Original Manuscript

Efficient CRISPR/Cas9-based genome editing and its application to conditional genetic analysis in *Marchantia polymorpha*

Shigeo S. Sugano[1, 2], Ryuichi Nishihama[3], Makoto Shirakawa[3], Junpei Takagi[4], Yoriko Matsuda[3], Sakiko Ishida[3], Tomoo Shimada[4], Ikuko Hara-Nishimura[4], Keishi Osakabe[5], Takayuki Kohchi[3,*]

1 R-GIRO, Ritsumeikan University, Noji-Higashi 1-1-1, Kusatsu, Shiga, Japan

2 JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, Japan

3 Graduate School of Biostudies, Kyoto University, Kitashirakawa-Oiwake-cho, Kyoto, Japan

4 Graduate School of Science, Kyoto University, Kitashirakawa-Oiwake-cho, Kyoto, Japan

5 Faculty of Bioscience and Bioindustry, Tokushima University, 2-1 Josanjima, Tokushima, Tokushima, Japan

* tkohchi@lif.kyoto-u.ac.jp

Abstract

Marchantia polymorpha is one of the model species of basal land plants. Although CRISPR/Cas9-based genome editing has already been demonstrated for this plant, the efficiency was too low to apply to functional analysis. In this study, we show the establishment of CRISPR/Cas9 genome editing vectors with high efficiency for both construction and genome editing. Codon optimization of Cas9 to Arabidopsis achieved over 70% genome editing efficiency at two loci tested. Systematic assessment revealed that guide sequences of 17 nt or shorter dramatically decreased this efficiency. We also demonstrated that a combinatorial use of this system and a floxed complementation construct enabled conditional analysis of a nearly essential gene. This study reports that simple, rapid, and efficient genome editing is feasible with the series of developed vectors.

Abbreviations: ARF1, *AUXIN RESPONSE FACTOR1*; Cas9, CRISPR-associated endonuclease 9; CRISPR, clustered regularly interspaced short palindromic repeats; DSB, double-strand break; *EF*, *ELONGATION FACTOR1α*; *HPT*, *hygromycin phosphotransferase*; gRNA, single guide RNA; NAA, 1-naphthalene acetic acid; NHEJ, non-homologous end joining; NLS, nuclear localization signal; *NOP1*, *NOPPERABO1*; *mALS*, *mutated acetolactate synthase*; MMEJ, microhomology-mediated end joining; PAM, protospacer adjacent motif; PCR, polymerase chain reaction; RT-PCR, reverse transcription polymerase chain reaction.

Introduction

The clustered regularly interspaced short palindromic repeats (CRISPR)/ CRISPR-associated endonuclease 9 (Cas9)-based genome editing system is a groundbreaking technology in molecular genetics, which enables alterations of target sequences in the genome [1, 2]. The system consists of two components: a Cas9 protein, which has an RNA-guided endonuclease activity, and a single guide RNA (gRNA), which specifies a target sequence within the genome. The Cas9 protein from *Streptococcus pyogenes* binds to the DNA sequence "NGG," which is known as the protospacer adjacent motif (PAM) sequence. The interaction between a Cas9 protein and a PAM sequence induces the interaction between a gRNA and its target DNA sequence. If a sufficient length of the gRNA matches with the target sequence, the nuclease domains of Cas9 become capable of cutting the phosphodiester bonds on both sides of the strands, which are located 3 bp upstream of the PAM sequence [3]. Once a double-strand break (DSB) occurs, the error-prone non-homologous end joining (NHEJ) repair pathway is activated and sometimes introduces indels or base substitutions randomly at the target site, which could result in disruption of the target locus with various alleles.

This simple CRISPR/Cas9 system has been reconstructed in a wide range of eukaryotes, and researchers are now able to use molecular genetics in the species that are suitable for the purposes of their specific biological research [4]. Previous studies demonstrated that the CRISPR/Cas9 system works in a variety of plant species, from algae to crops [5-7], which has greatly changed functional genetics in basic and applied plant research. Recent studies on the moss *Physcomitrella patens* [8, 9] and the green alga *Chlamydomonas reinhardtii* [10, 11] reported highly efficient genome editing methods using CRISPR-based genome editing. Compared to flowering plants,

such haploid generation-dominant plant species are free from the transheterozygosity issues associated with diploidy or polyploidy [12-14], allowing isolation of pure mutant lines for analysis with relative ease. In the meanwhile, regardless of the ploidy, but especially for haploid species, genome editing techniques cannot be simply applied to essential genes as this leads to lethality; conditional approaches are required.

The liverwort *Marchantia polymorpha* is an emerging model species of land plants for studying plant evolution and gene function [15]. *M. polymorpha* has good features for the application of reverse genetics. Most vascular plants and mosses are known to have experienced two or more whole genome duplication events, which makes it difficult to analyze gene functions due to the presence of paralogous genes. Sequencing of the *M. polymorpha* genome revealed no sign of a whole genome duplication and accordingly there is low genetic redundancy in most regulatory genes, such as transcription factors and signaling components [16]. In addition, non-chimeric individuals can be easily obtained and propagated via gemmae that are derived from single cells by asexual reproduction in *M. polymorpha* [17], which accelerates transgenic experiments [18]. A variety of tools for molecular genetic experiments have been developed for *M. polymorpha* [18], such as high-efficiency transformation methods [19-21], a homologous recombination-mediated gene targeting method [22], a systematic set of vectors [18], and a conditional gene expression/deletion system [23].

Recently, a transcription activator-like effector nuclease (TALEN)-based genome editing technology was established in *M. polymorpha* [24]. We have previously demonstrated that a CRISPR/Cas9-based knockout system, which exploited human-codon-optimized Cas9, can operate in *M. polymorpha* [25].

However, the efficiency of genome editing was so low that the identification of plants with a mutation at the target locus required selection by a phenotype attributed to the mutation. Systematic optimization of Cas9 and gRNAs with rice, tobacco, and Arabidopsis showed that their expression levels greatly affected genome editing efficiencies [26, 27], suggesting that there is room to improve genome editing efficiencies in *M. polymorpha*.

Here, we report remarkable improvement in genome editing efficiency using Arabidopsis-codon-optimized (Atco) Cas9 to a degree that simple direct sequencing analysis of few number of transformants is sufficient to obtain genome-edited plants. In this efficient genome editing system, we assessed off-target effects and evaluated the influence of gRNAs length. Occurrence of large deletions using two gRNAs was also demonstrated. In addition, we provide a simple CRISPR/Cas9-based method to generate conditional knockout mutants. Our improved CRISPR/Cas9-based genome editing system can be used as a powerful molecular genetic tool in *M. polymorpha*.

Materials and Methods

Accessions, growth conditions and transformation of *M. polymorpha*

M. polymorpha Takaragaike-1 (Tak-1, male accession) and Takaragaike-2 (Tak-2, female accession) were used as wild types [19]. F1 spores were generated as previously described [25]. *M. polymorpha* was cultured axenically under 50–60 μmol m^{-2} sec^{-1} continuous white light at 22 °C. *Agrobacterium*-mediated transformation of F1 sporelings was performed as described previously [19]. Transformants were selected on half-strength B5 medium [28] containing 1% agar with 0.5 μM chlorsulfuron (kindly provided by DuPont; in case of the assay in Fig. 6, Wako Pure

Chemical Industries) or 10 mg L^{-1} hygromycin (Wako Pure Chemical Industries) depending on the transformation vector.

Vector construction

All the DNA sequences of the vectors were deposited to DDBJ and Addgene: pMpGE_En01(LC090754; 71534), pMpGE_En03 (LC090755; 71535), pMpGE010 (LC090756; 71536), pMpGE011 (LC090757; 71537), pMpGE006 (LC375817; 108722), pMpGE013 (LC375815; 108681), pMpGE014 (LC375816; 108682), pMpGWB337tdTN (LC375949; 108717), pMpGWB337Cit (LC375950; 108718), pMpGWB337tdT (LC375951; 108719), pMpGWB337TR (LC375952; 108720), pMpGWB337mT (LC375953; 108721). The construction of these vectors was performed as follows:

-pMpGE010, pMpGE011. Firstly, a DNA fragment of nuclear localization signal (NLS)-tagged *Atco-Cas9* with the *Pisum sativum rbcS3A* terminator (Pea3ter) was PCR amplified with the primers cacc_AtCas9_F and Pea3Ter_R using the pDe-CAS9 vector [12] as a template and subcloned into pENTR/D-TOPO (Invitrogen). Using the entry clone and LR Clonase II (Invitrogen), LR reactions with pMpGWB103 [29] and pMpGWB303 [29] were conducted to express Cas9 under the *M. polymorpha* *ELONGATION FACTOR1*α promoter (MpEF_{pro}) [18]. The LR reaction product using pMpGWB103 was designated pMpGE006 and used for the two vector system experiments (Fig. 1A). Next, a Gateway attR1-attR2 cassette amplified from pMpGWB303 using the primers Infusion_GW_A51_F and Infusion_GW_A51_R was subcloned into the AorHI51 restriction enzyme site of the vectors produced by the LR reaction above to produce pMpGE010 and pMpGE011.

-pMpGE_En01. The 2 kbp promoter region of Mp*U6-1* was amplified from pENTR D-TOPO/MpU6-1pro:gRNA_ARF1 [25] using the primers Mp-U6_38003_F and Mp-U6_38003_R. The CmR-ccdB-gRNA fragment with SacI and PstI sites was also amplified from pENTR D-TOPO/AtU6pro:CmR-ccdB-gRNA (unpublished data) using the primers OE-MpU6-CmRccdB-F2 and gRNA-R3. These two amplified fragments were combined using overlap extension PCR and the combined fragment was cloned into pENTR/D-TOPO vector to produce pMpGE_En01.

-pMpGE_En02. PCR was conducted using pMpGE_En01 as a template and with the primers BsaI-Sp-sgRNA_F and gRNA_R to amplify a gRNA backbone fragment. Additionally, a 500 bp region of the Mp*U6-1* promoter was PCR-amplified using the MpU6-1_500_F and BsaI_MpU6_1R primers and plasmid pMpGE_En01 as a template. Amplified gRNA and Mp*U6-1* promoter fragments were conjugated by overlap extension PCR and cloned into pENTR/D-TOPO to produce pMpGE_En02.

-MpGE_En03. pMpGE_En02 was digested by BsaI. Two oligo DNAs, Mp_oligo6BsaI_Gf and Mp_oligo6BsaI_Gr, were annealed and cloned into BsaI-digested pMpGE_En02 to produce the pMpGE_En03 plasmid (see Fig. S1).

-pMpGE013 and pMpGE014 vectors. Briefly, annealed oligos harboring AarI recognition sites (Mp_oligo5AarI_Gf and Mp_oligo5AarI_Gr) were subjected to ligation reactions with the pMpGE_En02 vector linearized with BsaI. The resulting vector was subjected to LR reaction with pMpGE010 and pMpGE011 to produce pMpGE013 and pMpGE014, respectively.

-pMpGE vectors harboring *ARF1_1*, *NOP1_1*, and *NOP1_2* gRNA. pMpGWB301, harboring ARF1_1 gRNA, was the same construct as in the previous report [25]. Construction of other vectors was by the methods described in Fig. S1. Briefly, annealed oligos of ARF1_1 and NOP1_1 for pMpGE_En01 were subjected to the In-

Fusion™ HD Cloning Kit (TaKaRa) to the pMpGE_En01 vector linearized with PstI and SacI (TaKaRa). Conversely, annealed oligos of NOP1_2 gRNAs were subjected to ligation reactions with pMpGE_En03 linearized with BsaI (NEB). These constructs were subjected to the LR reaction to be introduced into pMpGE010 or pMpGE011. Oligos for gRNAs are listed in Supplemental Table S1.

-pMpGWB337 series. The plasmid pMp301-EFp:loxGW:Tlox:CitNLS:T[18] was digested with SalI and SacI and ligated with a SalI-SacI fragment containing an NruI site, which was generated by PCR using pMp301-EFp:loxGW:Tlox:CitNLS:T as template with the primer set ccdB_236F and loxP_NruI_Sac_R and subsequent enzyme digestion, to generate pMp301-EFp:loxGW:Tlox-NruI. Coding sequences of various fluorescent proteins were PCR amplified using a common reverse primer (NOSt_head_R_SacI) and the following forward primers and templates: TagRFP (TagRFP_CAGC_F; pMpGWB126 [29]), tdTomato (tdTomato_CAGC_F; pMpGWB129 [29]), tdTomato-NLS (tdTomato_CAGC_F; pMpGWB116 [29]), Atco-mTurquoise2 (mTurq_CAGC_F, pUGW2-mTurq2 [see below]), and Citrine (ccdB_236F, pMpGWB337 [23]). These amplified fragments (except for Citrine) were digested with SacI and ligated with NruI/SacI-digested pMp301-EFp:loxGW:Tlox-NruI to generate a pMp301-EFp:loxGW:Tlox:FP:T series (FP: TR for TagRFP; tdT for tdTomato; tdTN for tdTomato-NLS; mT for mTurquoise2). For Citrine, the amplified fragment was digested with SalI and SacI and ligated with SalI/SacI-digested pMp301-EFp:loxGW:Tlox:CitNLS:T to generate pMp301-EFp:loxGW:Tlox:Cit:T. A fragment consisting of the Mp*HSP17.8A1* promoter, the Cre-GR coding sequence, and the NOS terminator was amplified and cloned into the AscI site of the pMp301-EFp:loxGW:Tlox:FP:T series, as described previously [29], to generate a series of pMpGWB337-FP vectors (Fig. S7). For construction of

pUGW2-mTurq2, an Atco-mTurquoise2 DNA fragment, which was synthesized and amplified by PCR with the primer set pUGW_Aor_mTurq_IF_F and pUGW_Aor_mTurq_Stp_IF_R, was cloned into the unique Aor51HI site of pUGW2 [30] using the In-Fusion Cloning Kit. The primers used are listed in Supplementary Table S1.

Mutation analyses in on-target and off-target sites

Transformed sporelings were selected with antibiotics for 18 days, and selected lines, referred to as T1 plants, were transferred to fresh medium with the same antibiotics for approximately 2 weeks. The genomic DNAs of T1 thalli were extracted. The Mp*ARF1* target locus was PCR amplified using the primers ARF1_Seq_F3 and ARF1_Seq_R3 and subjected to direct sequencing. The Mp*NOP1* target locus was PCR amplified using the primers CRISPR_NOP1_F and CRISPR_NOP1_R then subjected to direct sequencing. Using genomic DNAs harboring mutations in the on-target sites of ARF1_1 or NOP1_1, corresponding off-target sites were PCR amplified using the primers listed in Supplementary Table S1, and subjected to direct sequencing. Off-target sites of ARF1_1 and NOP1_1 were searched using the *M. polymorpha* genome ver3.1 [16] and CasOT software [31].

Assessment of gRNA guide lengths

pMpGE010_NOP1_2 variants harboring various lengths of gRNAs were constructed using pMpGE_En02. For the addition of an “extra initial G,” the pMpGE_En03 vector was used (oligos are listed in Supplementary Table S1). T1 plants in petri dishes were placed onto a white-light-emitting display (iPad, Apple) and photos of the whole body of T1 plants were taken by a digital camera (EOS KissX3). Digital

images of T1 plants were analysed by ImageJ. Using Threshold Colour program, ratio of transparent area to whole plant were measured. The images were classified into three types; Class I: over 90% of the thallus area was transparent; Class II: less than 90% and over 20% transparent area; and Class III: less than 20% transparent.

Analysis of *de novo* mutations

Three pMpGE010_NOP1_1 transformants that had transparent (mutant) and non-transparent (wild-type) sectors in one individual were cultured. Four gemmae from a gemma cup formed on each of the two sectors were transplanted to new media and cultured to conduct genome analysis.

Large deletion induction by co-transformation

pMpGE013 and pMpGE014 were digested by the AarI restriction enzyme. Annealed oligos for NOP1_3 to NOP1_6 (their sequences are described in Supplemental Table S1) were ligated into the linearized vectors. Vectors were introduced into regenerating thalli via *Agrobacterium* [20]. T1 transformants selected by appropriate antibiotics were cultured at for least 2 weeks and subjected to DNA extraction. The extracted DNAs from T1 thalli were analyzed by PCR using the appropriate primers described in Supplementary Table S1. The PCR products were analyzed by direct sequencing to examine genome-editing events.

Generation of conditional knockout mutants.

Annealed oligos for an Mp*MPK1*-targeting gRNA (sgRNA_Bsa_MpMPK1_ex1t_F and -R) were ligated into BsaI-digested pMpGE_En02. The resulting plasmid was subjected to LR reaction with pMpGE010 to generate pMpGE010-MPK1ex1t.

Mp*MPK1* cDNA was amplified by RT-PCR using RNA from Tak-1 and the primer set MpMPK1_1F_TOPO and MpMPK1_c1131R_STP and cloned into pENTR/D-TOPO. The resulting entry vector was subjected to LR reaction with pMpGWB337 and pMpGWB337tdTN to generate pMpGWB337-cMPK1 and pMpGWB337tdTN-cMPK1, respectively [19]. Sporelings were transformed with pMpGE010-MPK1ex1t alone or together with pMpGWB337-cMPK1 or pMpGWB337tdTN-cMPK1 using the *Agrobacterium*-based method. T1 transformants selected by hygromycin only or by hygromycin and chlorsulfuron together, respectively, were genotyped by PCR with the primer set MpMPK1_-549F and MpMPK1_g540R, and the amplified fragments were directly sequenced. G1 gemmae of the pMpGWB337tdTN-cMPK1-harboring genome-edited lines were further confirmed to have the same mutations as those in the respective parental T1 lines. G2 gemmae of line #4 were planted on a 9-cm plastic plate containing half-strength B5 medium with 1 μM dexamethasone (DEX) and further treated with a drop of 1 μM DEX. The plate was incubated in a 37°C air incubator for 80 min and then moved to a 22°C growth room. In 24 h, the same procedures were repeated, and the plate was incubated in the 22°C growth room for 13 days. Observation of tdTomato fluorescence was performed using a stereoscope (M205C, Leica) with the Ds-Red2 filter set (Leica). Induction of cDNA deletion was examined by genomic PCR using the primer sets 1 (MpEF-P_seqL1 and MpMPK1_c1131R_STP), 2 (MpEF-P_seqL1 and tdTomato_753R), and 3 (MpMPK1_-549F and MpMPK1_g1010R). The primers used are listed in Supplementary Table S1.

Results

Improved genome editing efficiency in *M. polymorpha* by using Arabidopsis codon optimized Cas9

In the previous report, we co-introduced a construct for gRNA expression with a 2 kbp Mp*U6-1* promoter and another for the expression of human-codon-optimized Cas9-NLS (*hCas9-NLS*) to reconstruct a CRISPR/Cas9 system in *M. polymorpha*, and succeeded in obtaining plants whose target locus was edited [25]. However, the genome editing efficiency was very low, under 0.5%; that is, only a few plants with genome editing events were obtained for each co-transformation of sporelings derived from two sporangia. To improve the genome editing efficiency to a practical level, we examined the effects of codon optimization by replacing *hCas9-NLS* with *Atco-Cas9-NLS* (although the *35S* terminator was also replaced by the *Pisum sativum rbcS3A* terminator [12], this difference seems negligible (unpublished data)). In this new Cas9 expression plasmid, designated pMpGE006, *Atco-Cas9* was driven by the constitutive MpEF_{pro} promoter, which is preferentially expressed in meristematic tissues in *M. polymorpha* (Fig. 1A) [32]. For the expression of a gRNA, the same gRNA expression vector used in the previous study was again utilized [25].

To evaluate genome-editing efficiencies, we firstly chose the Mp*ARF1* (*AUXIN RESPONSE FACTOR1*) transcription factor as a target, which had been used in the previous report [25 , 33]. Since Mp*arf1* mutants are known to show an NAA resistant phenotype [34, 35], auxin-resistant T1 transformants with mutations in the Mp*ARF1* locus can be positively selected. The gRNA expression plasmid for Mp*ARF1*, pMpGWB301_ARF1_1, which had previously been proved to be effective [25], was also used in this experiment (Figs. 1A, S2). A single co-transformation of sporelings with pMpGE006 and pMpGWB301_ARF1_1 yielded hundreds of NAA-resistant plants (Figs. 1B, S3), whereas the same experiment with the *hCas9* plasmid,

instead of pMpGE006, yielded few NAA-resistant lines (Fig. S3). We analyzed the target sequence of Mp*ARF1* in 16 NAA-resistant T1 co-transformants of pMpGE006 and pMpGWB301_ARF1_1 by direct sequencing, all of which harbored indels and/or base substitutions in the target site (Fig. 1C). Therefore, we conclude that the Atco-Cas9 product has much higher efficiency than hCas9.

Next, we examined whether mutants could be isolated without the NAA-based phenotypic selection. T1 co-transformants were selected only by using hygromycin and chlorsulfuron, which are selection markers for pMpGE006 and pMpGWB301_ARF1_1, respectively. Direct sequencing analyses showed that 75% of the randomly selected T1 plants (12 of 16) had some mutations in the target sequence of Mp*ARF1* (Fig. 1D, E). Although one plant showed a mosaic sequence pattern (Fig. 1D, E), the majority was non-mosaic, suggesting that the genome editing events had occurred in an early phase of transformation. From these data, we concluded that the efficiency of genome editing using the Atco-Cas9 expression cassette is high enough to isolate mutants by direct sequencing analysis of target genes without any target-gene-dependent phenotypic selections.

Construction of genome editing vector series in *M. polymorpha*

In addition to pMpGE006, we constructed a series of genome editing vectors for addressing the necessity of various applications of genome editing. We constructed a single binary vector system with both gRNA and Cas9 expression cassettes to make more easy-to-handle vectors (Fig. 2). pMpGE010 and pMpGE011 are based on the Gateway cloning system (Fig. 2A). The entry clone vector harbors gRNA expression cassettes in pENTR/D-TOPO, designated as pMpGE_En01 to pMpGE_En03 (Fig. 2A). pMpGE_En01 contains a 2 kbp Mp*U6-1*$_{pro}$ for gRNA expression and the cloning

313 site for double stranded oligos with a guide sequence using the In-Fusion reaction
314 (Fig. S1A). pMpGE_En02 contains a 500 bp Mp*U6-1*$_{pro}$ sequence for gRNA
315 expression and is designed for restriction enzyme-based cloning (Fig. S1B). This
316 vector does not contain a purine nucleotide at the 5' end of the gRNA cloning site,
317 which is used to start transcription by RNA polymerase III [36]. pMpGE_En03 is
318 identical to pMpGE_En02 except for a built-in "extra initial G" (Figs. 2A, S1B). The
319 "extra initial G" also exists in pMpGE_En01 (Fig. S1A). For binary vectors,
320 pMpGE010 was constructed using pMpGWB103 [18] as a backbone vector, which
321 harbors a hygromycin resistance cassette. Likewise, pMpGE011 was constructed
322 using pMpGWB303, which harbors a chlorsulfuron resistance cassette.
323 For easier construction, binary vectors based on restriction-enzyme-based
324 cloning, pMpGE013 and pMpGE014, were also constructed using pMpGE010 and
325 pMpGE011 as backbones, respectively. Double stranded oligos harboring the target
326 sequence of gRNAs can be directly cloned into pMpGE013 and pMpGE014, which
327 contain the "extra initial G," at the AarI restriction enzyme sites (Fig. 2B). These
328 vectors provide a low-cost alternative for gRNA cloning.
329
330 **Evaluation of the efficiency of genome editing vectors**
331 To evaluate genome-editing efficiencies in the new system, we exploited targeted
332 mutagenesis at two loci (Fig. 3). We cloned into pMpGE_En01 a gRNA harboring a
333 guide sequence identical to that of pMpGWB301_ARF1_1 and transferred it to
334 pMpGE010 (Fig. S2). The constructed vector was used for transformation of
335 sporelings and the transformants were selected with hygromycin only on auxin-free
336 media. Direct sequencing of the ARF1_1 gRNA target site revealed that 75% of T1
337 plants (24 of 32) had mutations at the target site (Fig. 3A, B). A similar result was

obtained with pMpGE011 (Fig. S4A). Thus, the efficiency of targeted mutagenesis with the single vector system was comparable to that with the double vector system.

To examine whether genome editing could be applied at other loci, we chose a gene, *NOPPERABO1* (Mp*NOP1*), which encodes a plant U-box E3 ubiquitin ligase that is responsible for air-chamber formation in *M. polymorpha* [37]. Since Mp*nop1* mutants form transparent thalli due to the lack of air chambers, they can be easily distinguished from wild-type plants with the naked eye. An Mp*NOP1*-targeting gRNA, NOP1_1 (Fig. S2), was introduced into sporelings with pMpGE010. Direct sequencing analysis revealed that 87.5% of T1 plants (28 of 32) had some mutation in the target sequence (Fig. 3C, D). Many of the mutant lines exhibited the transparent phenotype in the entire body, reflecting the frequency of non-mosaic sequence reads (Fig. 3E).

We also assessed a shorter (0.5 kbp) Mp*U6-1* promoter by comparing pMpGE_En01 and pMpGE_En02. Genome-editing efficiency with pMpGE_En02 was comparable to that with pMpGE_En01, when examined with the gRNA NOP1_1 (Fig. S4B). This result suggests that the 0.5 kbp Mp*U6-1* promoter is sufficient to drive gRNA expression for efficient genome editing.

Influence of gRNA lengths to genome editing efficiency

Next, we assessed the length of the gRNA guide sequence. It has previously been reported that truncated gRNA guide sequences (17 nt and 18 nt) show low off-target activities in mammalian cells [38]. We chose a guide sequence (5'-CAAACCGGAATGAGTCAGCT-3'), which targets the exon of the Mp*NOP1* gene (NOP1_2; Fig. S2) and cloned various lengths (20 nt, 18 nt, 17 nt, and 16 nt) of the sequence into pMpGE_En02, which does not have the “extra initial G.” Genome-

363 editing efficiencies were scored by classifying the penetrance of the transparent
364 phenotype due to mutations in the Mp*NOP1* gene: class I, plants with the transparent
365 phenotype observed throughout; class II, plants with the transparent phenotype
366 observed in a mosaic fashion; and class III, plants with no obvious phenotype (Fig.
367 4A). The results clearly showed that guide sequences of 17 nt or fewer had much
368 lower genome-editing efficiencies than those of 18 nt or more (Fig. 4B).
369 The low genome editing efficiency when using the gRNAs with 17-nt and
370 16-nt guide lengths could have been caused by their reduced expression levels due to
371 the lack of an initial guanine. Thus, we investigated the effects of addition of an
372 guanine to the 5' end of gRNAs (using pMpGE_En03), which should facilitate
373 transcription by pol III [39]. However, no clear improvement in genome editing
374 efficiency was observed (Fig. S5). These results suggest that the occurrence of lower
375 genome editing events in *M. polymorpha* may strongly depend on the length of a
376 guide sequence that perfectly matches the target genome, which should be 18 nt or
377 longer.
378
379 **Assessments of off target effects**
380 As plants obtained by transformation with pMpGE010/011 stably express Cas9 and a
381 gRNA, genome sites with sequences similar to that of a gRNA are always at risk of
382 genome editing. Sequencing analysis of the genome of more than 30 T1 plants that
383 harbored on-target mutations in the ARF1_1 or NOP1_1 target locus revealed no
384 mutation at any of the three most potential off-target sites (Table 1). Collectively, it is
385 suggested that genome editing efficiencies in off-target sites are much lower than
386 those in on-target sites in *M. polymorpha*.
387

388 **_De novo_ mutations after prolonged culture**

389 The stable expression of the CRISPR/Cas9 system should provide continuous
390 opportunities for targeted mutagenesis in transformants until a mutation has been
391 introduced. We analyzed gemmae (G1 generation [18]) derived from
392 pMpGE010_NOP1_1 T1 transformants that had shown sectors of transparent
393 (mutant) and non-transparent (wild-type) thallus regions in one individual (Fig. 5A).
394 As expected, G1 gammalings obtained from transparent sectors basically inherited the
395 same mutations as those found in their corresponding T1 sectors (Fig. 5B).
396 Concurrently, some gammalings from non-transparent sectors were found to contain
397 *de novo* mutations, which were different from those in the transparent sectors of the
398 same parent individuals (Fig. 5B, C). These results indicate that various allelic
399 mutants can be isolated from single non-mutated transformants.
400
401 **Induction of large deletion using two gRNAs**
402 Previous studies have reported that CRISPR/Cas9 system could induce deletions
403 between two gRNA target sites in mosses [8]. Accordingly, induction of large
404 deletions using two gRNAs was tested in *M. polymorpha*. We designed four gRNAs
405 to the Mp*NOP1* gene, NOP1_3, NOP1_4, NOP1_5, and NOP1_6 (Fig. 6A).
406 Simultaneous introduction of pMpGE013 and pMpGE014 harboring different
407 combinations of gRNAs was conducted, and transformants were selected with both
408 hygromycin and chlorsulfuron. Deletions of expected sizes from the gRNA
409 combinations, that is, a 0.5 kbp deletion with NOP1_3 and NOP1_4, a 1.5 kbp
410 deletion with NOP1_3 and NOP1_5, and a 4.5 kbp deletion with NOP1_3 and
411 NOP1_6, were detected, respectively (Fig. 6B). The efficiencies of induction of large
412 deletions were almost comparable regardless of the deletion size: 9/26 for 0.5 kbp,

4/20 for 1.5 kbp, and 6/23 for 4.5 kbp (Fig. 6C). These large-deletion lines displayed the Mp*nop1* phenotype. (Fig. S6). Collectively, the induction of large deletions with this system would also be applicable for functional analysis of genes in general in *M. polymorpha*.

One-step generation of conditional knockout mutants

Mutants that exhibit phenotypes under certain conditions are useful when a gene of interest has multiple roles in the life cycle and/or, in particular, essential functions. In this study, we provide methodology for obtaining conditional knockout mutants in a single step using the CRISPR/Cas9 system for mutagenesis and an inducible Cre-*lox*P site-specific recombination for complementation. As our model, we chose one of the three mitogen-activated protein kinase (MAPK) genes in *M. polymorpha*, Mp*MPK1* [16], which was predicted to be an essential gene because attempts to obtain knockout mutants by the homologous-recombination-based gene targeting method [22] had failed.

We constructed pMpGE010 harboring a gRNA that was designed to target the first exon-intron junction in Mp*MPK1* (Fig. 7A). Because the conjunction between exons 1 and 2 does not reconstitute the PAM sequence for this gRNA (Fig. 7A), it is not supposed to target the Mp*MPK1* cDNA. Thus, for complementation, we cloned an Mp*MPK1* cDNA into the binary vector pMpGWB337 [23] or its derivative pMpGWB337tdTN (Fig. S7), both of which normally drive expression of the cDNA but inducibly allow deletion of the cDNA and expression of a fluorescent protein after application of heat shock and DEX. Transformation of sporelings with only the Mp*MPK1* CRISPR vector yielded a small number of mosaic mutants but no monoclonal frameshift mutants (Fig. 7B), which is indicative of possible lethality for

genome editing at the target site. However, simultaneous transformation with the same CRISPR vector and either of the complementation vectors described above gave rise to monoclonal frameshift mutants at high frequencies (Fig. 7B, C). These data suggest that the cDNA resistant to the gRNA allowed complementation of deleterious mutations in the endogenous Mp*MPK1* locus.

These complemented mutants grew normally and were fluorescent negative under mock conditions, whereas upon heat shock and DEX treatment, they grew extremely slowly and became fluorescent positive (Fig. 7D). Induction of cDNA deletion was confirmed by genomic PCR analysis (Fig. 7E). These results suggest that Mp*MPK1* is indeed a nearly essential gene and, more importantly, demonstrated that conditional knockout mutants of an essential gene can be generated by a simple procedure using a CRISPR/Cas9 vector and a pMpGWB337 derivative in *M. polymorpha*.

Discussion

Using the *Atco-Cas9* expression cassette, we successfully optimized the CRISPR/Cas9-based genome editing system for *M. polymorpha* and improved the efficiency to a degree that does not require target gene-based phenotypic selection. This is consistent with a previous report that codon optimization of Cas9 lead to significant improvement in efficiency [40]. In the pMpGE010/011 system, over 70% of transformants underwent targeted mutagenesis, shown by the two gRNAs ARF1_1 and NOP1_1 (Figs. 1, 2). These results indicate that the pMpGE010/011 vectors are highly reliable for obtaining genome-edited lines in *M. polymorpha*. Since the GC content of the *M. polymorpha* genome is 49.8% [16], Cas9 target sites with NGG PAM sequences can be found at high probability, and the selection of gRNA target sites with small numbers of off-target sites is possible. Taken together, we conclude

463 that the pMpGE010/pMpGE011 system is feasible to use for functional genetics in *M.*
464 *polymorpha* with likely avoidance of off-target effects.

465 As the system presented here allows constant expression of both Cas9 and a
466 gRNA, two risks are conceivable: (i) a side effect of the overexpression of Cas9
467 protein on plant development and (ii) genome editing events at off-target sites.
468 However, both risks appear to be negligible. Firstly, Mp*nop1* mutants obtained with
469 our CRISPR system were indistinguishable in terms of growth and morphology from
470 those obtained by T-DNA tagging [22] or gene targeting [37] (Fig. 3A). In addition,
471 the conditional Mp*mpk1* mutants generated in this study did not show any noticeable
472 developmental defects (Fig. 7D). These observations suggest that there is no
473 detrimental effect of Cas9 overexpression on plant development.

474 Sequencing of several potential off-target sites of 20-nt guide sequences
475 revealed remarkably low off-target effects: no mutation was observed at any of the
476 sites in over 30 lines that had on-target mutations. In plants, a similar low-level off-
477 target effect was reported for Arabidopsis [14]. It was reported that truncated gRNA
478 guide sequences (17 nt and 18 nt) show low off-target activities in mammalian cells
479 [38]. In *M. polymorpha*, 17-nt and shorter gRNA guide sequences were not effective
480 (Fig. 3). Taken together with the observed low off-target effects, to avoid unstable
481 outcomes in isolating genome-edited lines, it is recommended to use gRNAs with 18-
482 nt or longer guide sequences in *M. polymorpha*.

483 Constant expression of the CRISPR system allows transformants that have not
484 had any alteration to potentially acquire mutations later at the target site in random
485 cells during growth. Indeed, we observed that pMpGE010-transformed thalli with no
486 target mutation in the T1 generation produced gammae with *de novo* mutations (Fig.
487 5). This feature would be convenient for isolating mutant alleles without further

transformation and could also be used for mosaic analyses. Conversely, even if a monoclonal mutant pattern is detected in genotyping using a portion of T1 plants, it does not guarantee that the whole plant has the same genotype. For genotyping, it is recommended to use a piece of T1 tissue from the basal side of the apical notch, as G1 gemmae derived from its apical-side tissues are most likely to be clones, with an identical genotype to that found in T1 [17].

Diploid or polyploid plants can bear lethal mutations heterozygously if recessive, whereas *M. polymorpha*, a haploid-dominant plant, cannot. Therefore, functional analyses of essential genes require alternate strategies. Flores-Sandoval (2016) reported an inducible system for artificial microRNA expression in *M. polymorpha*. Another strategy would be to create conditional knockout mutants. In mice, this is usually achieved by inserting *lox*P sites into two different introns in the same direction with homologous-recombination-mediated gene targeting, and then expressing Cre recombinase at a specific time and/or location to remove an essential exon [41]. For *M. polymorpha*, we avoided using a laborious gene-targeting-based strategy. Instead, we established a method to generate a conditional knockout mutant by simultaneous introduction of a mutation in the endogenous target gene by the CRISPR/Cas9 system and of a conditionally removable complementation gene as a transgene.

In this strategy, the complementation gene cassette must have a structure that cannot be targeted by the gRNA used for knocking out the target gene. Although this "gRNA-resistant" complementation cassette can be prepared by introducing synonymous substitutions in the matching sequence, extra time is required. Thus, if possible, we recommend complementation, by expressing a non-modified cDNA, of mutations caused by a non-cDNA-targeting gRNA, which can be designed at exon-

intron junctions. A DNA fragment for complementation can be inserted between two *lox*P sites in pMpGWB337 [23] or its derivatives with different fluorescent protein markers (Fig. S7), all-in-one vectors equipped with a floxed Gateway cassette for introducing a complementing gene and with a heat-shock- and DEX-inducible Cre recombinase expression cassette.

Using this method, we successfully isolated transformants with a mutation in the Mp*MPK1* locus, which was expected to be essential. Mp*MPK1* is one of the three genes encoding canonical MAPKs and is most closely related to those categorized as groups A and B for plant MAPKs [16, 42]. Major Arabidopsis MAPKs in these groups (MPK3/MPK6 and MPK4, respectively) have been shown to regulate various aspects of growth and development [43]: *mpk3 mpk6* double mutants are embryonic lethal [44], and *mpk4* mutants show growth retardation with a cytokinesis defect [45]. The co-transformation-generated transgenic plants that had a mutation in the endogenous Mp*MPK1* locus exhibited no growth defect due to complementation by the expression of an Mp*MPK1* cDNA (Fig. 7). Induction of Cre-*lox*P recombination by heat shock and DEX treatment resulted in severe growth defects, indicating that MpMPK1 plays a critical role in the regulation of growth and development in *M. polymorpha* and that the conditional induction of the mutant phenotypes successfully occurred. Sectors of non-abnormal tissues sometimes arose, probably due to occasional failure of the Cre-*lox*P recombination. Thus, it is important to select conditional lines with highly efficient recombination induction from independently isolated candidates. The conditional knockout system developed in this study should not only facilitate functional analyses of essential genes, but also be useful for uncovering functions of non-essential genes in specific locations or timings.

DSBs created by CRISPR/Cas9-based genome editing are usually repaired by the NHEJ pathway. Although the error-prone NHEJ repair pathway randomly inserts or deletes bases at the DSB sites, some tendencies were observed in *M. polymorpha*. For example, mutants with a >20 bp deletion were frequently obtained (Figs. 1, 2), while such mutations are rare in genome editing in Arabidopsis [12, 46]. This suggests that the NHEJ-based repair activity in *M. polymorpha* is relatively weaker than that in Arabidopsis. Consistent with this idea, homologous recombination-based gene targeting is possible in *M. polymorpha* with the average rate of 2-3%, which is much higher than in Arabidopsis [22]. Since DSBs induced by the CRISPR/Cas9 system were reported to increase the efficiency of homologous recombination [47], it would be possible to improve the gene targeting efficiency in *M. polymorpha* by combination with the vectors presented in this study. Another frequently observed feature was the repair of DSBs by the microhomology-mediated end joining (MMEJ) pathway (Fig. S8). Thus, precise integration of a DNA construct into target loci assisted by MMEJ, which is available for animals [48, 49], would also be possible in *M. polymorpha*.

Construction of the genome editing vectors in the present study was very simple and inexpensive. Together with the fact that *M. polymorpha* has low redundancy and small numbers of regulatory genes, our highly efficient CRISPR/Cas9 system would allow genome-editing-based genetic screens targeting all genes in a large subset of the genome, e.g., protein kinases, transcription factors, and miRNAs [16, 50, 51]. Since the “pMpGE” CRISPR/Cas9-based genome editing vectors reported in this study possess 35S promoter-based marker cassettes, the vectors can be utilized for genome editing in other plants in which MpEF_{pro} and Mp$U6_{pro}$ are operable. Highly efficient genome editing should facilitate uncovering

562 gene regulatory networks that evolved for the land adaptation of plants and that
563 underlie subsequent successful expansion of land plants.
564

565 **Supporting information**

566 S1 Supplementary Figures S1-S8 (PDF)
567 S2 Supplementary Table S1(.xlsx file)
568

569 **Data availability**

570 The complete nucleotide sequences of the pMpGEs have been deposited in the
571 GenBank/EMBL/DDBJ databases under the accessions nos. LC090754 to LC090757.
572 pMpGE plasmids can be obtained from Addgene (www.addgene.org; plasmid
573 numbers 71534–71537).
574

575 **Acknowledgements**

576 We thank H. Puchta for the gifts of the pDe-CAS9 material, including Atco-Cas9. We
577 thank S. Yamaoka, K.T. Yamato, S. Zachgo, Y. Osakabe, M. Endo, and S. Toki for
578 helpful discussion. We also thank J. Haseloff and B. Pollak for the gRNA sequence of
579 the Mp*NOP1* gene. We thank M. Fukuhara, Y. Hatta, and Y. Koumoto for their
580 technical assistance.
581

582 **Author contribution**

583 TK, RN, and SSS conceived of the study and participated in its design and
584 coordination. SSS, KO, and IHN designed the pMpGE vectors, and SSS, MS, JT, and
585 YM constructed these vectors. RN designed the pMpGWB337 derivative vectors, and
586 SI constructed these vectors. RN designed and RN, YM, and SI conducted the

587 conditional knockout experiments. SSS and TS conducted the large deletion assays.
588 YM conducted most of the other experiments. TK, RN, and SSS wrote the manuscript
589 with inputs from co-authors.
590
591 **References**
592 1. Doudna JA, Charpentier E. Genome editing. The new frontier of genome
593 engineering with CRISPR-Cas9. Science. 2014;346(6213):1258096. Epub 2014/11/29.
594 doi: 10.1126/science.1258096
595 1258096 [pii]
596 346/6213/1258096 [pii]. PubMed PMID: 25430774.
597 2. Zhang F. CRISPR-Cas9: Prospects and Challenges. Hum Gene Ther.
598 2015;26(7):409-10. Epub 2015/07/16. doi: 10.1089/hum.2015.29002.fzh. PubMed
599 PMID: 26176430; PubMed Central PMCID: PMC4939442.
600 3. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A
601 programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.
602 Science. 2012;337(6096):816-21. Epub 2012/06/30. doi: 10.1126/science.1225829.
603 PubMed PMID: 22745249.
604 4. Gaj T, Sirk SJ, Shui SL, Liu J. Genome-Editing Technologies: Principles and
605 Applications. Cold Spring Harb Perspect Biol. 2016;8(12). Epub 2016/12/03. doi:
606 a023754 [pii]
607 10.1101/cshperspect.a023754
608 cshperspect.a023754 [pii]. PubMed PMID: 27908936.
609 5. Puchta H. Applying CRISPR/Cas for genome engineering in plants: the best is
610 yet to come. Curr Opin Plant Biol. 2016;36:1-8. Epub 2016/12/04. doi: S1369-
611 5266(16)30125-X [pii]

612 10.1016/j.pbi.2016.11.011. PubMed PMID: 27914284.

613 6. Yin K, Gao C, Qiu JL. Progress and prospects in plant genome editing. Nat

614 Plants. 2017;3:17107. Epub 2017/08/02. doi: 10.1038/nplants.2017.107. PubMed

615 PMID: 28758991.

616 7. Osakabe Y, Osakabe K. Genome editing with engineered nucleases in plants.

617 Plant Cell Physiol. 2015;56(3):389-400. Epub 2014/11/25. doi: 10.1093/pcp/pcu170.

618 PubMed PMID: 25416289.

619 8. Nomura T, Sakurai T, Osakabe Y, Osakabe K, Sakakibara H. Efficient and

620 Heritable Targeted Mutagenesis in Mosses Using the CRISPR/Cas9 System. Plant

621 Cell Physiol. 2016;57(12):2600-10. Epub 2016/12/18. doi: 10.1093/pcp/pcw173.

622 PubMed PMID: 27986915.

623 9. Lopez-Obando M, Hoffmann B, Gery C, Guyon-Debast A, Teoule E, Rameau

624 C, et al. Simple and Efficient Targeting of Multiple Genes Through CRISPR-Cas9 in

625 Physcomitrella patens. G3 (Bethesda). 2016. Epub 2016/09/11. doi:

626 10.1534/g3.116.033266. PubMed PMID: 27613750; PubMed Central PMCID:

627 PMCPMC5100863.

628 10. Greiner A, Kelterborn S, Evers H, Kreimer G, Sizova I, Hegemann P.

629 Targeting of Photoreceptor Genes in Chlamydomonas reinhardtii via Zinc-Finger

630 Nucleases and CRISPR/Cas9. Plant Cell. 2017;29(10):2498-518. Epub 2017/10/06.

631 doi: 10.1105/tpc.17.00659. PubMed PMID: 28978758.

632 11. Ferenczi A, Pyott DE, Xipnitou A, Molnar A. Efficient targeted DNA editing

633 and replacement in Chlamydomonas reinhardtii using Cpf1 ribonucleoproteins and

634 single-stranded DNA. Proc Natl Acad Sci U S A. 2017;114(51):13567-72. Epub

635 2017/12/07. doi: 10.1073/pnas.1710597114. PubMed PMID: 29208717; PubMed

636 Central PMCID: PMCPMC5754772.

637 12. Fauser F, Schiml S, Puchta H. Both CRISPR/Cas-based nucleases and
638 nickases can be used efficiently for genome engineering in Arabidopsis thaliana. Plant
639 J. 2014;79(2):348-59. Epub 2014/05/20. doi: 10.1111/tpj.12554. PubMed PMID:
640 24836556.
641 13. Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, et al. Simultaneous
642 editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance
643 to powdery mildew. Nat Biotechnol. 2014;32(9):947-51. Epub 2014/07/21. doi:
644 10.1038/nbt.2969
645 nbt.2969 [pii]. PubMed PMID: 25038773.
646 14. Feng Z, Mao Y, Xu N, Zhang B, Wei P, Yang DL, et al. Multigeneration
647 analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced
648 gene modifications in Arabidopsis. Proc Natl Acad Sci U S A. 2014;111(12):4632-7.
649 Epub 2014/02/20. doi: 10.1073/pnas.1400822111
650 1400822111 [pii]. PubMed PMID: 24550464; PubMed Central PMCID:
651 PMC3970504.
652 15. Chang C, Bowman JL, Meyerowitz EM. Field Guide to Plant Model Systems.
653 Cell. 2016;167(2):325-39. Epub 2016/10/08. doi: S0092-8674(16)31082-0 [pii]
654 10.1016/j.cell.2016.08.031. PubMed PMID: 27716506; PubMed Central PMCID:
655 PMC5068971.
656 16. Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, et al.
657 Insights into Land Plant Evolution Garnered from the Marchantia polymorpha
658 Genome. Cell. 2017;171(2):287-304 e15. Epub 2017/10/07. doi:
659 10.1016/j.cell.2017.09.030. PubMed PMID: 28985561.

17. Barnes CR, Land WJG. Bryological papers II. The origin of the cupule of marchantia - Contributions from the hull botanical laboratory 120. Bot Gaz. 1908;46:401-9. PubMed PMID: ISI:000202146600029.

18. Ishizaki K, Nishihama R, Yamato KT, Kohchi T. Molecular Genetic Tools and Techniques for Marchantia polymorpha Research. Plant Cell Physiol. 2016;57(2):262-70. Epub 2015/06/28. doi: 10.1093/pcp/pcv097 pcv097 [pii]. PubMed PMID: 26116421.

19. Ishizaki K, Chiyoda S, Yamato KT, Kohchi T. Agrobacterium-mediated transformation of the haploid liverwort Marchantia polymorpha L., an emerging model for plant biology. Plant Cell Physiol. 2008;49(7):1084-91. Epub 2008/06/07. doi: 10.1093/pcp/pcn085 pcn085 [pii]. PubMed PMID: 18535011.

20. Kubota A, Ishizaki K, Hosaka M, Kohchi T. Efficient Agrobacterium-mediated transformation of the liverwort Marchantia polymorpha using regenerating thalli. Biosci Biotechnol Biochem. 2013;77(1):167-72. Epub 2013/01/08. doi: 10.1271/bbb.120700. PubMed PMID: 23291762.

21. Tsuboyama-Tanaka S, Kodama Y. AgarTrap-mediated genetic transformation using intact gemmae/gemmalings of the liverwort Marchantia polymorpha L. J Plant Res. 2015;128(2):337-44. Epub 2015/02/11. doi: 10.1007/s10265-014-0695-2. PubMed PMID: 25663453.

22. Ishizaki K, Johzuka-Hisatomi Y, Ishida S, Iida S, Kohchi T. Homologous recombination-mediated gene targeting in the liverwort Marchantia polymorpha L. Sci Rep. 2013;3:1532. Epub 2013/03/26. doi: 10.1038/srep01532 srep01532 [pii]. PubMed PMID: 23524944; PubMed Central PMCID: PMC3607118.

684 23. Nishihama R, Ishida S, Urawa H, Kamei Y, Kohchi T. Conditional Gene
685 Expression/Deletion Systems for Marchantia polymorpha Using its Own Heat-Shock
686 Promoter and Cre/loxP-Mediated Site-Specific Recombination. Plant Cell Physiol.
687 2016;57(2):271-80. Epub 2015/07/08. doi: 10.1093/pcp/pcv102
688 pcv102 [pii]. PubMed PMID: 26148498.
689 24. Kopischke S, Schussler E, Althoff F, Zachgo S. TALEN-mediated genome-
690 editing approaches in the liverwort Marchantia polymorpha yield high efficiencies for
691 targeted mutagenesis. Plant Methods. 2017;13:20. Epub 2017/04/01. doi:
692 10.1186/s13007-017-0167-5. PubMed PMID: 28360929; PubMed Central PMCID:
693 PMCPMC5370431.
694 25. Sugano SS, Shirakawa M, Takagi J, Matsuda Y, Shimada T, Hara-Nishimura I,
695 et al. CRISPR/Cas9-mediated targeted mutagenesis in the liverwort Marchantia
696 polymorpha L. Plant Cell Physiol. 2014;55(3):475-81. Epub 2014/01/21. doi:
697 10.1093/pcp/pcu014
698 pcu014 [pii]. PubMed PMID: 24443494.
699 26. Mikami M, Toki S, Endo M. Comparison of CRISPR/Cas9 expression
700 constructs for efficient targeted mutagenesis in rice. Plant Mol Biol. 2015;88(6):561-
701 72. Epub 2015/07/21. doi: 10.1007/s11103-015-0342-x. PubMed PMID: 26188471;
702 PubMed Central PMCID: PMC4523696.
703 27. Johnson RA, Gurevich V, Filler S, Samach A, Levy AA. Comparative
704 assessments of CRISPR-Cas nucleases' cleavage efficiency in planta. Plant Mol Biol.
705 2015;87(1-2):143-56. Epub 2014/11/19. doi: 10.1007/s11103-014-0266-x. PubMed
706 PMID: 25403732.

707 28. Gamborg OL, Miller RA, Ojima K. Nutrient requirements of suspension
708 cultures of soybean root cells. Exp Cell Res. 1968;50(1):151-8. Epub 1968/04/01.
709 PubMed PMID: 5650857.
710 29. Ishizaki K, Nishihama R, Ueda M, Inoue K, Ishida S, Nishimura Y, et al.
711 Development of Gateway Binary Vector Series with Four Different Selection Markers
712 for the Liverwort Marchantia polymorpha. PLoS One. 2015;10(9):e0138876. Epub
713 2015/09/26. doi: 10.1371/journal.pone.0138876. PubMed PMID: 26406247; PubMed
714 Central PMCID: PMCPMC4583185.
715 30. Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, et al.
716 Development of series of gateway binary vectors, pGWBs, for realizing efficient
717 construction of fusion genes for plant transformation. J Biosci Bioeng.
718 2007;104(1):34-41. Epub 2007/08/19. doi: 10.1263/jbb.104.34. PubMed PMID:
719 17697981.
720 31. Xiao A, Cheng Z, Kong L, Zhu Z, Lin S, Gao G, et al. CasOT: a genome-wide
721 Cas9/gRNA off-target searching tool. Bioinformatics. 2014;30(8):1180-2. Epub
722 2014/04/15. doi: btt764 [pii]
723 10.1093/bioinformatics/btt764. PubMed PMID: 24389662.
724 32. Althoff F, Kopischke S, Zobell O, Ide K, Ishizaki K, Kohchi T, et al.
725 Comparison of the MpEF1alpha and CaMV35 promoters for application in
726 Marchantia polymorpha overexpression studies. Transgenic Res. 2013. Epub
727 2013/09/17. doi: 10.1007/s11248-013-9746-z. PubMed PMID: 24036909.
728 33. Kato H, Ishizaki K, Kouno M, Shirakawa M, Bowman JL, Nishihama R, et al.
729 Auxin-Mediated Transcriptional System with a Minimal Set of Components Is
730 Critical for Morphogenesis through the Life Cycle in Marchantia polymorpha. PLoS
731 Genet. 2015;11(5):e1005084. Epub 2015/05/29. doi: 10.1371/journal.pgen.1005084

732 PGENETICS-D-14-02665 [pii]. PubMed PMID: 26020919; PubMed Central PMCID:
733 PMC4447296.
734 34. Flores-Sandoval E, Eklund DM, Bowman JL. A Simple Auxin Transcriptional
735 Response System Regulates Multiple Morphogenetic Processes in the Liverwort
736 Marchantia polymorpha. PLoS Genet. 2015;11(5):e1005207. Epub 2015/05/29. doi:
737 10.1371/journal.pgen.1005207
738 PGENETICS-D-14-02661 [pii]. PubMed PMID: 26020649; PubMed Central PMCID:
739 PMC4447368.
740 35. Kato H, Kouno M, Takeda M, Suzuki H, Ishizaki K, Nishihama R, et al. The
741 Roles of the Sole Activator-Type Auxin Response Factor in Pattern Formation of
742 Marchantia polymorpha. Plant Cell Physiol. 2017;58(10):1642-51. Epub 2017/10/11.
743 doi: 10.1093/pcp/pcx095. PubMed PMID: 29016901.
744 36. Ma H, Wu Y, Dang Y, Choi JG, Zhang J, Wu H. Pol III Promoters to Express
745 Small RNAs: Delineation of Transcription Initiation. Mol Ther Nucleic Acids.
746 2014;3:e161. Epub 2014/05/08. doi: 10.1038/mtna.2014.12. PubMed PMID:
747 24803291; PubMed Central PMCID: PMCPMC4040628.
748 37. Ishizaki K, Mizutani M, Shimamura M, Masuda A, Nishihama R, Kohchi T.
749 Essential role of the E3 ubiquitin ligase nopperabo1 in schizogenous intercellular
750 space formation in the liverwort Marchantia polymorpha. Plant Cell.
751 2013;25(10):4075-84. Epub 2013/10/31. doi: 10.1105/tpc.113.117051
752 tpc.113.117051 [pii]. PubMed PMID: 24170128; PubMed Central PMCID:
753 PMC3877802.
754 38. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas
755 nuclease specificity using truncated guide RNAs. Nat Biotechnol. 2014;32(3):279-84.

doi: 10.1038/nbt.2808. PubMed PMID: 24463574; PubMed Central PMCID: PMCPMC3988262.

39. Ma J, Wang MD. RNA polymerase is a powerful torsional motor. Cell Cycle. 2014;13(3):337-8. Epub 2013/12/18. doi: 10.4161/cc.27508. PubMed PMID: 24335432; PubMed Central PMCID: PMCPMC3956522.

40. Endo M, Mikami M, Toki S. Multigene knockout utilizing off-target mutations of the CRISPR/Cas9 system in rice. Plant Cell Physiol. 2015;56(1):41-7. Epub 2014/11/14. doi: 10.1093/pcp/pcu154 pcu154 [pii]. PubMed PMID: 25392068; PubMed Central PMCID: PMC4301742.

41. Hall B, Limaye A, Kulkarni AB. Overview: generation of gene knockout mice. Curr Protoc Cell Biol. 2009;Chapter 19:Unit 19 2 2 1-7. Epub 2009/09/05. doi: 10.1002/0471143030.cb1912s44. PubMed PMID: 19731224; PubMed Central PMCID: PMCPMC2782548.

42. Group M. Mitogen-activated protein kinase cascades in plants: a new nomenclature. Trends Plant Sci. 2002;7(7):301-8. Epub 2002/07/18. PubMed PMID: 12119167.

43. Xu J, Zhang S. Mitogen-activated protein kinase cascades in signaling plant growth and development. Trends Plant Sci. 2015;20(1):56-64. Epub 2014/12/03. doi: 10.1016/j.tplants.2014.10.001. PubMed PMID: 25457109.

44. Wang H, Ngwenyama N, Liu Y, Walker JC, Zhang S. Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in Arabidopsis. Plant Cell. 2007;19(1):63-73. Epub 2007/01/30. doi: 10.1105/tpc.106.048298. PubMed PMID: 17259259; PubMed Central PMCID: PMCPMC1820971.

780 45. Kosetsu K, Matsunaga S, Nakagami H, Colcombet J, Sasabe M, Soyano T, et
781 al. The MAP kinase MPK4 is required for cytokinesis in Arabidopsis thaliana. Plant
782 Cell. 2010;22(11):3778-90. Epub 2010/11/26. doi: 10.1105/tpc.110.077164. PubMed
783 PMID: 21098735; PubMed Central PMCID: PMCPMC3015120.
784 46. Shen H, Strunks GD, Klemann BJ, Hooykaas PJ, de Pater S. CRISPR/Cas9-
785 Induced Double-Strand Break Repair in Arabidopsis Nonhomologous End-Joining
786 Mutants. G3 (Bethesda). 2017;7(1):193-202. Epub 2016/11/21. doi:
787 10.1534/g3.116.035204. PubMed PMID: 27866150; PubMed Central PMCID:
788 PMCPMC5217109.
789 47. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome
790 engineering using CRISPR/Cas systems. Science. 2013;339(6121):819-23. Epub
791 2013/01/05. doi: 10.1126/science.1231143
792 science.1231143 [pii]. PubMed PMID: 23287718; PubMed Central PMCID:
793 PMC3795411.
794 48. Aida T, Nakade S, Sakuma T, Izu Y, Oishi A, Mochida K, et al. Gene cassette
795 knock-in in mammalian cells and zygotes by enhanced MMEJ. BMC Genomics.
796 2016;17(1):979. Epub 2016/11/30. doi: 10.1186/s12864-016-3331-9. PubMed PMID:
797 27894274; PubMed Central PMCID: PMCPMC5126809.
798 49. Sakuma T, Nakade S, Sakane Y, Suzuki KT, Yamamoto T. MMEJ-assisted
799 gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. Nat Protoc.
800 2016;11(1):118-33. Epub 2015/12/19. doi: 10.1038/nprot.2015.140
801 nprot.2015.140 [pii]. PubMed PMID: 26678082.
802 50. Lin PC, Lu CW, Shen BN, Lee GZ, Bowman JL, Arteaga-Vazquez MA, et al.
803 Identification of miRNAs and Their Targets in the Liverwort Marchantia polymorpha
804 by Integrating RNA-Seq and Degradome Analyses. Plant Cell Physiol.

2016;57(2):339-58. Epub 2016/02/11. doi: 10.1093/pcp/pcw020. PubMed PMID: 26861787; PubMed Central PMCID: PMCPMC4788410.

51. Tsuzuki M, Nishihama R, Ishizaki K, Kurihara Y, Matsui M, Bowman JL, et al. Profiling and Characterization of Small RNAs in the Liverwort, Marchantia polymorpha, Belonging to the First Diverged Land Plants. Plant Cell Physiol. 2016;57(2):359-72. Epub 2015/11/22. doi: 10.1093/pcp/pcv182. PubMed PMID: 26589267.

Figure legends

Fig 1. Improvement of the *M. polymorpha* CRISPR/Cas9 system by codon optimization.

(A) Diagrams of the vectors used. pMpGE006 contains a cassette for the expression of Atco-Cas9 fused with an NLS under the control of MpEF_{pro}. pMpGWB301_ARF1_1 contains a cassette for the expression of the gRNA ARF1_1 under the control of Mp$U6\text{-}1_{pro}$ [25]. (B) Photograph of auxin-insensitive co-transformants of the two vectors described in (A). *Agrobacterium*-co-cultured sporelings that corresponded to one eighth of sporangium were plated on a medium in containing 3 μM NAA and two vector selection substances, 10 mg/L hygromycin and 0.5 μM chlorsulfuron. Diameter of the circle dish shows 10 cm. (C) Direct sequencing analysis of the target locus of Mp*ARF1* in auxin-selected T1 co-transformants. Inserted or substituted bases are colored in magenta. The target guide sequence of ARF1_1 is shown in bold face with the PAM sequence in blue. (D and E) Proportions of genome editing patterns (D) and sequences of the target site (E) in T1 co-transformants, which were obtained with no auxin selection. Sixteen independent transformants were analyzed for the target-site sequence by direct sequencing. “Monoclonal” and “Mosaic” indicate direct sequencing read patterns with mutated sequence peaks only and those with mixed sequence peaks, respectively. “WT” indicates read patterns identical to the original target sequence.

Fig 2. All-in-one vector systems for genome editing in *M. polymorpha*.

(A) Designs of Gateway-based all-in-one binary vectors and entry plasmids for gRNA cloning. pMpGE010 and pMpGE011 contain a cassette for the expression of Atco-Cas9 fused with an NLS under the control of MpEF_{pro}, a Gateway cassette, and a

838 cassette for the expression of hygromycin phosphotransferase (*HPT*) in pMpGE010
839 and mutated acetolactate synthase (*mALS*) in pMpGE011. pMpGE_En01 contains
840 recognition sites for two restriction enzymes, SacI and PstI, upstream of a gRNA
841 backbone for the insertion of a guide sequence by In-Fusion/Gibson cloning, which
842 automatically places a G nucleotide for transcription initiation by RNA polymerase III
843 (extra initial G). Expression of single guide RNAs is controlled by a 2 kbp fragment
844 of Mp*U6-1*$_{pro}$. pMpGE_En02 and pMpGE_En03 contain two BsaI recognition sites
845 upstream of the gRNA backbone for the insertion of a guide sequence by ligation
846 without or with an "extra initial G," whose expression is under the control of a 500 bp
847 Mp*U6-1*$_{pro}$ fragment. For all the entry vectors, the gRNA cassette is flanked by the
848 attL1 and attL2 sequences and is thus transferrable to the Gateway cassette in
849 pMpGE010 or pMpGWB011 by the LR reaction. (B) Designs of all-in-one binary
850 vectors for direct gRNA cloning. pMpGE013 (*HPT* marker) and pMpGE014 (*mALS*
851 marker) contain the Atco-Cas9-NLS expression cassette, a unique AarI site in the
852 upstream of the gRNA backbone for insertion of a guide sequence by ligation with an
853 "extra initial G," whose expression is under the control of a 500 bp Mp*U6-1*$_{pro}$
854 fragment.
855
856 **Fig 3.** High-efficiency genome editing by the gateway-based system in *M.*
857 *polymorpha*.
858 Sporelings were transformed with pMpGE010 harboring ARF1_1 gRNA (A and B) or
859 NOP1_1 gRNA (C-E). Randomly selected transformants were genotyped. Proportions
860 of genome editing patterns (A and C) and sequences of the target sites (B and D) in 32
861 T1 transformants are shown as in Fig. 1. Panel E shows photographs of a wild-type

plant (Tak-1) and a T1 transformant that exhibited the typical *NOP1*-defective "transparent" phenotype (Mp*nop1*^{ge}).

Fig 4. Effects of guide sequence lengths

(A) Classification of phenotypes in Mp*NOP1* genome editing lines by the appearance patterns of transparent portions. Class I, entirely transparent (Mutant); class II, mosaic of transparent and non-transparent sectors (Mosaic); class III, entirely non-transparent (wild-type, WT). Scale bars = 1 cm. (B) Proportions of mutant phenotype classes in T1 plants transformed with Mp*NOP1*-targeting gRNAs (NOP1_2) of different lengths. The shortest gRNA guide sequence tested was 16 nt. The numbers of T1 plants inspected are shown on the right hand side.

Fig 5. *De novo* mutations found in the G1 generation

(A) Delayed generation of new mutants from non-mutated sectors. G1 gemmae formed on a transparent or non-transparent sector of a class-II Mp*nop1*^{ge} plant, described in Fig. 4, were grown. Arrowhead shows a transparent sector due to a *de novo* mutation. Scale bar = 1 cm. (B) Proportions of genotypes in G1 populations derived from three independent class-II Mp*nop1*^{ge} T1 lines (#7, #11, and #13). "Inherited mutation" and "*de novo* mutation" indicate mutations identical to or different from those identified in individual T1 lines, respectively. (C) Target-site sequences in lines with "Inherited mutation" and "*de novo* mutation" by direct sequencing analysis. Inserted or substituted bases are colored in magenta. The target guide sequence of NOP1_1 is shown in bold face with the PAM sequence in blue.

886 **Fig 6.** Induction of large deletions using co-transformation of two genome editing
887 vectors
888 (A) Design of gRNAs to the Mp*NOP1* locus to dissect efficiencies of induction of
889 large deletions. Grey boxes and lines indicate exons and introns. The dotted line
890 indicates the downstream region of Mp*NOP1*. gRNA positions are shown in red lines.
891 Primer sets for PCR-based genotyping are also shown. (B) Representative images of
892 electrophoresis of PCR-based genotyping. The gRNAs and primer sets used in the
893 genotyping are shown. Expected sizes of the PCR products from the wild-type
894 genome are colored in blue. Expected sizes of the PCR products from the genome in
895 which inductions of large deletion occurred are colored in magenta. PCR products
896 from non-specific amplifications are indicated by asterisks. (C) Summary table of the
897 co-transformation experiment using pMpGE013 and pMpGE014. The ratio in the
898 "Large deletion" row shows the number of T1 plants harboring the expected large
899 deletion out of all the T1 plants inspected.
900
901 **Fig 7.** Generation of conditional knockout mutants of an essential gene
902 (A) Gene model of the Mp*MPK1* gene and the gRNA target site. Grey boxes and lines
903 indicate exons and introns. "ATG" and "Stop" denote the predicated initiation and
904 termination codons. PAM, target, and intron sequences are shown in blue, in bold, and
905 by small letters, respectively. Note that the third base (g) of the PAM sequence is in
906 intron 1 and that the first base of exon 2 is not G. (B) Targeted mutagenesis rate of the
907 endogenous Mp*MPK1* locus by CRISPR/Cas9. Bar graphs show proportions of clonal
908 frameshift (magenta), clonal in-frame (green), mosaic (yellow), and no mutations
909 (grey) that were identified in plants transformed with pMpGE010 containing the
910 gRNA shown in (A) only (Exp. 1) or together with Mp*MPK1* cDNA-containing

911 pMpGWB337 (Exp. 2) or pMpGWB337tdTN (Exp. 3). (C) Target-site sequences of
912 plants obtained in Exp. 3. Sequences of wild type (WT) and transgenic lines (#1 to #8
913 except #4) are shown as in (A). Inserted or deleted bases are colored in magenta with
914 their numbers in parentheses. In line #4, a 2 bp insertion and a large 970 bp deletion
915 that covers the predicted initiation codon were identified (magenta). (D) Growth
916 defect manifested by conditional deletion of the Mp*MPK1* cDNA. Gemmae of
917 transgenic line #4 were (+) or were not (–) subjected to heat shock (HS) and
918 dexamethasone (DEX) treatment on day 0 and day 1 and grown at 22°C for 2 weeks
919 (top) before observation by fluorescent microscopy for tdTomato-NLS (bottom).
920 Scale bars = 5 mm (top); 1 mm (bottom). (E) Confirmation of deletion events by
921 genomic PCR. From HS/DEX-treated plants (line #4), tdTomato-positive,
922 disorganized tissues ("KO") and tdTomato-negative, thallus-looking sectors [such as
923 the arrow in (D); "non"] were collected from two different individuals and analyzed
924 by genomic PCR using primer sets 1 and 2 shown in the schematic illustration of the
925 construct used (predicted product sizes are indicated) and primer set 3 shown in panel
926 (C).
927

928 **Table 1 Off target analysis of T1 transformants.**

Off target name	Sequence	Mutations detected	Sequenced
ARF1_1_OT1	GAGACCTTCATGATCAG**t**AG_TGG	0	39
ARF1_1_OT2	G**t**GACC**a**TCAT**ag**TCAGGAG_GGG	0	39
ARF1_1_OT3	**t**AGA**t**CTTCATGATCA**a**GA**t**_TGG	0	39
NOP1_1_OT1	G**ga**AGT**ac**TTGTGAGAGAAT_GGG	0	32
NOP1_1_OT2	GAT**ca**TCTTTGT**a**AGAGAA**a**_GGG	0	32
NOP1_1_OT3	**a**ATAG**g**CTTT**a**TGA**a**AGAAT_GGG	0	32

929 Each sequencing analysis was conducted using genomic DNA of lines harboring
930 mutations in on-target sites. Mismatches to the guide sequences are shown in lower
931 bold case.

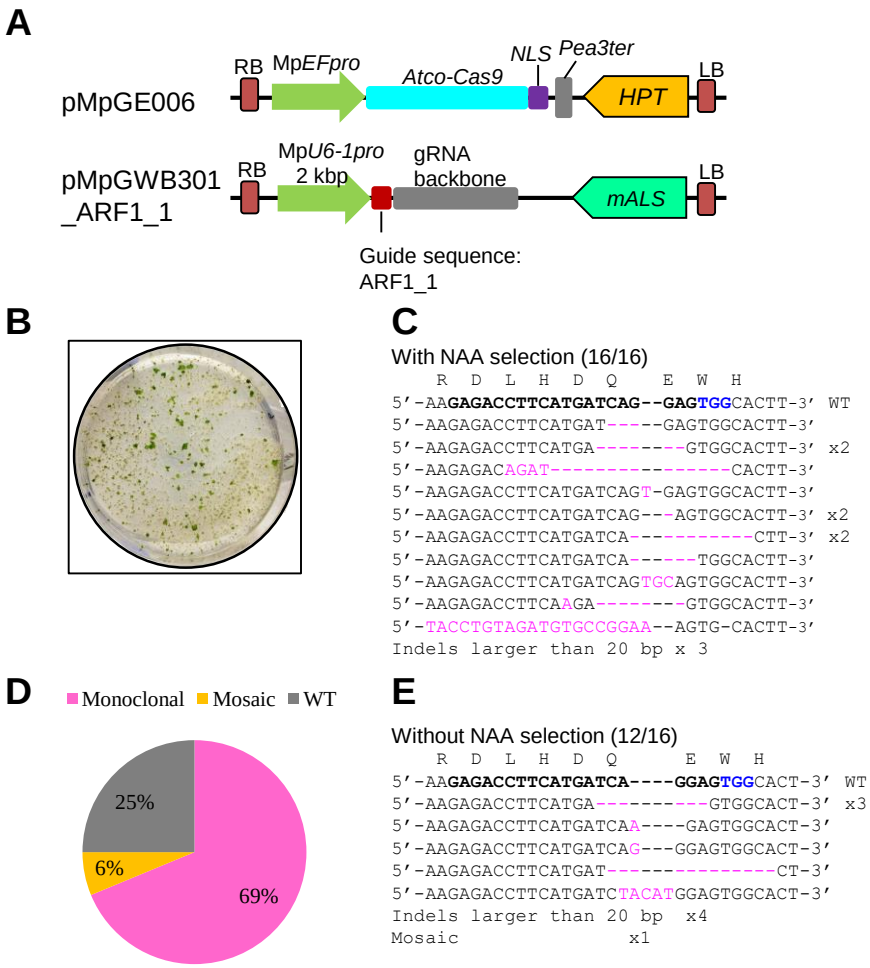


Figure 1.

Shigeo S. Sugano, Ryuichi Nishihama, Makoto Shirakawa, Junpei Takagi, Yoriko Matsuda, Sakiko Ishida, Tomoo Shimada, Ikuko Hara-Nishimura, Keishi Osakabe, and Takayuki Kohchi

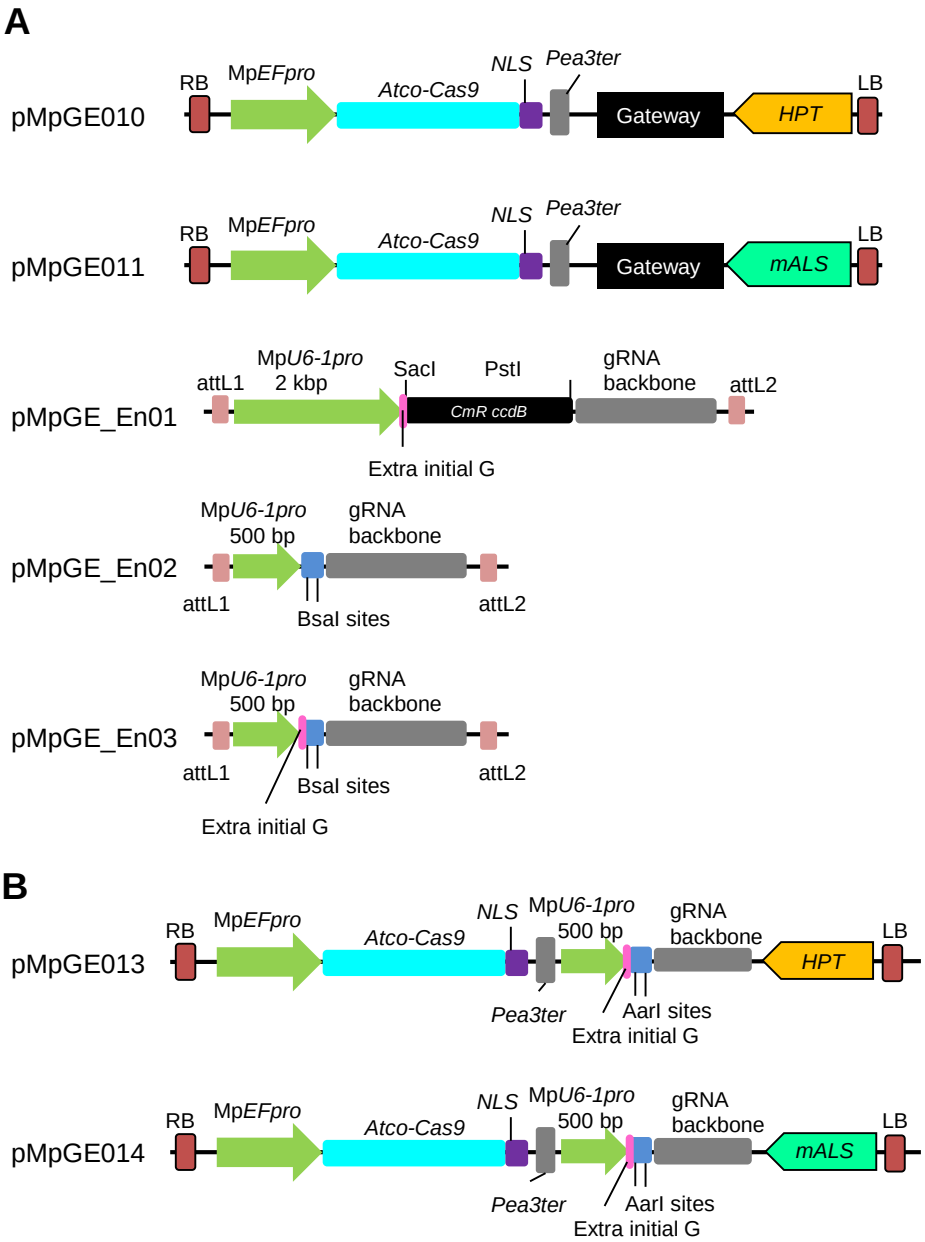


Figure 2.

Shigeo S. Sugano, Ryuichi Nishihama, Makoto Shirakawa, Junpei Takagi, Yoriko Matsuda, Sakiko Ishida, Tomoo Shimada, Ikuko Hara-Nishimura, Keishi Osakabe, and Takayuki Kohchi

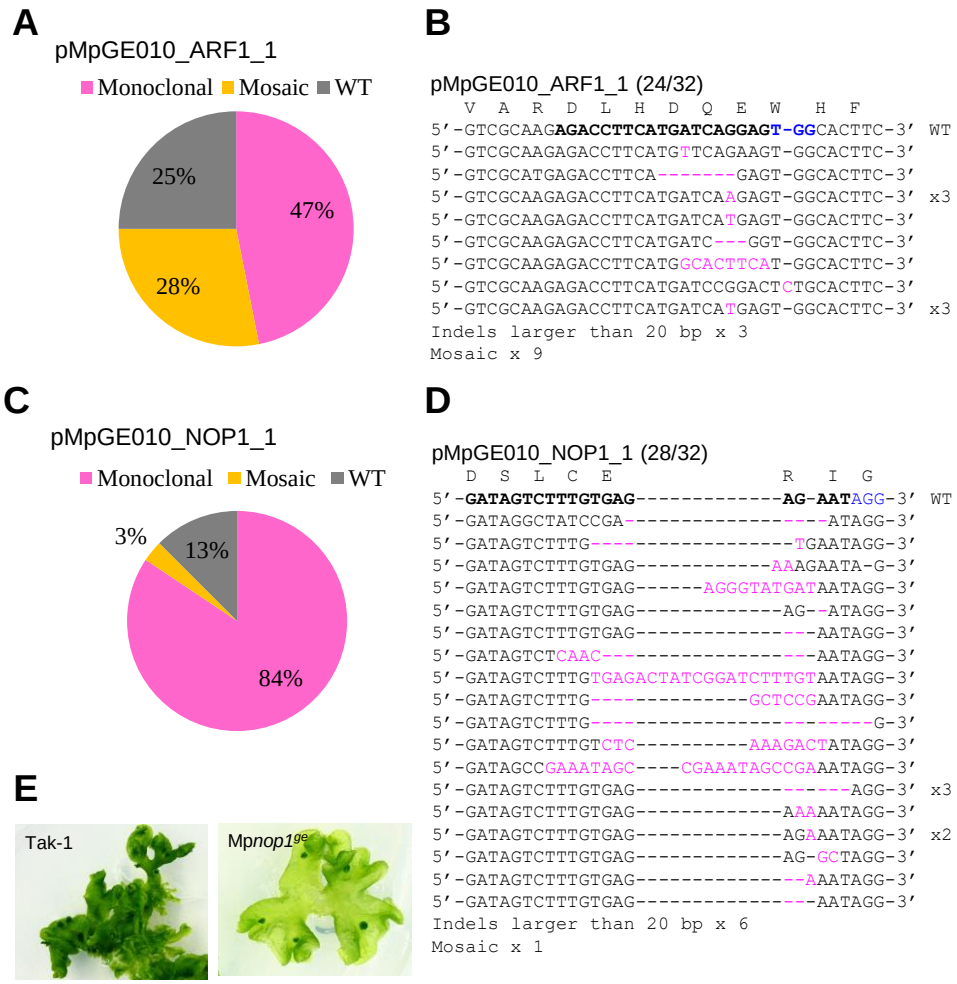


Figure 3.

Shigeo S. Sugano, Ryuichi Nishihama, Makoto Shirakawa, Junpei Takagi, Yoriko Matsuda, Sakiko Ishida, Tomoo Shimada, Ikuko Hara-Nishimura, Keishi Osakabe, and Takayuki Kohchi

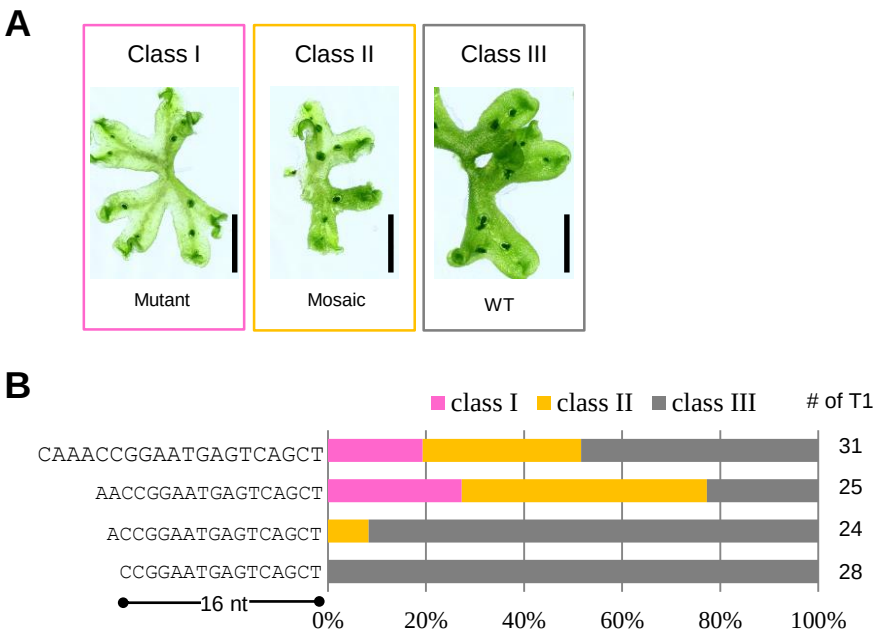


Figure 4.

Shigeo S. Sugano, Ryuichi Nishihama, Makoto Shirakawa, Junpei Takagi, Yoriko Matsuda, Sakiko Ishida, Tomoo Shimada, Ikuko Hara-Nishimura, Keishi Osakabe, and Takayuki Kohchi

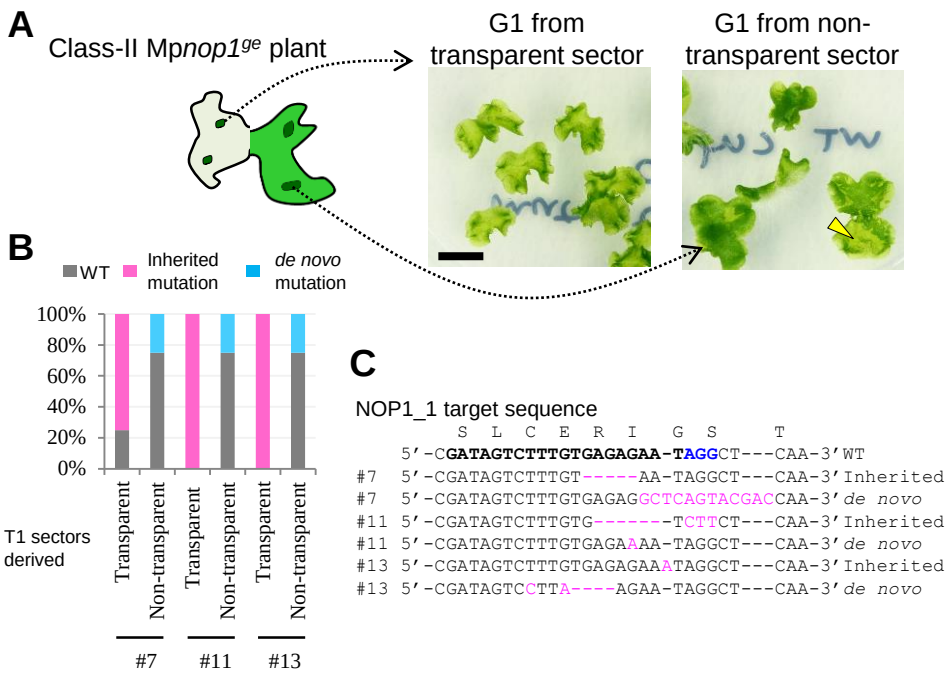
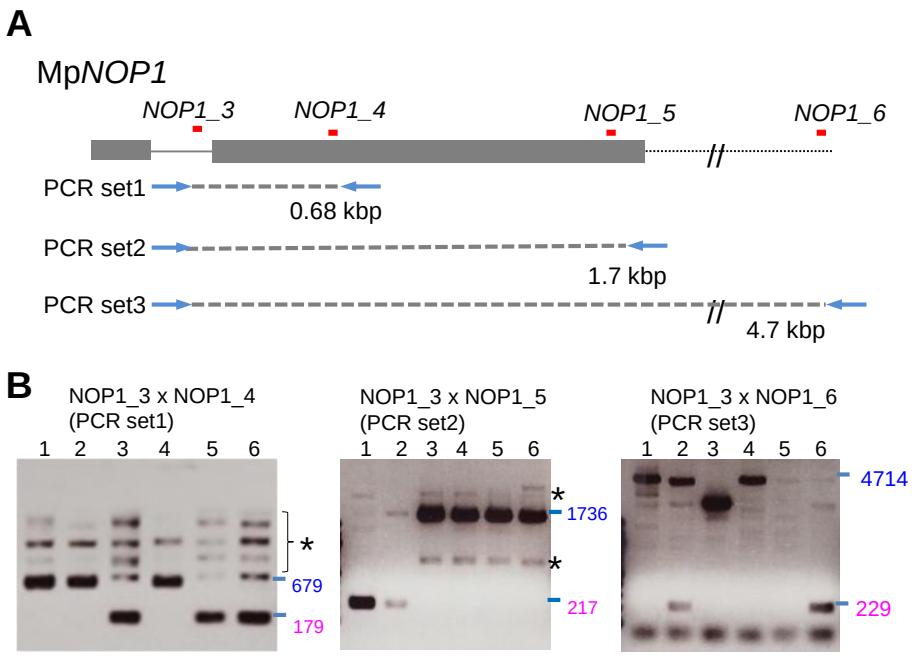


Figure 5.

Shigeo S. Sugano, Ryuichi Nishihama, Makoto Shirakawa, Junpei Takagi, Yoriko Matsuda, Sakiko Ishida, Tomoo Shimada, Ikuko Hara-Nishimura, Keishi Osakabe, and Takayuki Kohchi



C

Vector combination	#1	#2	#3
gRNA in pMpGE013	NOP1_3	NOP1_3	NOP1_3
gRNA in pMpGE014	NOP1_4	NOP1_5	NOP1_6
Large deletion	9/26	4/20	6/23
Expected deletion size	500 bp	1519 bp	4485 bp

Figure 6.

Shigeo S. Sugano, Ryuichi Nishihama, Makoto Shirakawa, Junpei Takagi, Yoriko Matsuda, Sakiko Ishida, Tomoo Shimada, Ikuko Hara-Nishimura, Keishi Osakabe, and Takayuki Kohchi

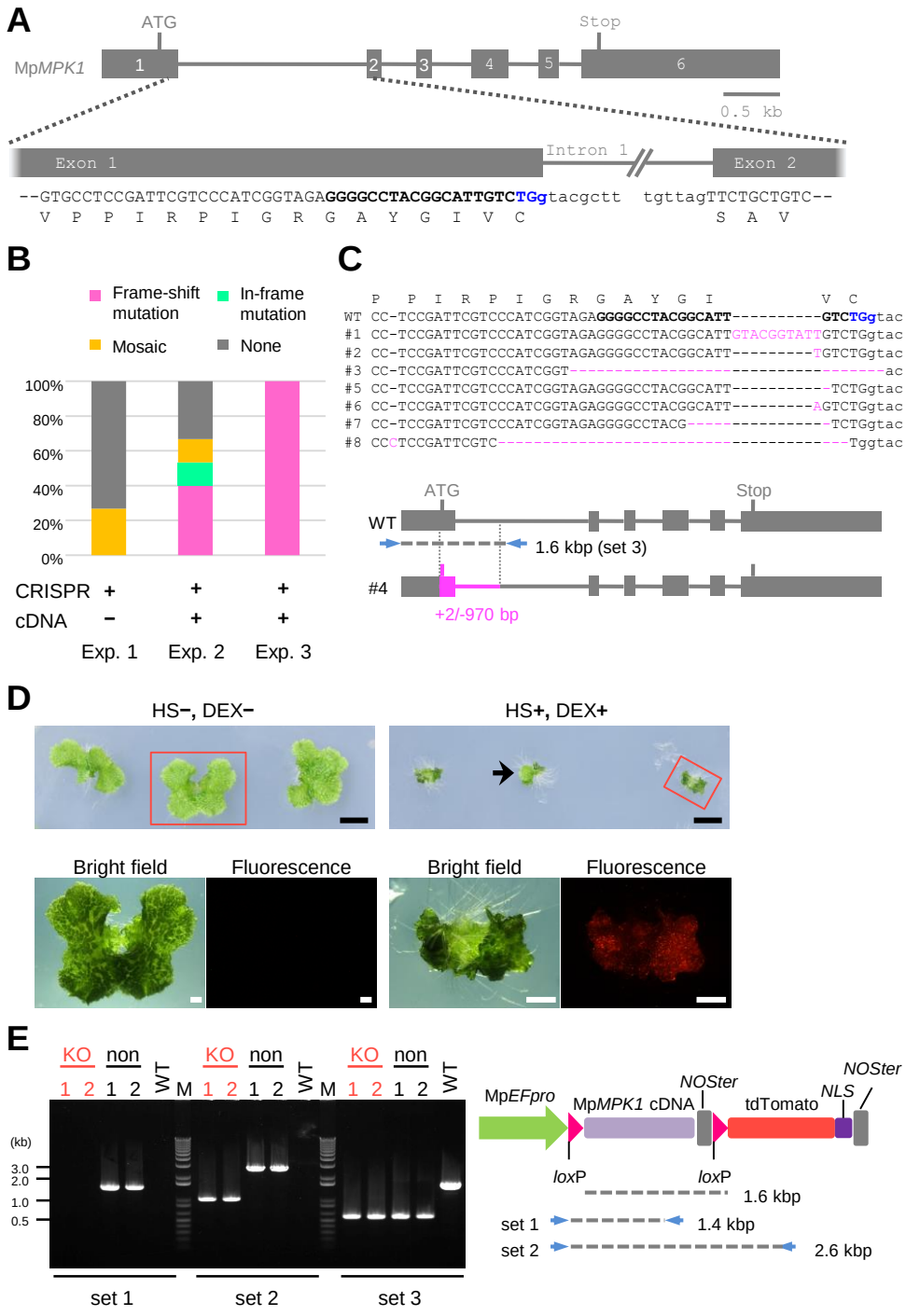


Figure 7.

Shigeo S. Sugano, Ryuichi Nishihama, Makoto Shirakawa, Junpei Takagi, Yoriko Matsuda, Sakiko Ishida, Tomoo Shimada, Ikuko Hara-Nishimura, Keishi Osakabe, and Takayuki Kohchi

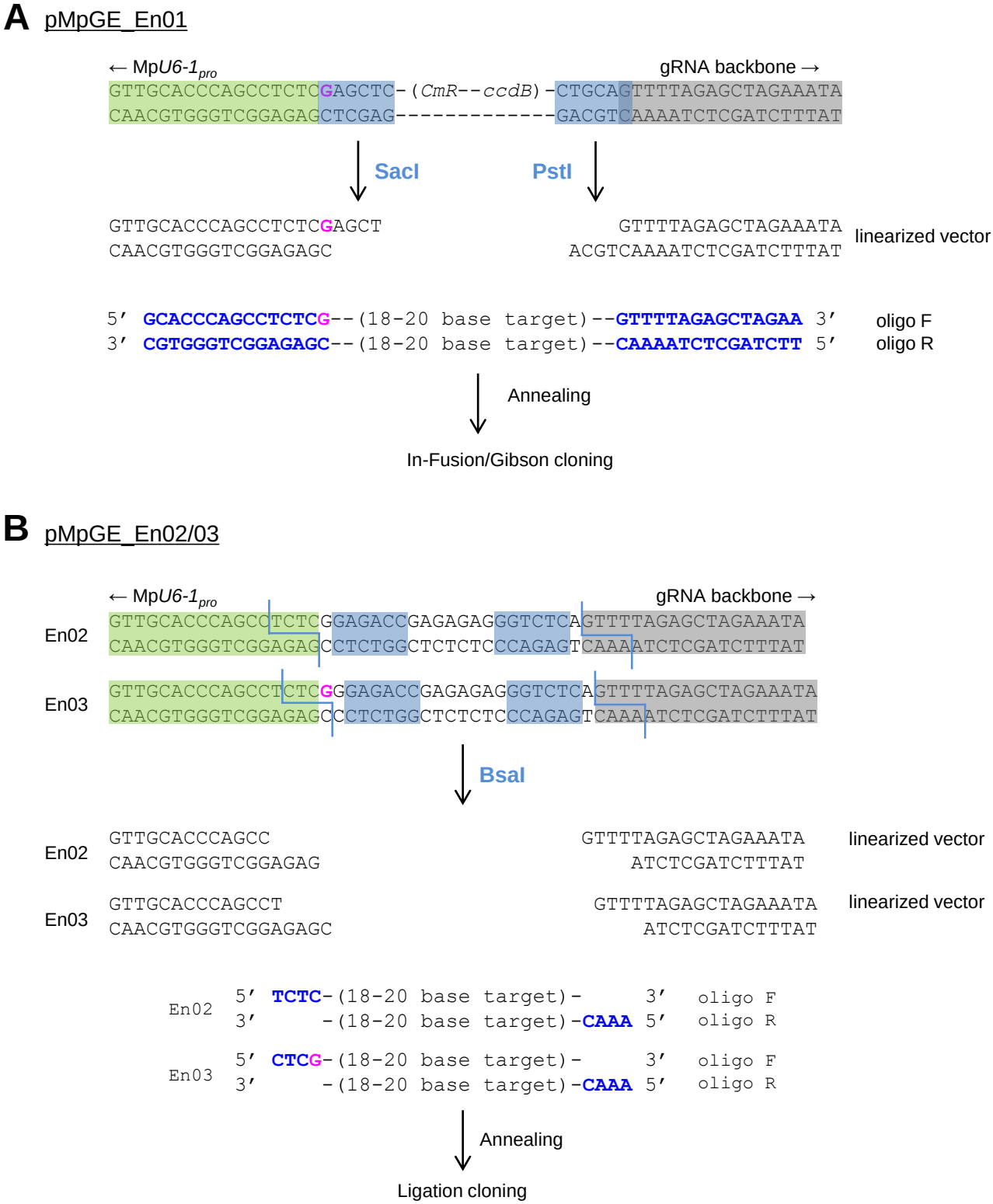


Figure S1. Protocols for gRNA cloning in entry vectors.
(A) pMpGE_En01 is designed to use the In-Fusion/Gibson cloning methods. pMpGE_En01 is digested by SacI and PstI. Two entirely complementary oligo DNAs, which contain at both ends the 15-bp sequences identical to each end of the digested vector and a guide sequence without PAM sequence in between, are annealed and cloned by use the In-Fusion/Gibson reaction. The sense strand of gRNAs should be coded in oligo F. The 'extra initial G' is colored in magenta.
(B) pMpGE_En02/03 are designed to use ligation reactions. pMpGE_En02 or pMpGE_En03 is digested by BsaI, which digests outside of its recognition sites. Oligo F, which contains a sense-strand guide sequence with TCTC at its 5' end, and oligo R, which contains the reverse-complement guide sequence with AAAC at its 5' end, are annealed and cloned by ligation reaction to pMpGE_En02. pMpGE_En02 does not contain an 'extra initial G' and thus requires G or A in the first nucleotide position of a guide sequence for efficient expression. In case of pMpGE_En03, the 'extra initial G' exists in the vector. Therefore, oligo F should contain a sense-strand guide sequence with CTCG at its 5' end for the construction of pMpGE_En03.

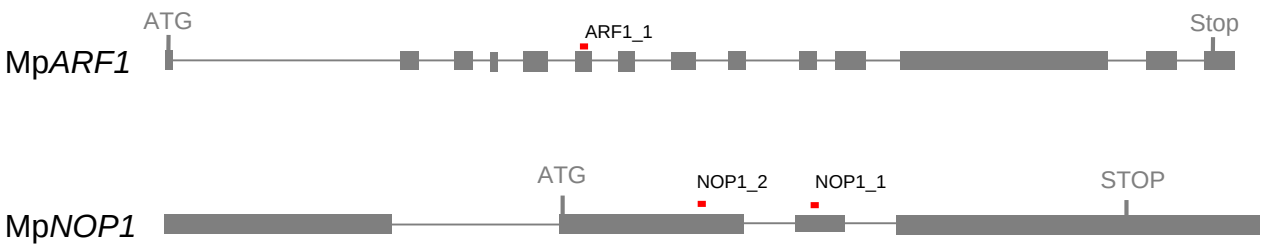


Figure S2. Gene structures of Mp*ARF1* and Mp*NOP1*.
Target sites of the gRNAs used are shown as red lines. Boxes and lines show exons and introns, respectively. "ATG" and "Stop" denote the predicated initiation and termination codons.

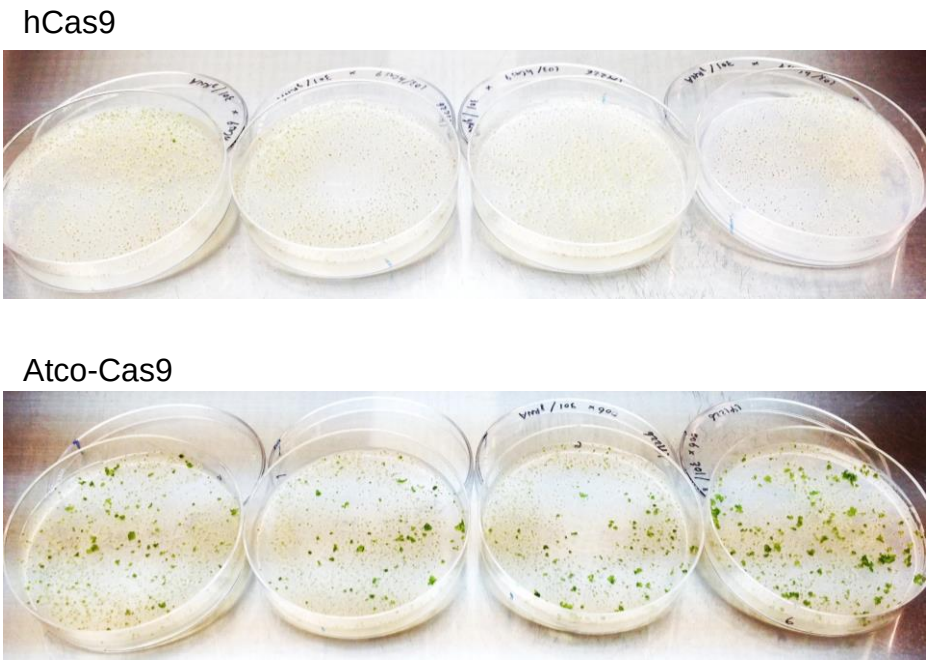


Figure S3. Comparison between *hCas9* and *Atco-Cas9* for genome editing efficiency.
The same Mp*ARF1*-trageting gRNA expression vector (pMpGWB301_ARF1_1) was introduced into sporelings together with either hCas9 expression vector (pMpGWB103-hCas9; top) or Atco-Cas9 expression vector (pMpGE006; bottom) and selected on media containing 10 μM NAA for three weeks. Mutations in the Mp*ARF1* gene are known to cause NAA resistance.

A

pMpGE011_ARF1_1 using pMpGE_En01 (28/32)

```
    H  P  A Q  E  L  V  A  R  D  L  H  D  Q    E  W  H  F  R
5′-CATCCTGCCCAGGAGCTTGTCGCAAGAGACCTTCATGATCA--GGAGTGGCACTTCCGG-3′(WT)
5′-CATCCTGCCCAGGAGCTTGTCGCAAGAGACCTTCAT-------GAGTGGCACTTCCGG-3′ x2
5′-CATCCTGCCCAGGAGCTTGTCGCAAGAGACCTTCATGAT----GAGTGGCACTTCCGG-3′ x2
5′-CATCCTGCCCAGGAGCTTGTCGCAAGAGACCTTCATGATCAGTGGAGTGGCACTTCCGG-3′
5′-CATCCTGCCCAGGAGCTTGTCGCAAGAGACCTTCATGATC-------TGGCACTTCCGG-3′
5′-CATCCTGCCCAGGAGCTTGTCGCAAGTGACCTTCATGATCG--TGAGTGGCACTTCCGG-3′
5′-CATCCTGCCCAGGAGCTTGTCGCAAGAGACCTTCATGATGA--GGAGTGGCACTTCCGG-3′ x2
5′-CATCCTGCCCAGGAGCTTGTCGCAAGAGACCTTCATGATCA--TGAGTGGCACTTCCGG-3′ x8
5′-CATCCTGCCCAGGAGCTTGTCGCAAGAGACCTTCATGATCA---GAGTGGCACTTCCGG-3′
5′-CATCCTGCCCAGGAGCTTGTCGCAAGAGACCTTCATGATCT--TGAGTGGCACTTCCGG-3′
5′-CATCCTGCCCAGGAGCTTGTCGCAAGAGACCTTCATGAT----------CACTTCCGG-3′
Indels larger than 20 bp x 2
Mosaic x 6
```

B

pMpGE010_NOP1_1 using pMpGE_En02 (12/16)

```
   E  L  S  K  S  D  S  L  C  E  R          I  G  S  T  N  G  A
5′-GAGCTCTCGAAATCCGATAGTCTTTGTGAGAG-------AATAGGCTCAACAAACGGAGCTA-3′(WT)
5′-GAGCTCTCGAAATCCGATAGTCTTTGTGAGAG--------CTAGGCTCAACAAACGGAGCTA-3′
5′-GAGCTCTCGAAATCCGATAGTCTTTGTGAGAGGCTCAACAATAGGCTCAACAAACGGAGCTA-3′
5′-GAGCTCTCGAAATCCGATAGTCTTTGTGA----------TAGGCTCAACAAACGGAGCTA-3′
5′-GAGCTCTCGAAATCCGATAGTCTTTGTAA----------TAGGCTCAACAAACGGAGCTA-3′
5′-GAGCTCTCGAAATCCGATAGTCTTTGTGA----------GAGGCTCAACAAACGGAGCTA-3′
5′-GAGCTCTCGAAATCCGATAGTCTTTGTGA----------------------GAGCTA-3′
5′-GAGCTCTCGAAATCCGATAGTCTTTGTGAGAG-------AATAGGCTCAACAAACGGAGCTA-3′ x4
Indels larger than 20 bp x 4
Mosaic x 2
```

Figure S4. Genome editing with different vector combinations.
(A) Mp*ARF1*-targeted mutagenesis with pMpGE011 containing the *gRNA* expression cassette for ARF1_1 derived from pMpGE_En01.
(B) Mp*NOP1*-targeted mutagenesis with MpGE010 containing the *gRNA* expression cassette for NOP1_1 derived from pMpGE_En02.
Inserted or substituted bases are colored in magenta. The target guide sequences are shown in bold face with their PAM sequences in blue.

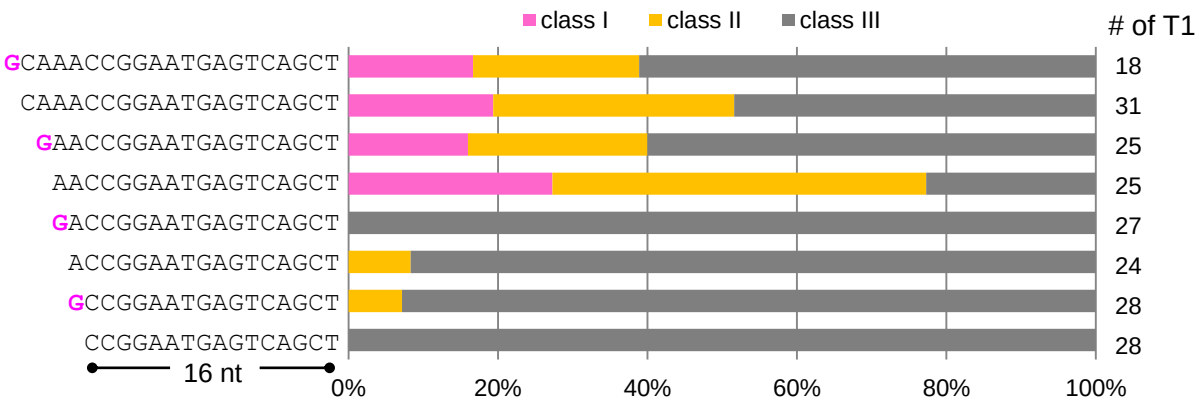


Figure S5. Effects of addition of an 'extra initial G'.
Proportions of mutant phenotype classes (see Fig. 4) in T1 plants transformed with Mp*NOP1*-targeting gRNAs (NOP1_2) of the indicated lengths with or without the 'extra initial G' (magenta). The numbers of T1 plants inspected are shown on the right side of the graph.

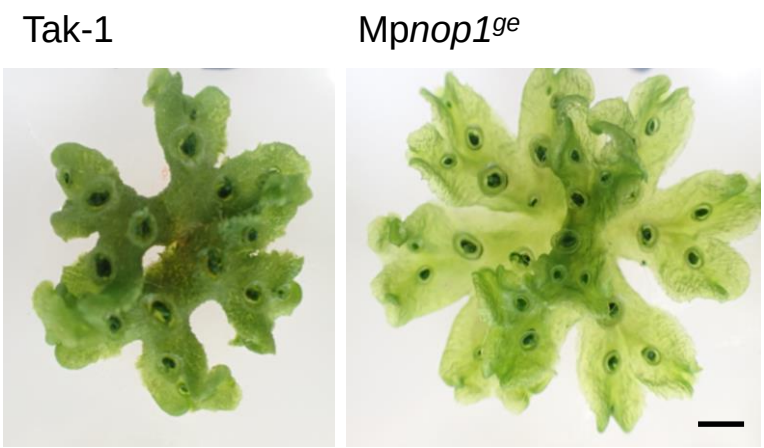


Figure S6. Representative photo of Mp*nop1*^{ge} mutant which harbors 4.5 kbp large deletion
Photos of Tak-1 control plant (Left) and the double transformant, which pMpGE013 with NOP1_3 gRNA and pMpGE014 with NOP1_6 gRNA were transfected (Right). Scale bar = 2 mm.

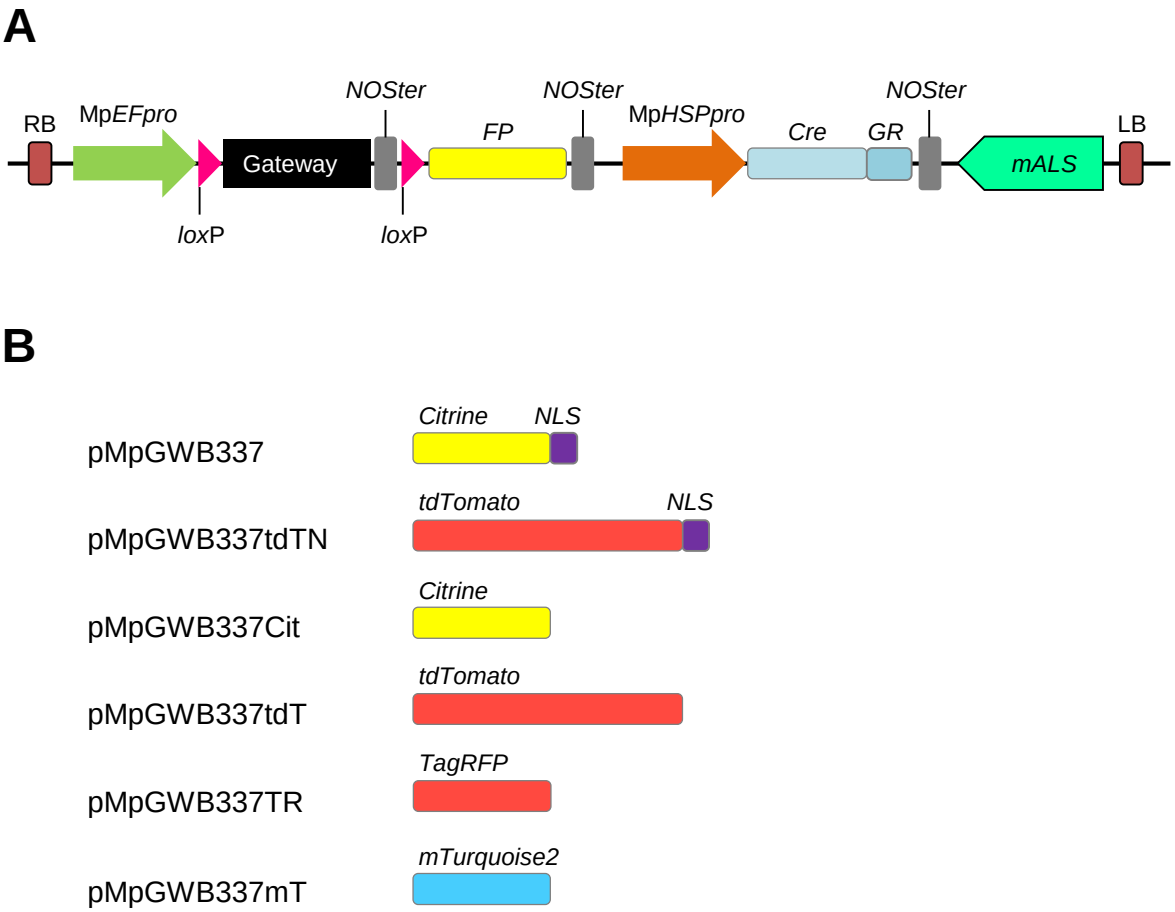


Figure S7. Derivatives of pMpGWB337 with various fluorescent protein markers.
(A) Structure of pMpGWB337 derivatives. Genes for complementation (either cDNA or genomic fragment) can be expressed under the control of MpEF_{pro} by introduction into the Gateway cassette and deleted in plants by heat shock and DEX treatment by virtue of the cassette expressing Cre recombinase fused to the rat glucocorticoid receptor domain (GR) under the control of the Mp*HSP17.8A1* promoter [23]. FP, fluorescent protein coding sequence.
(B) List of fluorescent protein sequences in pMpGWB337 derivatives. NLS, nuclear localization signal.

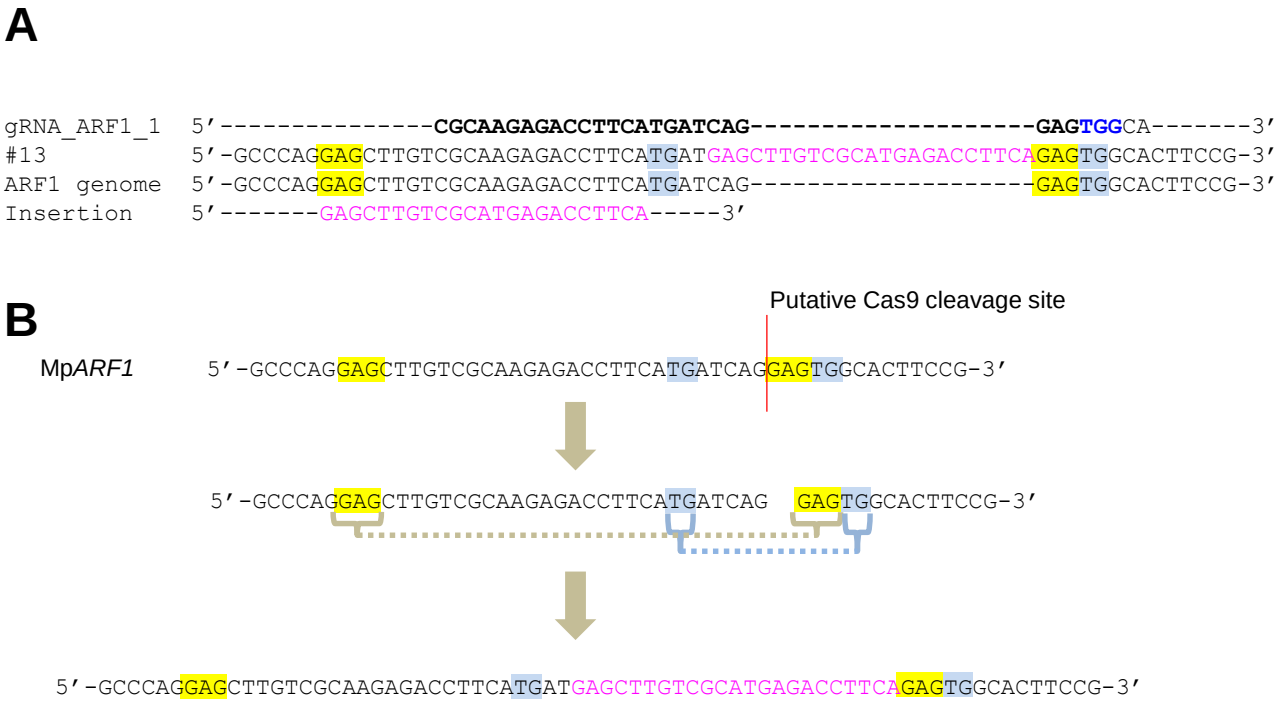


Figure S8. One of the example of microhomology-based repair in *M. polymorpha*.
(A) Alignment of gDNA sequence of #13. Microhomology was highlighted with yellow and blue. (B) Schematic putative repair pathway in the mutant # 13 in pMpGE011_ARF1_1. The sequence was the same as Fig. S4

Supplementary Table S1. Oligos used in this study

Purpose	Name	Sequence	Related figures
Atco-Cas9 cloning	cacc_AtCas9_F	5'-CACCATGGATAAGAAGTACTCTATCGG-3'	Fig. 1
Atco-Cas9 cloning	Pea3Ter_R	5'-AAGCCTATACTGTACTTAACTTGATT-3'	
Gateway cloning site amplification	Infusion_GW_A51_F	5'-GTGGTTGATAACAGCGGTTGACTAGAGTTATCA-3'	Fig. 2
Gateway cloning site amplification	Infusion_GW_A51_R	5'-ATTCGAGCTCTAAGCCTCTAAGCGCTGTTATCA-3'	
MpU6-1 cloning	Mp-U6_38003_F	5'-CACCTATTCATTCAAAAGAGATTTTTAAAGATC-3'	
MpU6-1 cloning	Mp-U6_38003_R	5'-GAGAGGCTGGGTGCAAC-3'	
ccdB site amplification	OE-MpU6-CmRccdB-F2	5'-GCAGAGTTGCACCCAGCCTCTCgagctcATTAGGCACCCCAGGCTTT-3'	
ccdB site amplification	gRNA-R3	5'-tagAAAAAAAGCACCGACTCGGTG-3'	
pMpGE_En02 construction	BsaI-Sp-sgRNA_F	5'-GACCGAGAGAGGGTCTCAGTTTTAGAGCTAGAAAT-3'	
pMpGE_En02 construction	gRNA_R	5'- GTGGCACCGAGTCGGTGCTTTTTTTCTA-3'	
pMpGE_En02 construction	MpU6-1_500_F	5'-GTAACGTGAGACTACTAC-3'	
pMpGE_En02 construction	BsaI_MpU6_1R	5'-GACCCTCTCTCGGTCTCCGAGAGGCTGGGTGCAAC-3'	
pMpGE_En03 construction	Mp_oligo6BsaI_Gf	5'-TCTCGGGAGACCGAGAGAGGGTCTCA-3'	
pMpGE_En03 construction	Mp_oligo6BsaI_Gr	5'-AAACTGAGACCCTCTCTCGGTCTCCC-3'	
pMpGE013 construction	Mp_oligo5AarI_Gf	5'-TCTCGAAATGCAGGTGATGACTCACCTGCATAA-3'	
pMpGE014 construction	Mp_oligo5AarI_Gr	5'-AAACTTATGCAGGTGAGTCATCACCTGCATTTC-3'	
ARF1_1 gRNA	ARF1_sgRNA_F	5'-GCACCCAGCCTCTCGAGACCTTCATGATCAGGAGGTTTTAGAGCTAGAA-3'	Figs. 3, S4
	ARF1_sgRNA_R	5'-TTCTAGCTCTAAAACCTCCTGATCATGAAGGTCTCGAGAGGCTGGGTGC-3'	
NOP1_1 gRNA for pMpGE_En01	NOP1_sgRNA_F	5'-GCACCCAGCCTCTCGATAGTCTTTGTGAGAGAATGTTTTAGAGCTAGAA-3'	Fig. 3
	NOP1_sgRNA_R	5'-TTCTAGCTCTAAAACATTCTCTCACAAAGACTATCGAGAGGCTGGGTGC-3'	
NOP1_1 gRNA for pMpGE_En02	Mp_oligo2_F	5'-TCTCGATAGTCTTTGTGAGAGAAT-3'	Fig. S4
	Mp_oligo2_R	5'-AAACATTCTCTCACAAAGACTATC-3'	
Mp*ARF1* genotyping	ARF1_Seq_F3	5'-GCCGATGTGCATATACCCAGCTATCCCAGT-3'	Figs. 1, 3
	ARF1_Seq_R3	5'-ATGTTATATCCTCGGTTGATTCTCGTACGA-3'	
Mp*NOP1* genotyping	CRISPR_NOP1_F	5'-ATGGAGCAAGTGCGGTTGAGGGCTCTCG-3'	Figs. 1, 3
	CRISPR_NOP1_R	5'-CGTGAGGTGACGATGCCAGTCCGACCAG-3'	
Mp*ARF1* off targets genotyping_OT1	ARF1_1_OT1_F2	5'-CAAACAATGACAGGTGAACAGCG-3'	Table 1
	ARF1_1_OT1_R2	5'-CTTCAATGAGCGTTAGTGCGAGC-3'	
Mp*ARF1* off targets genotyping_OT2	ARF1_1_OT2_F	5'-TCTCGTGACTCGATCAAGATGGG-3'	
	ARF1_1_OT2_R	5'-TTCACTCCCCTCGGCATGGTTTC-3'	
Mp*ARF1* off targets genotyping_OT3	ARF1_1_OT3_F	5'-CATCCACTCCCAGCTGTACCATC-3'	
	ARF1_1_OT3_R	5'-TCCATGGGCCTTAGATTAGGAGG-3'	
Mp*NOP1* off targets genotyping_OT1	NOP1_1_OT1_F	5'-CTCAACCATTGAACCAGCGTCGG-3'	Table 1
	NOP1_1_OT1_R	5'-TTTGTCCAGTTAGGCTGCTCATC-3'	
Mp*NOP1* off targets genotyping_OT2	NOP1_1_OT2_F	5'-TCGGGGCCTGCTCATCGGGCTCC-3'	
	NOP1_1_OT2_R	5'-GCTCAAGCTATGCCGAGGTCGTC-3'	
Mp*NOP1* off targets genotyping_OT3	NOP1_1_OT3_F	5'-TGGATTCGCCGCGTCCTCGTTCC-3'	
	NOP1_1_OT3_R	5'-TATTGGAATCGGCAGGACCGCGG-3'	
gRNA length assessment	NOP1-2_20ntF	5'-tctcCAAACCGGAATGAGTCAGCT-3'	Figs. 4, 5, S5
	NOP1-2_20ntR	5'-aaacAGCTGACTCATTCCGGTTTG-3'	
	NOP1-2_19ntF	5'-tctcAAACCGGAATGAGTCAGCT-3'	
	NOP1-2_19ntR	5'-aaacAGCTGACTCATTCCGGTTT-3'	
	NOP1-2_18ntF	5'-tctcAACCGGAATGAGTCAGCT-3'	
	NOP1-2_18ntR	5'-aaacAGCTGACTCATTCCGGTT-3'	
	NOP1-2_17ntF	5'-tctcACCGGAATGAGTCAGCT-3'	
	NOP1-2_17ntR	5'-aaacAGCTGACTCATTCCGGT-3'	
	NOP1-2_16ntF	5'-tctcCCGGAATGAGTCAGCT-3'	
	NOP1-2_16ntR	5'-aaacAGCTGACTCATTCCGG-3'	
For addition of extra intial G	NOP1-2_20ntFg	5'-ctcgCAAACCGGAATGAGTCAGCT-3'	
	NOP1-2_19ntFg	5'-ctcgAAACCGGAATGAGTCAGCT-3'	
	NOP1-2_18ntFg	5'-ctcgAACCGGAATGAGTCAGCT-3'	
	NOP1-2_17ntFg	5'-ctcgACCGGAATGAGTCAGCT-3'	
	NOP1-2_16ntFg	5'-ctcgCCGGAATGAGTCAGCT-3'	
Larger deletion analysis genotyping	NOP1_05_F	5'-CCTCATGGATTTTATCGC-3'	Figs. 6, S6
	NOP1_1k_R	5'-ACGATGGCACCAGCACT-3'	
	NOP1_2k_R	5'-TCCGACCTTTTGAAACAC-3'	
	NOP1_5k_R	5'-CCAACAATATTCAGCGAC-3'	
For ds oligo DNA	NOP1_3_F	5'-CTCGATTAAGAGTGGAAGTTGCTT-3'	
	NOP1_3_R	5'-AAACAAGCAACTTCCACTCTTAAT-3'	
	NOP1_4_F	5'-CTCGAGCTTCTCCAAGTTCTGGTC-3'	
	NOP1_4_R	5'-AAACGACCAGAACTTGGAGAAGCT-3'	
	NOP1_5_F	5'-CTCGCACGTTCACACGGCCATGGT-3'	
	NOP1_5_R	5'-AAACACCATGGCCGTGTGAACGTG-3'	
	NOP1_6_F	5'-CTCGGAAGATCAAGCATGAATCAA-3'	
	NOP1_6_R	5'-AAACTTGATTCATGCTTGATCTTC-3'	
Mp*MPK1* gRNA	sgRNA_Bsa_MpMPK1_ex1t_F	5'-tctcGGGGCCTACGGCATTGTC-3'	Fig. 7
Mp*MPK1* gRNA	sgRNA_Bsa_MpMPK1_ex1t_R	5'-aaacGACAATGCCGTAGGCCCC-3'	
Mp*MPK1* CRISPR genotyping and set 3	MpMPK1_-549F	5'-GGTCGAACGCACCCTTGCAGC-3'	
Mp*MPK1* CRISPR genotyping	MpMPK1_g540R	5'-CCAAAGTAGAGGCCATGCATGTG-3'	
Mp*MPK1* cDNA	MpMPK1_1F_TOPO	5'-caccATGGATTCCGCAGCAGCTGCCG-3'	
Mp*MPK1* cDNA and set 1	MpMPK1_c1131R_STP	5'-CTATTGCATCATGTCTGGTAGGGG-3'	
Sets 1 and 2	MpEF-P_seqL1	5'-CCCACTTTGGTCAGTCCTGT-3'	
Set 2	tdTomato_753R	5'-GATGACGGCCATGTTGTTG-3'	
Set 3	MpMPK1_g1010R	5'-CGAGACATCAGGGACACGGAAG-3'	
For pMpGWB337 series	ccdB_236F	5'-AAGTGGCTGATCTCAGCCACC-3'	Fig. S7
	loxP_NruI_Sac_R	5'-TTCGAGCTCTTTCGCGATAACTTCGTATAATGTATGC-3'	
	NOSt_head_R_SacI	5'-CAGCATGAGCGAGCTGATTAAG-3'	
	mCherry_CAGC_F	5'-GATCGGGGAAATTCGAGCTC-3'	
	mTurq_CAGC_F	5'-CAGCATGGTGAGCAAGGGCG-3'	
	TagRFP_CAGC_F	5'-CAGCATGGTGAGCAAGGGCGAGG-3'	
	tdTomato_CAGC_F	5'-CAGCATGGTGTCTAAGGGTGAGGAAC-3'	
	pUGW_Aor_mTurq_IF_F	5'-GTGGTTGATAACAGCATGGTGTCTAAGGGTGAGGAAC-3'	
	pUGW_Aor_mTurq_Stp_IF_R	5'-ATTCGAGCTCTAAGCCTATTTGTAAAGCTCATCCATTCCG-3'	