Homology-directed repair of a defective *glabrous* gene in *Arabidopsis* with Cas9-based gene targeting

Florian Hahn[1], Marion Eisenhut[1], Otho Mantegazza[1], Andreas P.M. Weber[1]*

[1]Institute of Plant Biochemistry, Cluster of Excellence on Plant Sciences (CEPLAS), Center for Synthetic Life Sciences (CSL), Heinrich Heine University, Düsseldorf, Germany

*Correspondence:
 Andreas P.M. Weber
 Andreas.Weber@uni-duesseldorf.de

Keywords: CRISPR/Cas9, homologous recombination, Glabrous1, trichomes, gene editing, marker, *in planta* gene targeting, viral replicons, gene targeting

ABSTRACT

The CRISPR/Cas9 system has emerged as a powerful tool for targeted genome editing in plants and beyond. Double-strand breaks induced by the Cas9 enzyme are repaired by the cell's own repair machinery either by the non-homologous end joining pathway or by homologous recombination. While the first repair mechanism results in random mutations at the double-strand break site, homologous recombination uses the genetic information from a highly homologous repair template as blueprint for repair of the break. By offering an artificial repair template, this pathway can be exploited to introduce specific changes at a site of choice in the genome. However, frequencies of double-strand break repair by homologous recombination are very low. In this study, we compared two methods that have been reported to enhance frequencies of homologous recombination in plants. The first method boosts the repair template availability through the formation of viral replicons, the second method makes use of an *in planta* gene targeting approach. Additionally, we comparatively applied a nickase instead of a nuclease for target strand priming. To allow easy, visual detection of homologous recombination events, we aimed at restoring trichome formation in a glabrous *Arabidopsis* mutant by repairing a defective *glabrous1* gene. Using this efficient visual marker, we were able to regenerate plants repaired by homologous recombination at frequencies of 0.12% using the *in planta* gene targeting approach, while both approaches using viral replicons did not yield any trichome-bearing plants.

INTRODUCTION

Gene targeting (GT) means integration of foreign DNA into a cell's genome by homologous recombination (HR; Paszkowski et al., 1988). GT has been a long-term goal for plant scientists since it allows modifying an endogenous gene *in planta* or integrating a transgene at a specific position in the genome (Puchta, 2002). GT is easily achieved in lower eukaryotes, such as yeast, or the moss *Physcomitrella patens,* because they include foreign DNA predominantly through a highly efficient HR pathways. However, GT is not easily achieved in higher plants, since they integrate foreign DNA via illegitimate recombination through the non-homologous end joining (NHEJ) pathway (Puchta, 2005). A simple introduction of a DNA donor molecule with homology arms (HA) only led to frequencies of 10^{-4} to 10^{-6} GT events per transformation event in higher plants, such as tobacco and *Arabidopsis* (Lee et al., 1990; Offringa et al., 1990; Miao and Lam, 1995; Risseeuw et al., 1995)*,* making GT impractical for most applications. Nevertheless, GT frequencies in higher plants can be increased by two orders of magnitude by inducing double-strand breaks (DSB) in the target

1

52 locus with sequence specific nucleases (SSN, Puchta et al., 1996). SSN are proteins that can
53 cleave DNA at predefined sequence motives. The first generation of SSN, including
54 meganucleases, Zinc-finger nucleases (ZFN), and Transcription Activator-like Effector
55 Nucleases (TALENs) recognized DNA by protein-DNA interactions, which required rational
56 protein design for every target sequence (Voytas, 2013). The most recent addition to the SSN
57 has been the CRISPR/Cas9 system (Cas9 system).
58 Cas9 is an endonuclease of the bacterium *Streptococcus pyogenes* that generates DSB in DNA
59 (Doudna and Charpentier, 2014) and that recently has become the SSN of choice for genome
60 editing and GT. Cas9 became so popular because, in contrast to other SSNs that recognize
61 DNA sequences through protein-DNA interactions, Cas9 is guided to its target site on DNA
62 by complementarity matching of a single guide RNA (sgRNA, Jinek et al., 2012). The RNA-
63 based targeting system makes Cas9 easily programmable towards a selected target. Therefore,
64 the Cas9 system has been used extensively by numerous laboratories to edit plant genomes
65 (Zhang et al., 2017). Additionally, mutation of one of the two endonuclease sites of Cas9
66 resulted into a Cas9 nickase, which introduces only single-strand nicks in the plant genome
67 (Fauser et al., 2014).
68 SSN-generated DSB are lethal for the cell and must be repaired by dedicated protein
69 machineries. Two major repair pathways can be distinguished: NHEJ and HR (Puchta, 2005).
70 Most DSB in higher eukaryotes are repaired by NHEJ (Sargent et al., 1997), a pathway that
71 involves direct re-ligation of the DSB ends (Gorbunova and Levy, 1997; Dueva and Iliakis,
72 2013), making it an error-prone process. This process can be exploited to introduce random
73 small indel mutations at a DSB site and has been used extensively to knock-out gene
74 functions with SSN (Bortesi and Fischer, 2015). In contrast, HR repair pathways use a
75 homology template (HT) to repair a DSB and are therefore of high potential for GT. In the
76 conservative synthesis-dependent strand annealing pathway, which is the most prominent HR
77 pathway in plant somatic tissues (Puchta, 1998), DNA ends are resected to create 3'
78 overhangs. One 3'-end can invade into a homologous sequence forming a loop-structure and
79 copy the sequence from the HT DNA. After release from the loop structure, the single strand
80 can reanneal with homologous sequences on the opposite side of the DSB, leading to
81 conservative repair without sequence information loss (Steinert et al., 2016). Artificial HT
82 with HA have already been used for GT experiments to integrate foreign DNA sequences into
83 a DSB generated by SSN (Puchta et al., 1996; Baltes et al., 2014; Schiml et al., 2014; Li et al.,
84 2015).
85 While GT frequencies were successfully enhanced by cutting the target site with SSN, this is
86 not sufficient for streamlined application of GT in plants. GT frequencies can be further
87 enhanced of approximately two orders of magnitudes (compared to conventional T-DNA
88 delivery) by mobilizing the HT after T-DNA transformation. This can be achieved with an *in
89 planta* gene targeting (IPGT) approach and with a viral replicon approach.
90 The IPGT approach (Fauser et al., 2012) employs SSN not only to cut the target locus but also
91 to excise the HT from the chromosomal T-DNA backbone. The liberation from the genomic
92 context of the T-DNA allows the activated HT to be available at the site of the DSB where it
93 can function as repair template. This method has been used successfully using various SSN in
94 *Arabidopsis* (Fauser et al., 2012; Schiml et al., 2014; Zhao et al., 2016), maize (Ayar et al.,
95 2013; Kumar et al., 2016), and rice (Sun et al., 2016).
96 An alternative approach has been developed by Baltes et al. in tobacco. It activates the HT by
97 including it in a *Geminivirus* based replicon (Baltes et al., 2014). Therefore, the HT is flanked
98 with three viral elements from the *Bean yellow dwarf virus* (BeYDV) that are required to
99 establish the replicon. These regions are known as the *cis*-acting long intergenic region (LIR),
100 the short interacting region (SIR) and the gene for the *trans*-acting viral replicase *Rep/RepA*.
101 The Rep protein initiates a rolling circle endoreplication at the LIR site, leading to a
102 replicational release of the HT in form of up to several thousand circular replicons (Chen et

2

103 al., 2011). This method has been successfully applied using components from various virus
104 strains and with different SSN for GT in wheat protoplast (Gil-Humanes et al., 2017), tomato
105 (Čermák et al., 2015), potato (Butler et al., 2016), and rice (Wang et al., 2017). While IPGT
106 has already been shown to induce stable GT events in *Arabidopsis* (Fauser et al., 2012;
107 Schiml et al., 2014), reports about successful applications of viral replicons based GT
108 approaches in *Arabidopsis* have not been published so far.
109 In this report, we aimed to evaluate and compare the efficiencies of the viral replicon system
110 and the IPGT system to enhance GT frequencies in the model plant *Arabidopsis thaliana*.
111 Therefore, we made use of the R2R3-MYB transcriptional master regulator of trichome
112 formation GLABROUS1 (GL1), which we established previously as non-invasive,
113 endogenous visual marker for Cas9 activity in *Arabidopsis* (Hahn et al., 2017). Null-
114 mutations of the *GL1* gene lead to loss of trichome formation (Koornneef et al., 1982) on
115 stems and leaves as GL1 is part of the signaling cascade which allows epidermal cells to
116 differentiate into trichomes (Pattanaik et al., 2014). Using the IPGT approach and the viral
117 replicon-based approach, we aimed at repairing a non-functional *gl1* gene with a 10 bp
118 deletion in the coding region (Hahn et al., 2017) by HR and by this, restoring trichome
119 formation. Using IPGT, we achieved successful GT and regenerated three stable non-glabrous
120 plants out of ~2,500 plants screened over three generations. However, we were not able to
121 regenerate non-glabrous plants using viral replicons. The *gl1* phenotype provided a simple
122 and efficient visual marker for tracking mutagenic activity *in planta*.
123
124 **METHODS**
125
126 **Plant growth conditions**
127 Surface-sterilized seeds of an *Arabidopsis gl1* mutant with a 10 bp deletion in the *GL1* gene
128 (Hahn et al., 2017) were stratified at 4 °C for three days. Seeds were germinated on 0.8%
129 (w/v) agar-solidified half-strength Murashige and Skoog medium (Duchefa) in growth
130 chambers with a light intensity of 100 μmol photons m^{-2} s^{-1} under long day conditions (16-h-
131 light/8-h-dark cycle, 22/18 °C). After 14 days, seedlings were transferred to soil and further
132 grown under the same conditions. For large-scale screening of plants in the T2 generation,
133 seeds were directly sowed on soil and the plants were grown in the green house after three
134 days of stratification at 4 °C.
135
136 **Generation of a homology repair template for reparation of the ten nucleotides deletion**
137 **in the *gl1* gene**
138 We designed a homology repair template (HT) containing a codon-modified version of the
139 missing 10 bp. The 10 bp (CTGCCGTTTA) were flanked by homology arms (HA)
140 complementary to the genomic regions adjacent to the cutting site of Cas9 in the *gl1* gene.
141 The HA consisted of 813 bp and 830 bp, respectively. Both homology arms contained a single
142 nucleotide exchange within the homology region to distinguish two-sided true GT events
143 from one-sided GT events combined with NHEJ mediated repair pathways (Schiml et al.,
144 2014). After a short linker sequence, the sequence of the sgRNA recognition site including the
145 protospacer adjacent motif (PAM) sequence followed. The HT was synthesized at GeneART
146 (ThermoFisher Scientific) and integrated in the vector backbone pMA-RQ (GeneART). The
147 resulting vector was named pFH95.
148
149 **Cloning of a binary T-DNA repair vector for *in planta* gene targeting repair approach**
150 We made use of the binary T-DNA vector pDE-Cas9, containing an *Arabidopsis*-codon-
151 optimized Cas9 controlled by the constitutive *Ubiquitin* promoter (PcUBp) from *Petroselinum*
152 *crispum* (Fauser et al., 2014) and the sgRNA subcloning vector pEN-C.1.1 containing the
153 *Arabidopsis* U6-26 promoter for sgRNA expression, the sgRNA scaffold and BbsI-sites for

3

154 subcloning of the protospacer sequence (Schiml et al., 2014). After digestion of pEN-C1.1
155 with BbsI-HF (NewEngland Biolabs), we inserted the target sequence
156 (AAGAGATGTGGGAAAAGAGA plus an additional guanine as first bp for transcription by
157 the U6-26 promoter) using two annealed primers FH210/FH211 (Table S1) resulting in vector
158 pFH89. We then digested pFH89 with MluI-HF (NEB) to extract the sgRNA cassette and
159 ligated it into MluI-HF-digested pDE-Cas9, thereby creating the vector pFH99. We amplified
160 the HT from pFH95 using primers FH278/FH279 (Table S1) and performed Gateway® BP-
161 reaction with pDONR 207 (Invitrogen) resulting in vector pFH122. LR reaction (Invitrogen)
162 of pFH99 with pFH122 led to the assembly of the IPGT repair vector pIPGT-Nuc.
163

Generation of binary T-DNA repair vectors for viral replicon repair approach
165 For the viral replicon repair approach, we amplified the viral components from BeYDV
166 necessary for replicon formation from the vector pZLSLD.R (Baltes et al., 2014) using
167 primers FH217/FH218 (Table S1) and cloned the PCR product into pDONR 207 (Invitrogen)
168 via BP reaction. This vector (pFH83) was then digested with PmII (NewEngland Biolabs) and
169 BsaXI (NewEngland Biolabs) to remove the original HT from Baltes et al. (2014). We PCR-
170 amplified the *GL1* HT from pFH95 using primers FH240/FH241 (Table S1) and inserted it
171 into the digested pFH83 vector by Gibson® cloning (NewEngland Biolabs), resulting in the
172 vector pFH84. The final viral replicon nuclease repair construct pVIR-Nuc containing Cas9
173 nuclease, the sgRNA cassette directed against the *gl1* gene and the HT was then cloned by
174 Gateway® LR reaction (Invitrogen) of pFH99 with pFH84.
175 For the viral replicon repair approach using the Cas9 nickase, we used the MluI-HF-extracted
176 sgRNA cassette from pFH89 and ligated it into MluI-HF-digested pDE-Cas9-D10A (Fauser et
177 al., 2014) resulting in the vector pFH103. LR reaction of pFH103 with pFH84 led to the final
178 viral replicon nickase repair construct pVIR-Nick.
179 The sequences of the vectors pIPGT-Nuc (accession number xxx), pVIR-Nuc (accession
180 number xxx) and pVIR-Nick (accession number xxx) are accessible on GenBank.
181

Plant transformation
183 *Agrobacterium tumefaciens*, strain GV3101::pMP90 was transformed with the vectors
184 pIPGT-Nuc, pVIR-Nuc, and pVIR-Nick. Floral dipping (Clough and Bent, 1998) was used to
185 transform *Arabidopsis gl1* plants with the three constructs. Kanamycin (pFH136) and
186 glufosinate (pFH114, pFH135) were used for selection of primary transformants.
187 Furthermore, leaf genomic DNA was isolated from surviving plants via isopropanol
188 precipitation (Weigel and Glazebrook, 2009). The presence of T-DNA insertion was verified
189 by PCR using primer pairs FH190/FH223 (pIPGT-Nuc) or FH223/FH265 (pVIR-Nuc, pVIR-
190 Nick, Table S1).
191

Detection of the *Cas9* gene in plants
193 Presence of the *Cas9* gene in plant genomes was verified by PCR on leaf genomic DNA using
194 primer pairs amplifying specifically the *Cas9* gene (FH313/FH314, Table S1).
195

Verification of viral replicon formation
197 Formation of circular viral replicons was verified in T1 generation plants by PCR using leaf
198 genomic DNA and the primers FH266/FH271 (Table S1).
199

Detection of Cas9-mediated gene repair
201 For verification of HDR events on genomic level, leaf genomic DNA was isolated via
202 isopropanol precipitation (Weigel and Glazebrook, 2009) and the region of interest in the *gl1*
203 gene was PCR-amplified with proofreading Q5© polymerase (NewEngland Biolabs) using
204 primers FH214/FH215 (Table S1). The obtained PCR amplicons were analyzed directly by

4

205 Sanger sequencing (Macrogen) or subcloned into pJET1.2 (ThermoFisher Scientific) and then
206 sequenced. 4Peaks software (Nucleobytes) and SnapGene Viewer (GSL Biotech) were used
207 to analyze sequencing histograms.
208
209 **RESULTS**
210
211 **Selection of a visual marker for testing gene targeting events**
212 In a previous study, we established trichome formation as efficient visual marker to monitor
213 Cas9 mutagenic activity in *Arabidopsis* (Hahn et al., 2017). We used a CRISPR/Cas9
214 approach to generate a T-DNA free glabrous *Arabidopsis* mutant that contains a deletion of
215 10 bp in the master regulator gene for trichome formation, *GL1*. This 10 bp deletion disrupts
216 the reading frame of *GL1* and leads to a premature stop codon. Homozygous mutant plants
217 show almost no leaf and stem trichomes (Fig. S1). We decided to use this mutant line as a
218 model system to test Cas9-based GT in *Arabidopsis*. Cas9-induced cleavage in combination
219 with homology directed repair can be used to integrate sequences of choice into a genomic
220 region. If we successfully reintegrated the missing 10 bp into this mutant line´s genome, we
221 would expect restoration of leaf trichome formation. Since the mutant line was created by
222 CRISPR/Cas9, a PAM sequence was already present at the site of mutation and the predicted
223 Cas9 cleavage site (3 bp upstream of the PAM sequence) matched the desired integration site
224 for the missing 10 bp (Fig. S1).
225
226 **Selection of a gene targeting system and design of related constructs**
227 HR is more efficient in plants if the HT is available at the site of the DSB (Baltes et al., 2014).
228 If the HT is delivered as T-DNA, which integrates randomly into the genome, it might not be
229 available at the DSB site for the DNA repair machinery. Thus, we comparatively analyzed
230 two methods to enhance the availability of the HT at the DSB site.
231 The first method was developed by Baltes et al (2014) in tobacco calli. This method increases
232 the availability of the HT by including it in small circular geminiviral replicons.
233 For the repair of the *gl1* gene, we accordingly created a single T-DNA vector, (Fig. 1A), that
234 contained an *Arabidopsis* codon-optimized *Cas9* gene under the control of the constitutive
235 PcUB promoter (Fauser et al., 2014), a sgRNA cassette targeting the *gl1* gene, and the HT
236 with ~ 800 bp of HA on both sides, flanked by LIRs, SIR, and a *Replicase A* (*RepA*) gene. For
237 an easy nomenclature of the plasmids, we decided to include the name of the HT activation
238 method (viral replicons = vir; *in planta* gene targeting = IPGT) and the used Cas9 variant
239 (nuclease = Nuc; nickase = Nick) in the plasmid name and therefore named this construct
240 pVIR-Nuc.
241 After expression, RepA binds to the LIR sites and initiates rolling circle replication, leading to
242 the formation of up to several thousand copies of the HT in the nucleus (Timmermans et al.,
243 1992; Baltes et al., 2014), which can then be used as repair template for the DSB created by
244 Cas9 nuclease (Fig. 2A).
245 To further explore the potential of this method, we additionally generated and tested a
246 complementary vector pVIR-Nick in which the Cas9 nuclease gene was replaced by a nickase
247 gene (Fig. 1B). DNA nicks trigger HR without triggering the competing NHEJ repair (Fauser
248 et al., 2014). NHEJ repair interferes with HR because it induces random indel mutations that
249 destroy the sgRNA target site and make it unsuitable for homology directed repair. By using
250 the Cas9 nickase, we expected to increase HR efficiency.
251 The second method, developed by Fauser et al. (2012), is called IPGT. This method increases
252 the availability of the HT by releasing it from the T-DNA backbone using the same SSN that
253 creates the DSB in the target region. Therefore, we designed the vector pIPGT-Nuc (Fig. 1C),
254 in which the HT is flanked by the same sgRNA target site that is used to create the DSB in the
255 *gl1* gene (Fig. 2B).

5

Plants transformed with pVIR-constructs form circular viral replicons

We transformed Arabidopsis *gl1* plants with either pVIR-Nuc, pVIR-Nick, or pIPGT-Nuc and selected three independent T1 lines per construct. Though we did not detect trichomes on the leaf surface of any T1 plant, we wanted to verify that the viral components were indeed producing circular HT in the T1 plants. Thus, we performed a PCR using primers that only yield a product in the circularized version of the HT but not in the linear T-DNA (Fig. 2A). We detected PCR products in the plants transformed with pVIR-Nuc and pVIR-Nick. PCR fragments for plants transformed with pIPGT-Nuc, which do not generate replicons (Fig. S2), were not obtained.

T2 generation plants show spots of trichomes on leaves

We analyzed several hundred plants from the progeny of each T1 plant for trichome reappearance. While we could not detect plants that fully recovered trichome production on leaves, we found several plants from the IPGT approach that harbored spots of trichomes (Tab. 1). These spots ranged from two trichomes on a single leaf of the plant to large spots on several leaves (Fig. 3A). In contrast, using the viral replicon method, we only detected one plant with a single trichome in plants transformed with pVIR-Nuc, which implements a full Cas9 nuclease to trigger HR. Plants transformed with pVIR-Nick (Fig. 3A), which implements the Cas9 nickase, did not form any trichomes on their leaves.

We isolated gDNA from leaf trichome spots of three IPGT plants (the single trichome on the pVIR-Nuc plant provided not enough plant material for gDNA isolation), amplified the *gl1* gene, subcloned the PCR product and sequenced 42 clones. While most clones still contained the *gl1* sequence with the 10 bp deletion, we detected various indel mutations but also the integration of the codon-modified 10 bp HT, which restored the wildtype amino acid sequence of the *GL1* gene (Fig. 3B). This chimerism on genotype level was in line with the highly chimeric phenotype of the leaves.

Fully regenerated trichomes were detected in plants of the T3 generation

We harvested seeds from the chimeric T2 plants and analyzed ~ 200 progeny T3 plants of each T2 plant. In most cases, the progeny plants were glabrous or just contained single trichome spots. However, we detected three plants (frequency: 1.5 %) in the progeny of the highly chimeric IPGT plant (Fig. 3A, left) with fully regenerated trichomes (Fig. 4A). We amplified the *gl1* gene from those plants and sequenced the PCR product. Two of the three plants showed clear peaks corresponding to the 10 bp insertion but also peaks corresponding to small indel mutations (Fig. 4B, Fig. S3), indicating that only one of the two alleles was repaired by HR. In the third plant, we detected the expected 10 bp insertion not in the sequencing histogram but in 5 out of 15 sequenced clones of subcloned PCR product (Fig. S3D).

The repaired *GL1* allele segregates in a mendelian way

In genome edited plants, while the editing nuclease (Cas9) is active, it is hard to distinguish between somatic and heritable mutations. The three plants that fully recovered trichome production were still carrying the *Cas9* gene (Fig. 4C), which was likely still active. Thus, we confirmed in those plants that the 10 bp insertion is indeed heritable and not somatic by Mendelian segregation analysis. We analyzed the offspring of the three T3 plants displaying trichomes and counted the amounts of trichome-bearing and glabrous plants in the T4 generation. If one allele was indeed repaired by HR in the whole plant and the other one not, we would expect a Mendelian distribution of the trichome phenotype with a ratio of non-glabrous:glabrous plants of 3:1. Indeed, the progeny of all three plants did not deviate significantly from the expected 3:1 ratio as calculated by chi-squared test (χ^2; $P > 0.5$, Tab. 2).

6

307 Sequencing of 125 progeny plants from all three T3 plants with or without trichomes revealed
308 that the trichome phenotype was always associated with at least one repaired *GL1* allele,
309 while the glabrous plants mostly contained various small indel mutations (Fig. S3). Among
310 the 125 sequenced plants, we detected 30 plants that showed a homozygous insertion of the
311 10 nucleotides according to the sequencing histogram (Fig. 5). The small indel mutations
312 found in glabrous plants or in heterozygous non-glabrous plants were in many cases not
313 consistent with the indel alleles found in the progenitor plants, indicating that only the
314 repaired *GL1* allele was already fixed in the T3 generation and that the other sequenced allele
315 in the T3 plants was caused by somatic mutations in the leaf used for gDNA extraction. This
316 putative chimerism in the T3 generation plants corresponds to the presence of more than two
317 sequence peaks in some of the sequencing histograms of the T3 parental plants (Fig. S3).
318 In summary, we screened between 2,200 to 2,500 plants containing the *gl1* repair constructs
319 pIPGT-Nuc, pVIR-Nuc, and pVIR-Nick, respectively, over three generations. Using the viral
320 replicon approach, we could not regenerate non-glabrous plants. Using the IPGT approach,
321 we were able to detect three stable GT events in the T3 generation with one functional *GL1*
322 allele.
323
324 **DISCUSSION**
325
326 **Trichome formation is a marker for mutagenic efficiency**
327 The Cas9 system has become a valuable and efficient tool for genome modifications in plants
328 (Arora and Narula, 2017). However, while random mutagenesis in plants by error-prone DSB
329 repair through the NHEJ-pathway has been successfully applied in many studies (Liu et al.,
330 2017), introduction of targeted mutations and gene knock-ins by GT remains challenging.
331 To explore the efficiency of two systems that have been described to induce HR in plants, we
332 used a Cas9-generated trichomeless *Arabidopsis* mutant (Hahn et al., 2017). This mutant
333 contains a 10 bp deletion (Fig S1) in the *GL1* gene, which leads to a premature STOP codon.
334 The trichome phenotype (Fig. 3) allowed easy selection of transformed plants, which were
335 likely to produce completely repaired offspring (Fig. 4).
336 An important aspect of our marker system is that DSB repair has to occur via HR and not via
337 NHEJ to induce trichome formation. Small indels of 1-2 nucleotides generated by error-prone
338 NHEJ might restore the open reading frame (ORF) of the gene, but at least three amino acid
339 residues are still missing in the MYB-binding domain of the GL1 protein, which are critical
340 for protein function (Hauser et al., 2001). As NHEJ mediated repair of DSB rarely results in
341 insertions of more than one nucleotide (Fauser et al., 2014), it is highly unlikely that NHEJ
342 mediated repair can restore trichome formation. Consequently, in a large-scale screening of
343 glabrous and non-glabrous T4 plants (Fig. S4), we detected a clear correlation of trichome
344 formation and repair via HR, while NHEJ mutations did not result in non-glabrous plants even
345 though the ORF of the gene was partly restored in some cases. Previously used markers, such
346 as ß-glucuronidase (GUS) or fluorescent proteins need first to be introduced into the plant and
347 GT events can only be detected by staining or fluorescence microscopy. In contrast, our
348 endogenous marker gene has two advantages: First, it allows to follow the mutagenesis
349 progress by eye and second, trichome absence does not have pleiotropic effects on plant
350 development under greenhouse conditions (Oppenheimer et al., 1991).
351
352 **Activation of the homology template by viral replicons did not result in gene targeting**
353 **events**
354 Baltes et al. (2014) developed a GT approach in tobacco that employs geminivirus-based
355 replicons to active the HT (Fig. 2). In a first approach in *Arabidopsis*, they aimed at inserting
356 a short DNA sequence into an endogenous gene by GT using a HT, which was activated by
357 replication in recombinant bipartite cabbage leaf curl virus genomes. While GT events could

7

358   be detected on somatic level by PCR, the authors did not report inheritance of the GT events
359   to progeny plants (Baltes et al., 2014). In further experiments, several groups used
360   deconstructed virus replicons of BeYDV (or other viruses) depleted of the coat and movement
361   proteins. These can be stably transformed into plants and have an increased vector cargo
362   capacity, allowing the use of larger HT. This approach allowed efficient GT in various crop
363   plants, including tobacco, tomato, potato, rice and wheat (Baltes et al., 2014; Butler et al.,
364   2016; Čermák et al., 2015; Gil-Humanes et al., 2017; Wang et al., 2017).
365   While the initial experiments with the full virus demonstrated that GT using viral replicons
366   can be achieved in *Arabidopsis*, no further studies reported the use of deconstructed replicons
367   in *Arabidopsis*. We were also unable to detect GT events in form of trichome regeneration in
368   2,334 analyzed plants from three independent lines transformed with pVIR-Nuc except a
369   single trichome (Tab. 1, Fig. 3). One possible explanation could be that the sgRNA that we
370   used did not allow efficient DSB induction in the target locus. Since we generated successful
371   GT events with the IPGT approach, using the same sgRNA, which requires efficient Cas9
372   cutting for HT activation, we do not consider this as likely. As we detected high frequencies
373   of NHEJ mutations next to GT events while sequencing T4 generation plants (Fig. S4), we
374   conclude that our sgRNA is highly active. It could be possible that replicon formation is not
375   efficient in *Arabidopsis* in contrast to other plant species. However, BeYDV infects
376   *Arabidopsis* (Liu et al., 1997). Thus, replicon formation should be possible. We furthermore
377   detected efficient replicon formation on DNA level in T1 plants (Fig. S2). Another option
378   could be an *Arabidopsis*-specific discrimination of the repair machinery for circular HT as
379   these do not occur naturally in the plant nucleus.
380   Differences between plant species with handling of foreign DNA molecules have already
381   been observed in *Arabidopsis* and tobacco (Orel and Puchta, 2003). It has been suggested that
382   the stimulatory effect of viral replicons on GT is not only caused by the amount of HT but
383   also by the impact of RepA on the plant cell cycle. RepA stimulates entry of cells into the S-
384   phase by interaction with plant proteins that control cell cycle so that the virus can use the
385   necessary factors for replicon replication that the cell produces for DNA synthesis (Baltes et
386   al., 2014). Cells in S-Phase show an increased HR efficiency in eukaryotes (Saleh-Gohari and
387   Helleday, 2004; Mathiasen and Lisby, 2014). Possibly, this effect is less pronounced in
388   *Arabidopsis* than in other plants.
389   In contrast to most of the studies using viral replicons, we did not include the SSN in the
390   replicating unit. Baltes and colleagues (2014) reported comparable GT efficiencies if the ZFN
391   was part of the replicating unit or not, and similar results were found in a Cas9-based
392   approach in tomato (Čermák et al., 2015). However, it might be possible that a high amount
393   of SSN is more critical in *Arabidopsis*.
394   Donor architecture might also account for variable efficiencies in GT. We used HA with a
395   length of app. 800 bp, which is in the range of the most successful GT studies so far (Fauser et
396   al., 2012; Baltes et al., 2014; Schiml et al., 2014; Čermák et al., 2015; Butler et al., 2016; Gil-
397   Humanes et al., 2017; Wang et al., 2017). Additionally, we used the viral components
398   described in the original report of Baltes et al. (2014) and just integrated our HT. Even though
399   we used a different vector backbone (Fauser et al., 2014), the replicons should therefore be
400   comparable to the ones described in successful GT studies by Baltes et al. (2014) and Čermák
401   et al. (2015).
402   While we could not detect considerable trichome formation in our experiments, it has to be
403   taken into account that GT events might occur within cells, which are not determined to
404   develop into trichomes. That is, using trichome formation as readout for mutagenic
405   efficiencies, we actually underestimate the amount of successful GT events on the somatic
406   cell level. PCR-based methods, as used for GT detection in *Arabidopsis* using cabbage leaf
407   curl virus genomes (Baltes et al., 2014) will therefore allow detection of more GT events. In
408   other experiments of our group with a different target gene (unpublished data), we actually

8

409 could detect GT events using PCR-based approaches on somatic level at low frequencies but
410 we were also unable to regenerate inheritable events Taken together, these data indicate that
411 viral replicons allow GT only at low frequencies in *Arabidopsis*, which contrasts with the
412 considerable success in many crop species. Further experiments are needed to unravel the
413 reasons for these differences so that this promising technique might also be applicable in this
414 important model organism.
415

Cas9 nickase is not improving HR events in combination with viral replicons
416
417 We tested whether Cas9 nickase can induce GT at higher efficiencies than Cas9 nuclease as
418 proposed by Fauser et al. (2014). We only applied the Cas9 nickase in combination with the
419 viral replicon approach, since IPGT needs a nuclease to release the HR template. We did not
420 detect GT events in form of trichome formation on any plant (Tab. 1). However, when we
421 combined   the   viral   replicon   approach   with   a   Cas9   nuclease,
422 we also detected only a single putative GT event in ~2,300 plants screened (Tab. 1, Fig. 3).
423 Therefore, it is likely that the viral replicon approach does not trigger GT in
424 *Arabidopsis*, independent from its Cas9 variant.
425 However, besides the initial report of Fauser et al. (2014), reports on GT using nickases in
426 plants are scarce. For their GT experiments, Fauser and colleagues used a reporter construct
427 (Orel et al., 2003) containing a disrupted *GUS* ORF as well as a homologous sequence, which
428 can function as HT for repair of the *GUS* ORF upon induction of a DSB by Cas9. Up to 4
429 times more HR events were detected compared to experiments with Cas9 nuclease.
430 Importantly, only somatic mutations were quantified and information on the regeneration of
431 stably mutated plants was not provided. It has been observed earlier, that GT efficiencies
432 increase if target locus and donor construct are located on the same chromosome (Fauser et
433 al., 2012). In the GUS experiments described above, target locus and HT are directly adjacent,
434 which possibly explains the successful HR events. In our case, the HT integrated randomly
435 into the genome and even after replicon formation, the replicons still had to "find" the nick
436 site to be available for HR. Supporting this hypothesis, other studies detected lower GT
437 efficiency when using nickases compared to nucleases (Ran et al., 2013; Čermák et al., 2017).
438 In a recent study, Cas9 nickase was used in combination with viral replicons to repair a
439 dysfunctional *GUS* gene in tobacco. Using two sgRNAs to create adjacent nicks, a 2-fold
440 increase in GT was observed (Čermák et al., 2017). However, paired nicks on different DNA
441 strands in close proximity are considered to induce a DSB with single strand overhangs
442 (Schiml et al., 2014). Thus, it is questionable if nick HR repair pathways were indeed
443 responsible for the successful GT in that study. In human cells, DNA nick repair via HR
444 occurs 8 times more frequently on the transcribed (non-coding) DNA strand, probably due to
445 stimulation by the transcription machinery (Davis and Maizels, 2014). The Cas9 nickase used
446 in our study only contains a functional HNH-cleavage domain so that the nick is created on
447 the non-coding strand. However, whether the strand preference in plants is comparable to that
448 in human cells remains to be elucidated and might provide a possibility to enhance HR
449 frequencies.
450 In general, HR repair mechanisms on DNA nicks in plants are not well studied. In humans,
451 two competing HR mechanisms are proposed. The canonical HR pathway resembles the
452 synthesis-dependent strand annealing repair pathway for DSB, while the alternative one
453 preferentially employs single stranded oligonucleotides or nicked dsDNA HT for repair
454 (Davis and Maizels, 2014). The alternative HR pathway shares features with another HR
455 pathway, the single strand annealing pathway, which is induced by micro-homologies at the
456 break ends. As HR via the single strand annealing pathway seems to be more efficient in
457 plants than the synthesis-dependent strand annealing repair pathway (Orel et al., 2003; Fauser
458 et al., 2014), it might be interesting to include nickase target site in the repair template to
459 induce target locus repair via the alternative HR nick repair pathway. However, the detailed

9

460 mechanisms underlying DNA nick repair in plants via HR needs to be elucidated to verify if
461 the same mechanisms as in human cells apply. As the Cas9 nickase has already been shown to
462 induce gene knock in in human embryonic stem cells using dsDNA as template (Rong et al.,
463 2014), it remains an interesting option for future HR experiments *in planta*.
464

465 **The *in planta* gene targeting approach allows regeneration of stable gene targeting**
466 **events in Arabidopsis**
467 The IPGT approach, which excises the HT from the T-DNA backbone by a SSN (Fig. 2), has
468 been used successfully in *Arabidopsis*. By the application of the I-*Sce*I meganuclease a *GUS*
469 marker gene was restored (Fauser et al., 2012), and by the application of the Cas9 system a
470 resistance cassette was integrated into the *ADH* locus (Schiml et al., 2014) and a *GFP* gene
471 introduced into the *TFL1* locus (Zhao et al., 2016). In contrast to these previous studies, we
472 did not exploit an artificial marker gene but restored the gene function of the endogenous *GL1*
473 gene, using trichome formation as direct read-out for GT events.
474 We detected spots of trichomes ranging from single trichomes to fully covered leaves in T2
475 generation plants (Fig. 3, Tab. 1). Other studies reported similar chimeric GT spots using
476 GUS staining (Fauser et al., 2012). The T2 plant, which showed the highest level of
477 chimerism (Fig. 3) produced offspring plants with WT-like trichome patterning. We therefore
478 conclude, that the trichome phenotype can be used to select for plants, in which GT
479 efficiencies are enhanced and which are candidate plants for production of stably mutated
480 progeny. This enhancement can have several reasons. The location of the T-DNA, which
481 contains HR template and SSN gene, in the plant genome likely has influence on GT and
482 close proximity of repair template and DSB locus seems favorable (Fauser et al., 2012).
483 Similarly, we found differences in the frequency of chimeric T2 plants depending on the
484 parental T1 line (Tab. 1). It is therefore highly recommended to start with as many
485 independent transformation events as possible, in contrast to screening a lot of progeny plants
486 from few primary transformants (Schiml et al., 2017).
487 T2 plants might be homozygous or heterozygous for the T-DNA insertion. It has been shown
488 that the heterozygosity state of the HR template influences recombination frequencies (Puchta
489 et al., 1995; Orel et al., 2003). Indeed, since all three plants, which completely restored
490 trichome formation in the T3 generation (Fig. 4, Fig. S3), contained the Cas9-T-DNA cassette
491 and also produced only T-DNA containing offspring, it is likely, that the parental T2 plant
492 contained already two copies of the repair template. Crop improvement usually requires the
493 creation of marker-free plants. Thus, more screening effort might be needed to detect T-DNA
494 free GT plants.
495 It has been reported that one-sided GT events can occur during repair (Schiml et al., 2014), in
496 which one side of the of the DSB is repaired by HR and the other strand is repaired using the
497 stably integrated T-DNA as template. This can lead to integration of large parts of the HT,
498 which finally results in a NHEJ-mediated ligation of the repaired strand and the loose end of
499 the DSB. To detect such events, we had added single bp exchanges in both HA. Our
500 sequencing results of our repaired *GL1* alleles did not contain these bp exchanges, which
501 indicates that we achieved true GT events.
502 In previous studies the frequencies of GT events using IPGT in *Arabidopsis* ranged from
503 0.14-0.83 % (Fauser et al., 2012; Schiml et al., 2014; Zhao et al., 2016) which corresponds
504 well to our GT frequency of 0.12 %. The single bp exchanges in the HA, which allowed us to
505 detect true GT events, might explain the slightly lower efficiencies in our experimental
506 approach. Additionally, efficiencies in Cas9 experiments are difficult to compare due to use of
507 different sgRNAs (Peng et al., 2016).
508

509 **Further options to enhance GT**

510   In this report, we identified three stable GT event out of ~2,500 *Arabidopsis* plants screened
511   across three generations using an IPGT approach. In principle, a GT event at such frequency
512   could be detected exclusively by PCR and without a phenotype based screening. However, we
513   achieved such a high frequency because we were able to select parental lines prone to GT by
514   phenotype. Therefore, it might still be desirable to increase GT efficiencies if no visual
515   marker is used for GT detection. In general, it is advisable to test sgRNA efficiencies in
516   protoplast systems before using them in stable plant transformations as efficient induction of
517   DSB is a prerequisite for successful GT (Puchta et al., 1993).
518   Higher GT efficiencies of ~5.5 % have been achieved in *Arabidopsis* protoplasts using single
519   stranded oligonucleotides as repair template (Sauer et al., 2016) to exchange specific
520   nucleotides in a fluorescent protein to change its emission spectrum. However, single stranded
521   oligonucleotides are not suitable for knock-in of larger DNA fragments or whole genes due to
522   size limitations and are not transformed via Agrobacteria. For targeted exchange of a single
523   bp, the use of base changer Cas9 provides an attractive alternative to HR-mediated repair
524   processes. In this approach, an inactive Cas9 is fused to cytidine deaminase, which converts
525   cytosines to uracil without DNA cutting (Komor et al., 2016; Zong et al., 2017). Recently, an
526   adenine base editor Cas9 has also been described (Gaudelli et al., 2017). While mutation
527   frequencies of up to 50% can be achieved using base changer Cas9, this approach cannot be
528   applied for gene knock-ins.
529   In general, protoplast transformation seems to yield higher mutation frequencies compared to
530   Agrobacterium-mediated transformation (Li et al., 2013; Sauer et al., 2016), since more DNA
531   donor molecules are integrated in the plant cells. However, regeneration of *Arabidopsis* plants
532   from protoplasts is challenging and is therefore not an attractive alternative to floral dipping.
533   Focusing research on improved regeneration protocols is therefore desirable. Several groups
534   tried to enhance HR efficiencies by disrupting either the NHEJ pathway (Maruyama et al.,
535   2015; Zhu et al., 2015) or introducing HR stimulating enzymes from other organisms (Reiss
536   et al., 2000; Shaked et al., 2005; Even-Faitelson et al., 2011). These approaches can
537   destabilize the genome as HR frequencies are enhanced between all repetitive sequences in
538   the genome (Puchta and Fauser, 2013).
539   In the IPGT approach, the HT is excised from the genome. If this happens at an early stage of
540   plant development, the HT might not be passed on during cell division. This could reduce the
541   efficiency of this approach. Using germline-specific promoters (Wang et al., 2015; Mao et al.,
542   2016) for the SSN might increase chances of simultaneous cleavage of target locus and
543   release of the HT specifically in the cells, where GT needs to occur to generate stable
544   offspring.
545   A similar problem is target locus inactivation through error-prone NHEJ repair. Besides the
546   before mentioned nickase approach, the Cas9 variant CpfI (CRISPR from *Prevotella* and
547   *Francisella*) might provide an attractive alternative. CpfI generates DSB more distal to the
548   PAM than Cas9 (Zetsche et al., 2015), thus making the target locus accessible to a second
549   round of cleavage even after NHEJ mediated repair. An interesting approach would be to
550   combine the viral replicon approach with the IPGT approach. Therefore, the HT would be
551   flanked by SSN cutting site and the viral components. By this, large amounts of linear dsDNA
552   molecules would be generated that can be used as repair templates. Since more target sites are
553   present in this approach, we suggest including the gene for the SSN and the sgRNA within the
554   replicon unit, so that sufficient amounts of all genome editing components are available.
555

556   **CONCLUSION**

557   In this report, we applied the Cas9 system to induce GT events in *Arabidopsis*. The HT
558   needed to be activated by an IPGT approach. We were able to stably introduce a short DNA
559   sequence into an endogenous gene and the inserted sequence was transmitted through the

11

560 germline into the next generation. The high GT efficiency allows to use GT approaches also
561 in marker-free approaches in which GT events need to be detected by PCR. Thus, we
562 conclude that Cas9 mediated GT in combination with IPGT is a powerful tool that has high
563 potential to become routinely applied in plant research in the future. Trichome formation as
564 marker for editing efficiency proved to be a reliable system to compare different GT
565 approaches. We suggest that the *GL1* gene can be used for optimization of Cas9-based
566 mutagenesis approaches in future experiments.

567 **AUTHOR CONTRIBUTIONS**
568 FH, OM, ME, and AW designed the study. FH carried out all laboratory experiments. FH,
569 OM, ME, and AW interpreted the data and wrote the manuscript.
570
575 **ABBREVIATIONS**
576 BeYDV = *bean yellow dwarf virus*; bp = base pair(s); Cas9 system = CRISPR/Cas9 system;
577 DSB = double strand break; GT = gene targeting; GUS = ß-glucuronidase; HA = homology
578 arm(s); HR = homologous recombination; HT = homology template; IPGT = *in planta* gene
579 targeting; LIR = long intergenic region; NHEJ = non-homologous end joining; ORF = open
580 reading frame; PAM = protospacer adjacent motif; sgRNA = single guide RNA; SIR = short
581 intergenic region; SSN = sequence specific nuclease(s); TALEN = Transcription Activator-
582 like Effector Nuclease; ZFN = Zinc finger nuclease(s)
583
591 **CONFLICT OF INTERREREST STATEMENT**
592 The authors declare that the research was conducted in the absence of any commercial or
593 financial relationships that could be construed as a potential conflict of interest.
594
595 **REFERENCES**
596 Arora, L., and Narula, A. (2017). Gene Editing and Crop Improvement Using CRISPR-Cas9
597 System. *Front. Plant Sci.* 8, 1932. doi:10.3389/fpls.2017.01932.
598 Ayar, A., Wehrkamp-Richter, S., Laffaire, J.-B., Le Goff, S., Levy, J., Chaignon, S., et al.
599 (2013). Gene targeting in maize by somatic ectopic recombination. *Plant Biotechnol. J.* 11,
600 305–314. doi:10.1111/pbi.12014.
601 Baltes, N. J., Gil-Humanes, J., Cermak, T., Atkins, P. A., and Voytas, D. F. (2014). DNA
602 replicons for plant genome engineering. *Plant Cell* 26, 151–163. doi:10.1105/tpc.113.119792.
603 Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and
604 beyond. *Biotechnol. Adv.* 33, 41–52. doi:10.1016/j.biotechadv.2014.12.006.
605 Butler, N. M., Baltes, N. J., Voytas, D. F., and Douches, D. S. (2016). Geminivirus-Mediated
606 Genome Editing in Potato (Solanum tuberosum L.) Using Sequence-Specific Nucleases.
607 *Front. Plant Sci.* 7, 1045. doi:10.3389/fpls.2016.01045.
608 Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., and Voytas, D. F. (2015). High-frequency,
609 precise modification of the tomato genome. *Genome Biol.* 16, 232. doi:10.1186/s13059-015-

610   0796-9.

611   Čermák, T., Curtin, S. J., Gil-Humanes, J., Čegan, R., Kono, T. J. Y., Konečná, E., et al.
612   (2017). A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. *Plant
613   Cell* 29, 1196–1217. doi:10.1105/tpc.16.00922.

614   Chen, Q., He, J., Phoolcharoen, W., and Mason, H. S. (2011). Geminiviral vectors based on
615   bean yellow dwarf virus for production of vaccine antigens and monoclonal antibodies in
616   plants. *Hum. Vaccin.* 7, 331–338.

617   Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-
618   mediated transformation of Arabidopsis thaliana. *Plant J.* 16, 735–743. doi:10.1046/j.1365-
619   313x.1998.00343.x.

620   Davis, L., and Maizels, N. (2014). Homology-directed repair of DNA nicks via pathways
621   distinct from canonical double-strand break repair. *Proc. Natl. Acad. Sci. U. S. A.* 111, E924-
622   932. doi:10.1073/pnas.1400236111.

623   Doudna, J. A., and Charpentier, E. (2014). Genome editing. The new frontier of genome
624   engineering with CRISPR-Cas9. *Science* 346, 1258096. doi:10.1126/science.1258096.

625   Dueva, R., and Iliakis, G. (2013). Alternative pathways of non-homologous end joining
626   (NHEJ) in genomic instability and cancer. *Transl. Cancer Res.* 2, 163–177.
627   doi:10.21037/1152.

628   Even-Faitelson, L., Samach, A., Melamed-Bessudo, C., Avivi-Ragolsky, N., and Levy, A. A.
629   (2011). Localized egg-cell expression of effector proteins for targeted modification of the
630   Arabidopsis genome. *Plant J.* 68, 929–937. doi:10.1111/j.1365-313X.2011.04741.x.

631   Fauser, F., Roth, N., Pacher, M., Ilg, G., Sánchez-Fernández, R., Biesgen, C., et al. (2012). In
632   planta gene targeting. *Proc. Natl. Acad. Sci. U. S. A.* 109, 7535–7540.
633   doi:10.1073/pnas.1202191109.

634   Fauser, F., Schiml, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and
635   nickases can be used efficiently for genome engineering in Arabidopsis thaliana. *Plant J. Cell
636   Mol. Biol.* 79, 348–359. doi:10.1111/tpj.12554.

637   Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al.
638   (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage.
639   *Nature* 551, 464. doi:10.1038/nature24644.

640   Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C. V., Sánchez-León, S., et al.
641   (2017). High-efficiency gene targeting in hexaploid wheat using DNA replicons and
642   CRISPR/Cas9. *Plant J.* 89, 1251–1262. doi:10.1111/tpj.13446.

643   Gorbunova, V., and Levy, A. A. (1997). Non-homologous DNA end joining in plant cells is
644   associated with deletions and filler DNA insertions. *Nucleic Acids Res.* 25, 4650–4657.

645   Hahn, F., Mantegazza, O., Greiner, A., Hegemann, P., Eisenhut, M., and Weber, A. P. M.
646   (2017). An Efficient Visual Screen for CRISPR/Cas9 Activity in Arabidopsis thaliana. *Front.
647   Plant Sci.* 8. doi:10.3389/fpls.2017.00039.

648   Hauser, M. T., Harr, B., and Schlötterer, C. (2001). Trichome distribution in Arabidopsis
649   thaliana and its close relative Arabidopsis lyrata: molecular analysis of the candidate gene
650   GLABROUS1. *Mol. Biol. Evol.* 18, 1754–1763.

651   Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A
652   programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*
653   337, 816–821. doi:10.1126/science.1225829.

654   Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable
655   editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533,
656   420–424. doi:10.1038/nature17946.

657   Koornneef, M., Dellaert, L. W., and van der Veen, J. H. (1982). EMS- and radiation-induced
658   mutation frequencies at individual loci in Arabidopsis thaliana (L.) Heynh. *Mutat. Res.* 93,
659   109–123.

660   Kumar, S., Worden, A., Novak, S., Lee, R., and Petolino, J. F. (2016). A trait stacking system

661   via intra-genomic homologous recombination. *Planta* 244, 1157–1166. doi:10.1007/s00425-
662   016-2595-2.
663   Lee, K. Y., Lund, P., Lowe, K., and Dunsmuir, P. (1990). Homologous recombination in plant
664   cells after Agrobacterium-mediated transformation. *Plant Cell* 2, 415–425.
665   doi:10.1105/tpc.2.5.415.
666   Li, J.-F., Aach, J., Norville, J. E., McCormack, M., Zhang, D., Bush, J., et al. (2013).
667   Multiplex and homologous recombination-mediated plant genome editing via guide
668   RNA/Cas9. *Nat. Biotechnol.* 31, 688–691. doi:10.1038/nbt.2654.
669   Li, Z., Liu, Z.-B., Xing, A., Moon, B. P., Koellhoffer, J. P., Huang, L., et al. (2015). Cas9-
670   Guide RNA Directed Genome Editing in Soybean. *Plant Physiol.* 169, 960–970.
671   doi:10.1104/pp.15.00783.
672   Liu, L., van Tonder, T., Pietersen, G., Davies, J. W., and Stanley, J. (1997). Molecular
673   characterization of a subgroup I geminivirus from a legume in South Africa. *J. Gen. Virol.* 78
674   (Pt 8), 2113–2117. doi:10.1099/0022-1317-78-8-2113.
675   Liu, X., Wu, S., Xu, J., Sui, C., and Wei, J. (2017). Application of CRISPR/Cas9 in plant
676   biology. *Acta Pharm. Sin. B* 7, 292–302. doi:10.1016/j.apsb.2017.01.002.
677   Mao, Y., Zhang, Z., Feng, Z., Wei, P., Zhang, H., Botella, J. R., et al. (2016). Development of
678   germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene
679   modifications in Arabidopsis. *Plant Biotechnol. J.* 14, 519–532. doi:10.1111/pbi.12468.
680   Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., and Ploegh, H.
681   L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by
682   inhibition of nonhomologous end joining. *Nat. Biotechnol.* 33, 538. doi:10.1038/nbt.3190.
683   Mathiasen, D. P., and Lisby, M. (2014). Cell cycle regulation of homologous recombination
684   in Saccharomyces cerevisiae. *FEMS Microbiol. Rev.* 38, 172–184. doi:10.1111/1574-
685   6976.12066.
686   Miao, Z. H., and Lam, E. (1995). Targeted disruption of the TGA3 locus in Arabidopsis
687   thaliana. *Plant J. Cell Mol. Biol.* 7, 359–365.
688   Offringa, R., de Groot, M. J., Haagsman, H. J., Does, M. P., van den Elzen, P. J., and
689   Hooykaas, P. J. (1990). Extrachromosomal homologous recombination and gene targeting in
690   plant cells after Agrobacterium mediated transformation. *EMBO J.* 9, 3077–3084.
691   Oppenheimer, D. G., Herman, P. L., Sivakumaran, S., Esch, J., and Marks, M. D. (1991). A
692   myb gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules.
693   *Cell* 67, 483–493.
694   Orel, N., Kyryk, A., and Puchta, H. (2003). Different pathways of homologous recombination
695   are used for the repair of double-strand breaks within tandemly arranged sequences in the
696   plant genome. *Plant J.* 35, 604–612. doi:10.1046/j.1365-313X.2003.01832.x.
697   Orel, N., and Puchta, H. (2003). Differences in the processing of DNA ends in Arabidopsis
698   thaliana and tobacco: possible implications for genome evolution. *Plant Mol. Biol.* 51, 523–
699   531.
700   Paszkowski, J., Baur, M., Bogucki, A., and Potrykus, I. (1988). Gene targeting in plants.
701   *EMBO J.* 7, 4021–4026.
702   Pattanaik, S., Patra, B., Singh, S. K., and Yuan, L. (2014). An overview of the gene regulatory
703   network controlling trichome development in the model plant, Arabidopsis. *Front. Plant Sci.*
704   5. doi:10.3389/fpls.2014.00259.
705   Peng, R., Lin, G., and Li, J. (2016). Potential pitfalls of CRISPR/Cas9-mediated genome
706   editing. *FEBS J.* 283, 1218–1231. doi:10.1111/febs.13586.
707   Puchta, H. (1998). Repair of genomic double-strand breaks in somatic plant cells by one-sided
708   invasion of homologous sequences. *Plant J.* 13, 331–339. doi:10.1046/j.1365-
709   313X.1998.00035.x.
710   Puchta, H. (2002). Gene replacement by homologous recombination in plants. *Plant Mol.*
711   *Biol.* 48, 173–182.

712 Puchta, H. (2005). The repair of double-strand breaks in plants: mechanisms and
713 consequences for genome evolution. *J. Exp. Bot.* 56, 1–14. doi:10.1093/jxb/eri025.
714 Puchta, H., Dujon, B., and Hohn, B. (1993). Homologous recombination in plant cells is
715 enhanced by in vivo induction of double strand breaks into DNA by a site-specific
716 endonuclease. *Nucleic Acids Res.* 21, 5034–5040.
717 Puchta, H., Dujon, B., and Hohn, B. (1996). Two different but related mechanisms are used in
718 plants for the repair of genomic double-strand breaks by homologous recombination. *Proc.*
719 *Natl. Acad. Sci. U. S. A.* 93, 5055–5060.
720 Puchta, H., and Fauser, F. (2013). Gene targeting in plants: 25 years later. *Int. J. Dev. Biol.*
721 57, 629–637. doi:10.1387/ijdb.130194hp.
722 Puchta, H., Swoboda, P., Gal, S., Blot, M., and Hohn, B. (1995). Somatic intrachromosomal
723 homologous recombination events in populations of plant siblings. *Plant Mol. Biol.* 28, 281–
724 292.
725 Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., et al.
726 (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing
727 Specificity. *Cell* 154, 1380–1389. doi:10.1016/j.cell.2013.08.021.
728 Reiss, B., Schubert, I., Köpchen, K., Wendeler, E., Schell, J., and Puchta, H. (2000). RecA
729 stimulates sister chromatid exchange and the fidelity of double-strand break repair, but not
730 gene targeting, in plants transformed by Agrobacterium. *Proc. Natl. Acad. Sci. U. S. A.* 97,
731 3358–3363.
732 Risseeuw, E., Offringa, R., Franke-van Dijk, M. E., and Hooykaas, P. J. (1995). Targeted
733 recombination in plants using Agrobacterium coincides with additional rearrangements at the
734 target locus. *Plant J. Cell Mol. Biol.* 7, 109–119.
735 Rong, Z., Zhu, S., Xu, Y., and Fu, X. (2014). Homologous recombination in human
736 embryonic stem cells using CRISPR/Cas9 nickase and a long DNA donor template. *Protein*
737 *Cell* 5, 258–260. doi:10.1007/s13238-014-0032-5.
738 Saleh-Gohari, N., and Helleday, T. (2004). Conservative homologous recombination
739 preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human
740 cells. *Nucleic Acids Res.* 32, 3683–3688. doi:10.1093/nar/gkh703.
741 Sargent, R. G., Brenneman, M. A., and Wilson, J. H. (1997). Repair of site-specific double-
742 strand breaks in a mammalian chromosome by homologous and illegitimate recombination.
743 *Mol. Cell. Biol.* 17, 267–277.
744 Sauer, N. J., Narváez-Vásquez, J., Mozoruk, J., Miller, R. B., Warburg, Z. J., Woodward, M.
745 J., et al. (2016). Oligonucleotide-Mediated Genome Editing Provides Precision and Function
746 to Engineered Nucleases and Antibiotics in Plants. *Plant Physiol.* 170, 1917–1928.
747 doi:10.1104/pp.15.01696.
748 Schiml, S., Fauser, F., and Puchta, H. (2014). The CRISPR/Cas system can be used as
749 nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in
750 Arabidopsis resulting in heritable progeny. *Plant J. Cell Mol. Biol.* 80, 1139–1150.
751 doi:10.1111/tpj.12704.
752 Schiml, S., Fauser, F., and Puchta, H. (2017). CRISPR/Cas-Mediated In Planta Gene
753 Targeting. *Methods Mol. Biol. Clifton NJ* 1610, 3–11. doi:10.1007/978-1-4939-7003-2_1.
754 Shaked, H., Melamed-Bessudo, C., and Levy, A. A. (2005). High-frequency gene targeting in
755 Arabidopsis plants expressing the yeast RAD54 gene. *Proc. Natl. Acad. Sci. U. S. A.* 102,
756 12265–12269. doi:10.1073/pnas.0502601102.
757 Steinert, J., Schiml, S., and Puchta, H. (2016). Homology-based double-strand break-induced
758 genome engineering in plants. *Plant Cell Rep.* 35, 1429–1438. doi:10.1007/s00299-016-1981-
759 3.
760 Sun, Y., Zhang, X., Wu, C., He, Y., Ma, Y., Hou, H., et al. (2016). Engineering Herbicide-
761 Resistant Rice Plants through CRISPR/Cas9-Mediated Homologous Recombination of
762 Acetolactate Synthase. *Mol. Plant* 9, 628–631. doi:10.1016/j.molp.2016.01.001.

15

763 Timmermans, M. C., Das, O. P., and Messing, J. (1992). Trans replication and high copy
764 numbers of wheat dwarf virus vectors in maize cells. *Nucleic Acids Res.* 20, 4047–4054.
765 Voytas, D. F. (2013). Plant genome engineering with sequence-specific nucleases. *Annu. Rev.*
766 *Plant Biol.* 64, 327–350. doi:10.1146/annurev-arplant-042811-105552.
767 Wang, M., Lu, Y., Botella, J. R., Mao, Y., Hua, K., and Zhu, J.-K. (2017). Gene Targeting by
768 Homology-Directed Repair in Rice Using a Geminivirus-Based CRISPR/Cas9 System. *Mol.*
769 *Plant* 10, 1007–1010. doi:10.1016/j.molp.2017.03.002.
770 Wang, Z.-P., Xing, H.-L., Dong, L., Zhang, H.-Y., Han, C.-Y., Wang, X.-C., et al. (2015).
771 Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous
772 mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biol.* 16.
773 doi:10.1186/s13059-015-0715-0.
774 Weigel, D., and Glazebrook, J. (2009). Quick miniprep for plant DNA isolation. *Cold Spring*
775 *Harb. Protoc.* 2009, pdb.prot5179. doi:10.1101/pdb.prot5179.
776 Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S.,
777 Essletzbichler, P., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a Class 2
778 CRISPR-Cas system. *Cell* 163, 759–771. doi:10.1016/j.cell.2015.09.038.
779 Zhang, K., Raboanatahiry, N., Zhu, B., and Li, M. (2017). Progress in Genome Editing
780 Technology and Its Application in Plants. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.00177.
781 Zhao, Y., Zhang, C., Liu, W., Gao, W., Liu, C., Song, G., et al. (2016). An alternative strategy
782 for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Sci. Rep.* 6, 23890.
783 doi:10.1038/srep23890.
784 Zhu, L., Mon, H., Xu, J., Lee, J. M., and Kusakabe, T. (2015). CRISPR/Cas9-mediated
785 knockout of factors in non-homologous end joining pathway enhances gene targeting in
786 silkworm cells. *Sci. Rep.* 5. doi:10.1038/srep18103.
787 Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise base editing in
788 rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* 35, 438.
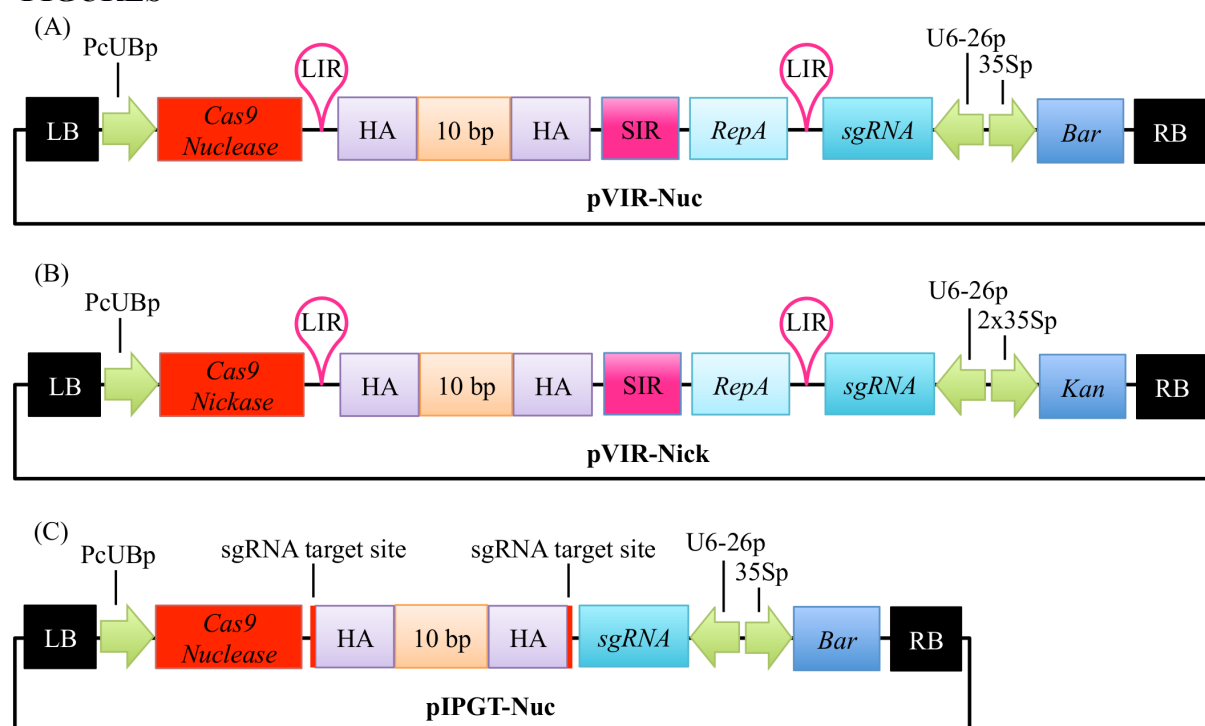789 doi:10.1038/nbt.3811.

790
791 **FIGURES**

792
793 **Figure 1: Vector design for repair of the *gl1* gene**. Single T-DNA vectors were designed to
794 repair the dysfunctional *gl1* gene. (A) pVIR-Nuc contains the Cas9 Nuclease gene (red box)

795 under a constitutive *Ubiquitin6* promoter from parsley (PcUBp) and the sgRNA cassette
796 (petrol box) controlled by the *U6-26* promoter (U6-26p). A Basta resistance cassette (*Bar*,
797 dark blue box) is included as selection marker. The repair template itself consists of 10
798 nucleotides (10 bp, CTGCCGTTTA, orange box), which should restore the reading frame of
799 the *gl1* gene flanked by homology arms (HA, purple box). Rolling circle replication of the HT
800 is ensured by the flanking long intergenic region (LIR, pink hairpin), short intergenic region
801 (pink box, SIR), and the gene encoding the replicase *RepA* (light blue box). (B) pVIR-Nick
802 contains similar features but harbors the gene for the Cas9 nickase instead of the nuclease and
803 a kanamycin resistance cassette (*Kan*, dark blue box) as selection marker. (C) pIPGT-Nuc
804 contains the gene for the Cas9 nuclease, the sgRNA cassette, the Basta resistance cassette and
805 the repair template with HA flanked by target sites for the Cas9 nuclease (red lines). RB =
806 right T-DNA border, 35Sp = *Cauliflower Mosaic Virus* 35S promoter, LB = left T-DNA
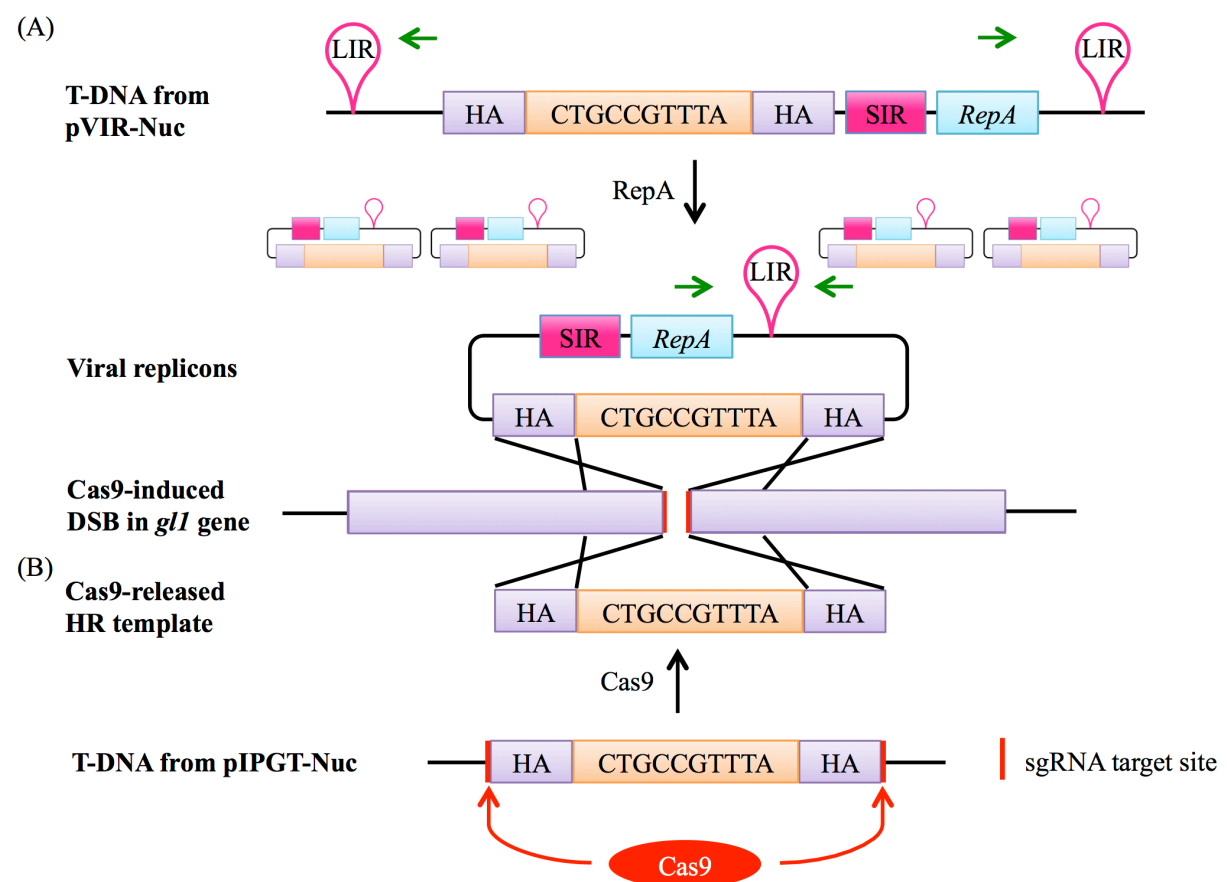807 border. Size not to scale.
808

809
810 **Figure 2: Two different mechanisms enhance the availability of the HT in the plant cell**
811 **nucleus**. (A) If the T-DNA from pVIR-Nuc is integrated into the plant genome, RepA is
812 expressed and initiates rolling circle endoreplication of the HT leading to thousands of viral
813 replicons. The presence of viral replicons can be detected by PCR using primers that only
814 generate an amplicon on the circular replicons but not on the T-DNA (green arrows). The HT
815 can then attach to the DSB site in the *gl1* gene induced by Cas9 due to homology between the
816 HA and the genomic regions next to the DSB (purple). Repair of the DSB by HR will then
817 lead to integration of the 10 bp, which restores the ORF. pVIR-Nick functions accordingly,
818 except that Cas9 nickase only induces a DNA nick in the *gl1* gene. (B) The HT in the T-DNA
819 of pIPGT-Nuc is flanked by the same sgRNA target site that is present in the *gl1* gene.
820 Expression of Cas9 will therefore simultaneously target the *gl1* gene and release the HT,
821 which can then attach to the DSB and function as repair template.
822
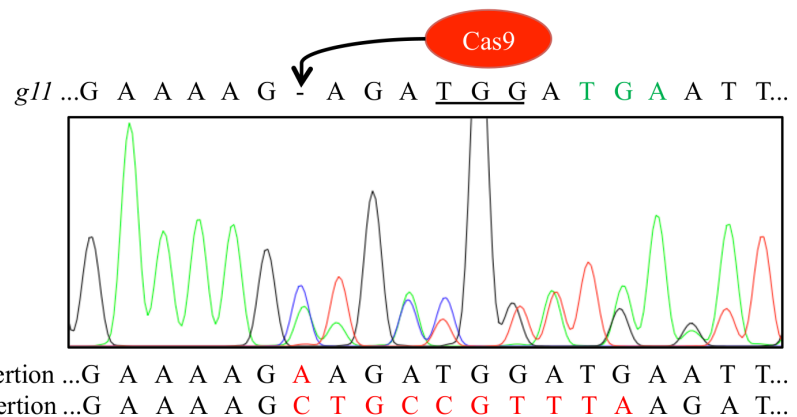
17

(A)

(B)

gl1	... CTA[N$_{53}$]GGCTAAAGAGATGTGGGAAAAG――――――――――AGATGGATGAA...	19/42 clones
+G	... CTA[N$_{53}$]GGCTAAAGAGATGTGGGAAAAG――――――――GAGATGGATGAA...	2/42 clones
+T	... CTA[N$_{53}$]GGCTAAAGAGATGTGGGAAAAG――――――――TAGATGGATGAA...	6/42 clones
+A	... CTA[N$_{53}$]GGCTAAAGAGATGTGGGAAAAG――――――――AAGATGGATGAA...	4/42 clones
-AG	... CTA[N$_{53}$]GGCTAAAGAGATGTGGGAAAAG―――――――――ATGGATGAA...	4/42 clones
-14 bp	... CTA[N$_{53}$]GGCTAAAGAGATG――――――――――――――――――GATGAA...	1/42 clones
Indel	... CTA―――――――AA―――――T―――――AAAG―――――――AGATGGATGAA...	1/42 clones
+10 bp	... CTA[N$_{53}$]GGCTAAAGAGATGTGGGAAAAGCTGCCGTTTAAGATGGATGAA...	4/42 clones

Figure 3: T2 generation plants show spots of trichomes. (A) Representative images of T2 plants transformed with pIPGT-Nuc (red frame) or pVIR-Nuc (blue frame) showing spots of trichomes (yellow circles) ranging from one single trichome to fully covered leaves. (B) The *gl1* gene was amplified from leaves with trichomes, subcloned and sequenced. Sequence analysis revealed clones with various indel mutations next to clones containing the desired 10 bp insertion (red letters). PAM sequence is underlined, STOP codon in *gl1* line is marked in green letters, Cas9 target site is marked in blue letters.

(A)

(B)

gl1 ...G A A A A G - A G A T G G A T G A A T T...

+A insertion ...G A A A A G A A G A T G G A T G A A T T...
+10 bp insertion ...G A A A A G C T G C C G T T T A G A T...
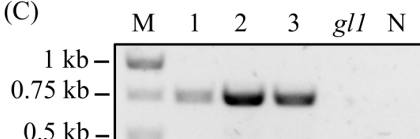
(C)

Figure 4: Wildtype-like trichome patterning was detected in three T3 generation plants. (A) Exemplary picture of one of the three T3 plants (#1) with full trichome covering. (B) We verified the repair of the *gl1* gene by amplifying the *gl1* gene and sequencing the PCR amplicon. Double peaks appeared at the site of Cas9 cleavage hinting at a chimeric or biallelic mutation, the peaks corresponded to an insertion of the ten bps (red) from the HT and to an adenine insertion. Sequence of the dysfunctional *gl1* gene is given on top of the sequencing histogram for comparison. Peak colors: Adenine = green, cytosine = blue, guanine = black, thymine = red; PAM sequence underlined, premature STOP codon marked in green letters. (C) Presence of the *Cas9* gene in the three non-glabrous plants was detected by PCR using *Cas9*-specific primers. All three plants revealed PCR amplicons at the expected size of 749 bp. In contrast, DNA from the *gl1* background line did not yield a PCR Cas9 signal. Image of agarose gel was color inverted for better visibility. M = Marker, N = water control.

gl1 ...G A A A A G A G A T G G A T G A A T T A...
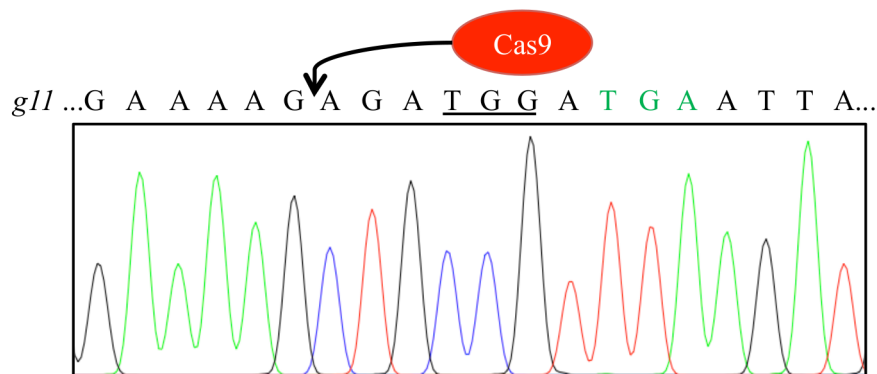
+10 bp insertion ...G A A A A G C T G C C G T T T A A G A T...

Figure 5. Sequencing histogram of T4 generation plant reveals homozygous integration of the 10 nucleotides thereby restoring the *GL1* reading frame. Sequence of the dysfunctional *gl1* gene is given on top of the sequencing histogram for comparison. Peak colors: Adenine = green, cytosine = blue, guanine = black, thymine = red; PAM sequence underlined, premature STOP codon marked in green letters.

TABLES

Table 1: Trichome appearance in the T2 generation

Repair construct	Transformant line	Number of analyzed T2 plants	Number of plants with trichome spots (% of T2 plants)
pIPGT-Nuc	#1	378	4 (1.1)
	#2	368	2 (0.5)
	#3	683	0 (0)
pVIR-Nuc	#1	728	0 (0)
	#2	884	1 (0.1)
	#3	514	0 (0)
pVIR-Nick	#1	925	0 (0)
	#2	528	0 (0)

19

	#3	826	0 (0)

857

858

859 **Table 2: Segregation pattern of trichome phenotype in T4 generation plants**
860 **transformed with pIPGT-Nuc.** Progenitor T3 plant number according to Fig. 4 and Fig. S3.

Progenitor T3 plant	Number of plants with trichomes	Number of plants without trichomes	Ratio of plants with/without trichomes	χ^2 value	P value
#1	164	55	2.98/1	0.0015	> 0.9
#2	150	52	2.88/1	0.0594	> 0.5
#3	210	75	2.80/1	0.2632	> 0.5

861