Optimized Cas9 expression systems for highly efficient Arabidopsis genome editing facilitate isolation of complex alleles in a single generation

Jana Ordon, Mauro Bressan, Carola Kretschmer, Luca Dall'osto, Sylvestre Marillonnet, Roberto Bassi, Johannes Stuttmann

Title: Optimized Cas9 expression systems for highly efficient Arabidopsis genome editing facilitate isolation of complex alleles in a single generation

Jana Ordon[1], Mauro Bressan[2], Carola Kretschmer[1], Luca Dall'Osto[2], Sylvestre Marillonnet[3], Roberto Bassi[2], Johannes Stuttmann[1]*

1 – Institute for Biology, Department of Plant Genetics, Martin Luther University Halle-Wittenberg, Weinbergweg 10, 06120 Halle (Saale), Germany

2 - Dipartimento di Biotecnologie, Università di Verona, Strada Le Grazie 15, 37134 Verona, Italy

3 - Department of Cell and Metabolic Biology, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle (Saale), Germany

* Corresponding author: Johannes Stuttmann, Martin Luther University of Halle (Saale), Weinbergweg 10, 06120 Halle, Germany, phone +49-345-5526345, fax +49-345- 5527151, email, ORCID ID orcid.org/0000-0002-6207-094X

Jana Ordon	jana.ordon@gmx.de
Mauro Bressan	mauro.bressan@univr.it
Carola Kretschmer	carola.kretschmer@genetik.uni-halle.de
Sylvestre Marillonnet	smarillo@ipb-halle.de
Luca Dall'Osto	luca.dallosto@univr.it
Roberto Bassi	roberto.bassi@univr.it
Johannes Stuttmann	johannes.stuttmann@genetik.uni-halle.de

Running title: Genome Editing Efficiency in Arabidopsis

Key Words: *Sp*Cas9; Arabidopsis; Promoter; FAST marker; Lhcb1; RPS5a

Acknowledgements

We acknowledge Bianca Rosinsky for taking care of plant growth facilities and growing plants. Ulla Bonas is acknowledged for generous support. This work was funded by GRC grant STU 642-1/1 (Deutsche Forschungsgemeinschaft, DFG) and seed funding by the CRC 648 (DFG) to Johannes Stuttmann. Mauro Bressan and Luca Dall'Osto were supported by University of Verona (Program CooperInt2017 and HuntingLight Ricerca di Base 2015).

Abstract

Genetic resources for the model plant Arabidopsis comprise mutant lines defective in almost any single gene in reference accession Columbia. However, gene redundancy and/or close linkage often render it extremely laborious or even impossible to isolate a desired line lacking a specific function or set of genes from segregating populations. Therefore, we here evaluated strategies and efficiencies for the inactivation of multiple genes by Cas9-based nucleases and multiplexing. In first attempts, we succeeded in isolating a mutant line carrying a 70 kb deletion, which occurred at a frequency of ~1.6% in the T_2 generation, through PCR-based screening of numerous individuals. However, we failed to isolate a line lacking *Lhcb1* genes, which are present in five copies organized at two loci in the Arabidopsis genome. To improve efficiency of our Cas9-based nuclease system, regulatory sequences controlling Cas9 expression levels and timing were systematically compared. Indeed, use of *DD45* and *RPS5a* promoters improved efficiency of our genome editing system by approximately 25-30-fold in comparison to the previous ubiquitin promoter. Using an optimized genome editing system with *RPS5a* promoter-driven Cas9, putatively quintuple mutant lines lacking detectable amounts of Lhcb1 protein represented approximately 30% of T_1 transformants. These results show how improved genome editing systems facilitate the isolation of complex mutant alleles, previously considered impossible to generate, at high frequency even in a single (T_1) generation.

Introduction

Sequence-specific nucleases (SSNs) based on *Sp*Cas9, which derives from the *Streptococcus pyogenes* CRISPR/Cas system, are currently the most commonly used tool for genome editing in plants and animals (reviewed in Ceasar et al., 2016). *Sp*Cas9 (hereafter termed Cas9) and related RNA-guided nucleases (RGNs) are directed to DNA target sequences by a guide RNA incorporating into the nuclease protein. The guide RNA may consist of a chimera from base-pairing between a CRISPR RNA (crRNA) and a trans-activating RNA (tracrRNA), or crRNA and tracrRNA may be collapsed into a single guide RNA (sgRNA; Jinek et al., 2012). In either case, the variable stretch of crRNA/sgRNA base-pairs with complementary target DNA sequences flanked by a protospacer-adjacent motif (PAM, NGG for unmodified Cas9). Cas9 functions in its native configuration as a nuclease and cleaves target sequences, but can be considered as a programmable DNA-binding scaffold for tethering diverse activities or functionalities to precise chromatin positions. As such, e.g. transcriptional regulators, chromatin modifiers or base editors were constructed on the basis of catalytically inactive Cas9 (*d*Cas9), or fluorescent protein fusions were exploited for live cell imaging (Chavez et al., 2016; Chen et al., 2016; Dominguez et al., 2016; Komor et al., 2016; Ren et al., 2018).

In the nuclease mode, Cas9 generates either blunt-ended double-strand breaks (preferentially three base pairs upstream of the PAM sequence), or single-strand breaks when converted to a nickase (Jinek et al., 2012; Ran et al., 2013). Gene targeting may be achieved from both types of lesions, but remains technically challenging, at least in plants (Fauser et al., 2014; Shi et al., 2017). In contrast, the disruption of genes through error-prone repair of double strand breaks by non-homologous end-joining (NHEJ) is now routinely used for reverse genetics approaches in many different plant systems (reviewed in Malzahn et al., 2017). When using RGNs for precision gene editing in crop improvement, the specificity of the enzyme may be of major importance. Indeed, delivery of ribonucleoprotein complexes has been reported to minimize RGN off-target activity, and is also preferable to avoid regulation of the final product (Woo et al., 2015; Huang et al., 2016; Wolt et al., 2016; Zhang et al., 2016; Liang et al., 2017). In contrast, *Agrobacterium*-facilitated transformation and *in planta* expression yet remains the most important approach for RGN delivery in basic research. Efficiency in this case mainly depends on timing and levels of expression of nuclease and sgRNA and efficient nuclear import. Although high expression levels may increase off-target activity, these effects can be mitigated, e.g., through analysis of multiple alleles, and efficiency is generally at prime. In some species, as e.g. rice, high efficiencies for genome editing regularly approaching 100% in T_0 plants (with Cas9 or Cpf1) were reported, suggesting that further optimization is not required in this respect (Mikami et al., 2015; Tang et al., 2017). However, especially in Arabidopsis, the genetic analyses workhorse, genome editing efficiencies are often comparably low, and severely vary between different studies. In the Arabidopsis system, transformation does not depend on somatic embryogenesis, as T-DNAs are directly delivered to female ovules during floral dip transformation (Ye et al., 1999; Desfeux et al., 2000). Accordingly, T-DNA-encoded SSNs are subsequently expressed (or not) in different cell types of the developing embryo, and expression levels and timing will be decisive for efficiency and germ line entry of genome modifications. Indeed, exceptionally high genome editing efficiencies were reported when *RPS5a* or *DD45* promoters (or derivatives) were used for Cas9 expression (Wang et al., 2015; Mao et al., 2016; Tsutsui and Higashiyama, 2017). These effects were mainly attributed to activity of these promoters in early embryogenesis. However, the cross-study comparison of genome editing efficiencies is of limited validity, as differences between constructs go beyond the use of a particular promoter. Especially the sgRNA/target site represents a major variable affecting genome editing efficiency, and also further differences between vectors may have profound and unexpected consequences.

Here, we systematically compared regulatory elements in order to determine improved Cas9 expression systems for Arabidopsis genome editing. While a previously used *ubiquitin* promoter-driven Cas9 produced mutants in the T_2 generation and at moderate frequencies only, optimized expression systems were highly efficient in the T_1 generation. Indeed, this enabled us to isolate a putative quintuple mutant lacking detectable amounts of Lhcb1 protein in a single generation and at high frequencies, while we had failed to

isolate this mutant line prior to system optimization. This shall guide further optimization of Arabidopsis genome editing systems, and also researchers in their future choice of system.

Material and Methods

Plant material and transformation

Arabidopsis accession Landsberg *erecta* (L*er*), the *old3-1* mutant line (Tahir et al., 2013), accession Columbia and the NoMxB3 quadruple mutant line were used. The NoM line is published (Dall'Osto et al., 2017), and a T-DNA insertion in At5g54270 (*Lhcb3*; NASC N520342) was introgressed to generate the NoMxB3 quadruple mutant. Plants were grown in soil at a 19°C: 22°C night: day cycle (200 µE/m^2*s, 60% relative humidity) under short (16h dark, 8h light) or long (8h dark, 16h light) day conditions. To suppress autoimmunity, plants were grown at a 26°C: 28°C night: day cycle in a growth cabinet (200µE/m^2*s, 60% relative humidity) with either short day or long day conditions. Plants were transformed by floral dip as described (Logemann et al., 2006), and *old3-1* plants were cultivated at 28°C to suppress autoimmunity for transformation.

Molecular cloning

Golden Gate technology (Engler et al., 2008) was used for all cloning tasks, and the Modular Cloning and Plant Parts toolkits were used (pICH/pAGM/pICSL vectors; Engler et al., 2014). Generally, 20 fmol of DNA modules were used for Golden Gate reactions with either *Bsa*I, *BsmB*I or *Bpi*I and T4 DNA Ligase. Reactions were carried out in a PCR cycler (2 min 37°C, 5 min 16°C, 10-55 cycles; terminated by 10 min 50°C and 10 min 80°C steps), and transformed either into *E. coli* TopTen or ccdB survival II cells (Invitrogen; distributed by Thermo Fisher). To generate the adaptable nuclease activity reporter (pJOG367), a linker sequence was appended to a ccdB cassette (lacking *BsmB*I and *Bsa*I sites) by PCR amplification, and subcloned in a custom cloning vector (pJOG397) to yield pJOG395. Similarly, an amplicon of a *GUS* gene with introns (Engler et al., 2014) and a linker sequence was subcloned to yield pJOG396. These modules were subsequently assembled together with a 35S promoter (pICH51277) and an *ocs* terminator (pICH41432) in a Level 1 recipient (pICH47732) to yield pJOG367. Target sequences are inserted in this adaptable reporter scaffold as hybridized oligonucleotides by a *BsmB*I Golden Gate reaction (Online Resource 1). Promoter fragments were amplified by PCR, internal *Bsa*I and *Bpi*I restriction sites domesticated and subcloned into pICH41295. The previously reported *rbcS E9* terminator fragment (Wang et al., 2015) was amplified from *Pisum sativum* genomic DNA and subcloned into pICH41276. All genome editing constructs were assembled as previously described (Ordon et al., 2017) and essentially following the Modular Cloning strategy (Weber et al., 2011). A

previously published *h*Cas9 coding sequence (Belhaj et al., 2013) was used for assemblies, and sgRNAs were driven by a 90 nt promoter fragment of Arabidopsis U6-26, as previously described (Ordon et al., 2017). Further details and oligonucleotide sequences are available upon request. Genbank files for most important modules and nuclease vectors are provided in Online Resource 2.

sgRNA selection and deletion screening

A local instance of CasOT (Xiao et al., 2014) was used to identify specific sgRNAs for generation of the Δ*dm2a-g* mutant. Sequence windows flanking the *DM2a* and *DM2g* genes, respectively, were defined for selection of targets, and a PacBio assembly of the Landsberg *erecta* genome was used as reference (https://www.pacb.com). Specific sgRNAs were subsequently evaluated with the sgRNA designer tool (Doench et al., 2014; Doench et al., 2016) to select those with highest predicted activity. For editing of the *DM2h* gene (promoter comparison), the target sites TGATTTCTGCTAATTCATCAAGG and TTATTGATAATAATATAGAGAGG were selected. ChopChop (Labun et al., 2016) was used for selection of target sites within *Lhcb1* genes, and potential sites were further manually curated and selected. The following target sites were used for sgRNA design: CGCGGCAGTTCGGTCCGCCAAGG [1], GCCGACCTGCCGCCTAATTGTGG [2], CACTGCAGAGATATTGAACGAGG [3], GTTATATAATGCTTGATGGATGG [4] for the Δ*dm2a-g* deletion, and GGTTCACAGATCTTCAGCGACGG [1], ATGGACCCAAGTACTTGACTCGG [2], TGTGGATAACTTCTAGCTCACGG [3], GGCTACTCAAGTTATCCTCATGG [4], GAAGCGGCCGTGTGACAATGAGG [5], AGAAGTTATCCACAGCAGGTGGG [6], GAGGACTTGCTTTACCCCGGTGG [7], AGGGGAGGAGAGAGCCATTGTGG [8] for *Lhcb1* genes. Oligonucleotides JS1382/83 (TGCAGCTGAAGATCATGGC/GACTAGCGATTGTGTCCATC) were used for detection of a *Δdm2a-g* deletion allele. Oligonucleotides MB1/MB4 ("PCR *Lhcb1.1/.2./.3*"; AAAGCCTCTGGGTCGGTAGCA/TCTGGGTCGGTAGCCAAACCC) and MB6/MB7 ("PCR *Lhcb1.4/.5"*; CCGGCGACTCTGTAGCCCTCA/TCCGGCGACTCTGTAGCCTTC) were used to screen for deletions at *Lhcb1* loci.

Transgenic plant selection and estimation of genome editing efficiencies

For seed fluorescence-based selection using the FAST marker (Shimada et al., 2010), T_1 or T_2 seeds were spread on moist Whatman paper, and observed under a motorized SteREO Discovery.V12 microscope (Zeiss) connected to an AxioCam MRc camera. Pictures were taken under bright field conditions or UV illumination and using an RFP filter set. For selection of transgenic plants from *old3-1* transformations, seedlings were grown at 28°C, treated 3 – 4 times with BASTA and resistant plants transferred to new soil. For promoter comparison, T_1 seedlings were transferred to 22°C, onset of autoimmunity was scored after 8d, and

transgenic plants transferred back to 28°C to obtain T_2 seeds. T_2 seeds were directly grown at 22°C in short day conditions, and phenotypes scored 20 dag to calculate editing frequencies in individual T_2 families. Plants were transformed with a construct conferring Hygromycin resistance (pDGE277) for generation of an *lhcb1* mutant line. Seeds were surface-sterilized, grown on MS 1/10 plates containing 0,5% sucrose, Hygromycin B (25 µg/ml) and Carbenicillin (100 µg/ml), and Hygromycin-resistant seedlings selected 15 dag.

Gel electrophoresis, immunoblotting and sample preparation

SDS-PAGE analysis was performed with the Tris-Tricine buffer system (Schagger and von Jagow, 1987). For immunodetection, samples corresponding to 0.5 µg of Chlorophyll were loaded for each sample and electroblotted on nitrocellulose membranes. Proteins were detected with alkaline phosphatase-conjugated secondary antibodies purchased from Sigma-Aldrich (A3687). Primary antibodies used were α-Lhcb1 (AS01 004), α-PsaA (AS06 172), α-PsbB/CP47 (AS04 038) from Agrisera.

Transient genome editing efficiency assays

For GUS-based nuclease assays, an sgRNA directed against the target site TATATAAACCCCCTCCAACCAGG was used. This target site was inserted into the adaptable reporter plasmid (pJOG367). *Agrobacterium* strains containing reporter or nuclease-encoding constructs were infiltrated alone or in a 1:1 ratio at an OD_{600} = 0.6 into leaves of four different *N. benthamiana* plants. Tissue samples of individual plants were treated as replicates, and GUS activity was determined as previously described (Ordon et al., 2017). GUS activity was normalized to the reporter alone, which was arbitrarily set to 1.

Results

Generation of complex alleles with *ubiquitin* promoter-driven Cas9 in Arabidopsis

We had previously developed a Golden Gate cloning-based toolkit for highly multiplexed genome editing in dicotyledonous plants (Ordon et al., 2017). In this system, expression of sgRNAs is driven by the Arabidopsis U6-26 promoter, and several different promoters were provided for Cas9 expression. Using *ubiquitin* promoter-driven Cas9 (p*PcUbi4-2*, from parsley), a 120 kb deletion encompassing the *DM2* cluster of *Resistance* genes in accession Landsberg *erecta* (L*er*) was previously generated (Ordon et al., 2017). The respective *⊿dm2* deletion allele could be phenotypically selected, and occurred at low frequencies only (~

0.5%) among T_2 individuals. To evaluate strategies for and feasibility of generating large deletion alleles not linked to a phenotype, we attempted deletion of ~ 70 kb within the *DM2*^{Ler} cluster containing all genes of the cluster except *DM2h* (Fig. 1a). The *DM2*^{Ler} cluster encodes 7-8 complete or truncated *Resistance* genes (*DM2a-h*, or *RPP1-like*^{Ler} *R1-R8*) most similar to *RPP1*, conferring resistance to the obligate biotrophic oomycete *Hyaloperonospora arabidopsidis* in accession Wassilijewska (Botella et al., 1998; Alcazar et al., 2009; Chae et al., 2014). The function(s) of the *DM2*^{Ler} locus remain yet unknown, but one or several *DM2*^{Ler} genes provoke constitutive activation of immune responses (autoimmunity) when combined in a single genetic background with different "inducers": Alleles of *STRUBBELIG-RECEPTOR FAMILY 3* originating from the South Asian accessions Kashmir and Kondara (*SRF3*$^{Kas/Kond}$; Alcazar et al., 2010), a transgene encoding for ENHANCED DISEASE SUSCEPTIBILITY1 fused to YFP and an SV40 nuclear localization signal (EDS1-YFP^{NLS}; Stuttmann et al., 2016) or the *onset of leaf death3-1* allele affecting a cytosolic O-acetylserine(thiol)lyase (*old3-1;* Tahir et al., 2013). *The DM2h* gene is required for all cases of autoimmunity, but the contribution of additional *DM2* genes remains unknown (Alcazar et al., 2014; Stuttmann et al., 2016). The *Δdm2a-g* deletion was conducted in the L*er old3-1* background, and plants were grown at 28°C to suppress temperature-sensitive autoimmunity and seedling necrosis in this line (Tahir et al., 2013).

For deletion of *DM2a-g*, a construct containing Cas9 driven by the pPcUbi4-2 promoter and four sgRNAs directed against sites flanking the targeted region was transformed into L*er old3-1* plants (Fig. 1a). T_1 plants were selected by BASTA resistance, and T_2 populations consisting of 4-5 T_1 plants assembled. 96 DNA pools containing 7-11 T_2 plants each (equivalent of ~ 850 T_2 plants) were screened by PCR with oligonucleotides flanking the targeted region for occurrence of a *Δdm2a-g* deletion allele (Figs. 1a,b). A clear signal of the expected size was detected in 14/96 DNA pools, representing 15% of pools or approximately 1.6% of T_2 plants considering a single line with a deletion allele in each PCR-positive pool. Selected pools were deconvoluted, and single plants screened (Fig. 1c). For some pools, only a weak signal corresponding to the deletion allele was detected among single plants (pool #8 in Fig. 1c), but most pools contained 1-2 plants positive for the deletion allele (pool #89 in Fig. 1c). T_3 seeds were obtained for these single plants, and PCR-screened for segregation of the *Δdm2a-g* allele and Cas9. From analysis of 96 T_3 plants, several lines homozygous for the *Δdm2a-g* allele (absence of *DM2c*, boxed lanes in Fig. 1d) could be isolated, but all still contained the genome editing transgene, as detected by presence of Cas9. Outcrossing of the Cas9 construct is most likely not required for most experimental settings. However, sequencing of the *Δdm2a-g* deletion allele in one of the isolated homozygous lines revealed that it still contained an intact sgRNA target site (Fig. 1e), suggesting it might not be fully stable in subsequent generations. Summarizing, PCR-based isolation of large deletion alleles is feasible with the used genome editing system (containing pPcUbi-driven Cas9), but deletions occur at low frequency, and isolation requires extensive screening.

Furthermore, we attempted to generate a mutant lacking Lhcb1 (chlorophyll a/b binding protein 1, CAB), a major subunit of light-harvesting complex II (LHCII) and the most abundant membrane protein in nature (Galka et al., 2012; Su et al., 2017). The *Arabidopsis thaliana* genome contains five *Lhcb1* genes, which are organized in tight linkage groups on chromosomes 1 and 2 (Fig. 1f). For inactivation of *Lhcb1* genes, a genome editing construct (with pPcUbi4-2-driven Cas9) containing eight sgRNAs was constructed, and transformed into Columbia (Col) wild type plants. *Lhcb1* genes in each linkage group share high sequence homology, and sgRNAs were designed to target four sites with perfect match in each *Lhcb1* gene (Fig. 1f). Thus, genome editing activity might produce SNPs within individual genes, or larger deletions between target sites. As for *⊿dm2a-g*, T_2 plants were PCR-screened for occurrence of larger deletions with flanking oligonucleotides. From screening ~ 200 T_2 plants, no line PCR-positive for a deletion in either of the linkage groups could be detected. Also, none of ~ 400 plants visually inspected showed the pale green phenotype expected for lines with reduced Lhcb1 levels (Pietrzykowska et al., 2014). We concluded that efficiency of our genome editing system was not sufficient for convenient isolation of complex or highly multiplexed alleles.

Improved genome editing efficiencies through optimized Cas9 expression systems

To improve efficiency of our genome editing system, we focused on regulatory elements controlling Cas9 expression. The *rbcS E9* terminator from pea (*Pisum sativum*) was previously described to positively affect genome editing efficiencies in comparison to the *nos* terminator (*nopaline synthase*; *Agrobacterium tumefaciens*) in several independent constructs and transformations (Wang et al., 2015), suggesting stabilization of the respective mRNA. To more generally test the importance of transcriptional terminators for Cas9 activity, we wanted to compare genome editing efficiencies of nuclease constructs differing only in 3'UTR sequences and transcriptional terminators for Cas9 expression. Since mRNA stabilization should not be strictly limited to a particular plant system, genome editing efficiencies were compared in quantitative, transient assays in *Nicotiana benthamiana* (*N. benth.*). For this, an adaptable genome editing efficiency reporter was first constructed (Online Resource 1). This adaptable reporter allows insertion of user-defined target sequences in a linker region of a *ß-glucuronidase* (*uidA*) gene. Insertion of a target sequence disrupts the *GUS* reading frame, which can be restored by RGN-mediated cleavage and repair through the NHEJ pathway (Online Resource 1). Transient co-expression of reporter constructs and corresponding nucleases in *N. benth.* faithfully restored GUS activity, and reporter/nuclease combinations showed variable GUS activities potentially reflecting sgRNA efficacy (Online Resource 1). Nine minimalistic nuclease constructs containing only the sgRNA transcriptional unit and 35S promoter-driven Cas9 with different transcriptional terminators were quantitatively compared for genome editing efficiency (Fig. 2a). Across multiple biological

replicates, no significant and reproducible differences were measured (Fig. 2b). However, high genome editing efficiencies were obtained in all replicates when using the *rbcS E9* terminator. This corroborates the previous notion that the *rbcS E9* terminator is well-suited for Cas9 expression even when combined with *h*Cas9 (human codon-optimized), in contrast to previously used *z*Cas9 (*Zea mays* codon-optimized; Wang et al., 2015). Furthermore, this is in line with potential masking of the beneficial effects of this terminator by expression of Cas9 by strong constitutive promoters (Wang et al., 2015). Since no other terminator out-competed *rbcS E9*, this was used for further experiments.

Next, different promoters were tested for their effect on genome editing efficiency. As far as we are aware, the 35S promoter (e.g. Feng et al., 2013; Ordon et al., 2017), several different *ubiquitin* promoters (e.g. Mao et al., 2013; Fauser et al., 2014; Peterson et al., 2016), egg cell-specific promoters or derivatives (*DD45* and EC1.2-EC1.1; Wang et al., 2015; Mao et al., 2016), promoters of *APETALA1* (*AP1*; Gao et al., 2015), *INCURVATA2* (*ICU2*; Hyun et al., 2015), *YAO* (Yan et al., 2015), *SPOROCYTELESS* and *LAT52* (Mao et al., 2016), MGE1/2/3 (Eid et al., 2016), *HISTONE H4* and *EF1α* (Osakabe et al., 2016) and *RIBOSOMAL PROTEIN S5a* and *WUSCHEL RELATED HOMEOBOX 2* (*RPS5a* and *WOX2*; Tsutsui and Higashiyama, 2017) were previously used to drive Cas9 expression for Arabidopsis genome editing. The occurrence of homozygous mutants in the T_1 generation at high frequencies was mainly reported for the *RPS5a* promoter and the egg cell-specific *DD45* promoter or derivatives (Wang et al., 2015; Mao et al., 2016; Tsutsui and Higashiyama, 2017), and also for the meiosis-specific MGE1 promoter (Eid et al., 2016). We decided to focus on the *RPS5a* and *DD45* regulatory elements, and to compare these with several popular promoters not reported to generate mutations in T_1 generation (p35S, pPcUbi, p*AP1*, p*ICU2*) and two further promoters not previously employed for genome editing (p*GILT*, p*ALB*; see Online Resource 2 for sequence details for used promoter fragments).

Promoter fragments were used for driving Cas9 expression (terminated by t*rbcS E9*) in genome editing vectors containing two sgRNA transcriptional units, a BASTA resistance cassette and also the FAST marker for seed coat fluorescence-based identification of transgenic plants (Fig. 3a; Shimada et al., 2010). The positioning and orientation of the FAST cassette was altered in comparison to vectors we had previously generated and containing this element (Ordon et al., 2017). Although transgenic plants could be selected by monitoring seed fluorescence with the previous vector architecture, the antibiotic/herbicide resistance markers neighboring the FAST cassette were not functional, and no genome editing activity was observed in several independent experiments and with different sgRNAs in Arabidopsis. This shall act as a cautionary note for use of the FAST marker in novel assemblies, and exemplifies the synthetic biology crux that a system's performance is not simply the sum of its components. With the novel vector architecture, transgenic plants could conveniently be selected by herbicide resistance in the T_1 generation (Fig. 3b), and also the FAST marker was functional for selection/counter-selection in T_1 and T_2 generations (Fig. 3c). The

eight constructs differing only in promoter fragments driving Cas9 expression contained sgRNAs for targeting of the *DM2h* gene of the *DM2*^{Ler} cluster, and were transformed into L*er old3-1* plants (cultivated at 28°C to suppress autoimmunity). Inactivation of *DM2h* rescues the autoimmune phenotype of the *old3-1* line in a dose-dependent manner, and both heterozygous (*DM2h/dm2h*) and homozygous (*dm2h/dm2h*) plants can be phenotypically identified by simple survival at different temperature regimes (Ordon et al., 2017).

Genome editing in the T_1 generation was first tested by shifting BASTA-resistant transformants from 28°C to 22°C. At this temperature, inactivation of a single *DM2h* allele is sufficient to suppress autoimmunity of the parental line. Eight days after shifting, all plants showed signs of autoimmunity, arguing against efficient genome editing with any of the constructs/promoters in this generation and with the used sgRNAs. Five to eleven BASTA-resistant T_1 plants of each transformation were further cultivated at 28°C to obtain T_2 seeds. T_2 plants were grown alongside control plants (L*er*, L*er old3-1*) at 22°C, and *old3-1* plants became necrotic after 20 d of growth (Fig. 3d). Although only limited numbers of T_2 pools were analyzed for constructs containing different promoters, obvious differences for the frequencies of phenotypically rescued (L*er*-like), and thus genome-edited, plants became evident: Rescued plants were not present upon expression of Cas9 by 35S, *ALB* or *GILT* promoters, were rare for *ICU2*, *AP1* and *Ubi* promoters, and frequent with p*DD45* or p*RPS5a* driving Cas9 expression (Fig. 3d). We further quantified efficiencies by counting rescued plants across T_2 populations (Fig. 3e). When comparing *Ubi*, *AP1* and *ICU2* promoters, similar genome editing efficiencies of approximately 1% were observed among T_2 plants. Nonetheless, the *Ubi* promoter might perform somewhat better, since all analyzed T_2 populations contained rescued, non-necrotic plants. In clear contrast, rescued plants occurred in all T_2 batches from transformation of p*RPS5a*- and *pDD45*-containing constructs at high frequencies averaging to 33 and 24%, respectively (Fig. 3e). Up to 70% of rescued plants were observed in some T_2 populations, but all still contained necrotic plants, confirming that no homozygous mutants were obtained in the T_1 generation in our experiments. The differences between *RPS5a* and *DD45* promoters observed in our comparison were not statistically significant. Summarizing, our promoter comparison clearly points out superior performance of the *RPS5a* and *DD45* promoters for Arabidopsis genome editing. It should be noted that performance of p*DD45* should be further enhanced by use of the derived “EC1.2-EC1.1 fusion promoter”, which was not tested here (Wang et al., 2015).

High frequency generation of quintuple mutants in T_1 generation with improved Cas9 expression system

Having identified regulatory elements suitable for high efficiency Arabidopsis genome editing, the generation of an *lhcb1* mutant line was reattempted. A genome editing construct similar to that shown in Fig. 3a, but containing p*RPS5a*-driven Cas9, a Hygromycin resistance marker cassette (p*nos*:*hpt*-t*nos*; instead of p*nos*:Bar-t*nos*) and the same eight sgRNAs previously used (Fig. 1f) was transformed into a NoMxB3 mutant line already defective in *Lhcb4.1*, *Lhcb4.2*, *Lhcb5* and *Lhcb3* genes. T_1 plants were selected by Hygromycin resistance, and further grown in soil. From 30 Hygromycin-resistant plants, 20 showed clearly reduced chlorophyll accumulation in comparison to the parental line, and thus the pale green phenotype expected for reduced *Lhcb1* function (Pietrzykowska et al., 2014). From the 20 plants phenotypically distinct to the NoMxB3 line, 10 had an intermediate phenotype, and the remaining 10 were severely pale (Fig. 4a). To evaluate remaining levels of Lhcb1, leaf tissue samples of three severely pale T_1 plants and the parental NoMxB3 line were used for immunoblot analysis (Fig. 4b). While the control proteins PsaA (PSI core subunit; Mazor et al., 2017) and PsbB/CP47 (PSII core subunit; Wei et al., 2016) were detected to similar levels in all lines, no signal was obtained for Lhcb1 in the genome edited T_1 individuals, suggesting that all five *Lhcb1* genes had been inactivated. The accumulation of light harvesting complex II (LHCII), of which Lhcb1 is a major subunit, was further analyzed by SDS-PAGE and Coomassie staining. Intensity of the major band corresponding to LHCII was strongly reduced in comparison to the parental NoMxB3 line, and we assume that the residual LHCII signal in *lhcb1* genome edited lines originates exclusively from Lhcb2, another LHCII subunit (Liu et al., 2004). PCR-screening with oligonucleotides flanking regions targeted for deletion or mutagenesis indicated complex and diverse rearrangements at the *Lhcb1* loci in putative mutant lines (Fig. 4d). In all lines, novel deletion alleles not present in the parental line (ctrl, NoMxB3) were detected. The amplification of multiple PCR products (> 2) from individual lines might result from detection of multiple somatic events (chimeric mutants), or also from low specificity of oligonucleotides due to high homology within *Lhcb1* genes. Analysis of transgene-free T_2 individuals will be required to reveal the precise molecular lesions in putative *lhcb1* lines, and this analysis is yet ongoing. Nonetheless, these results suggest that the improved Cas9 expression system not only enhances overall genome editing frequencies, but even allows the isolation of complex alleles, in this case a quintuple mutant, in a single step in the T_1 generation.

DISCUSSION

Until recently, higher order Arabidopsis mutants could be generated exclusively by crossing of lines harboring the respective lesions. A number of e.g. quintuple, hexuple or even up to decuple mutants were previously reported (e.g. Fujii et al., 2011; Maekawa et al., 2012; Wild et al., 2016), but their isolation is

extremely laborious due to segregation of multiple alleles and/or close linkage between loci of interest. Accordingly, it was previously stated that due to "close genetic linkage, loss-of-function (quintuple *Lhcb1* or triple *Lhcb2*) T-DNA KO mutants are almost impossible to generate" (Pietrzykowska et al., 2014). The application of SSNs as reverse genetics tools theoretically alleviated these limitations, but commonly suffered from low efficiency in first reports. Here, by using an optimized Cas9-based genome editing system with high multiplexing capacity, we show that complex alleles such as higher order (quintuple) mutants may be generated with high efficiency even in a single generation (T_1). This demonstrates how SSNs can, through current and future optimization steps, match the increasingly complex demands and requirements of basic research projects in the Arabidopsis model system. The finding that the egg-cell specific promoters (*DD45* and especially the EC1.2-EC1.1 fusion promoter) or the *RPS5a* promoter confer particularly high genome editing efficiencies readily in the T_1 generation (Wang et al., 2015; Tsutsui and Higashiyama, 2017) is by itself not new, but is shown here to withstand direct comparison with other promoter systems using identical vector architectures and sgRNAs/target sites. Vector maps and sequence details for optimized vectors also providing positive/negative selection used here (Fig. 3) are provided in Online Resource 2. Vectors allow simple Golden Gate-based assembly of multiplexing constructs containing up to eight sgRNAs in four days without any PCR steps as previously described (Ordon et al., 2017), and are available upon request.

The expression of Cas9 under control of the *DD45* or *RPS5a* promoters improved genome editing efficiencies at the *DM2h* locus approximately 25-fold in comparison to *ubiquitin*, *AP1* and *ICU2* promoters (Fig. 3). Nonetheless, we were successful in isolating a 70 kb deletion allele produced by the previous, non-optimized system containing *ubiquitin* promoter-controlled Cas9 (Fig. 1). This validates the used strategy of PCR-screening large numbers of pooled T_2 individuals and may act as guidance for future isolation of deletion alleles for functional interrogation of gene clusters or non-coding regions. However, this strategy is obviously hampered by lacking controls for functionality of the conducted PCR prior to detection of the desired deletion allele. Interestingly, we estimated the occurrence of the 70 kb deletion allele to approximately 1.6% of T_2 individuals, but detected editing at the *DM2h* locus (in promoter comparison experiments, Fig. 3) among only 1% of T_2 individuals under similar Cas9 expression conditions. Our previous data suggested that point mutations were the most frequent type of Cas9-induced alleles in Arabidopsis, and that the frequency of deletion alleles (between sgRNA target sites in multiplexing applications) was inversely correlated with deletion size, as also reported in at least some studies from animal systems (Canver et al., 2014; Ordon et al., 2017). Taken together, this suggests poor efficiency of the sgRNAs used for *DM2h* editing, which may also explain failure to isolate hetero- or homozygous *dm2h* mutants in the T_1 generation when using *DD45* or *RPS5a* promoters. Indeed, mutations in the T_1 generation were also rare when the *GLABRA2* locus was targeted with *DD45*-driven Cas9 (Mao et al., 2016), supporting that recovery of T_1-edited plants might strongly depend on sgRNAs and/or target sites. Notably, all of the *Δdm2a-g* deletion alleles detected by PCR

screening and further analyzed were heritable, and segregated at Mendelian ratios in respective T_3 generations (Fig. 1d), suggesting that detection of somatic genome editing events is not a major issue for isolation of deletions at least under the used conditions.

An optimal genome editing system for reverse genetics in Arabidopsis will produce homozygous mutants at near 100% efficiency in the T_1 generation at any given locus. Although the optimization of Cas9 expression conditions tremendously improved efficiencies, further modifications are required to obtain this goal. In respect to Cas9 regulatory sequences, we here focused on previously reported elements, and it is well conceivable that yet uncharacterized promoters and/or transcriptional terminators might further improve genome editing efficiencies. However, also all remaining components may be further optimized, and additional functionalities implemented into T-DNA constructs. To this end, e.g. nuclear import of Cas9 might be enhanced by different or additional nuclear localization signals, or its expression improved by codon optimization or addition of introns. Furthermore, especially sgRNA expression levels appear to have a major influence on a system's performance. To date, sgRNAs were mainly expressed directly from Polymerase III (Pol III)-transcribed U3/U6 promoters. Additionally, sgRNAs were expressed as polycistronic transcripts from Polymerase II (Pol II)-transcribed promoters, and mature sgRNAs are subsequently produced by cleavage through Csy4, self-cleaving ribozymes or the endogeneous tRNA processing system (Gao and Zhao, 2014; Nissim et al., 2014; Xie et al., 2015; Tang et al., 2016). Different Pol III promoters and Pol II-driven ribozyme, tRNA and Csy4 systems were recently systematically compared for genome editing in tomato protoplasts. The Csy4 and tRNA systems improved genome editing efficiencies approximately two-fold in comparison to Pol III-driven sgRNAs (Cermak et al., 2017). However, similar as for the nuclease itself, Pol II promoters providing strong and timely expression in the embryo will most likely be required to improve RGN efficiency by these approaches in Arabidopsis. Recently, also an optimized sgRNA scaffold developed for mammalian cells was reported to enhance editing efficiencies in rice, and in particular the abundance of biallelic and double mutants increased (Dang et al., 2015; Hu et al., 2018). We have now also implemented this improved sgRNA backbone to our system, but are awaiting experimental data to confirm improved efficiency. It should be noted that Cermak et al. (2017) did not observe improved efficiencies when employing a similar optimized sgRNA architecture (Chen et al., 2013). Additional to optimizing components or expression of the RGN system itself, e.g. co-expression of the exonuclease Trex2 was reported to improve genome editing efficiencies ~ two-fold in tomato and barley protoplasts (Cermak et al., 2017). Exonuclease co-expression concomitantly augmented average deletion size, which may facilitate initial mutation detection and simplify design of genetic markers. The described approaches provide ample opportunities to further boost genome editing efficiencies in the Arabidopsis systems towards the development of an optimal reverse genetic tool. Based on our findings that *RPS5a* and egg cell specific promoters confer highest genome editing efficiencies, we propose that one of the well-characterized vector systems incorporating these elements (e.g. Wang et

al., 2015; Tsutsui and Higashiyama, 2017; or as described here) should be included for any further benchmarking of Cas9-based RGNs in Arabidopsis.

AUTHOR CONTRIBUTIONS

JS and JO conceived the work, performed experiments and analyzed data. CK performed experiments. MB, LD'O and RB conceived *Lhcb1*-related experiments, and MB performed experiments. JS wrote the manuscript with contributions from all authors.

FIGURE LEGENDS

Fig. 1: *Ubiquitin* promoter-driven Cas9 for generation of complex alleles

(a) Schematic drawing of the *DM2* cluster from accession Landsberg *erecta* (not drawn to scale). The location of sgRNA target sites and PCR primers for screening (1382/1383) is indicated.
(b) PCR-screening of pooled DNAs for occurrence of a *Δdm2a-g* allele. Each pool contained 7-11 T_2 individuals from transformation of the *Δdm2a-g* genome editing construct. A size of ~ 500 bp is expected upon deletion of *DM2a-g*. PCR products were resolved on a 1% agarose gel and DNA visualized with ethidium bromide.
(c) Deconvolution of pools to identify single plants carrying the *Δdm2a-g* deletion allele. DNA was extracted from single plants of pools #8 and #89 from (b), and PCR-screened for a *Δdm2a-g* deletion as before. The parental line (ctrl) and a previously PCR-positive pool DNA were used as controls.
(d) Inheritability and segregation of the *Δdm2a-g* deletion allele in the T_3 generation. DNA was extracted from single T_3 plants derived from plant #36 in (c), and was used for genotyping: Presence of *Δdm2a*-g, presence of *DM2c*, and presence of Cas9. Results shown are representative for several independent T_3 populations analyzed in parallel. Two individuals homozygous for *Δdm2a-g* (absence of *DM2c*) are boxed.
(e) Molecular lesion in a *Δdm2a-g* deletion line. The amplicon from (d) representing the *Δdm2a-g* deletion was sequenced directly. The sgRNA target sites are indicated.
(f) Genetic organization of the two *Lhcb1* linkage groups on chromosome 1 and chromosome 2 of the Arabidopsis genome (drawn to scale). 1-8 indicate the positioning of sgRNA target sites.

Fig. 2: Effect of transcription terminators on Cas9 activity in transient reporter assays

(a) Architecture of minimalistic nuclease constructs used for systematic comparison of 3'UTR sequences and transcriptional terminators for Cas9 expression.
(b) Evaluation of genome editing efficiency of nuclease constructs differing only in Cas9 3'UTR/terminator sequences. A *GUS*-based nuclease activity reporter and different nuclease constructs were transiently co-transformed into *N. benth.* tissues by Agroinfiltration, and GUS activity was determined 3 dpi. Background activity of the reporter alone was arbitrarily set to 1, and GUS activities normalized. A nuclease construct with t35S-terminated Cas9 was included as control (nuclease). Error bars represent standard deviations from four replicates. The experiment was repeated four times with similar results. The following 3'UTR/terminator sequences were used: 35S - 35S terminator from cauliflower mosaic virus; Atug7 - Atug7 terminator from *Agrobacterium tumefaciens (A. tumefaciens)*; *nos* - *nos* terminator from *A. tumefaciens*; *act2* - *Actin2* terminator from *Arabidopsis thaliana*; *mas* - *mas* terminator from *A. tumefaciens*; *ATPase* - *ATPase* terminator from *Solanum lycopersicum (S. lycopersicum*); *rbcS C3* - from *S. lycopersicum*; H4 - *Histone H4* from *Solanum tuberosum* (all Engler et al., 2014, and references therein); *rbcS E9* - from *Pisum sativum* (Wang et al., 2015).

Fig. 3: Systematic comparison of promoters for driving Cas9 expression in *Arabidopsis thaliana*

(a) Schematic drawing of constructs used for Arabidopsis transformation. Both sgRNAs are driven by identical fragments of the pU6-26 promoter. Constructs differ only in Cas9 promoter/5'UTR sequences.
(b) Functionality of the BASTA resistance marker.
(c) Functionality of the FAST marker in T_1 and T_2 generations.
(d) Phenotypic survey of genome editing efficiencies with different promoters driving Cas9 expression. Representative pictures of T_2 pools and control plants (L*er*, L*er old3-1*) grown at 22°C are shown.
(e) Quantitative assessment of genome editing efficiencies. Necrotic/rescued plants from (d) were counted.

Fig. 4: Generation of *Lhcb1* mutant plants in a single generation

(a) Phenotype of putative *lhcb1* mutant plants. The NoMxB3 parental line is shown as control, in comparison to one of the severely pale T_1 lines recovered from Hygromycin selection and editing of *Lhcb1* genes.
(b) Immunoblot analysis of Lhcb1 protein accumulation. Protein extracts from the parental NoMxB3 line and three independent T_1 plants putatively deficient in *Lhcb1* genes were used for immunodetection of Lhcb1. PsaA and PsbB/CP47 were detected as control proteins and loading control.

(c) Tris-Tricine SDS-PAGE and Coomassie staining of total protein as in (b). The major signal corresponding to LHCII is marked, and strongly reduced in genome-edited T_1 individuals due to loss of Lhcb1.
(d) PCR interrogation at *Lhcb1* loci in T_1 genome-edited lines. DNA was extracted from T_1 lines and the parental NoMxB3 line (ctrl), and used for PCR with primers flanking outermost sgRNA target sites (Fig. 1f) in *Lhcb1* genes on chromosome 1 (upper panel, PCR *Lhcb1.1/.2./.3*) and chromosome 2 (lower panel, PCR *Lhcb1.4/.5*).

Online Resource 1: Architecture and functional verification of an adaptable, *GUS*-based nuclease activity reporter
(a) T-DNA region of the adaptable reporter plasmid, and cloning of user-defined target sequences. The "empty" plasmid contains a 35S-driven *GUS*, with a *BsmB*I-excisable ccdB cassette inserted between the initiating ATG and the *GUS* coding sequence. In a *BsmB*I Golden Gate reaction, the ccdB cassette is exchanged for a user-defined target sequence introduced as hybridzed oligonucleotides. A configuration in which the reporter detects a -1 nt repair event (or e.g. -4 or +2 nt events) is shown as example. The reporter may be adapted to detect different events by varying the length of the introduced target site. It should be noted that the introduced target site may not, after repair, contain an in-frame STOP codon.
(b) Functional verification of the *GUS*-based reporter. Two different target sites were introduced into the adaptable plasmid to obtain Reporters 1/2, reporters were co-expressed with respective nucleases in *N. benth*., and GUS activity visualized by X-Gluc 3 dpi. Nuclease/reporter combination 2 consistently showed stronger GUS activity.

Online Resource 2: Sequence details on nuclease and reporter constructs used in this study (annotated GenBank files)

REFERENCES

Alcazar, R., Garcia, A.V., Parker, J.E., and Reymond, M. (2009). Incremental steps toward incompatibility revealed by Arabidopsis epistatic interactions modulating salicylic acid pathway activation. PNAS **106,** 334-339.

Alcazar, R., Garcia, A.V., Kronholm, I., de Meaux, J., Koornneef, M., Parker, J.E., and Reymond, M. (2010). Natural variation at *Strubbelig Receptor Kinase 3* drives immune-triggered incompatibilities between *Arabidopsis thaliana* accessions. Nat Genet **42,** 1135-1139.

Alcazar, R., von Reth, M., Bautor, J., Chae, E., Weigel, D., Koornneef, M., and Parker, J.E. (2014). Analysis of a plant complex resistance gene locus underlying immune-related hybrid incompatibility and its occurrence in nature. PLoS Genet **10,** e1004848.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods **9,** 39.

Botella, M.A., Parker, J.E., Frost, L.N., Bittner-Eddy, P.D., Beynon, J.L., Daniels, M.J., Holub, E.B., and Jones, J.D. (1998). Three genes of the Arabidopsis *RPP1* complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. Plant Cell **10,** 1847-1860.

Canver, M.C., Bauer, D.E., Dass, A., Yien, Y.Y., Chung, J., Masuda, T., Maeda, T., Paw, B.H., and Orkin, S.H. (2014). Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. Journal of Biological Chemistry **289,** 21312-21324.

Ceasar, S.A., Rajan, V., Prykhozhij, S.V., Berman, J.N., and Ignacimuthu, S. (2016). Insert, remove or replace: A highly advanced genome editing system using CRISPR/Cas9. Biochimica et Biophysica Acta **1863,** 2333-2344.

Cermak, T., Curtin, S.J., Gil-Humanes, J., Cegan, R., Kono, T.J.Y., Konecna, E., Belanto, J.J., Starker, C.G., Mathre, J.W., Greenstein, R.L., and Voytas, D.F. (2017). A multi-purpose toolkit to enable advanced genome engineering in plants. Plant Cell **29**, 1196–1217.

Chae, E., Bomblies, K., Kim, S.T., Karelina, D., Zaidem, M., Ossowski, S., Martin-Pizarro, C., Laitinen, R.A., Rowan, B.A., Tenenboim, H., Lechner, S., Demar, M., Habring-Muller, A., Lanz, C., Ratsch, G., and Weigel, D. (2014). Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. Cell **159,** 1341-1351.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J., Buchthal, J., Housden, B.E., Perrimon, N., Collins, J.J., and Church, G. (2016). Comparison of Cas9 activators in multiple species. Nat Methods **13,** 563-567.

Chen, B., Hu, J., Almeida, R., Liu, H., Balakrishnan, S., Covill-Cooke, C., Lim, W.A., and Huang, B. (2016). Expanding the CRISPR imaging toolset with *Staphylococcus aureus* Cas9 for simultaneous imaging of multiple genomic loci. Nucleic Acids Res **44,** e75.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell **155,** 1479-1491.

Dall'Osto, L., Cazzaniga, S., Bressan, M., Palecek, D., Zidek, K., Niyogi, K.K., Fleming, G.R., Zigmantas, D., and Bassi, R. (2017). Two mechanisms for dissipation of excess light in monomeric and trimeric light-harvesting complexes. Nat Plants **3,** 17033.

Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., Shankar, P., and Wu, H. (2015). Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. Genome Biology **16,** 280.

Desfeux, C., Clough, S.J., and Bent, A.F. (2000). Female reproductive tissues are the primary target of Agrobacterium-mediated transformation by the Arabidopsis floral-dip method. Plant Phys **123,** 895-904.

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

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H.W., Listgarten, J., and Root, D.E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature Biotech **34,** 184-191.

Dominguez, A.A., Lim, W.A., and Qi, L.S. (2016). Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. Nature Reviews **17,** 5-15.

Eid, A., Ali, Z., and Mahfouz, M.M. (2016). High efficiency of targeted mutagenesis in Arabidopsis via meiotic promoter-driven expression of Cas9 endonuclease. Plant Cell Reports **35,** 1555-1558.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS ONE **3,** e3647.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.M., Werner, S., Jones, J.D., Patron, N.J., and Marillonnet, S. (2014). A Golden Gate Modular Cloning Toolbox for Plants. ACS Synthetic Biology.

Fauser, F., Schiml, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. Plant Journal **79,** 348-359.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.L., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y., and Zhu, J.K. (2013). Efficient genome editing in plants using a CRISPR/Cas system. Cell Res **23,** 1229-1232.

Fujii, H., Verslues, P.E., and Zhu, J.K. (2011). Arabidopsis decuple mutant reveals the importance of SnRK2 kinases in osmotic stress responses *in vivo*. PNAS **108,** 1717-1722.

Galka, P., Santabarbara, S., Thi, T.H.K., Degand, H., Morsomme, P., Jennings, R.C., Boekema, E.J., and Caffarri, S. (2012). Functional Analyses of the Plant Photosystem I-Light-Harvesting Complex II Supercomplex Reveal That Light-Harvesting Complex II Loosely Bound to Photosystem II Is a Very Efficient Antenna for Photosystem I in State II. Plant Cell **24,** 2963-2978.

Gao, Y., and Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing. Journal of Integr Plant Biol **56,** 343-349.

Gao, Y., Zhang, Y., Zhang, D., Dai, X., Estelle, M., and Zhao, Y. (2015). Auxin binding protein 1 (ABP1) is not required for either auxin signaling or Arabidopsis development. PNAS **112,** 2275-2280.

Hu, X., Meng, X., Liu, Q., Li, J., and Wang, K. (2018). Increasing the efficiency of CRISPR-Cas9-VQR precise genome editing in rice. Plant Biotechnol Journal **16,** 292-297.

Huang, S., Weigel, D., Beachy, R.N., and Li, J. (2016). A proposed regulatory framework for genome-edited crops. Nat Genet **48,** 109-111.

Hyun, Y., Kim, J., Cho, S.W., Choi, Y., Kim, J.S., and Coupland, G. (2015). Site-directed mutagenesis in *Arabidopsis thaliana* using dividing tissue-targeted RGEN of the CRISPR/Cas system to generate heritable null alleles. Planta **241,** 271-284.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337,** 816-821.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature **533,** 420-424.

Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B., and Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Res **44,** W272-276.

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., Liu, J., Zhang, H., Liu, C., Ran, Y., and Gao, C. (2017). Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nat Commun **8,** 14261.

Liu, Z., Yan, H., Wang, K., Kuang, T., Zhang, J., Gui, L., An, X., and Chang, W. (2004). Crystal structure of spinach major light-harvesting complex at 2.72 A resolution. Nature **428,** 287-292.

Logemann, E., Birkenbihl, R.P., Ulker, B., and Somssich, I.E. (2006). An improved method for preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol. Plant Methods **2,** 16.

Maekawa, T., Kracher, B., Vernaldi, S., Ver Loren van Themaat, E., and Schulze-Lefert, P. (2012). Conservation of NLR-triggered immunity across plant lineages. PNAS **109,** 20119-20123.

Malzahn, A., Lowder, L., and Qi, Y. (2017). Plant genome editing with TALEN and CRISPR. Cell Biosci **7,** 21.

Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F., and Zhu, J.K. (2013). Application of the CRISPR-Cas system for efficient genome engineering in plants. Mol Plant **6,** 2008-2011.

Mao, Y., Zhang, Z., Feng, Z., Wei, P., Zhang, H., Botella, J.R., and Zhu, J.K. (2016). Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in Arabidopsis. Plant Biotechnol Journal **14,** 519-532.

Mazor, Y., Borovikova, A., Caspy, I., and Nelson, N. (2017). Structure of the plant photosystem I supercomplex at 2.6 A resolution. Nat Plants **3,** 17014.

Mikami, M., Toki, S., and Endo, M. (2015). Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice. Plant Molecular Biology **88,** 561-572.

Nissim, L., Perli, S.D., Fridkin, A., Perez-Pinera, P., and Lu, T.K. (2014). Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. Molecular Cell **54,** 698-710.

Ordon, J., Gantner, J., Kemna, J., Schwalgun, L., Reschke, M., Streubel, J., Boch, J., and Stuttmann, J. (2017). Generation of chromosomal deletions in dicotyledonous plants employing a user-friendly genome editing toolkit. Plant Journal **89,** 155-168.

Osakabe, Y., Watanabe, T., Sugano, S.S., Ueta, R., Ishihara, R., Shinozaki, K., and Osakabe, K. (2016). Optimization of CRISPR/Cas9 genome editing to modify abiotic stress responses in plants. Scientific Reports **6,** 26685.

Peterson, B.A., Haak, D.C., Nishimura, M.T., Teixeira, P.J., James, S.R., Dangl, J.L., and Nimchuk, Z.L. (2016). Genome-Wide Assessment of Efficiency and Specificity in CRISPR/Cas9 Mediated Multiple Site Targeting in Arabidopsis. PLoS ONE **11,** e0162169.

Pietrzykowska, M., Suorsa, M., Semchonok, D.A., Tikkanen, M., Boekema, E.J., Aro, E.M., and Jansson, S. (2014). The light-harvesting chlorophyll a/b binding proteins Lhcb1 and Lhcb2 play complementary roles during state transitions in Arabidopsis. Plant Cell **26,** 3646-3660.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell **154,** 1380-1389.

Ren, B., Yan, F., Kuang, Y., Li, N., Zhang, D., Zhou, X., Lin, H., and Zhou, H. (2018). Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-guided hyperactive *hAID* mutant. Mol Plant.

Schagger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem **166,** 368-379.

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

Shimada, T.L., Shimada, T., and Hara-Nishimura, I. (2010). A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of *Arabidopsis thaliana*. Plant Journal **61,** 519-528.

Stuttmann, J., Peine, N., Garcia, A.V., Wagner, C., Choudhury, S.R., Wang, Y., James, G.V., Griebel, T., Alcazar, R., Tsuda, K., Schneeberger, K., and Parker, J.E. (2016). *Arabidopsis thaliana DM2h* (*R8*) within the Landsberg *RPP1-like* Resistance Locus Underlies Three Different Cases of EDS1-Conditioned Autoimmunity. PLoS Genet **12,** e1005990.

Su, X., Ma, J., Wei, X., Cao, P., Zhu, D., Chang, W., Liu, Z., Zhang, X., and Li, M. (2017). Structure and assembly mechanism of plant C2S2M2-type PSII-LHCII supercomplex. Science **357,** 815-820.

Tahir, J., Watanabe, M., Jing, H.C., Hunter, D.A., Tohge, T., Nunes-Nesi, A., Brotman, Y., Fernie, A.R., Hoefgen, R., and Dijkwel, P.P. (2013). Activation of R-mediated innate immunity and disease susceptibility is affected by mutations in a cytosolic O-acetylserine (thiol) lyase in Arabidopsis. Plant Journal **73,** 118-130.

Tang, X., Zheng, X., Qi, Y., Zhang, D., Cheng, Y., Tang, A., Voytas, D.F., and Zhang, Y. (2016). A Single Transcript CRISPR-Cas9 System for Efficient Genome Editing in Plants. Mol Plant **9,** 1088-1091.

Tang, X., Lowder, L.G., Zhang, T., Malzahn, A.A., Zheng, X., Voytas, D.F., Zhong, Z., Chen, Y., Ren, Q., Li, Q., Kirkland, E.R., Zhang, Y., and Qi, Y. (2017). A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. Nat Plants **3,** 17018.

Tsutsui, H., and Higashiyama, T. (2017). pKAMA-ITACHI Vectors for Highly Efficient CRISPR/Cas9-Mediated Gene Knockout in *Arabidopsis thaliana*. Plant Cell Physiol **58,** 46-56.

Wang, Z.P., Xing, H.L., Dong, L., Zhang, H.Y., Han, C.Y., Wang, X.C., and Chen, Q.J. (2015). Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. Genome Biology **16,** 144.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. PLoS ONE **6,** e16765.

Wei, X., Su, X., Cao, P., Liu, X., Chang, W., Li, M., Zhang, X., and Liu, Z. (2016). Structure of spinach photosystem II-LHCII supercomplex at 3.2 A resolution. Nature **534,** 69-74.

Wild, M., Daviere, J.M., Regnault, T., Sakvarelidze-Achard, L., Carrera, E., Lopez Diaz, I., Cayrel, A., Dubeaux, G., Vert, G., and Achard, P. (2016). Tissue-Specific Regulation of Gibberellin Signaling Fine-Tunes Arabidopsis Iron-Deficiency Responses. Developemental Cell **37,** 190-200.

Wolt, J.D., Wang, K., and Yang, B. (2016). The Regulatory Status of Genome-edited Crops. Plant Biotechnol Journal **14,** 510-518.

Woo, J.W., Kim, J., Kwon, S.I., Corvalan, C., Cho, S.W., Kim, H., Kim, S.G., Kim, S.T., Choe, S., and Kim, J.S. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nature Biotech **33,** 1162-1164.

Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G., and Zhang, B. (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. Bioinformatics.

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. PNAS **112,** 3570-3575.

Yan, L., Wei, S., Wu, Y., Hu, R., Li, H., Yang, W., and Xie, Q. (2015). High-Efficiency Genome Editing in Arabidopsis Using *YAO* Promoter-Driven CRISPR/Cas9 System. Mol Plant **8,** 1820-1823.

Ye, G.N., Stone, D., Pang, S.Z., Creely, W., Gonzalez, K., and Hinchee, M. (1999). Arabidopsis ovule is the target for Agrobacterium *in planta* vacuum infiltration transformation. Plant Journal **19,** 249-257.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., Qiu, J.L., and Gao, C. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. Nat Commun **7,** 12617.

Figure 1 Ordon et al.

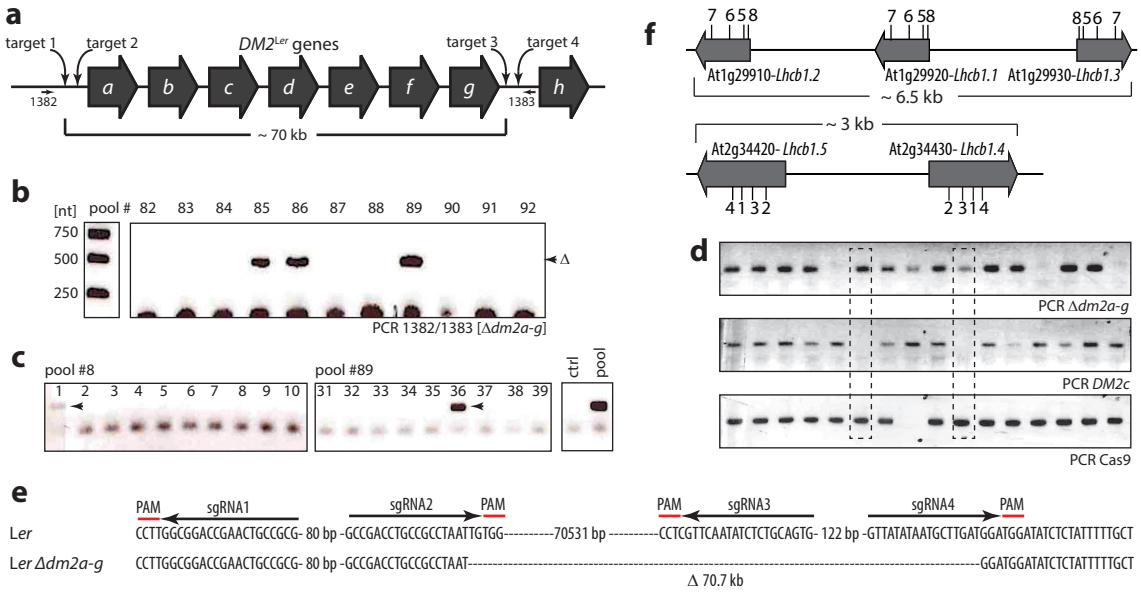


Fig. 1: Ubiquitin promoter-driven Cas9 for generation of complex alleles
(a) Schematic drawing of the *DM2* cluster from accession Landsberg *erecta* (not drawn to scale). The location of sgRNA target sites and PCR primers for screening (1382/1383) is indicated.
(b) PCR-screening of pooled DNAs for occurrence of a *Δdm2a-g* allele. Each pool contained 7-11 T_2 individuals from transformation of the *Δdm2a-g* genome editing construct. A size of ~ 500 bp is expected upon deletion of *DM2a-g*. PCR products were resolved on a 1% agarose gel and DNA visualized with ethidium bromide.
(c) Deconvolution of pools to identify single plants carrying the *Δdm2a-g* deletion allele. DNA was extracted from single plants of pools #8 and #89 from (b), and PCR-screened for a *Δdm2a-g* deletion as before. The parental line (ctrl) and a previously PCR-positive pool DNA were used as controls.
(d) Inheritability and segregation of the *Δdm2a-g* deletion allele in the T_3 generation. DNA was extracted from single T_3 plants derived from plant #36 in (c), and was used for genotyping: Presence of *Δdm2a-g*, presence of *DM2c*, and presence of Cas9. Results shown are representative for several independent T_3 populations analyzed in parallel. Two individuals homozygous for *Δdm2a-g* (absence of *DM2c*) are boxed.
(e) Molecular lesion in a *Δdm2a-g* deletion line. The amplicon from (d) representing the *Δdm2a-g* deletion was sequenced directly. The sgRNA target sites are indicated.
(f) Genetic organization of the two *Lhcb1* linkage groups on chromosome 1 and chromosome 2 of the Arabidopsis genome (drawn to scale). 1-8 indicate the positioning of sgRNA target sites.

Figure 2 Ordon et al.

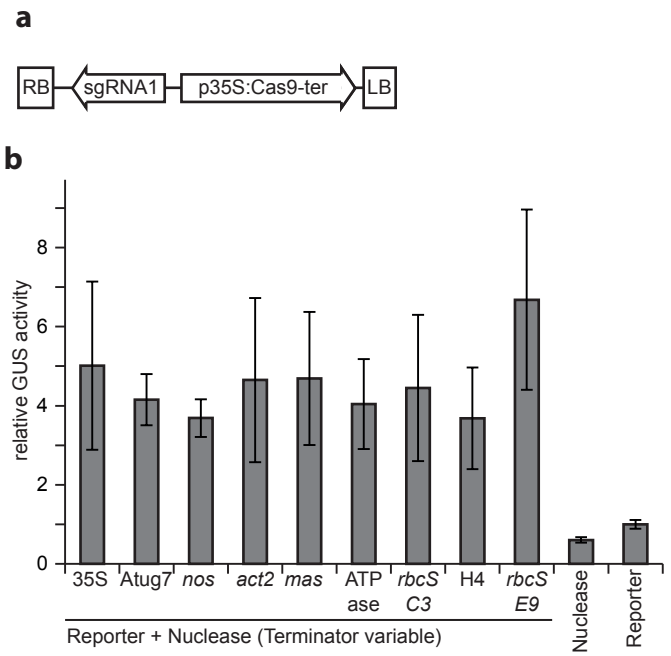
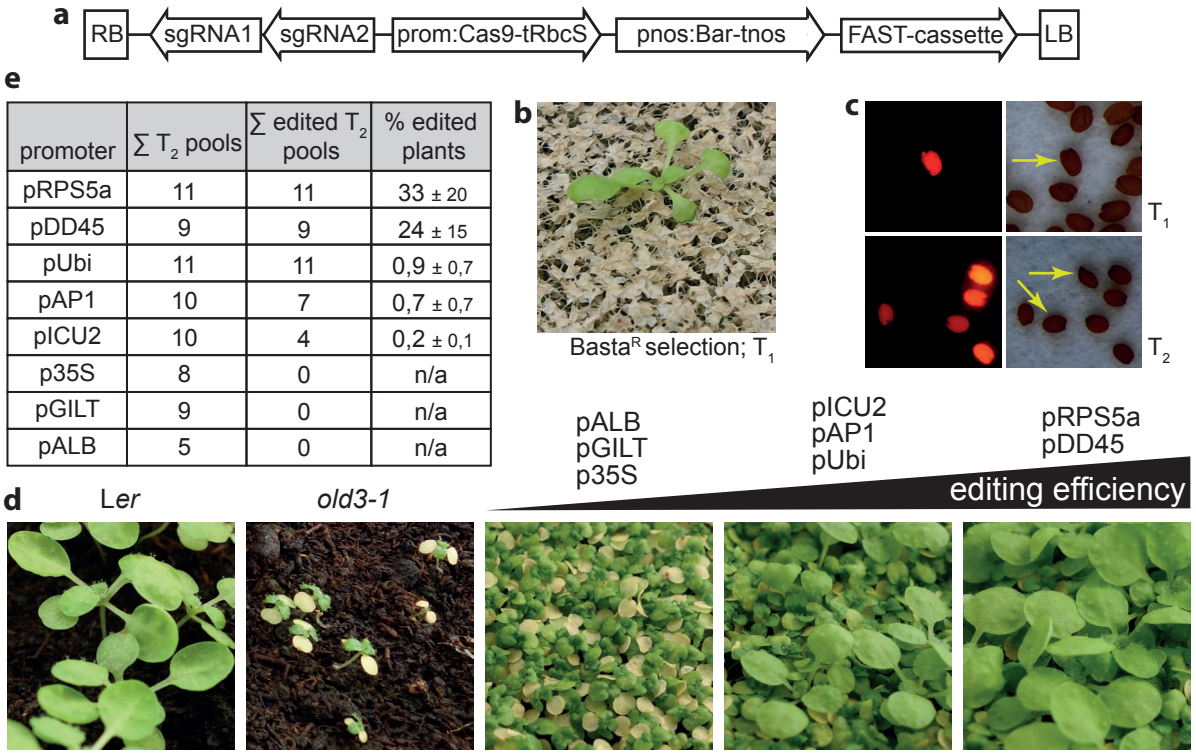


Fig. 2: Effect of transcription terminators on Cas9 activity in transient reporter assays
(a) Architecture of minimalistic nuclease constructs used for systematic comparison of 3'UTR sequences and transcriptional terminators for Cas9 expression.
(b) Evaluation of genome editing efficiency of nuclease constructs differing only in Cas9 3'UTR/terminator sequences. A GUS-based nuclease activity reporter and different nuclease constructs were transiently co-transformed into *N. benth*. tissues by Agroinfiltration, and GUS activity was determined 3 dpi. Background activity of the reporter alone was arbitrarily set to 1, and GUS activities normalized. A nuclease construct with t35S-terminated Cas9 was included as control (nuclease). Error bars represent standard deviations from four replicates. The experiment was repeated four times with similar results. The following 3'UTR/terminator sequences were used: 35S - 35S terminator from cauliflower mosaic virus; Atug7 - Atug7 terminator from *Agrobacterium tumefaciens (A. tumefaciens)*; *nos - nos* terminator from *A. tumefaciens*; *act2 - Actin2* terminator from *Arabidopsis thaliana; mas - mas* terminator from *A. tumefaciens*; *ATPase - ATPase* terminator from *Solanum lycopersicum (S. lycopersicum)*; *rbcS C3* - from *S. lycopersicum*; H4 - *Histone H4* from *Solanum tuberosum* (all Engler et al., 2014, and references therein); *rbcS E9* - from *Pisum sativum* (Wang et al., 2015).

Figure 3 Ordon et al.



e

promoter	∑ T_2 pools	∑ edited T_2 pools	% edited plants
pRPS5a	11	11	33 ± 20
pDD45	9	9	24 ± 15
pUbi	11	11	0,9 ± 0,7
pAP1	10	7	0,7 ± 0,7
pICU2	10	4	0,2 ± 0,1
p35S	8	0	n/a
pGILT	9	0	n/a
pALB	5	0	n/a

Fig. 3: Systematic comparison of promoters for driving Cas9 expression in *Arabidopsis thaliana*
(a) Schematic drawing of constructs used for Arabidopsis transformation. Both sgRNAs are driven by identical fragments of the pU6-26 promoter. Constructs differ only in Cas9 promoter/5'UTR sequences.
(b) Functionality of the BASTA resistance marker.
(c) Functionality of the FAST marker in T_1 and T_2 generations.
(d) Phenotypic survey of genome editing efficiencies with different promoters driving Cas9 expression. Representative pictures of T_2 pools and control plants (L*er*, L*er old3-1*) grown at 22°C are shown.
(e) Quantitative assessment of genome editing efficiencies. Necrotic/rescued plants from (d) were counted.

Figure 4 Ordon et al.

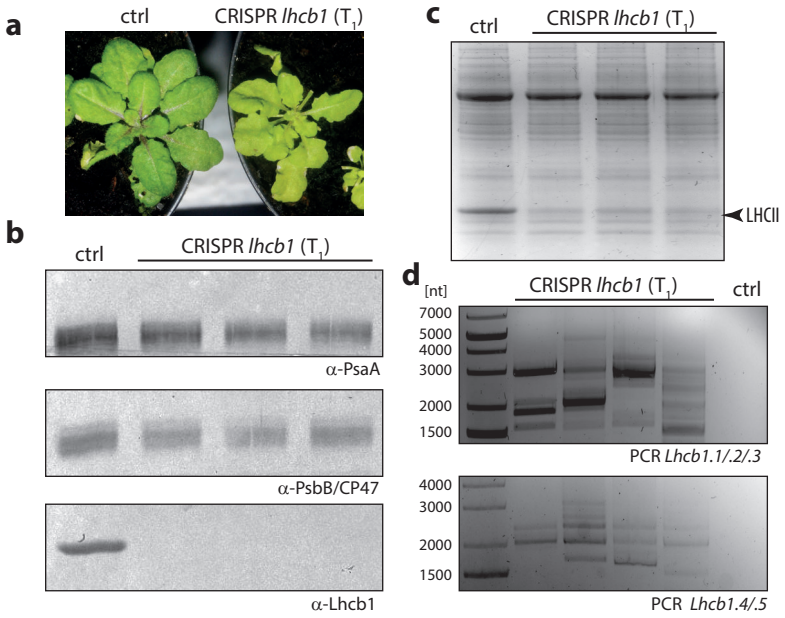


Fig. 4: Generation of *Lhcb1* mutant plants in a single generation

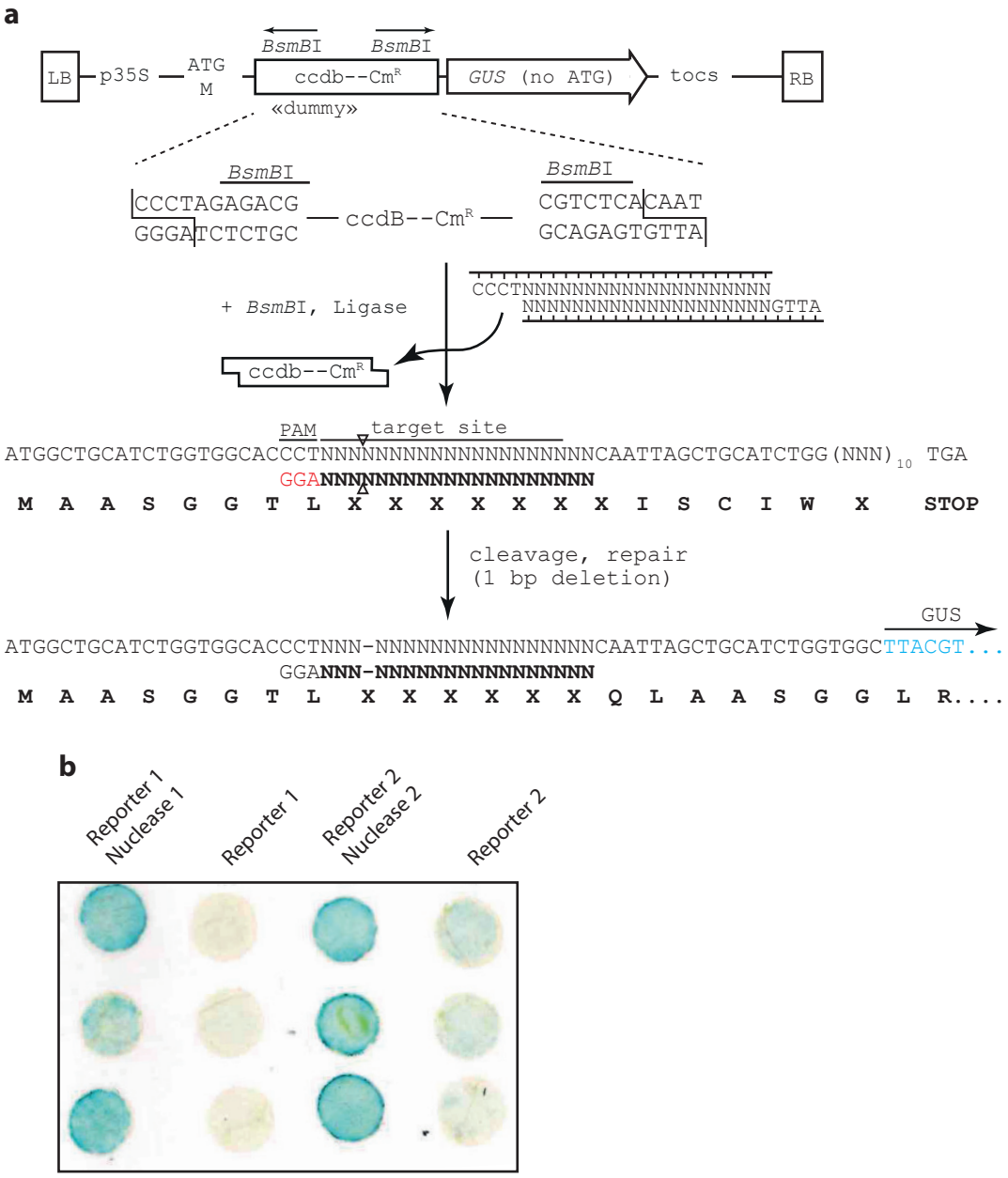
(a) Phenotype of putative *lhcb1* mutant plants. The NoMxB3 parental line is shown as control, in comparison to one of the severely pale T_1 lines recovered from Hygromycin selection and editing of *Lhcb1* genes.

(b) Immunoblot analysis of *Lhcb1* protein accumulation. Protein extracts from the parental NoMxB3 line and three independent T_1 plants putatively deficient in *Lhcb1* genes were used for immunodetection of Lhcb1. PsaA and PsbB/CP47 were detected as control proteins and loading control.

(c) Tris-Tricine SDS-PAGE and Coomassie staining of total protein as in (b). The major signal corresponding to LHCII is marked, and strongly reduced in genome-edited T_1 individuals due to loss of Lhcb1.

(d) PCR interrogation at *Lhcb1* loci in T_1 genome-edited lines. DNA was extracted from T_1 lines and the parental NoMxB3 line (ctrl), and used for PCR with primers flanking outermost sgRNA target sites (Fig. 1f) in *Lhcb1* genes on chromosome 1 (upper panel, PCR *Lhcb1.1/.2./.3*) and chromosome 2 (lower panel, PCR *Lhcb1.4/.5*).

Supplemental Figure 1 Ordon et al.



Supplemental Figure S1: Architecture and functional verification of an adaptable, *GUS*-based nuclease activity reporter

(a) T-DNA region of the adaptable reporter plasmid, and cloning of user-defined target sequences. The "empty" plasmid contains a 35S-driven *GUS*, with a *BsmB*I-excisable ccdB cassette inserted between the initiating ATG and the *GUS* coding sequence. In a *BsmB*I Golden Gate reaction, the ccdB cassette is exchanged for a user-defined target sequence introduced as hybridized oligonucleotides. A configuration in which the reporter detects a -1 nt repair event (or e.g. -4 or +2 nt events) is shown as example. The reporter may be adapted to detect different events by varying the length of the introduced target site. It should be noted that the introduced target site may not, after repair, contain an in-frame STOP codon.

(b) Functional verification of the *GUS*-based reporter. Two different target sites were introduced into the adaptable plasmid to obtain Reporters 1/2, reporters were co-expressed with respective nucleases in *N. benth.*, and GUS activity visualized by X-Gluc 3 dpi. Nuclease/reporter combination 2 consistently showed stronger GUS activity.