to elongate. Grown buds (Figure 2b-b) were then selected to convey on rooting media (Recovering) to sieve recovered events (Figure 2b-c).

We planned five different HDR experiments (Figure 2c). In the designed Experiment I (EXI), we tried to improve HDR efficiency by transferring only DDT into the plant cells (Figure 2c; Supplementary 1b). In this experiment, 34 individuals were regenerated on selection media (Including Zeocin). We then allowed them to grow and elongate. Only nine buds were grown on selection media. Grown buds were later transferred on rooting media to sieve recovered edited events. It was not observed any recovered in EXI (Figure 2c). We decided to design a plasmid that included a fused CtIP (Tran *et al.*, 2019) and Cas9 (pgCtIP) instead of pgRNA with a ratio of 4:1 pDDT/pgCtIP to promote HDR efficiency in poplars via Experiment II (ExII) (Figure 2c; Supplementary 5a). Only seventeen events were observed to be grown from a total of 42 regenerated buds. Also, only one recovered event was discerned after transferring on rooting

media. In continuous and to investigate the effect of MRE11 (Tran *et al.*, 2019) to improve HDR efficiency in poplars, we designed plasmid harboring a combined MRE11 and Cas9 (pgMR) instead of pgRNA with the same ratio of 4:1 pDDT/pgMR via Experiment III (ExIII) (Figure 2c; Supplementary 5b). In this experiment, we observed fifteen grown buds and only one recovered edited event. Because our experiments did not show significant recovered events in overcoming NHEJ to integrate *BleoR*, we determined to design Experiment IV (ExIV), and one plasmid harboring fused both MRE11 and CtIP with Cas9 (pgCtMR) (Figure 2c; Supplementary 5c). Recovered events were increased insignificantly to four. Therefore, we decided to target *XRCC4* as one key factor in the NHEJ pathway (Maruyama *et al.*, 2015) besides CtIP and MRE11 overexpressing. For this purpose, we designed Experiment V (ExV) and one plasmid harboring *XRCC4* gRNA and also fused both MRE11 and CtIP with Cas9 (pggCtMR) (Figure 2c; Supplementary 5d). We tried to transfer this plasmid into the plant cells with the same ratio of 4:1 pDDT/pggCtMR. In this experiment, recovered events were shown increased surprisingly to twelve events from thirty-one grown buds on selection media.

Confirmation of transformants by western blotting, RT-PCR, and Southern blotting

Western blotting has been carried out to confirm the exact transformation to conjugate the BleoR and MKK2 proteins. Using Western blotting, we also find out which grown buds on Zeocin have been genetically edited by integrating the BleoR. We fused 6xHis tag sequences in the C-terminal position of BleoR (Supplementary 1b). Regarding our results, we could not recognize any successful editions in events from ExI. Through screening of ExII events, one bond (event II#29) was observed about 54 KDa (Figure 3a), which led us to hypothesize the successful integration of BleoR (125 amino acids and ~13.7 KDa) into the MKK2 (365 amino acids and ~40.5 KDa) (Figure 3b). Within screening results of events from ExIII, we observed only one bond (III#6) about 14 KDa (Figure 3a). In this issue, we hypothesized that only BleoR could be expressed, and MKK2 may be knocked out by sudden Insertions and Deletions (InDels) throughout exon 7, 8, or 9 (Figure 3b). The results from ExIV showed an insufficient increase in a complete edition with three bonds (Events IV#17, #54, and #68) about 54 KDa and one bond (Event IV#92) about 14 KDa (Figure 3a). After that, we screened the results of ExV, and surprisingly, it was observed sufficient increase in successful editions with ten bonds (Events V#21, #25, #29, #32, #39, #59, #73, #88, #91, and #94) about 54 KDa and two bonds (Events V#37, and #53) about 14 kDa (Figure 3a).

To confirm western blotting issues and the expression of integrated BleoR and edited MKK2 in transformant poplars, we designed two RT-PCR assays (Figure 3c and d). Regarding the results, we did not observe any desired bond from ExI events. Then, we considered ExII and Ex III events respectively and found out 3 (events #24, #29, and #35) and 4 (events #10, #23, #36, and #45) 920 bp bonds (Figure 3e). Next, we considered the ExIV events and observed 9 (events #9, #17, #39, #45, #54, #60, #68, #72, and #83) increased desired bonds compared with events included in ExII and ExIII. We then considered RT-PCR resulted from ExV and surprisingly discovered 20 desired bonds. The second RT-PCR was also carried out, and results revealed that we could not achieve the desired 413 bp bond (Figure 3f). ExII events revealed only one 413 bp amplification (event #29), but ExIII revealed no desired bond (Figure 3f). We then considered ExIV and observed 3 (events #17, #54, and #68) 413 bp bonds, while ExV events revealed significantly increased 10 (events #21, #25, #29, #32, #39, #59, #73, #88, #91, and #94) bonds.

We decided to carry out southern blotting to confirm the achieved results by western blotting and RT-PCR assays. We selected all events that showed bonds in western blotting and designed probes that could bind with only exogenous *BleoR* sequences (Figure 3g). Regarding southern blotting (Figure 3h), all issued results from western blotting were confirmed by southern blotting. Several events (III#6, IV#90, V#37, and V#53), which have been resulted in southern blotting, were not amplified through the second RT-PCR.

Accurate investigation of edited events and HDR efficiency

We then used TaqMan real-time PCR to detect HDR probabilities more accurately and investigate *XRCC4* deficiency. It was necessary to have two probes FAM1 including exon 8 and exon 9 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 3' *BleoR* CDS (Figure 4a). In this strategy, the events that revealed fluorescent signals of FAM1 or FAM2 were assumed to be partially edited (Figure 4b). On the other hand, the events that revealed fluorescent signals of FAM1&2 were supposed to be successfully edited (Figure 4b). The events with no fluorescent signals of FAM1 and FAM2 were considered to be mutant or outlasted Wild-type (WT) (Figure 4b).

We analyzed all signals and used two-dimensional kernel density plots (Figure 4 c-g) and also one-dimensional Box and whisker (Supplementary 13a-d) to illustrate them. The analyzing

data from ExI events exhibited no edited event, while we got all signals as the partial FAM1 or partial FAM2 and much more signals as the mutant or WT poplars (Figure 4c). In ExI, the averages of fluorescent signals of FAM1 ΔΔCt and FAM2 ΔΔCt were shown proximal to 0 (Supplementary 13a). Therefore, we performed Sanger sequencing to confirm these results, and our analysis interpreted that we had not been able to achieve an edited event through ExI (Supplementary 7).

We then considered events involved in ExII and -III and noticed an increase in the signal densities of the FAM1 and FAM2 compared with ExI (Figure 4d, and e). The analysis of fluorescent signals proved the increase in the average of FAM1 ΔΔCt and FAM2 ΔΔCt about 14.5 and 13.5 from ExII events, while it was determined more FAM1 ΔΔCt about 16 and a lesser FAM2 ΔΔCt about 10 from ExIII events (Supplementary 13b, and c). The alignment of Sanger sequencing proved our findings confidently. In ExII, we found four edited events (II#7, II#19, II#53, and II#59), four partial FAM1 events (II#13, II#21, II#35, and II#41), and four partial FAM2 events (II#3, II#11, II#14, and II#23) (Supplementary 8; Supplementary 14a). In ExIII, we found three edited events (III#21, III#45, and III#61), five partial FAM1 events (III#10, III#23, III#27, III#32, and III#53), three partial FAM2 events (III#6, III17, and III#36) (Supplementary 9; Supplementary 14b). Regarding the analysis, the signal density of edited events from ExIV was increased, while the signal densities of Partial FAM1, partial FAM2, and WT or mutant events were significantly decreased (Figure 4f). The mean fluorescent signals of FAM1 ΔΔCt and FAM2 ΔΔCt from ExIV events also were observed with an increase of about 19 and 15, respectively (Supplementary 13d). We found nine edited events (IV#9, IV#27, IV#39, IV#45, IV#54, IV#68, IV#79, IV#83, and IV#90), seven partial FAM1 events (IV#13, IV#17, IV#19, IV#46, IV#60, IV#75, and IV#85), four partial FAM2 events (IV#13, IV#76, IV#80, and IV#92)(Supplementary 10; Supplementary 14c). Finally, the signal density of ExV edited events was meaningfully increased, while the partial, WT, and mutant signal densities were surprisingly reduced (Figure 4g). Moreover, the mean of fluorescent signals of FAM1 ΔΔCt and FAM2 ΔΔCt was increased about 21.5 and 18, respectively (Supplementary 13e). Therefore, we consider to analyze the related alignment and discovered 15 edited events (V#3, V#9, V#21, V#25, V#29, V#33, V#39, V#67, V#73, V#79, V#88, V#91, V#92, V#94, and V#101) (Supplementary 11; Supplementary 14d; Supplementary 17). We then decided to analyze the total achieved FAM fluorescent signals to show *XRCC4* deficiency affecting enhancing HDR based on the promising

results. We analyzed all FAM signals (FAM1, FAM2, and FAM1&2) achieved among real-time PCR and compared them through each experiment (Figure 4h). Our interpretation showed more FAM signals remarkably measured in ExV than ExI, II, and-III. According to the analysis, we detected the highest FAM signals from ExV events significantly more than ExI,-II, -III events (Figure 4h). Moreover, CtIP and MRE11 (ExIV) overexpression simultaneously increased these signals and promoted HDR not as big as ExV events (Figure 4h). We also observed that the only CtIP (ExII) or MRE11 (ExIII) were not able to improve HDR occurring significantly (Figure 4h).

The expression of exogenous *BleoR* integrated into the poplar genome was used as the authority of HDR efficiency. We performed real-time PCR to evaluate the percentage of delta-delta Ct mean (Supplementary 15a) and then analyzed the mean achieved to compare and illustrate the bar plot supported by standard distribution curves (Figure 4i; Supplementary 15b). Our analysis revealed the BleoR expression a mean of -1.2287 from ExI events. This mean was increased through EXII and EXIII events with 4.40787 and 6.11543. We then considered that mean within EXIV events and discovered one increase of 19.06057. Finally, despite all our previous observations, we found one significant development in BleoR expression integrated into the poplar genome throughout EXV events of 48.90032. Our analysis proved this increase of HDR efficiency significantly more than EXIV, EXIII, EXII, and EXI (Figure 4i).

The effect of efficient HDR on the expression of NHEJ and HDR factors

Regarding the new edition of the poplar genome in our study by integrating exogenous *BleoR* fused with *MKK2*, we decided to investigate these two genes' expression and their interdependence. The complete and exact expression of each of these genes indicates the efficiency of HDR. It could also show the conventional functioning of these genes in the new version of the poplar genome. So, we analyzed the achieved data of RT-qPCR from all events and used Violon plots to describe their distributed expressions (Figure 5) and Column plots to show all the gene expressions from each event separately (Supplementary 16). While our analysis revealed distributed expressions of *MKK2* and *BleoR* about +1 and -1 among ExI events, they were shown between about 100 and zero with medians about zero (Figure 5a and b). Within the ExII, we got three *MKK2* expressions (Events #21, #29, and #35) and only one *BleoR* expression (Event #29) (Supplementary 16). We then analyzed these gene expressions from ExIII and discovered promoted distributed expression of *MKK2* (Figure 5c). In this experiment, we found four *MKK2* expressions (Events #10, #23, #36, and #45) with

one *BleoR* expression (Event #6) (Supplementary 16). RT-qPCR results for ExV events revealed enhanced distributions of *BleoR* and *MKK2* expressions (Figure 5d). In this experiment, we achieved nine *MKK2* expressions (Events #9, #17, #39, #45, #54, #60, #68, #79, and #83) and four *BleoR* expressions (Events #17, #54, #68, and #90) (Supplementary 16). Regarding the expressions of these genes from ExV events, we observed significant promotions in *MKK2* and *BleoR* distributed expressions with a median of about 100 (Figure 5e). Also, the column bar analysis confirmed these distributions with twenty MKK2 expressions (Events #3, #9, #18, #21, #25, #29, #32, #33, #39, #59, #67, #73, #79, #82, #86, #88, #91, #92, #94, and #101) and twelve BleoR expressions (Events #21, #25, #29, #32, #37, #39, #53, #59, #73, #88, #91, and #94) (Supplementary 16).

Moreover, we decided to assess gene expressions involved in HDR and NHEJ pathways affected by our methods for increasing HDR efficiency. We used a Heat-map plot to interpret the obtained data from RT-qPCR (Figure 5f). While *CtIP* (~116), *MRE11* (~115), *BRCA1* (~114), *Rad50* (~113), and *Rad51* (~116) expressions were increased via ExI compared to WT, it was much more in *Lig4* (~146) and *XRCC4* (~143) (Figure 5f). The expression of *CtIP* was increased impressively (~166) via ExII compared with WT, while the expression of *MRE11* (~129), *BRCA1* (~119), *Rad50* (~120), and *Rad51* (~121) were increased shortly. Through ExII, the expression of *Lig4* (~104) and *XRC44* (~105) were decreased compared to ExI (Figure 5f). Within ExIII, the expression of *MRE11* was increased impressively (~162), but the expression of *CtIP* (~134), *BRCA1* (~120), *Rad50* (~122), and *Rad51* (~119) were increased a little. Within this experiment, the expression of *Lig4* (~107) and *XRC44* (~103) were decreased contrasted with ExI (Figure 4f). All HDR factors revealed enhanced expressions among ExIV as *CtIP* (~165), *MRE11* (~164), *BRCA1* (~130), *Rad50* (~128), and *Rad51* (~129), but *Lig4* (~101) and *XRC44* (~99) revealed more decreased compared with ExI (Figure 5f). *XRCC4* deficiency in ExV and enhancing *CtIP* and *MRE11* expressions caused amplifying the expression of *CtIP* (~170) and *MRE11* (~165) much more than WT events. Also, the expression of *Lig4* (~87) revealed more decreased than WT, and *XRCC4* was knocked out. Through ExV, the other HDR factors also revealed more expression as *BRCA1* (~145), *Rad50* (~139), and *Rad51* (~142) (Figure 5f).

XRCC4 deficient dramatically enhanced HDR efficiency and decreased polymorphisms

To test whether the promotion of HDR efficiency affects the happened polymorphisms, we analyzed the polymorphism types, variants genotypes, protein effects, and nucleotide genotyping of homology arms and also integrated *BleoR* into recovered events. We firstly analyzed the homology arms for polymorphisms happened by HDR improvement experiments. Seven polymorphism varieties including deletions, deletion tandem repeats, insertions, insertion tandem repeats, SNP transitions with A to C or G to T and reversely, SNP transversions with Purines to pyrimidines or reversely, and substitutions were detected through these editions (Supplementary Table 3). We then analyzed all detected variant nucleotides together and found out that HDR happening through ExI events induced the highest polymorphisms significantly more than ExIV and -V events (Figure 6a). We also observed more happened polymorphisms through ExII and -III than ExIV and -V. Furthermore, happened polymorphisms through ExIV events were observed more than ExV events (Figure 6a).

We then decided to investigate all polymorphisms in more detail on the homology arms and the *BleoR* integrated into the recovered poplar genome (Supplementary Table 4). We detected the highest frequency of deletion nucleotides through ExI events and the least within ExV events (Figure 6b). We also observed that *XRCC4* deficiency revealed no SNP happening and the least of SNP transition. The overexpression of *CtIP* decremented deletion tandem repeats (ExII), and the overexpression of *MRE11* decremented SNPs, SNP transitions, and SNP transversions (Figure 6b). The overexpression of *CtIP* and *MRE11* simultaneously (ExIV) decremented substitution polymorphisms (Figure 6b).

Moreover, the whisker plot of total polymorphisms presented the maximum distribution of polymorphisms through ExI events and the minimum of those in ExV events (Figure 6b).

Discussion

Despite extensive research on the use of HDR factors to increase HDR efficiency in plants, no research has been reported using this system in haploid species such as poplar. We hypothesized that if we increase the HDR pathway efficiency and decrease the NHEJ pathway efficiency simultaneously, we might overcome it. To achieve this purpose, we must first increase DDR amount in the cell nucleus correctly and at the right concentration. Therefore, we tried to achieve this goal by increasing Agrobacterium concentration and increasing the plasmid ratio containing DDT (pDDT) to plasmid containing gRNA (pgRNA) (Figure 2a). Based on the

preliminary results obtained from transformation and grown buds on media containing Zeocin, we hypothesized that the grown buds were all transformed. Therefore, we investigated the grown buds and transferred them to the rooting medium, including Zeocin and Cefotaxime, for recovery. Many of the grown buds could not be recovered and died.

The western blotting was applied to verify the integration of the 6xHis tag fused with exogenous BleoR. The observations proved the successful combination of BleoR through edited events. All events that exposed 54.2 KDa were supposed to be successfully edited, and all events that revealed 13.7 KDa were assumed might be unsuccessfully edited. For more details, we used RT-PCR to investigate the precise integration of *BleoR*, followed by *MKK2*. Regarding our designed primers, the bonds achieved from the first RT-PCR exhibited the successfully *MKK2* edited events (Figure 3e) and the bonds from the second RT-PCR exhibited the successfully *MKK2+BleoR* edited events (Figure 3f). The comparison of these RT-PCR assays uncovered the WT or unsuccessfully edited (Only edited *MKK2*) events.

For more confirmation, we used primers to probe *BleoR* applying southern blotting. We discovered which events were edited successfully by our designed DDT regarding obtained issues from western blotting, RT-PCR, and southern blotting. Because only edited events integrated by *BleoR* or *MKK2+BleoR* could be recovered, therefore we figured out why some of the grown buds on Zeocin were drowned and not recovered after transferring to rooting media. For more investigation, we applied TaqMan real-time PCR for detecting designed FAM1 and FAM2 signals (Figure 4b-g). Using this method, we compared the occurred HDR through all grown buds involved in designed experiments by calculating the gained total FAM signals (Figure 4h). The highest total FAM from ExV events proved that *XRCC4* deficiency caused HDR development significantly more than overexpression of HDR factors. Moreover, the investigation of HDR efficiency confirmed that the NHEJ pathway deficient is meaningfully more efficient to HDR development compared to focus on only overexpression of HDR factors (Figure 4i).

RT-qPCR and compared the *BleoR* and *MKK2* expressions through all grown bud transcriptomes helped us discover more deeply and accurately the role of decreasing activities of the NHEJ and increasing activities of the HDR pathways in poplars.

The significant difference in the distribution of *BleoR* and *MKK2* gene expressions within ExI to -IV events confirmed that the overexpression of factors involved in the HDR pathway

alone could not have sufficient and appropriate ability to create homologous recombination in the target genome. In contrast, the similar distribution of *BleoR* and *MKK2* expressions through ExV events proved that by reducing the NHEJ and improving the HDR pathway, we could achieve appropriate and desired homologous recombination (Figure 5e).

We resulted in further *Lig4* and *XRCC4* expressions than the factors involved in the HDR pathway (Figure 5f, ExI). Increase the CtIP expression within ExII and the MRE11 expression within ExIII events caused to decrease the activities of *XRCC4* and *Lig4*, but could not able to significantly improve the other HDR factor activities and also to improve the development of desired distributed expressions of *BleoR* and *MKK2* (Figure 5b, c, and f, 5xII and III). The simultaneously *CtIP* and *MRE11* expressions caused to amplify the other HDR factor activities and also to decrease the activities of *XRCC4* and *Lig4*, but could not improve the development of desired distributed expressions of *BleoR* and *MKK2* (Figure 5d, and f; ExIV). The *XRCC4* deficiency caused not only to amplify the *CtIP* and *MRE11* expressions meaningfully but also caused a much more increase of the other HDR factor activities and a much more decrease of *Lig4* expression (Figure 5f, ExV). Furthermore, the *XRCC4* deficiency developed homologous recombinant and then desired distributed expressions of *BleoR* and *MKK2* (Figure 5e).

Moreover, we decreased the polymorphisms through *XRCC4* deficiency significantly compared with ExI (Figure 6a). *XRCC4* deficiency also decreased in kinds of happening polymorphisms compared to the overexpression of *CtIP* and *MRE11* throughout the homologous recombinant development (Figure 6a and b). Altogether, NHEJ deficiency caused to improve the HDR efficiency in poplar meaningfully.

In summary, we have proved that *XRCC4* deficiency can promote HDR, therefore greatly expanding our capacity to improve hereditary developments in poplar. This breakthrough technology is likely to encourage biotechnological researches, breeding programs, and forest conservation of tree species.

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, the Talent Funding Project of Nanjing Forestry University (No. 163108059).

Author contribution

AM: Conceptualization, Software, Formal analysis, Writing - Original Draft, Visualization, Project administration, and Funding acquisition; HW Methodology, Formal analysis, Writing - Review & Editing and Data Curation; ZHC: Conceptualization, Validation, Data Curation, Writing - Review & Editing; WS, JZ, DL: Validation, Writing - Review & Editing; LY: Conceptualization, Software, Formal analysis, Visualization, and Funding acquisition; QZ: Conceptualization, Software, Formal analysis, Visualization, Supervision, and Funding acquisition.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgment

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

References

Butler NM, Baltes NJ, Voytas DF, Douches DS. 2016. Geminivirus-Mediated Genome Editing in Potato (Solanum tuberosum L.) Using Sequence-Specific Nucleases. *Front Plant Sci* **7**: 1045.

Cermak T, Baltes NJ, Cegan R, Zhang Y, Voytas DF. 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, Levy AA. 2018. Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. *Plant J* **95**: 5–16.

Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32**: 1262–1267.

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

Gil-Humanes J, Wang Y, Liang Z, Shan Q, Ozuna CV, Sanchez-Leon S, Baltes NJ, 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, Lieber MR. 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, Lieber MR. 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 DW, Orooji Y, Ruan HH, Zhuge Q. 2019. Strategies to Increase On-Target and Reduce Off-Target Effects of the CRISPR/Cas9 System in Plants. *International Journal of Molecular Sciences* **20**.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O *et al.* 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**: 827–832.

Hummel AW, Chauhan RD, Cermak T, Mutka AM, Vijayaraghavan A, Boyher A, Starker CG, Bart R, Voytas DF, Taylor NJ. 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, Yang H. 2019. Advances in detecting and reducing off-target effects generated by CRISPR-mediated genome editing. *J Genet Genomics*. doi: 10.1016/j.jgg.2019.11.002.

Li Z, Liu ZB, Xing A, Moon BP, Koellhoffer JP, Huang L, Ward RT, Clifton E, Falco SC, Cigan AM. 2015. Cas9-Guide RNA Directed Genome Editing in Soybean. *Plant Physiol* **169**: 960–970.

Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR, Ploegh HL. 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 JX, Amirian R, Zhuge Q. 2014. An Efficient Agrobacterium-Mediated Transformation System for Poplar. *International Journal of Molecular Sciences* **15**: 10780–10793.

Movahedi A, Zhang JX, Gao PH, Yang Y, Wang LK, Yin TM, Kadkhodaei S, Ebrahimi M, Qiang ZG. 2015. Expression of the chickpea CarNAC3 gene enhances salinity and drought tolerance in transgenic poplars. *Plant Cell Tissue and Organ Culture* **120**: 141–154.

Panier S, Boulton SJ. 2014. Double-strand break repair. 53BP1 comes into focus. *Nat Rev Mol Cell Biol* **15**: 7–18.

Pierce AJ, Hu P, Han M, Ellis N, 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.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning. a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Shi J, Gao H, Wang H, Lafitte HR, Archibald RL, Yang M, Hakimi SM, Mo H, Habben JE. 2017. ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol J* **15**: 207–216.

Song F, 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, 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 JK, Cigan AM. 2016. Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nature Communications* **7**.

Svitashev S, Young JK, Schwartz C, Gao H, Falco SC, Cigan AM. 2015. Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA. *Plant Physiol* **169**: 931–945.

Tran NT, Bashir S, Li X, Rossius J, Chu VT, Rajewsky K, Kuhn R. 2019. Enhancement of Precise Gene Editing by the Association of Cas9 With Homologous Recombination Factors. *Frontiers in Genetics* **10**.

Wang MG, Lu YM, Botella JR, Mao YF, Hua K, Zhu JK. 2017. Gene Targeting by Homology-Directed Repair in Rice Using a Geminivirus-Based CRISPR/Cas9 System. *Molecular Plant* **10**: 1007–1010.

Wu J, Yin H. 2019. Engineering guide RNA to reduce the off-target effects of CRISPR. *J Genet Genomics*. doi: 10.1016/j.jgg.2019.11.003.

Xie KB, Yang YN. 2013. RNA-Guided Genome Editing in Plants Using a CRISPRCas System. *Molecular Plant* **6**: 1975–1983.

Yang D, Scavuzzo MA, Chmielowiec J, Sharp R, Bajic A, 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 JP, Li XL, Li GH, Chen W, Arakaki C, Botimer GD, Baylink D, Zhang L, Wen W, Fu YW *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: Schematics of *MKK2* locus before and after edition and integration of exogenous BleoR protein into the poplar genome **(a)** Schematic of the purpose of this research to integrate exogenous BleoR into the poplar genome. Dash line reveals the target site. **(b)** Protospacer Adjacent Motif (PAM) was detected at the end of exon 8 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 exon 6 and -7, exon 7, intron sequences between exon 7 and -8, and a part of exon 8. The 3' homology arm included intron sequences between exon 8 and -9 and 3' UTR of the *MKK2* locus up to 400 bp. Designed DDT included remained sequences of exon 8, exon 9, BleoR CDS, 6xHis, and PolyA sequences flanked by the 3'- and 5' homology arms. We added two special targets besides DDT. The DDT was then ligated into the pRGEB31 vector to form pDDT.

Figure 2: The transformation strategy and designing of experiments. **(a)** pDDT and pgRNA were mixed 4:1 and introduced to the *Agrobacterium tumefaciens* to form inoculator suspension. We condensed inoculator up to OD_{600}=2.5 and then dipped all cut off stems. **(b)** The putatively edited events were regenerated on Zeocin. We allowed putative edited events to bud. The elongated buds were then transferred on rooting media and allowed to be recovered. **(c)** Designed experiments for this study including (I) No HDR factors, (II) overloaded CtIP, (III) overloaded MRE11, (IV) Overloaded CtIP+MRE11, and (V) Overloaded CtIP+MRE11 with *XRCC4* deficiency.

Figure 3: Western blotting to reveal the fused 6xHis tag with BleoR integration into the poplar genome. **(a)** Different experiments exhibited different quantities of 6xHis tag fusion. **(b)** Schematic of fusion 6xHis tag with edited poplar genome triggered by different experiments. The shape **a** reveals successful fusion of *BleoR* and *MKK2* with about 54 kDa. The shape **b** reveals an unsuccessful combination of mentioned proteins with about 14 kDa **(c)** Schematic of right HDR happening caused to attach exon 8 and 9 in the edited genome. **(d)** Schematic of proper integration in edited genome caused to connect the BleoR to the C-terminal of MKK2. **(e)** RT-PCR exhibited the HDR in exon 8 and 9, revealing a 920 bp of transcribed *MKK2* RNA in triggered events from ExII to ExV. The β-actin was used as the control in all RT-PCR assays; WT was used as a positive control. **(f)** RT-PCR revealed that BleoR CDS was adequately inserted in the target region with amplifying 413 bp of transcribed RNA in the recovered events. The β-actin was used as the control in all RT-PCR assays; BleoR protein extracted from pDDT plasmid was used as the positive control. WT was used as the negative control. **(g)** Schematic of probing BleoR in edited events and WT as the control using Southern blotting. **(h)** Southern blot proved that *BleoR* CDS was integrated into the precise recombinant genome. Digested pDDT plasmid was used as positive control.

Figure 4: The 2D kernel density plot of TaqMan real-time PCR fluorescent intensities and HDR efficiency percentage. **(a)** The TaqMan real-time PCR assay designing to detect HDR happened, and evaluation included FAM1 and FAM2 DNA binding probes. **(b)** Strategy to classify edited events. **(c)** Experiment I revealed no density for the edited events. **(d)** The density plot of FAM1 and -2 intensities resulted from experiment II revealed an expansion in edited events against partial, mutant, and wild-types. **(e)** The density plot of FAM1 and -2 ΔΔCt resulted from experiment III revealed an increased intensity of partial FAM1 events. **(f)** Experiment IV revealed a remarkable increase of edited events signals in confronting with three earlier experiments. **(g)** The Density plot of experiment V revealed a significant increase of FAM1 and -2 intensities in edited events compared to the earlier experiments and a significant decrease in intensities in WT and mutated events. All samples were analyzed in quadruplicate. **(h)** Diamond box and whisker plot revealed the identification of all FAM signals visualized in the experiments and showed more signals remarkably measured in ExV than ExI, II, and-III; Error bars represent SE; Asterisks represent p-value as *≤0.05, **≤0.01, and ***≤0.001. **(i)** The bar plot represents the HDR efficiency in different experiments; The overlap data are shown as bin

bars, and the standard distribution curves are added. HDR efficiency plot revealed that *XRCC4* deficient (ExV) let to HDR happening significantly more than the fusion of CtIP (ExII), MRE11 (ExIII), and CtIP+MRE11 (ExV). Also, ExIV meaningfully revealed more HDR happening than ExII and -III.; Error bars represent SE; Asterisks represent p-value as **≤0.01, ***≤0.001, and ****≤0.0001; Triplicate technical repeats were considered for each sample.

Figure 5: Violin plots reveal the *BleoR* and *MKK2* expression and the success happening HDR via different experiments. **(a-e)** The differences between *BleoR* and *MKK2* expression. Three technical repeats were used for each event in this assay; Dash lines present quartiles; Solid lines present median. **(f)** Heat-map to show the effect of efficient HDR on the expression of NHEJ and HDR factors. 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* deficiency decreased the expression of NHEJ factor *Lig4* and intensified HDR efficiency. Triplicate technical repeats were considered for each sample.

Figure 6: Polymorphisms happened in this study. **(a)** Identification of the polymorphisms happened in homology arms through the 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; Asterisks represent p-value as *≤0.05; Error bars represent SE. **(b)** Stacked column plot of total polymorphisms happened in DDT integration into the poplar genome. Deletion and insertion occurred much more than the other types. SNP and substitution occurred less than the other types. Whisker and standard distribution curves exposed that the total polymorphisms caused by *XRCC4* deficiency were less than the other experiments.

Supplementary data

Supplementary 1: Schematic of pgRNA, DDT, and pDDT. **(a)** pgRNA included the *MKK2* target seed and Cas9. **(b)** pDDT included DDT ligated into pRGEB31 by restriction enzyme cloning method.

Supplementary 2: Schematic construction of DDT and pDDT fragments, primers, and oligos.

Supplementary 3: Schematic construction of CtIP, MRE11, CtIP+MRE11, and XRCC4 cassette primers and oligos.

Supplementary 4: Schematic construction of cassettes (CtIP and MR) and vectors (pgCtIP and pgMR) and their primers.

Supplementary 5 : Schematics of constructed cassettes and plasmids. **(a)** pgCtIP plasmid including CtIP cassette. **(b)** pgMR plasmid including MR cassette. **(c)** pgCtMR plasmid including CtMR cassette. **(d)** pggCtMR plasmid including XRCC4 cassette.

Supplementary 6: Schematic construction of cassettes (CtMR, XRCC4) and vectors (pgCtMR and pggCtMR) and their primers.

Supplementary 7: Alignment of events involved in experiment I.

Supplementary 8: Alignment of events involved in experiment II.

Supplementary 9: Alignment of events involved in experiment III.

Supplementary 10: Alignment of events involved in experiment IV.

Supplementary 11: Alignment of events involved in experiment V.

Supplementary 12: Schematic of TaqMan real-time PCR FAM and VIC target assays in this study. Yellow rectangles exhibited CDS.

Supplementary 13: Box-and-whisker (Min-Max) plots of one-dimensional FAM delta-delta Ct signals in designed experiments. All signals were calculated as quadruplicates.

Supplementary 14: Schematics of sequence analyzing of triggered events from different experiments. **(a)** Sequence analysis of triggered events included in EXII reveals one recovered event. **(b)** Sequence analysis of triggered events included in EXIII reveals one recovered event. **(c)** Sequence analysis of triggered events included in EXIV reveals four recovered events. **(d)** Sequence analysis of triggered events included in EXV reveals 12 recovered events.

Supplementary 15: The raw data of real-time PCR evaluates the percentage of delta-delta Ct mean from *BleoR* in all experiments. **(a)** Delta-delta Ct mean of *BleoR* expression from grown buds. Each sample was investigated with three technical repeats. **(b)** Descriptive statistic table of raw data calculated by ANOVA-One way.

Supplementary 16: Column plots of the expression of integrated *BleoR* and new recombinant *MKK2* genes via different designed experiments. Three technical repeats were used for each event in this assay; Error bars represent SD; WT and pDDT were used as the control for *MKK2* and *BleoR* expression, respectively.

Supplementary 17: Chromatogram alignments of events included in experiment V.

Supplementary Table

Supplementary Table 1: CRISPR sites located on 3' region of *MKK2*. The yellow highlight reveals the selected CRISPR target in this study.

Supplementary Table 2: Oligos and primers used in this study.

Supplementary Table 3: All polymorphisms detected in homology arms happened by HDR through experiments.

Supplementary Table 4: Variant nucleotides happened from experiments.

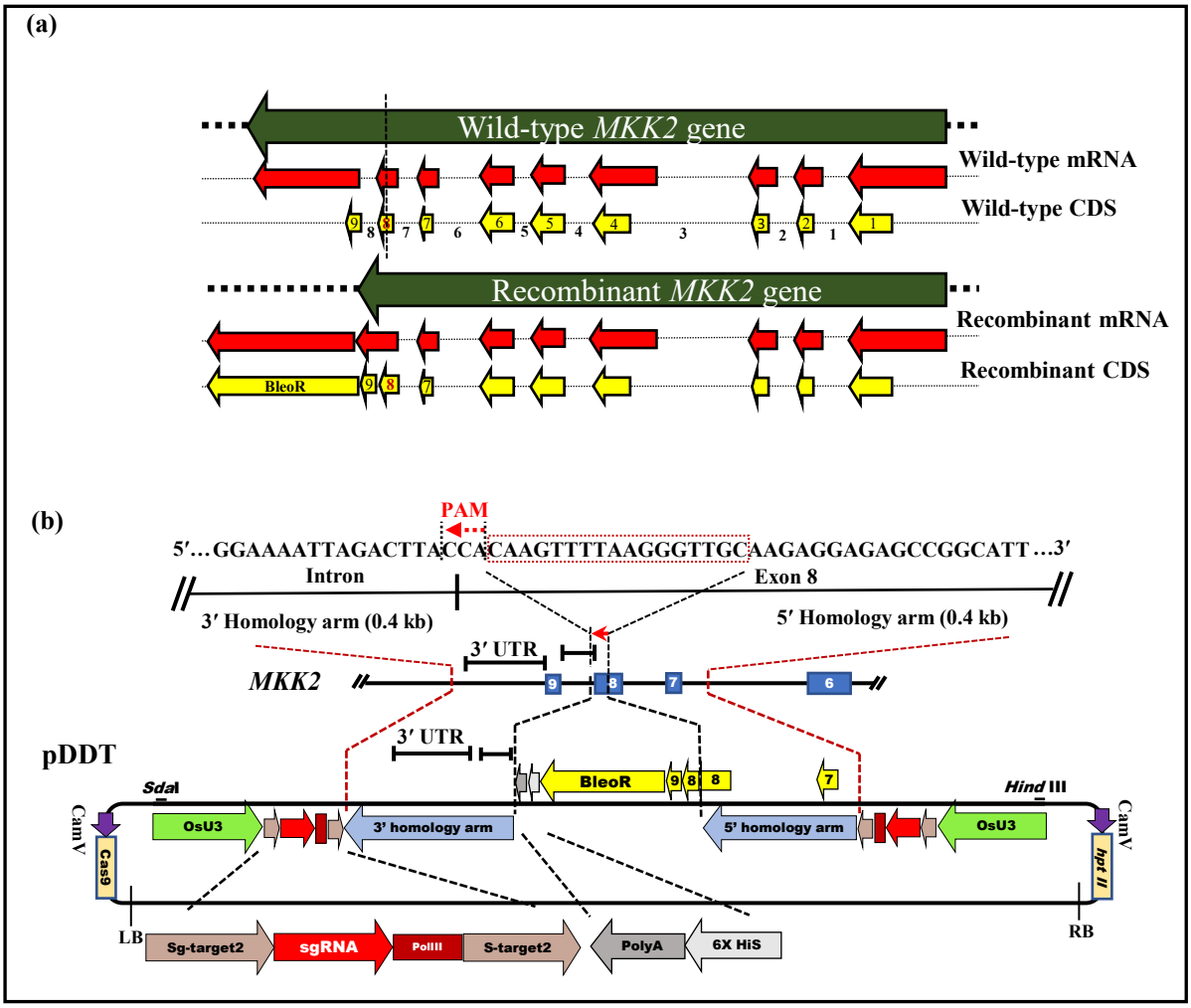
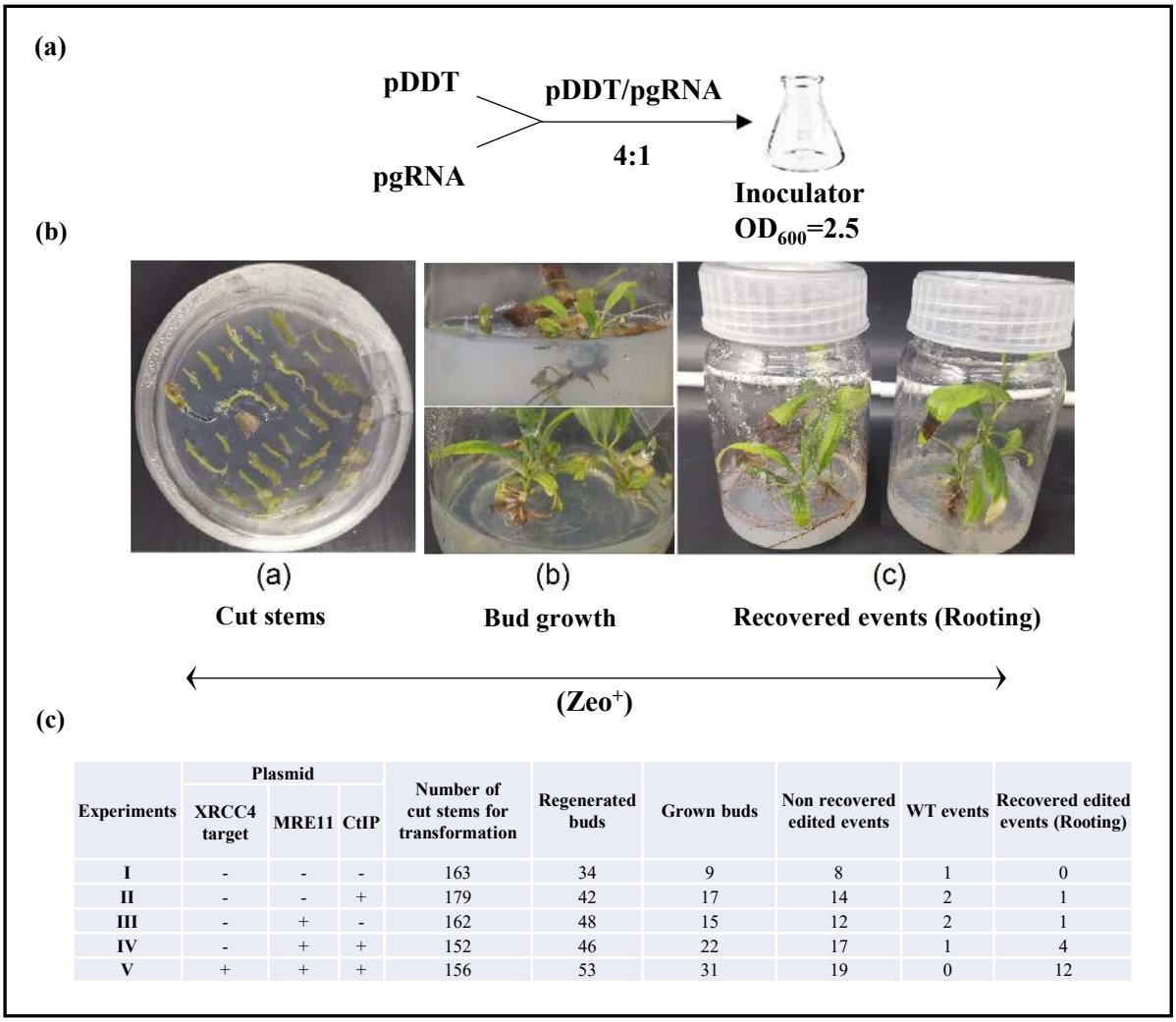


Figure 1: Schematics of *MKK2* locus before and after edition and integration of exogenous BleoR protein into the poplar genome **(a)** Schematic of the purpose of this research to integrate exogenous BleoR into the poplar genome. Dash line reveals the target site. **(b)** Protospacer Adjacent Motif (PAM) was detected at the end of exon 8 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 exon 6 and -7, exon 7, intron sequences between exon 7 and -8, and a part of exon 8. The 3' homology arm included intron sequences between exon 8 and -9 and 3' UTR of the *MKK2* locus up to 400 bp. Designed DDT included remained sequences of exon 8, exon 9, BleoR CDS, 6xHis, and PolyA sequences flanked by the 3'- and 5' homology arms. We added two special targets besides DDT. The DDT was then ligated into the pRGEB31 vector to form pDDT.



Experiments	Plasmid			Number of cut stems for transformation	Regenerated buds	Grown buds	Non recovered edited events	WT events	Recovered edited events (Rooting)
	XRCC4 target	MRE11	CtIP						
I	-	-	-	163	34	9	8	1	0
II	-	-	+	179	42	17	14	2	1
III	-	+	-	162	48	15	12	2	1
IV	-	+	+	152	46	22	17	1	4
V	+	+	+	156	53	31	19	0	12

Figure 2: The transformation strategy and designing of experiments. **(a)** pDDT and pgRNA were mixed 4:1 and introduced to the *Agrobacterium tumefaciens* to form inoculator suspension. We condensed inoculator up to OD_{600}=2.5 and then dipped all cut off stems. **(b)** The putatively edited events were regenerated on Zeocin. We allowed putative edited events to bud. The elongated buds were then transferred on rooting media and allowed to be recovered. **(c)** Designed experiments for this study including (I) No HDR factors, (II) overloaded CtIP, (III) overloaded MRE11, (IV) Overloaded CtIP+MRE11, and (V) Overloaded CtIP+MRE11 with *XRCC4* deficiency.

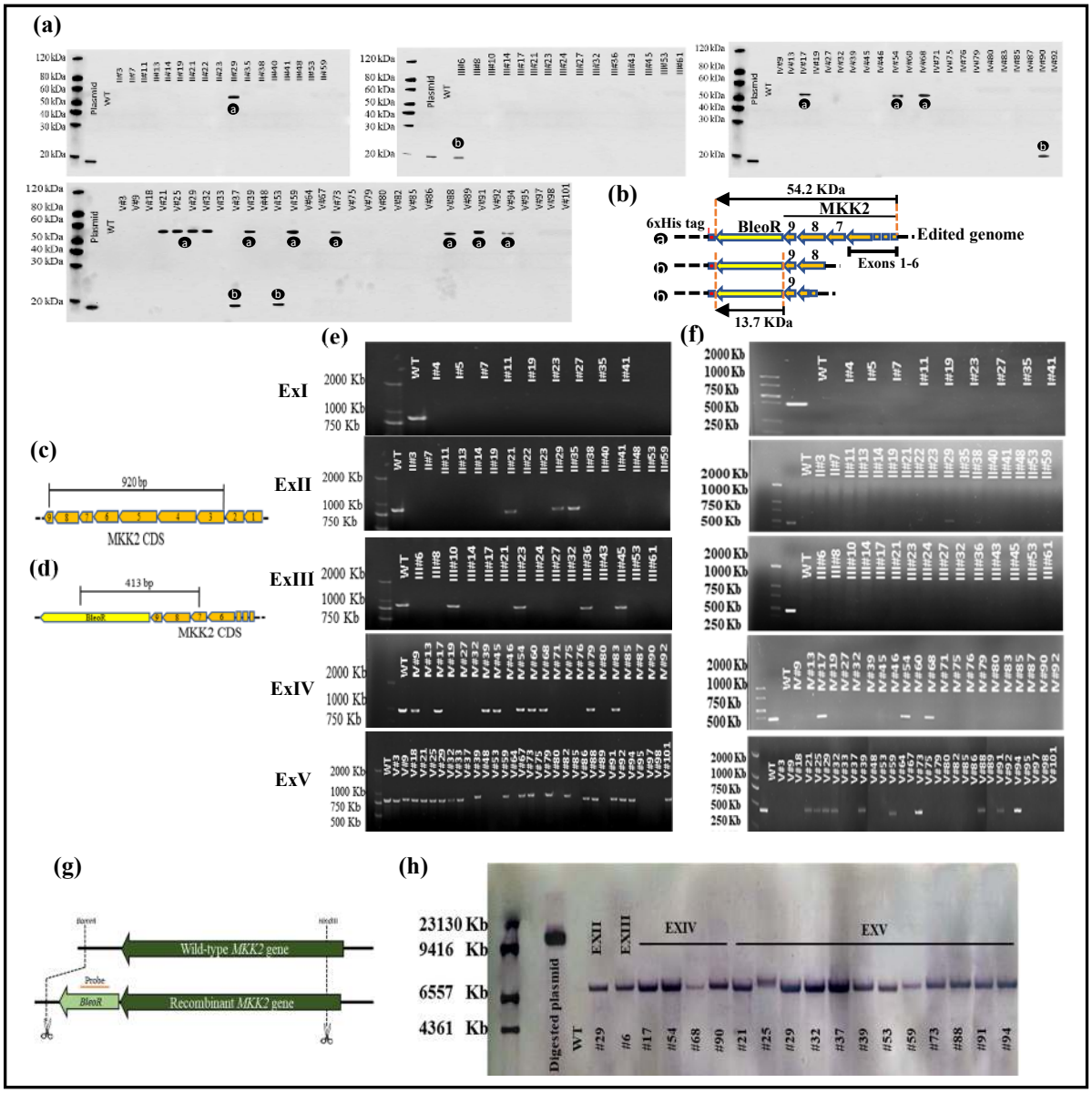


Figure 3: Western blotting to reveal the fused 6xHis tag with BleoR integration into the poplar genome. **(a)** Different experiments exhibited different quantities of 6xHis tag fusion. **(b)** Schematic of fusion 6xHis tag with edited poplar genome triggered by different experiments. The shape **a** reveals successful fusion of *BleoR* and *MKK2* with about 54 kDa. The shape **b** reveals an unsuccessful combination of mentioned proteins with about 14 kDa **(c)** Schematic of right HDR happening caused to attach exon 8 and 9 in the edited genome. **(d)** Schematic of proper integration in edited genome caused to connect the BleoR to the C-terminal of MKK2. **(e)** RT-PCR exhibited the HDR in exon 8 and 9, revealing a 920 bp of transcribed *MKK2* RNA in triggered events from ExII to ExV. The β-actin was used as the control in all RT-PCR assays; WT was used as a positive control. **(f)** RT-PCR revealed that BleoR CDS was adequately inserted in the target region with amplifying 413 bp of transcribed RNA in the recovered events. The *β-actin* was used as the control in all RT-PCR assays; BleoR protein extracted from pDDT plasmid was used as the positive control. WT was used as the negative control. **(g)** Schematic of probing BleoR in edited events and WT as the control using Southern blotting. **(h)** Southern blot proved that *BleoR* CDS was integrated into the precise recombinant genome. Digested pDDT plasmid was used as positive control.

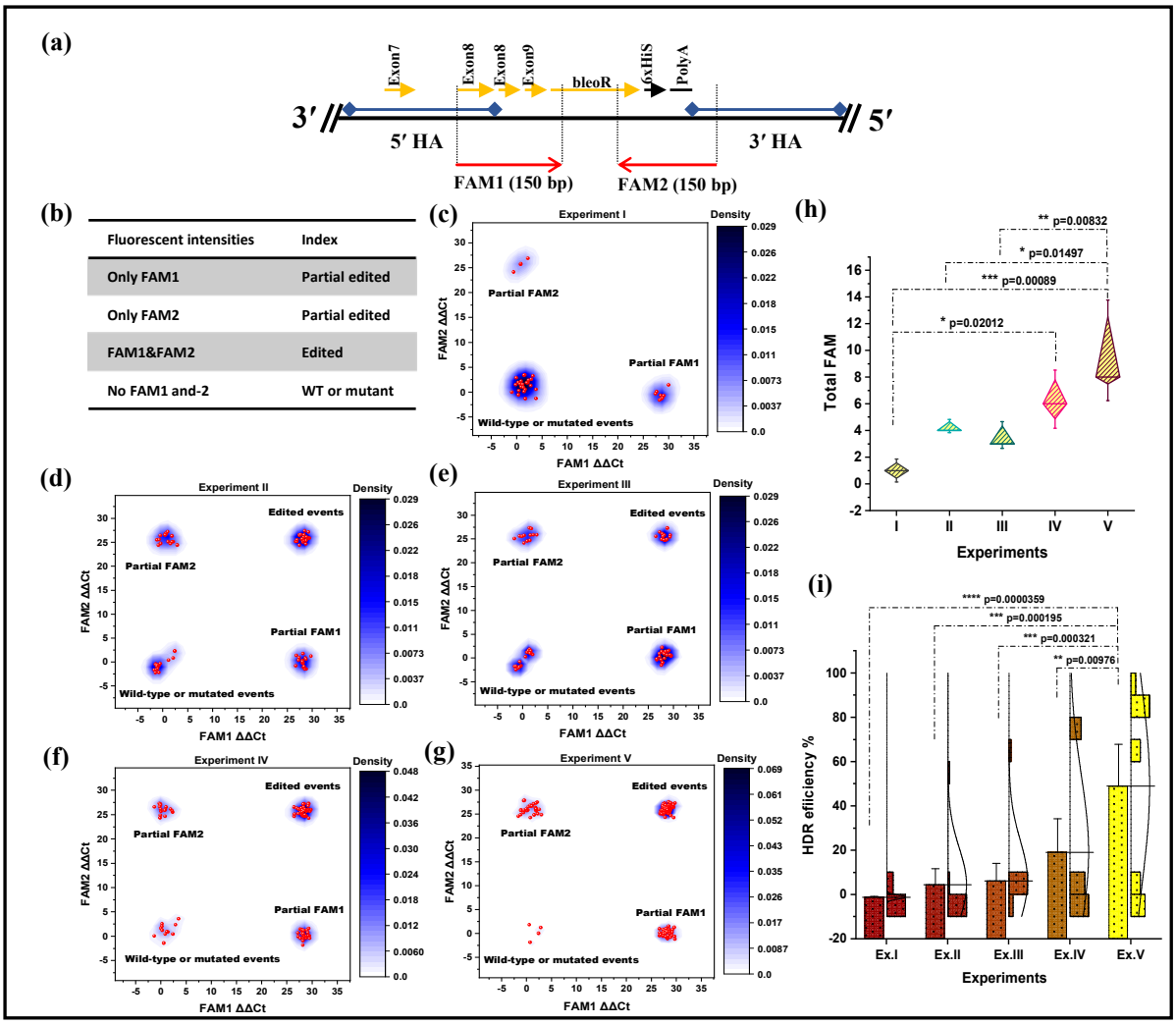


Figure 4: The 2D kernel density plot of TaqMan real-time PCR fluorescent intensities and HDR efficiency percentage. **(a)** The TaqMan real-time PCR assay designing to detect HDR happened, and evaluation included FAM1 and FAM2 DNA binding probes. **(b)** Strategy to classify edited events. **(c)** Experiment I revealed no density for the edited events. **(d)** The density plot of FAM1 and -2 intensities resulted from experiment II revealed an expansion in edited events against partial, mutant, and wild-types. **(e)** The density plot of FAM1 and -2 ΔΔCt resulted from experiment III revealed an increased intensity of partial FAM1 events. **(f)** Experiment IV revealed a remarkable increase of edited events signals in confronting with three earlier experiments. **(g)** The Density plot of experiment V revealed a significant increase of FAM1 and -2 intensities in edited events compared to the earlier experiments and a significant decrease in intensities in WT and mutated events. All samples were analyzed in quadruplicate. **(h)** Diamond box and whisker plot revealed the identification of all FAM signals visualized in the experiments and showed more signals remarkably measured in ExV than ExI, II, and-III; Error bars represent SE; Asterisks represent p-value as *≤0.05, **≤0.01, and ***≤0.001. **(i)** The bar plot represents the HDR efficiency in different experiments; The overlap data are shown as bin bars, and the standard distribution curves are added. HDR efficiency plot revealed that *XRCC4* deficient (ExV) let to HDR happening significantly more than the fusion of CtIP (ExII), MRE11 (ExIII), and CtIP+MRE11 (ExV). Also, ExIV meaningfully revealed more HDR happening than ExII and -III.; Error bars represent SE; Asterisks represent p-value as **≤0.01, ***≤0.001, and ****≤0.0001; Triplicate technical repeats were considered for each sample.

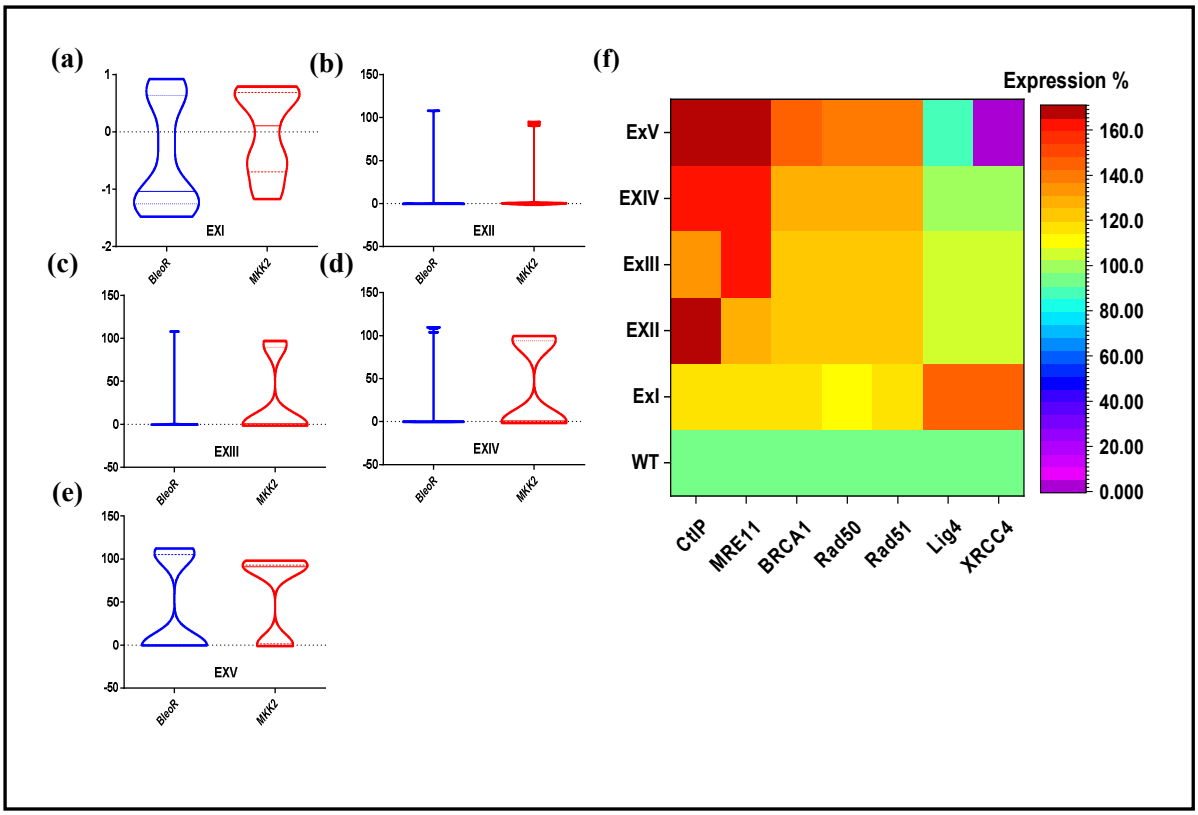


Figure 5: Violin plots reveal the *BleoR* and *MKK2* expression and the success happening HDR via different experiments. **(a-e)** The differences between *BleoR* and *MKK2* expression. Three technical repeats were used for each event in this assay; Dash lines present quartiles; Solid lines present median. **(f)** Heat-map to show the effect of efficient HDR on the expression of NHEJ and HDR factors. 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* deficiency decreased the expression of NHEJ factor *Lig4* and intensified HDR efficiency. Triplicate technical repeats were considered for each sample.

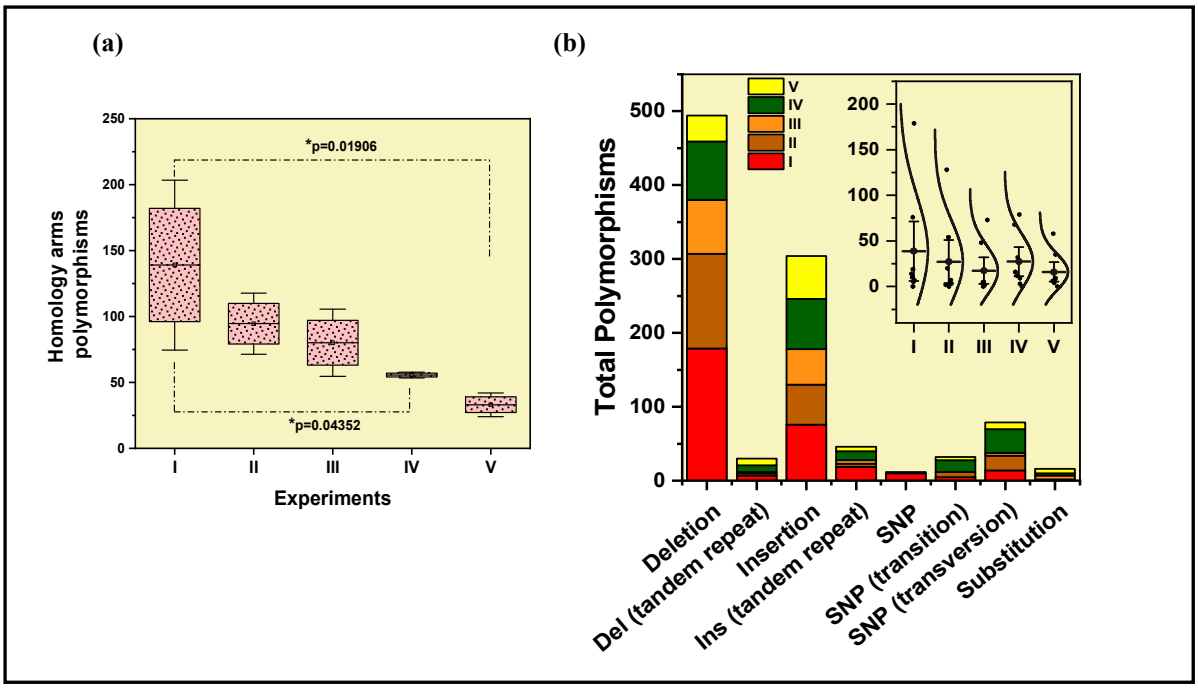


Figure 6: Polymorphisms happened in this study. **(a)** Identification of the polymorphisms happened in homology arms through the experiments. Box and Whisker plot revealed that most polymorphisms happened in homology arms by ExI, and it was significantly more than those in ExV and –IV; Asterisks represent p-value as *≤0.05; Error bars represent SE. **(b)** Stacked column plot of total polymorphisms happened in DDT integration into the poplar genome. Deletion and insertion occurred much more than the other types. SNP and substitution occurred less than the other types. Whisker and standard distribution curves exposed that the total polymorphisms caused by *XRCC4* deficiency were less than the other experiments.