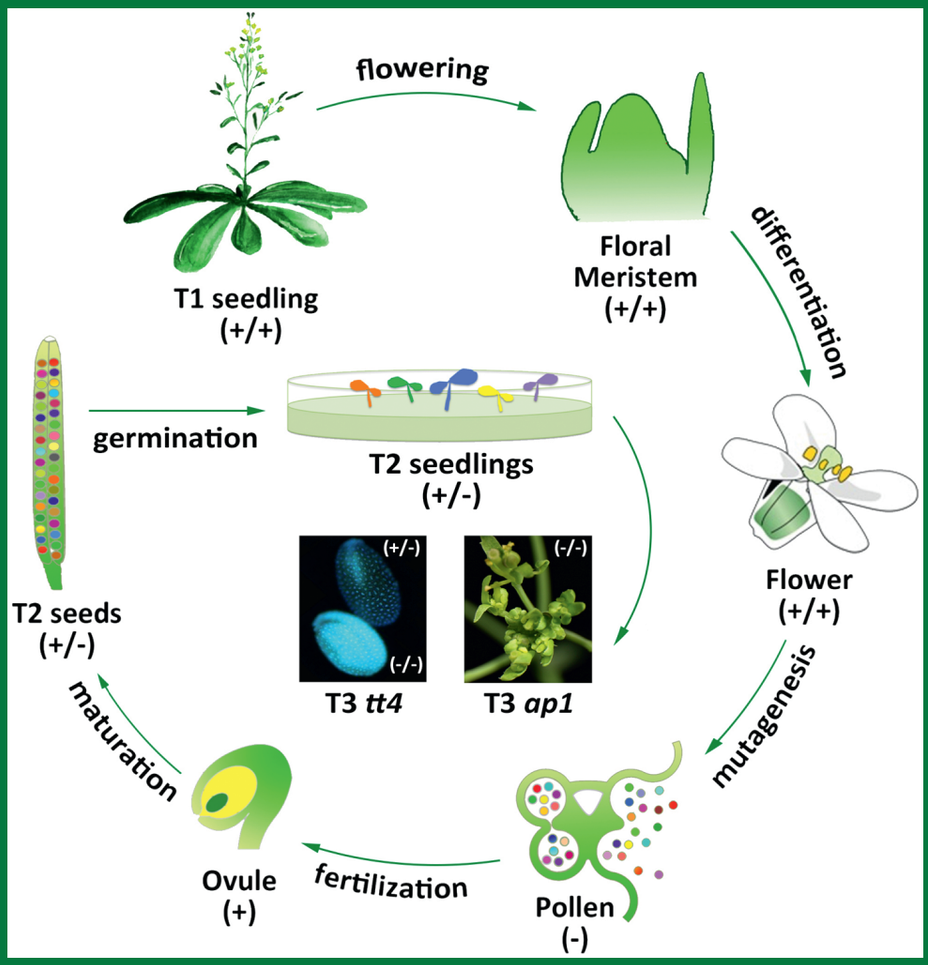
Plant Biotechnology Journal

Open Access

Volume 14 • Issue 2 • February 2016



www.plantbiotechjournal.com | ISSN 1467–7644

WILEY

Plant Biotechnology Journal

Plant Biotechnology Journal (2016) **14**, pp. 519–532 doi: 10.1111/pbi.12468

Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in *Arabidopsis*

Yanfei Mao[1], Zhengjing Zhang[1,2], Zhengyan Feng[1,2,3], Pengliang Wei[1,2], Hui Zhang[1], José Ramón Botella[4] and Jian-Kang Zhu[1,5,*]

[1]*Shanghai Center for Plant Stress Biology, Chinese Academy of Sciences, Shanghai, China*

[2]*University of Chinese Academy of Sciences (CAS), Shanghai, China*

[3]*Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China*

[4]*School of Agriculture and Food Sciences, University of Queensland, Brisbane, Qld, Australia*

[5]*Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN, USA*

Received 6 July 2015;
revised 30 July 2015;
accepted 5 August 2015.
**Correspondence* (Tel 1-765-496-7601; fax 1-765-494-0391; email jkzhu@purdue.edu)
Y.M., Z.Z and Z.F. contributed equally to this work.

Keywords: CRISPR-Cas9, germ-line-specific, gene modification, *Arabidopsis*.

Summary

The *Streptococcus*-derived CRISPR/Cas9 system is being widely used to perform targeted gene modifications in plants. This customized endonuclease system has two components, the single-guide RNA (sgRNA) for target DNA recognition and the CRISPR-associated protein 9 (Cas9) for DNA cleavage. Ubiquitously expressed CRISPR/Cas9 systems (UC) generate targeted gene modifications with high efficiency but only those produced in reproductive cells are transmitted to the next generation. We report the design and characterization of a germ-line-specific Cas9 system (GSC) for *Arabidopsis* gene modification in male gametocytes, constructed using a *SPOROCYTELESS* (*SPL*) genomic expression cassette. Four loci in two endogenous genes were targeted by both systems for comparative analysis. Mutations generated by the GSC system were rare in T1 plants but were abundant (30%) in the T2 generation. The vast majority (70%) of the T2 mutant population generated using the UC system were chimeras while the newly developed GSC system produced only 29% chimeras, with 70% of the T2 mutants being heterozygous. Analysis of two loci in the T2 population showed that the abundance of heritable gene mutations was 37% higher in the GSC system compared to the UC system and the level of polymorphism of the mutations was also dramatically increased with the GSC system. Two additional systems based on germ-line-specific promoters (pDD45-GT and pLAT52-GT) were also tested, and one of them was capable of generating heritable homozygous T1 mutant plants. Our results suggest that future application of the described GSC system will facilitate the screening for targeted gene modifications, especially lethal mutations in the T2 population.

Introduction

In the past decades, forward genetics approaches have been the method of choice to provide insights into gene function. Random mutagenesis using chemical or physical agents was extensively used to produce large mutant populations that were subsequently screened for phenotypes of interests (Meyerowitz, 1987; Page and Grossniklaus, 2002). Nevertheless, a large number of genes are still annotated as having 'unknown function' and many of the assigned gene functions have not been experimentally validated (Alonso *et al.*, 2003). For these genes, reverse genetics strategies such as gene silencing by RNA interference (RNAi) can be employed for loss-of-function studies (Alonso and Ecker, 2006; Waterhouse *et al.*, 1998). However, the level of silencing achieved by RNAi in individual lines is variable and the stability of silencing over several generations is not guaranteed. In the case of Arabidopsis, the easiness of transformation allowed the production of large collections of transgenic gene disruption lines by Agrobacterium-mediated T-DNA insertions that were initially used as a forward genetics resource. Some of these collections were subsequently characterized to determine the exact location of the T-DNA insertion providing a whole genome scale gene-indexed resource for reverse genetics studies (Alonso *et al.*, 2003). An inherent drawback of this approach is the vast amount of transgenic lines needed to account for the inability to target-specific loci.

Engineered endonuclease technologies offer a promising alternative to perform targeted gene mutation and other modifications in many species (Joung and Sander, 2012; Kim and Kim, 2014; Urnov *et al.*, 2010). Customized endonucleases can be programmed to cleave very specific DNA sequences in the genome and induce double-strand break repair either through the error-prone nonhomologous end-joining repair pathway or the error-free homologous recombination pathway (Symington and Gautier, 2011). Zinc-finger nucleases (ZFNs) and TALE nucleases (TALENs) recognize DNA sequences using their highly repetitive DNA-binding motifs (Kim *et al.*, 1996; Miller *et al.*, 2011). A drawback of these two technologies for popular use, especially the ZFN technology, is that the design and assembly of the genetic constructs encoding for the correct DNA-binding motifs is complicated and requires experience. In contrast, the *Streptococcus*-derived CRISPR/Cas9 system is relatively simple. This two-component system is composed of a single-guide RNA (sgRNA) for target recognition via RNA–DNA base pairing and the

CRISPR-associated protein9 (Cas9) endonuclease for DNA cleavage (Cong *et al.*, 2013; Jinek *et al.*, 2012; Mali *et al.*, 2013). Targeting of specific genomic sequences does not require any modifications in Cas9 and can be easily achieved by replacing the first 20 nucleotides of the sgRNA with the appropriate sequence. So far, this flexible endonuclease system has been used in many species, including various plants, to generate targeted gene modifications (Brooks *et al.*, 2014; Doudna and Charpentier, 2014; Fauser *et al.*, 2014; Feng *et al.*, 2013; Hsu *et al.*, 2014; Jia and Wang, 2014; Jiang *et al.*, 2013; Mao *et al.*, 2013; Nekrasov *et al.*, 2013; Shan *et al.*, 2013; Upadhyay *et al.*, 2013; Wang *et al.*, 2013; Xie and Yang, 2013; Xing *et al.*, 2014).

In animals, *in vitro* transcribed RNAs encoding Cas9 and gRNA can be directly injected into zygotes to generate biallelic mutants with high frequency (Gratz *et al.*, 2013; Hwang *et al.*, 2013; Wang *et al.*, 2013). However, the plant's cell wall provides a strong physical barrier that handicaps mechanical approaches such as microinjection and alternative delivery agents such as Agrobacterium have been used on Arabidopsis egg cells as well as rice and wheat embryonic calli (Feng *et al.*, 2013; Shan *et al.*, 2013). Thus far, strong constitutive promoters have been used to achieve high expression of the CRISPR/Cas9 system (sgRNA and Cas9) in plants (Belhaj *et al.*, 2013; Bortesi and Fischer, 2014). This strategy has been shown to be quite efficient for generating heritable gene mutations in reproductive cells, but at the same time, a large proportion of noninheritable somatic mutations are also created, which may affect the growth or fertility of T1 plants, and complicates the identification of heritable mutations (Feng *et al.*, 2014; Zhang *et al.*, 2014). Furthermore, only half of the mutations detected in the T1 generation of ubiquitiously expressed Cas9 systems can be transmitted to the T2 progeny, thus limiting the diversity of gene modifications in T2 populations (Feng *et al.*, 2014). A possible strategy to increase the heritability of CRISPR-induced gene mutations in Arabidopsis is to direct the expression of engineered endonucleases, such as *TALEN* and *Cas9,* to dividing cells using meristem-specific promoters, such as the *CLAVATA3*, *APETALA1* and *INCURVATA2* promoters (Forner *et al.*, 2015; Gao *et al.*, 2015; Hyun *et al.*, 2014). However, this strategy still results in the production of a large number of somatic mutations in transgenic plants.

To further improve the production of heritable mutations, especially those lethal and sterile ones, we designed three expression cassettes to constrain the expression of Cas9 to germ-line cells. The 5′ regulatory sequences of the Arabidopsis *SPOROCYTELESS* (*SPL*) and *DD45* genes (AT4G27330 & AT2G21740) and the tomato *LAT52* gene (GenBank: X15855.1) were used to drive Cas9 expression. The *SPL* gene is specifically expressed in sporogenous cells and microsporocytes, but not active in megasporocytes (Yang *et al.*, 1999); the *DD45* gene (At2g21740) encodes an egg-cell-specific protein expressed only in zygotes and early embryos (Steffen *et al.*, 2007), while the tomato *LAT52* gene is transcribed only in pollen cells (Eady *et al.*, 1994; Muschietti *et al.*, 1994). Recently published work has shown that the *DD45* promoter-controlled CRISPR/Cas9 system can generate multiple homozygous mutants in a single generation, but its efficiency depends on the presence of a suitable terminator (Wang *et al.*, 2015). For comparative analysis, we performed an in-depth characterization of the number, frequency, diversity and zygosity of the mutations generated in T1 and T2 Arabidopsis populations using each of the tissue-specific promoters to drive germ-line-specific Cas9 expression as well as the Arabidopsis *UBQ1* gene (AT3g52590) promoter to drive ubiquitous Cas9 expression. Our results show that directing Cas9 expression to germ-line cells with the appropriate strength and developmental timing can significantly increase the rates of heritable mutations, reduce the proportion of chimeras and increase the mutation diversity in the T2 generation, thus providing a specialized CRISPR/Cas9 system for genetic screening of lethal or other desired mutations in Arabidopsis.

Results

Design of a germ-line-specific Cas9 system for gene modification

The Arabidopsis *SPOROCYTELESS* gene encodes a MADS box transcription factor expressed during microsporogenesis and megasporogenesis to regulate the development of sporocytes. *In situ* hybridization in germ-line cells detected strong signals in early microsporocytes and weak signals in developing megasporocytes (Ito *et al.*, 2004; Yang *et al.*, 1999). We used an *SPL* expression cassette to achieve germ-line-specific expression of *Cas9* in Arabidopsis. To mimic as much as possible the native *SPL* gene structure and maintain its tissue and developmental specificity, the entire *SPL* genomic context was reproduced, from the 5′ regulatory sequences to the 3′ gene termination region, and the *Cas9* cDNA (with stop codon) was cloned to replace the first *SPL* exon (Figure S1a). The overall length of the *SPL* genomic sequence used was 5.7 kb, which contains the *SPL* promoter, 5′ UTR regions and the remaining *SPL* gene from 1st intron to the 3′ termination region (i.e. 1st *SPL* intron, 2nd *SPL* exon, 2nd *SPL* intron, 3rd *SPL* exon and *SPL* terminator). The resulting SPL/Cas9 module was then cloned into the pBIN19 binary vector along with a module containing the sgRNA driven by the PolIII-dependent *AtU6-26* promoter. In the pBIN19 binary vector, the selectable marker gene *NPTII* is under the control of the nopaline synthase promoter, which does not alter the expression level and patterns of adjacent tissue- and organ-specific gene promoters (Zheng *et al.*, 2007). The resulting pGSC vector allows one-step ligation of two annealed 20-nt oligos for target recognition at the 2 × BsaI sites (Figure S1a). For comparative analysis, the pUC vector was constructed using the Arabidopsis *UBQ1* gene (At4g05320) promoter and terminator regions to drive constitutive *Cas9* expression (Figure S1b). For simplicity, in this work, we will designate the pGSC and pUC vector constructs as 'germ-line-specific Cas9 system' ('GSC') and 'UC' systems, respectively. Two protein coding genes, *APETALA1* (*AP1*) (AT1G69120) and *TRANSPARENT TESTA 4* (*TT4*) (AT5G13930), were targeted, each at two different positions within their coding regions. The locations and sequences of these target sites are shown in Figure S1c,d. In total, eight different constructs were made to perform *in planta* gene modification studies in Arabidopsis using Agrobacterium-mediated floral dipping transformation.

The GSC system causes targeted gene mutations in T2 instead of T1 generation

For each of the eight CRISPR binary constructs, 32 individual Arabidopsis transgenic T1 lines were subjected to mutation detection by Sanger sequencing of PCR products amplified with primers flanking the sgRNA target sites from both leaf and floral tissue samples. As T1 mutants were usually chimeras, analysis was performed by examining their sequencing chromatograms. A typical chromatogram for a T1 chimera shows single peaks extending up to the sgRNA target site, but immediately after the

target site, the quality of the chromatogram dramatically decreases and multiple peaks start to appear in each nucleotide position (Figure 1). The level of 'chimerism' of a sequenced sample is therefore inversely correlated with the percentage of high-quality base calls or HQ% value, which indicates the proportion of high-quality signals in all the reads (Figure 1).

To evaluate the gene modification activity of the GSC system in the T1 generation, for each of the eight constructs, targeted gene mutations were detected using DNA samples of 32 individual T1 lines extracted from leaf and floral tissues separately. Compared to the UC system, which generated T1 chimeras at a frequency between 10% and 30% for all target sites in *AP1* and *TT4* (Figure 2a,b and Table S1), only two T1 chimeras were detected of a total of 128 plants analysed for the GSC system. In addition, the sequencing chromatograms of these two samples revealed only slight chimerism at the targeted loci in *AP1* (Figure 2c) and *TT4* (Figure 2d) according to their HQ% values, which were higher than any of the chimeras obtained using the UC system. It was also evident that the mutation efficiency of the UC system was higher in vegetative tissue vs. floral tissue (Figure 2a,b).

To comprehensively compare the gene mutagenesis efficiencies of both systems at the targeted gene loci in the T2 generation, for each of the 32 T1 lines, eight individual T2 seedlings were pooled for DNA extraction and mutation analysis by sequencing. Compared to the T1 generation, the mutagenesis frequencies of all the four target loci were increased for the UC system (Figure 2a,b and Table S1), suggesting that *de novo* mutations were induced in the T2 generation. In the case of the GSC-transformed populations, the increase in mutagenesis was quite dramatic, especially for the two *AP1* sites. Given that the GSC system did not induce large-scale somatic mutations in the T1 generation, we propose that these mutations were inherited from T1 germ-line cells.

The GSC system reduces the chimerism of CRISPR/Cas9-generated T2 mutants

To study in more detail the number and nature of the gene mutations present in the UC- and GSC-generated T2 populations, a large number of individual T2 plants from eight randomly selected T1 lines were sequenced for each transgenic population. The overall gene mutation ratios for each of the targeted sites generated by the two systems are shown in Figure 3a. Except for the AP1-178 locus, the UC system exhibited a higher efficiency of gene mutations compared to the GSC system. These results are consistent with preliminary data obtained using pooled seedlings from all the 32 T1 lines (Figure 2a,b). Overall, the mutation efficiency of the GSC system was lower than the UC system, but

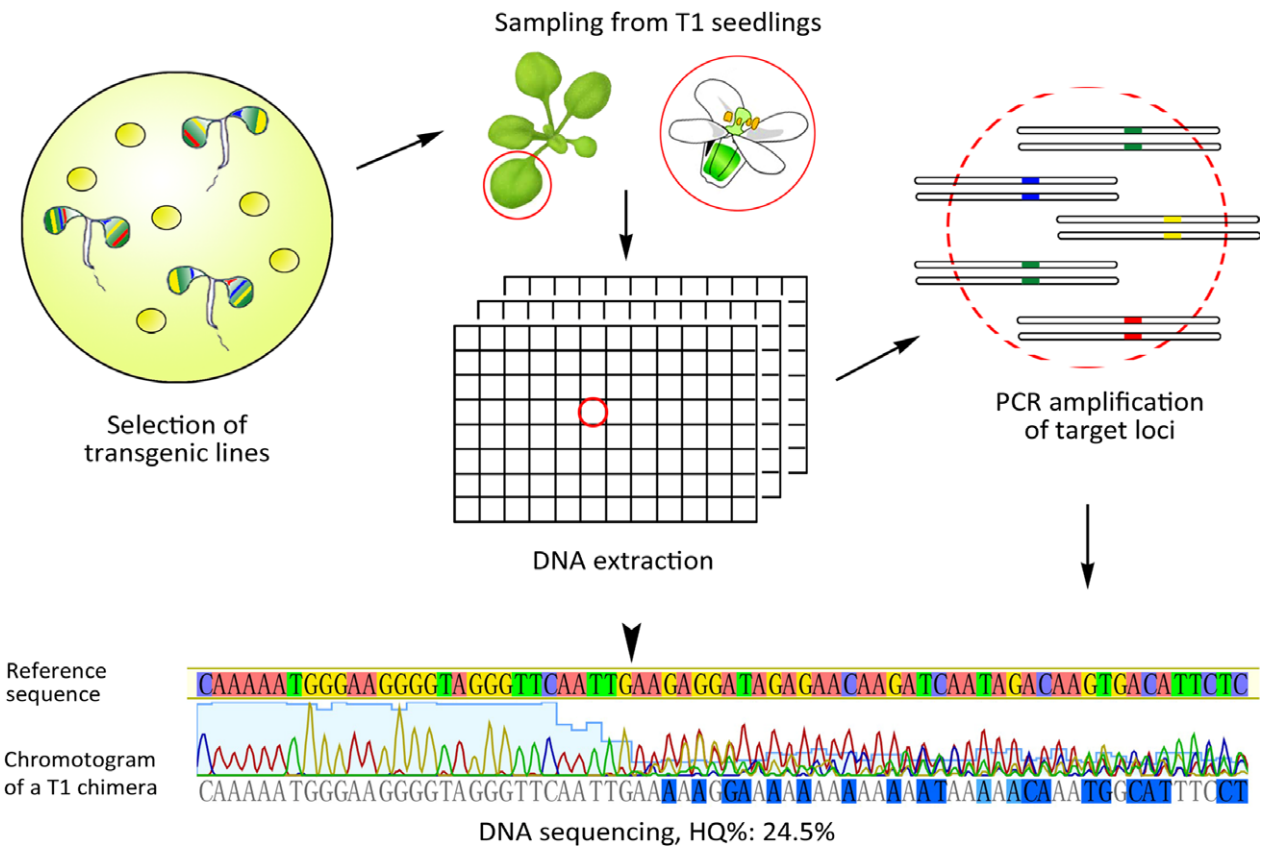


Figure 1 Schematic representation of the workflow designed to detect targeted gene mutations in the Arabidopsis T1 population. For each of the eight CRISPR binary constructs, 32 individual Arabidopsis transgenic T1 lines were subjected to mutation detection by Sanger sequencing of PCR products amplified with primers flanking the sgRNA target sites from both leaf and floral tissue samples. For T1 chimeras, analysis was performed by examining their sequencing chromatograms. A typical chromatogram for a T1 chimera shows single peaks extending up to the sgRNA target site but immediately after the target site the quality of the chromatogram dramatically decreases and multiple peaks start to appear in each nucleotide position. HQ%: proportion of high-quality reads. The arrow head indicates the position of sgRNA target site.

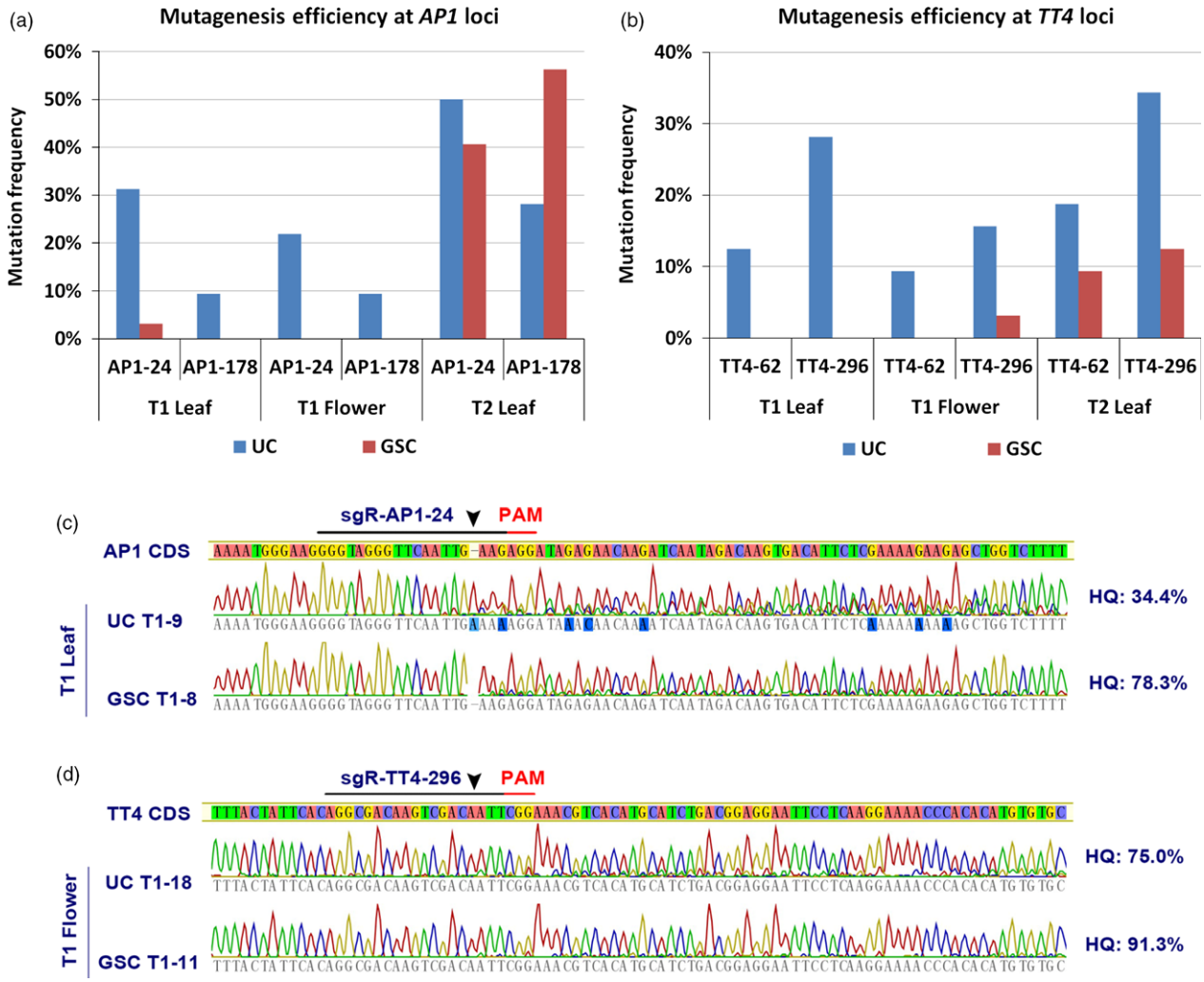


Figure 2 Gene mutagenesis efficiency of the GSC and UC CRISPR/Cas9 systems in T1 and T2 generations. (a), (b) Frequency of targeted gene mutations at the *AP1* (a) and *TT4* (b) loci in plants transformed with the two CRISPR/Cas9 systems. Thirty-two individual T1 lines were analysed for each transformation. DNA was extracted from leaf and floral tissues separately. For the T1 generation, 32 plants were analysed and mutation frequencies calculated for each locus. For the T2 generation, eight individuals from each T1 line were pooled for mutation detection by sequencing. (c), (d) Sequencing chromatograms of T1 chimeras generated by the UC system (upper line) and the GSC system (lower line) showing mutations at the AP1-24 (c) or the TT4-296 (d) site. The HQ % value of each chromatogram is on the right side. HQ%: proportion of high-quality reads. The arrows indicate the position of sgRNA target sites.

within the population of mutated plants, the overall proportion of chimeras was reduced by 40% in the GSC system (Figure 3b).

Detailed examination of the chromatograms flanking the *AP1* sites (Figure S2) and *TT4* sites (Figure S3) can identify the mutants of four different zygotic types. Chromatograms for homozygotes are evenly spaced and well-resolved showing indels at the sgRNA target sites in alignment with the wild-type sequences. Chromatograms for heterozygotes and bialleles show double peaks in the nucleotide positions starting at the sgRNA target site. In contrast, chromatograms for chimeras are very untidy from the position of the sgRNA target site with multiple peaks shown in each position as a consequence of high polymorphism and in extreme cases, only a baseline noise was observed after the target site. As explained above, a high degree of 'chimerism' was associated with low HQ% values.

Analysis of the T2 generation categorized by their zygotic types is shown in Figure 3c. No homozygotes or bialleles were detected in any of the four T2 populations transformed with the GSC system while a small but significant percentage of homozygotes were observed in the UC system-generated plants. The UC system generated a large majority of chimeras in all loci with the notable exception of the AP1-178 site where the majority of plants were heterozygotes or bialleles. Analysis of the GSC T2 population generated only heterozygous and chimeric chromatograms. In addition, the GSC system showed seemingly conflicting results, with a strong prevalence of heterozygotes at the two *AP1* sites (83%–89%) while, at the two *TT4* sites, chimeras prevailed (63%–75%) (Figure 3c and Table S2). Nevertheless, the small number of T2 plants containing mutations at the *TT4* sites makes statistical analysis on the ratios of heterozygotes over chimeras irrelevant for these loci and we therefore believe the results for the AP1 loci to be representative of the GSC system. Moreover, when the average HQ% values of chimeras derived from the different systems were determined, it was found that the sequencing fidelity of GSC-generated *TT4* chimeras was very high (Figure 3d). In other words, the level of 'chimerism' of

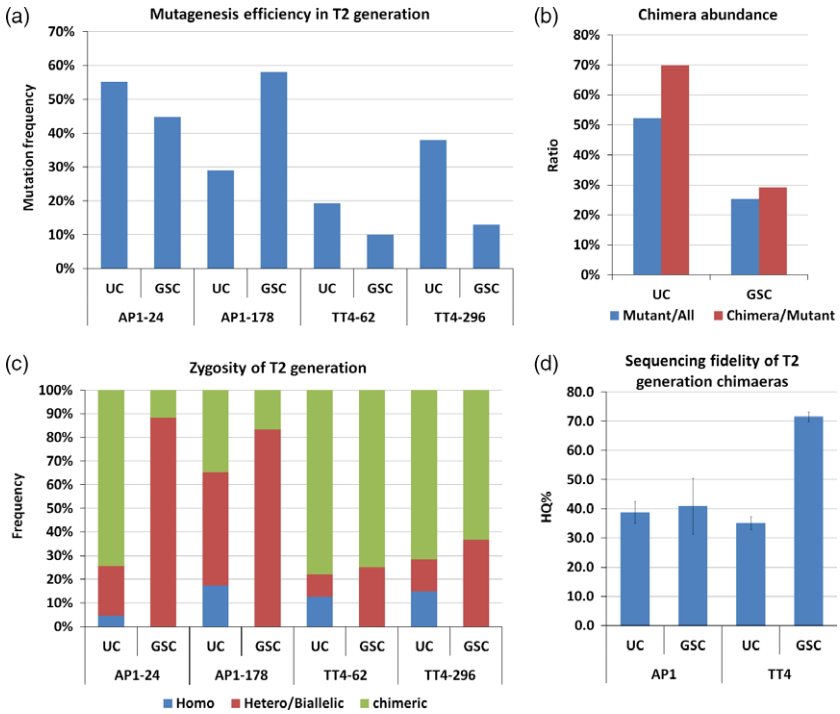


Figure 3 Genotyping of targeted gene mutations induced by the GSC and UC CRISPR/Cas9 systems in the T2 generation. (a) Mutagenesis frequencies for all four targeted loci. The mutagenesis frequency was calculated as the percentage of mutated plants in T2 progenies derived from eight individual T1 lines. (b) Combined data for all four loci showing mutagenesis frequency of both CRISPR/Cas9 systems and the proportion of chimeras present in the mutant population. (c) Zygosity of targeted gene mutations in T2 populations. The frequencies of each zygotic type are shown as percentages of the overall T2 mutant population. (d) Average HQ% value of the chromatograms for T2 chimera generated by the UC and GSC systems grouped by target gene. HQ% represents the proportion of high-quality reads. Error bar = SE.

chimeric T2 plants was very low for the *TT4* target sites. Our data suggest that the GSC system is less efficient than the UC system for generating somatic gene mutations.

The GSC system increases the diversity of targeted gene mutations in the T2 population

To determine the diversity of heritable gene mutations derived from different CRISPR/Cas9 systems, we genotyped all the inherited gene mutations at the two AP loci, including heterozygous, biallelic or homozygous mutations. For the UC system, only 11 of the 78 plants analysed were detected as bialleles or homozygotes at the AP1-24 locus. Within this population, we obtained 22 mutated alleles, among which 12 are unique (Table 1). Thus, the occurrence rate of each mutation type is 1.83 (=22/12), that is each generated mutation type was repeated 1.83 times on average. Comparatively, at the AP1-178 locus, 15 of the 100 plants were heterozygotes or homozygotes, while only four unique mutations were found among the 19 mutated alleles with an occurrence rate of 4.75 (=19/4). In contrast, for the GSC system, 23 of the 82 T2 plants were heterozygotes at the AP1-24 locus, although no homozygotes were detected. In this case, there were 19 unique mutations of a total of 23, resulting in an occurrence rate of 1.21 (=23/19). Similarly, at the AP1-178 locus, the total number of heritable mutations was 25, among which 19 were unique, showing an occurrence rate of 1.32 (=25/19). The detailed mutation types of each mutant are listed in Figure S3. Thus, the combined analysis of the two AP1 target sites clearly shows that the GSC system generated a higher number (48 vs 26) and more diverse (38 vs 16) of heritable gene mutations than the UC system (Table 1).

Table 1 Summary of the diversity of heritable T2 gene mutations generated by UC and GSC

Construct	Target site	T2 plants	Heritable mutants	Mutated alleles	Unique mutations	Occurrence rate
UC	AP1-24	78	11	22	12	1.83
	AP1-178	100	15	19	4	4.75
	Total	178	26	41	16	2.56
GSC	AP1-24	82	23	23	19	1.21
	AP1-178	93	25	25	19	1.32
	Total	175	48	48	38	1.26

Each of the two alleles of the targeted site is counted separately. Occurrence rate indicates the average number of times that a unique mutation type was present.

Figure S3 also shows that the T2 mutations derived from the same T1 line usually share similar mutation types. If we consider these common mutations as genetic markers of cell lineage, it suggests that the mutated alleles derived from the UC system were more likely accumulated from several ancestor cells instead of each independently triggered.

The GSC system is predominantly functional in male gametocytes

To study the lineage of the mutated alleles, 28 individual T2 mutants of the AP1-24 site generated by either the UC or the GSC system were randomly selected for genotype analysis in successive generations. We found that most of the T2 generation mutants produced by the UC system were chimeras containing multiple genotypes, among which only a small proportion can be detected in the progeny (Table 2). In contrast, almost all of the GSC-generated mutants were heterozygotes that segregated homozygotes in the T3 generation according to classic Mendelian law (Table 2).

As no homozygotes or bialleles were detected in the T2 population, we hypothesized that the GSC system was functional only in one of the two gametocytes. To test this hypothesis, we studied the expression pattern of the *Cas9* transgene in germ-line cells of T2 plants by *in situ* hybridization (Figure S4). During the development of anther cells, very weak hybridization signal was first detected in the sporogenous cells at stage 3 (Figure S4a), and the signal became more evident in cells of the tapetum layer than in microsporocytes during stage 5 (Figure S4b). Later in development, the strongest signal was observed in released microspores at stage 7 (Figure S4c). However, in developing ovules, the *Cas9* hybridization signal was mainly found in distal nucellus, but was very weak in mature megasporocytes (Figure S4d). No signal was observed in sections hybridized with sense probes as negative controls (Figure S4e–h). These results strongly suggest that the GSC system is far more active in male gametocytes than in female ones.

To further test the role of parental sex organs in generating heritable mutations, we performed reciprocal crosses between GSC T2 lines and wild-type plants. Four individual T2 GSC lines targeting the TT4-296 locus were used for this assay. Three of them did not have any detectable mutations in leaves (leaf-WT), and the remaining one was a weak chimera based on genotyping of the leaves. In the F1 progenies derived from the leaf-WT T2 lines, heterozygous mutants were generated only when the GSC lines were used as pollen donors. However, in the offspring of the chimeric line, one paternal- and one maternal-derived heterozygote were obtained (Table 3).

Modified UC and GSC systems for *in planta* targeted gene modification

We attempted to improve the performance of the GSC system of generating germ-line mutations by making some modifications to the original system. We suspected that even though the SPL expression cassette achieved the intended tissue specificity, the level of *Cas9* expression in germ-line cells was suboptimal. To enhance the Cas9 germ-line expression, a 2× enhancer element from the cauliflower mosaic virus (CaMV) 35S promoter was fused to the 5′ end of the *SPL* promoter (Figure S5a). Although this modification might reduce the expression specificity of the *Cas9* gene, it may increase *Cas9* level in gametocytes. In addition, the original pBIN19 binary vector used as backbone for T-DNA delivery was also replaced by the pCAMBIA1300 vector to generate the pGEC plasmid (Figure S5a). The architectures of these two binary vectors were quite different in many respects, including the transcriptional cassettes of the selection markers, the origins for plasmid replication in Agrobacterium and the relative orientation of the elements within the T-DNA (Figures S1 and S5), For comparative analysis, the ubiquitous Cas9 system was also constructed into the pCAMBIA1300 vector to produce the pIUC plasmid (Figure S5b).

To compare the gene modification activity of the modified systems with the original ones, 32 transgenic T1 lines were sequenced for each targeted site following the same procedure as described in Figure 1. Our results show that the improved UC (IUC) system was far more efficient for mutagenesis than the UC system in the T1 generation at the analysed target sites. The mutagenesis efficiencies achieved by the IUC system were between 2 and 4 times higher than those achieved by the UC system. Analysis of the T2 generation was performed using pooled samples from 16 of the 32 lines and shows a twofold increase in mutagenesis efficiency (Figure 4a, Table S4). The germ-line-enhanced (GEC) system showed an increase in the number of T1 chimeras compared to the original GSC system but the most dramatic improvement was observed in the T2 generation (Figure 4b, Table S4) where the GEC system achieved very high mutagenesis efficiencies ranging from 65% to 88% compared to 12%–40% by the original GSC system. These results strongly suggest that the addition of the CaMV 35S enhancers to the *SPL* promoter increased germ-line expression (Figure 4b). The sequencing chromatograms of the T1 chimeras generated by the GEC system displayed HQ% values ranging from 24.5% to 70.4% at the AP1-24 locus (Figure 4c) and 75.7%–86.7% at the TT4-296 locus (Figure 4d) indicating that chimerism in the T1 generation was increased in the GEC system compared to the GSC (Figure 1d).

The pCAMBIA vector increased the gene mutagenesis efficiency of the UC CRISPR/Cas9 system, but the addition of the 35S enhancers had a deleterious effect

To study the efficiency of the modified systems in T2 generation plants, 12 seedlings from eight randomly selected T1 lines were sequenced for genotyping at each locus. The IUC system showed higher mutation frequency than the UC system in three of the four transgenic populations (Figure 5a). Similarly, mutagenesis efficiency of the GEC system was also higher than that of the GSC system for all of the tested loci (Figure 5b). Moreover, according to the zygosity analysis of the T2 generation at the four targeted loci (Table S2), the proportion of chimeras was severely reduced in the four T2 populations generated with the IUC system compared to those generated with the UC system and, correspondingly, the proportion of homozygotes was dramatically increased (Figure 5c). As the only difference between the UC and IUC systems is the binary vector backbone, we propose that the pCAMBIA vector is more suitable than pBIN19 to perform CRISPR/Ca9-mediated *in planta* gene modification in Arabidopsis. However, compared to the original GSC system, the performance of the modified GEC system was not satisfactory. Although the overall mutation frequency in the T2 generation was enhanced as expected, the proportion of chimeras was also increased and correspondingly the frequency of heterozygotes decreased dramatically for all targeted loci (Figure 5d). Our results indicate that the presence of 35S enhancers hindered the activity of *SPL* promoter in germ-line cells.

Table 2 Genotyping of T2 mutants generated by the GSC and UC CRISPR/Cas9 systems at the AP1-24 locus

Target site	Construct	T2 line	Zygosity	T2 mutations	T3 mutations	% from T2
AP1-24	UC	1-2-5	Chimeric	3 × I1a, 1 × I1b, 1 × I1c, 1 × D24, 1 × D10	1 × D1, 2 × I1a, 13 × Bi/He, 6 × Ch, 1 × WT	20
		1-2-10	Chimeric	2 × D4, 1 × I1a, 1 × 1b, 1 × D11, 1 × D20, 1 × D2, 1 × D7	1 × I1a, 13 × Ch, 9 × WT	14
		1-13-1	Chimeric	1 × I1a, 1 × I1b, 1 × D1, 1 × D3, 1 × D6, 1 × D33, 1 × WT	4 × I211, 3 × Bi/He, 4 × Ch	0
		1-13-2	Chimeric	1 × I1a, 1 × R1, 1 × D3, 1 × R30, 1 × D8, 1 × D11, 1 × WT	4 × Bi/He, 5 × Ch, 3 × WT	0
		1-15-2	Chimeric	3 × I1a, 1 × R24, 1 × D20	1 × D1, 4 × D19, 11 × He, 3 × Ch, 5 × WT	0
		1-19-5	Chimeric	3 × I1a, 1 × I1b, 1 × I14, 1 × I16, 1 × D1, 1 × R88	1 × I1b, 18 × Bi/He, 5 × Ch	17
		1-19-8	Chimeric	6 × D20, 2 × D8, 1 × R23	4 × I1a, 2 × D8, 1 × D19, 16 × Bi/He, 1 × WT	33
		1-20-1	Chimeric	1 × I1d, 1 × D1, 5 × D3, 1 × D6, 1 × D9, 1 × D26	3 × D3, 19 × Bi/He, 2 × Ch	17
		1-20-2	Biallelic	3 × I1c, 3 × D19	2 × I1c, 2 × D19, 8 × Bi/He	100
		1-23-2	Biallelic	3 × D1, 6 × D26	1 × D1, 1 × D26, 10 × Bi/He	100
		1-23-6	Biallelic	4 × I1a, 5 × D26	3 × I1a, 2 × D26, 7 × Bi/He	100
		1-24-1	Chimeric	6 × D3, 1 × D2, 1 × D26, 2 × D33	6 × D3, 14 × Bi/He, 3 × Ch, 1 × WT	25
		1-24-2	Chimeric	1 × I1a, 1 × D1, 1 × I2, 1 × D9	1 × D1, 1 × I1a, 1 × I1b, 1 × D8, 1 × Bi/He, 8 × Ch	50
		1-24-10	Homo	8 × D1	12 × D1	100
	GSC	2-2-9	Hetero	1 × D19, 2 × WT	2 × D19, 8 × Bi/He, 2 × WT	100
		2-9-8	Chimeric	1 × I1a, 1 × D10, 7 × WT	3 × Bi/He, 11 × Ch, 10 × WT	0
		2-18-2	Hetero	2 × D19, 2 × WT	2 × D19, 8 × Bi/He, 2 × WT	100
		2-18-4	Hetero	5 × I1b, 4 × WT	2 × I1b, 9 × Bi/He, 1 × WT	100
		2-19-1	Hetero	4 × D10, 5 × WT	3 × D10, 5 × Bi/He, 4 × WT	100
		2-19-4	Hetero	3 × I33, 6 × WT	1 × I33, 5 × Bi/He, 6 × WT	100
		2-25-4	Hetero	5 × D1, 3 × WT	1 × D1, 10 × Bi/He, 1 × WT	100
		2-25-6	Hetero	5 × D3, 6 × WT	2 × D3, 4 × Bi/He, 6 × WT	100
		2-27-3	Hetero	6 × I1a, 5 × WT	3 × I1a, 6 × Bi/He, 1 × WT	100
		2-27-5	Hetero	6 × D9, 4 × WT	2 × D9, 7 × Bi/He, 2 × WT	100
		2-27-6	WT	ND	6 × WT, 5 × Bi/He	0
		2-28-5	Hetero	2 × D4, 2 × WT	1 × D4, 8 × Bi/He, 2 × WT	100
		2-30-2	Hetero	3 × I6, 8 × WT	2 × I6, 7 × Bi/He, 2 × WT	100
		2-30-7	Hetero	4 × D1, 5 × WT	5 × D1, 5 × Bi/He, 2 × WT	100

Mutation types were determined by cloning the PCR products into plasmid vectors followed by Sanger sequencing. Mutation types: D#z, I#z; D and I refer to deletion and insertion, respectively; '#'represents the number of bases mutated; 'z', which is a letter, indicates different mutations for the same target gene. For example I1a, I1b and I1c denote three different 1 base insertions. The number of times that the mutation was detected is indicated in front of the mutation type; that is, 5×D3 indicates that the five bases deletion was found five times. WT, wild type; Bi/He, biallelic/heterozygous; Ch, chimaera.

Table 3 Genotyping of F1 hybrids obtained from reciprocal crosses of Col seedlings with T2 lines of the GSC system targeting the TT4-296 locus

Construct	Parental line	Genotype	Hybridization combination (mother × father)	F1 seedlings	Number of mutants
GSC-TT4-296	T2-11-11	Chimeric	11-11 × Col	75	1
			Col × 11-11	29	1
	T2-11-13	Leaf-WT	11-13 × Col	45	0
			Col × 11-13	35	1
	T2-11-15	Leaf-WT	11-15 × Col	8	0
			Col × 11-15	21	1
	T2-11-18	Leaf-WT	11-18 × Col	31	0
			Col × 11-18	29	3

Four individual T2 lines were used for this assay. Except for one, all mutations were detected in F1 hybrids obtained using the transgenic lines as pollen donor.

Directing Cas9 expression to gametocytes using the *DD45* and *LAT52* gene promoters

In addition to the *SPL* gene, promoters of two other previously characterized gamete-specific genes were also used to direct the expression of Cas9 in germ-line cells. The Arabidopsis *DD45* gene was cloned by differential screening and identified as egg-cell specific. Promoter studies using GFP as a reporter gene showed GFP fluorescence in zygotes and early embryos (Steffen *et al.*, 2007). Although *LAT52* is a tomato gene, its promoter shows pollen-specific expression in Arabidopsis (Eady *et al.*, 1994). The *LAT52* promoter drives GUS expression in late microspores and is especially strong at all developmental stages occurring after the bicellular pollen stage (Eyal *et al.*, 1995). To mimic the architecture of those germ-line-specific gene reporter systems, each promoter was fused to the Cas9 gene in combination with a Nos terminator. The resulting expression cassettes were then introduced into the pCAMBIA1300 binary vector together with the AtU6-sgRNA modules to generate the pDD45-GT and pLAT52-GT system, respectively (Figure S6a,b). Two sites were targeted within the

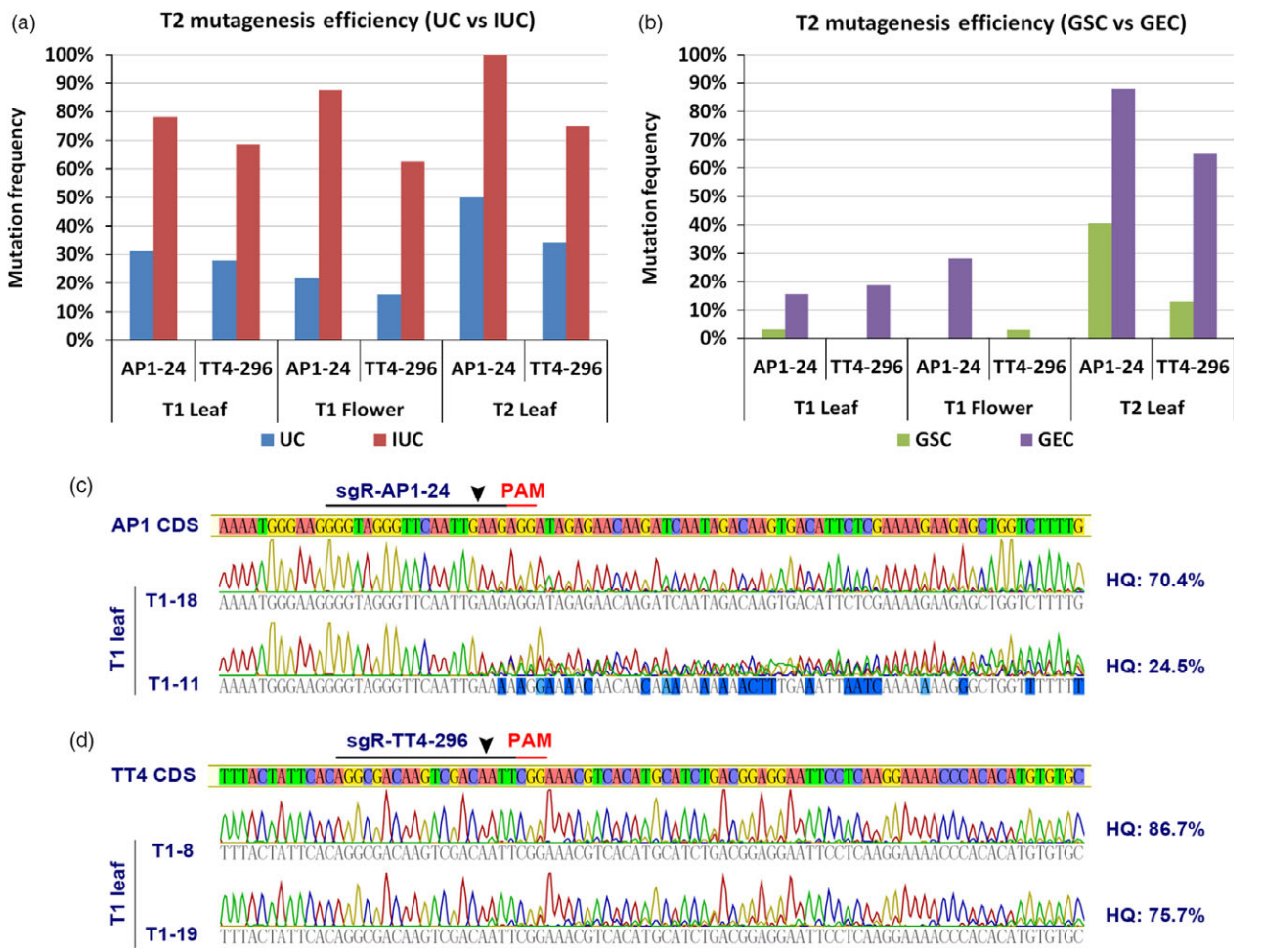


Figure 4 Gene mutagenesis efficiency of the modified CRISPR/Cas9 systems in T1 and T2 generations. (a) Frequency of targeted gene mutations at the AP1-24 and TT4-296 loci in plants transformed with the either the IUC system or the original UC system. (b) Comparative analysis of the GEC and the original GSC systems. (c), (d) Sequencing chromatograms of T1 chimeras generated by the GEC system showing the highest HQ% value (upper line) and the lowest HQ% value (lower line) at the AP1-24 locus (c) and the TT4-296 locus (d). Line numbers are shown on the left of the chromatograms and the HQ % values on the right side. The arrows indicate the position of sgRNA target sites.

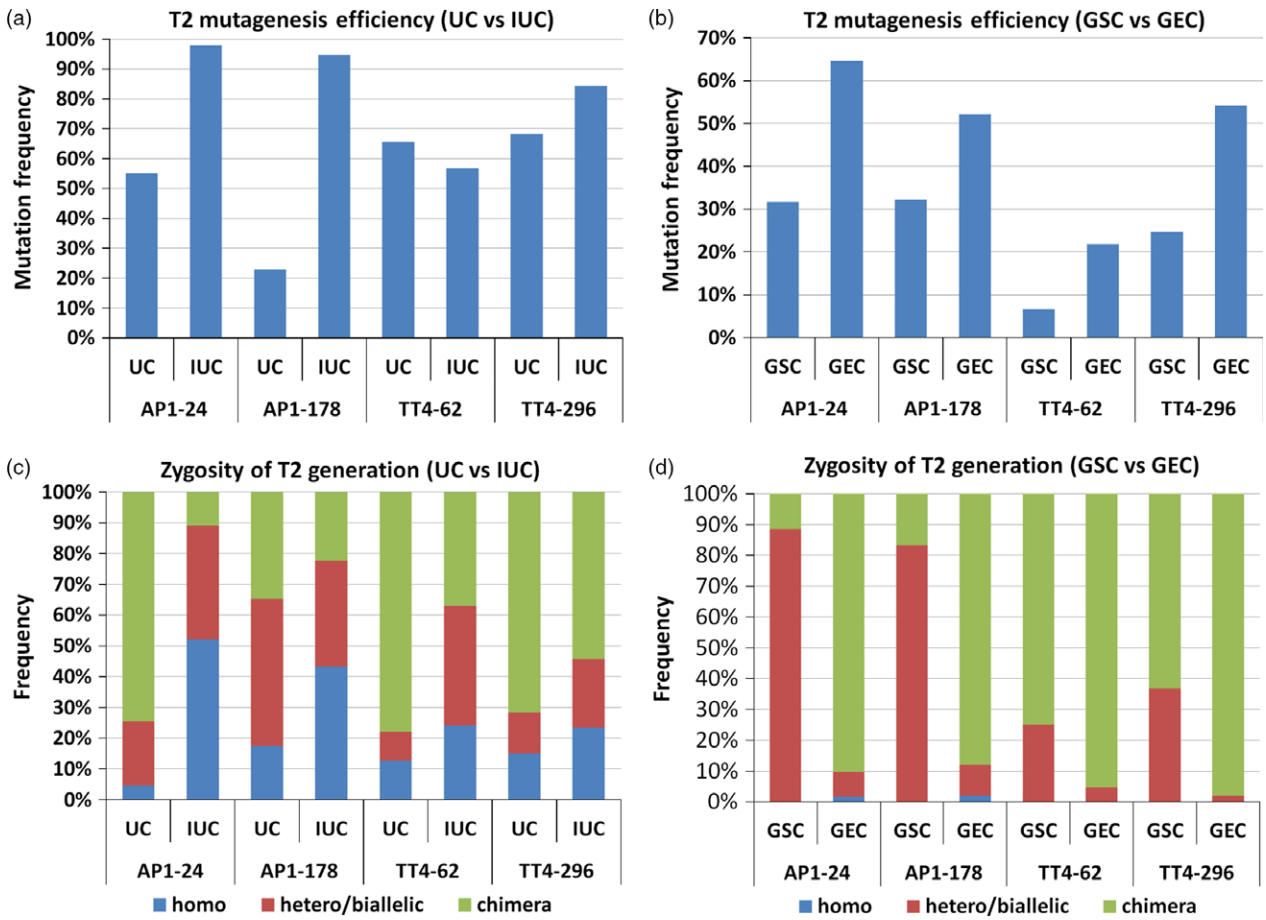


Figure 5 Genotyping for targeted gene mutations induced by the modified CRISPR/Cas9 systems in the T2 generation. (a) Mutagenesis frequencies for all four targeted loci in T2 progenies generated by the UC and IUC systems. The mutagenesis frequency was calculated as the percentage of mutated plants in T2 progenies derived from 8 individual T1 lines. (b) Similar analysis performed with T2 plants generated by the GSC and GEC systems. (c) Zygosity of targeted gene mutations in T2 populations generated by the UC and IUC systems. The frequencies of each zygotic type are shown as percentages of the overall T2 mutant population. (d) Zygosity of targeted gene mutations in T2 populations generated by the GSC and GEC systems. The frequencies of each zygotic type are shown as percentages of the overall T2 mutant population.

GLABRA2 (*GL2*) gene (At1G79840), which encodes a transcription factor required for trichome differentiation. The locations and sequences of the target sites are shown in Figure S6c.

For each of the four binary constructs, 36 T1 individuals were analysed as described in Figure 1. Targeted gene mutations were readily detected in leaf tissues for all constructs during vegetative growth suggesting that neither of the two candidate systems achieved the expected germ-line specificity in Arabidopsis. Analysis of floral samples identified only three mutants generated by the pDD45 system and none by the pLAT52 system (Figure 6a). The sequencing chromatograms of T1 mutants show that the majority of them were chimeras with HQ% values ranging from 24.3% to 80.7%. However, it is worth noting that three heritable gene mutations were detected within the two T1 populations transformed with the DD45 system in leaf and floral samples. A heterozygote was identified in the GL2-89 locus and a biallele and a homozygote in the GL2-97 locus (Figure 6b). These results suggest that the egg-cell-specific DD45 system is expressed and capable of generating targeted gene mutations in the one-cell stage embryo in T0 flowers, although with very low frequency.

Further analysis of targeted gene modification in T2 populations indicated that the overall mutagenesis frequency achieved by the pDD45 system was approximately two times higher than that achieved by the pLAT52 system (Figure 6c). To study the efficiency of these two CRISPR/Cas9 systems in generating heritable gene mutations, T2 mutants were categorized according to their zygosity. We observed dramatic variations in the distribution of zygotic types, perhaps reflecting that the affinity of different sgRNA to its target is different in embryos. Interestingly, in contrast with the SPL-derived GSC system, homozygotes were observed occasionally in T2 populations of the pDD45 and pLAT52 constructs, suggesting that both of them are active in early zygotes (Figure 6d). Even though the *DD45* and *LAT52* gene expression cassettes were not ideal to produce a germ-line-specific gene-targeting system, it is conceivable that improvements in the DD45 system could result in the production of homozygotes in as early as the T1 generation at a higher frequency.

Discussion

Design of germ-line-specific systems for Cas9 expression in Arabidopsis gametocytes

Many of the published CRISPR/Cas9 systems developed for gene modification in Arabidopsis have used constitutive promoters to

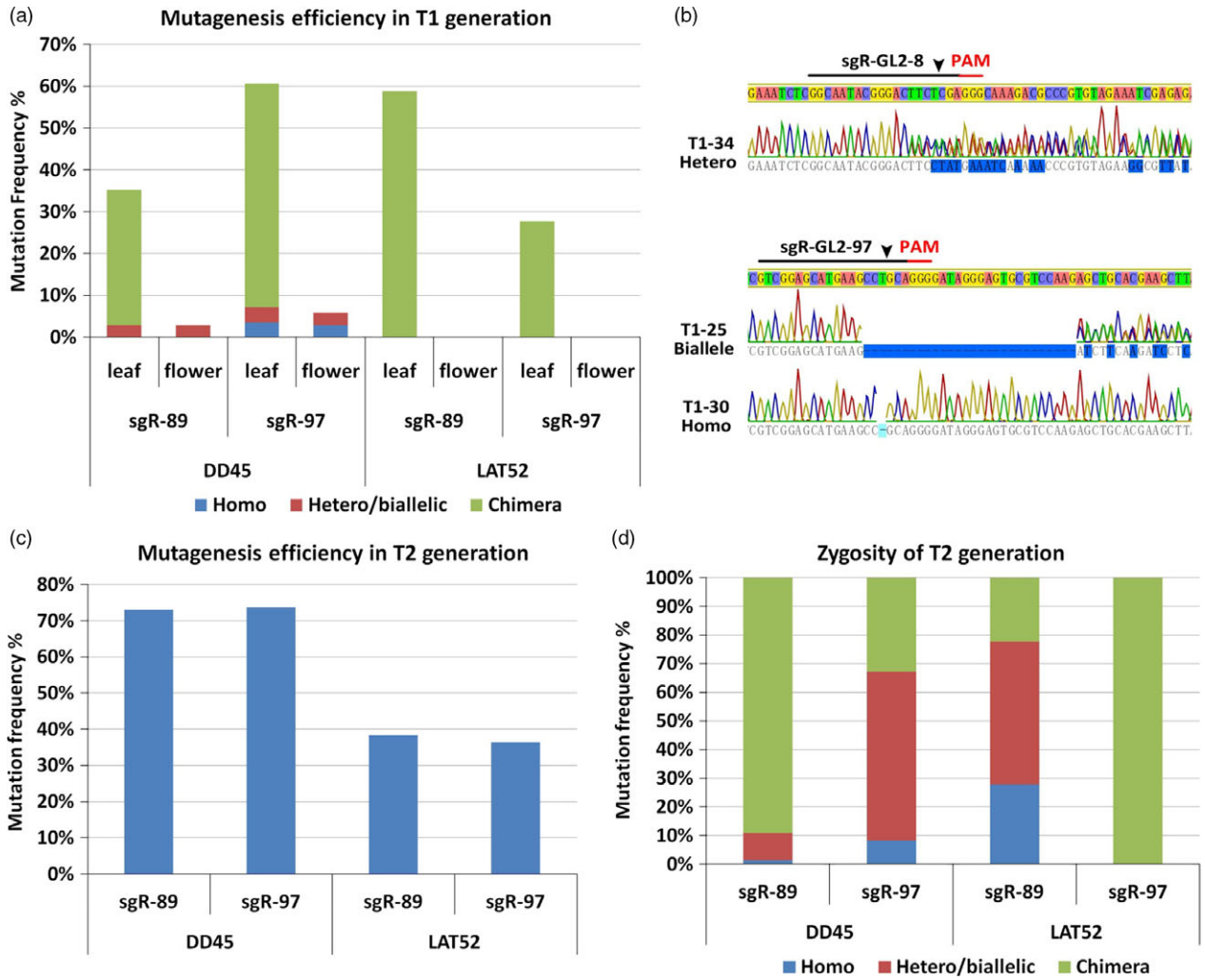


Figure 6 Gene mutagenesis efficiency of the gamete-specific DD45 and LAT52 CRISPR/Cas9 systems in T1 and T2 generations. (a) Frequency of targeted gene mutations at the GL2-89 and GL2-97 loci in T1 plants using the DD45 and LAT52 systems. Mutants of different zygotic types are shown in different colour. (b) Sequencing chromatograms of the three heritable T1 mutants generated by the DD45 system. The zygotic types and mutant line numbers are shown on the left side. The arrows indicate the position of sgRNA target sites. (c) Mutation frequency in 24 T2 individuals generated by the DD45 and LAT52 systems. (d) Zygosity of T2 progeny.

drive ubiquitous expression of Cas9. Although these ubiquitous systems were highly efficient for generating heritable gene mutations in T1 plants, a large proportion of nonheritable mutations were also created in somatic cells (Feng *et al.*, 2014). In some cases, this is an advantage for quick identification of targeted gene modifications. However, in cases when lethal or sterile mutations are anticipated, the growth and productivity of those T1 plants will be affected. In an attempt to avoid somatic mutations and increase heritable ones, we designed a set of germ-line-specific gene modification systems aiming to constrain Cas9 expression to sporocytes or egg cells. Germ-line-specific systems would reduce the proportion of unwanted chimeras in both T1 and T2 generations.

To achieve this purpose, three germ-line-specific promoters, *SPL*, *DD45* and *LAT52,* were used to drive the expression of Cas9 in Arabidopsis with very different outcomes. *Cas9* expression in the SPL-derived GSC system was predominantly observed in early microsporocytes instead of megasporocytes so that only heterozygotes were obtained within the T2 populations. In contrast, although the pDD45 and pLAT52 systems can generate homozygous mutants in the T2 generation, neither of them was specific to germ-line cells. According to previous studies, the DD45 promoter remains active in early embryos (Steffen *et al.*, 2007) so that mutations can also be induced in the multicellular stage in addition to the one-cell stage. As for the LAT52 promoter, although this gene's expression seems to be specific to pollen during reproductive growth (Eady *et al.*, 1994), we cannot exclude the possibility that it is also expressed at low levels in early embryos or young seedlings during vegetative growth.

Of the three systems that we tested, only the DD45 system was capable of generating heritable gene mutations in the T1 generation. As Arabidopsis ovules are the primary targets of Agrobacterium-mediated floral dipping transformation (Ye *et al.*, 1999), we speculate that the activation of the *DD45* promoter in zygotes will be synchronized with the infection of egg cells by Agrobacteria so that homozygotes can be generated as early as in T1 generation. However, *LAT52* promoter expression is maximal during the last stages of pollen development (Bate *et al.*, 1996), and as a consequence, the only chance to simultaneously deliver the CRISPR/Cas9 components to zygotes is during the fertilization

of T1 ovules. Thus, the transcription level and developmental timing of the different germ-line-specific promoters are vital to determine the gene modification activity of the corresponding CRISPR/Cas9 system.

According to previous studies, the *SPL* gene was mainly transcribed in early microsporocytes, but showed only weak expression in megasporocytes during ovule development (Yang *et al.*, 1999). Our results using the *SPL* expression cassette to drive the expression of Cas9 are quite similar, showing a much higher activity in male gametocytes than in female ones; however, the strongest Cas9 expression was detected in released microspores instead of in microsporocytes. This delayed gene expression might be caused by adjacent regulatory motifs either inside or outside the T-DNA regions.

A number of CRISPR/Cas9 systems have been now described using meristem-specific promoters (Forner *et al.*, 2015; Gao *et al.*, 2015; Hyun *et al.*, 2014). Even though this strategy aims to target the precursors of the pollen and egg cells on the parental T1 plant, it does also affect a large number of somatic cells during development. In addition, the diversity of targeted gene mutations in gamete cells derived from the modified precursors should be limited. Compared to the meristem-specific systems, the SPL-derived GSC system is highly specific to microspores, generating relatively few somatic mutations but increasing the diversity of the targeted gene mutations, as the NHEJ-based gene repair process is independently induced in each microspore that expresses CRISPR/Cas9.

Strategies to improve gene modification efficiency and specificity of the GSC system in Arabidopsis

Compared to the UC system, the average gene mutagenesis efficiency of the SPL-derived GSC system in the T2 population was lower. One possible explanation for this result is that the *Cas9* expression achieved by the GSC system in sporocytes was not strong enough or alternatively it was restricted to a short time span. In the UC system, continuous *Cas9* expression can allow the Cas9 protein to accumulate in floral meristems to higher levels than in the GSC system, where Cas9 expression occurs only during a late stage of microsporogenesis. In some cases, the CaMV 35S enhancers led primarily to an enhancement of the endogenous expression pattern rather than to constitutive ectopic expression (Weigel *et al.*, 2000). However, our attempt to increase *Cas9* expression in microspores by adding the CaMV 35S enhancers produced the opposite effect suggesting that the CaMV 35S enhancer somehow may have hindered the *SPL* promoter activity in germ-line cells. Alternative strategies, such as the GAL4-UAS gene activation system (Waki *et al.*, 2013), could be applied in the future to the GSC system to enhance the expression of Cas9 gene in germ-line cells without losing tissue specificity.

In this study, we found that the efficiency of the UC system was much higher when delivered by the pCAMBIA vector than it was when delivered by pBIN19. These two binary vectors use different promoters to drive the expression of the selectable marker genes in plants. Compared to the nopaline synthase promoter in the pBIN19 vector, the CaMV 35S promoters in pCAMBIA vector are 30-fold stronger on average (Sanders *et al.*, 1987). It is possible that the CaMV 35S promoter might have affected the expression levels and/or patterns of the adjacent UBQ or SPL promoters in the pIUC and pGEC systems as it has been reported in some cases (Zheng *et al.*, 2007). This might explain the increased efficiency for targeted gene modification shown by the pCAMBIA1300 vector compared to pBIN19. But we do not believe that the specificity of the pGSC system is compromised by the presence of the nopaline synthase promoter in pBIN as it has been shown that the nopaline synthase promoter does not have any detectable effects on the expression of nearby promoters (Zheng *et al.*, 2007). The plant resistance marker in pBIN19 is next to T-DNA 5′ end, a position that is more resistant to nucleolytic degradation; extensive deletion of the T-DNA can happen to the 3′ end where the inserted genes are located (Rossi *et al.*, 1996). As a result, dysfunctional plant transformants can be obtained exhibiting Kanamycin resistance but lacking the desired transgenes.

Application prospects for the GSC system

T1 mutants generated using ubiquitous CRISPR/Cas9 systems are usually chimeras combining multiple mutations (Figure 7). As most of the mutations occur in somatic cells, only a minority of them that happen to be located in the progenitor cells of the gametocytes will be transferred to the next generation. In the T2 generation, four different zygotic types of mutants are expected with variable segregation ratios. Our results showed that whenever the gene modification efficiency was low, most T2 mutants were chimeras of undetermined genotypes. However, if the gene modification efficiency was high in the T1 population, determined genotypes such as homozygotes, bialleles and heterozygotes were predominant in the T2 generation. As determined zygosity mutants are usually derived from a few ancestor cells, the level of polymorphism in their genotypes is inherently limited. In contrast, in T1 plants generated by the germ-line-specific system, almost no mutations were detected in somatic cells, presumably due to the tissue specificity of Cas9 expression (Figure 7). We hypothesize that efficient gene modifications most probably occur in microspores before meiosis. For each microspore, targeted gene mutagenesis is initiated as an independent event, therefore considering the large amount of microspores in anthers, the diversity of heritable gene mutations is expected to be high. However, for the SPL-derived GSC system, only one of the two gametocytes is mutated; therefore, most of the T2 progeny are expected to be heterozygotes instead of homozygotes or bialleles. Our results fully support this hypothesis.

To date, the most common application of the CRISPR/Cas9 system is to perform targeted gene mutagenesis regardless of the final mutation types. However, in many situations, such as to study the function of a given gene motif in either promoters or CDS regions, defined mutation types might be required, such as in-frame deletions, nucleotide replacements or fragment reversions. A possible strategy to accomplish this is to perform targeted gene modification in gametocytes via homologous recombination in the presence of donor DNA templates. Considering that this DSB repair pathway is most active in sporocytes during meiosis, such germ-line-specific systems might help to synchronize the generation of DSBs with HR-based DNA repair. However, a premise of this method is the effective delivery of suitable repair templates to the DSB sites. An alternative strategy is to identify the desired genotype by screening a large number of gene modification events. In this case, generation of a large number of T2 individuals is required to ensure the diversity of targeted gene mutations. The GSC system should be ideal to generate a large level of polymorphism in the T2 population from relatively few T1 lines.

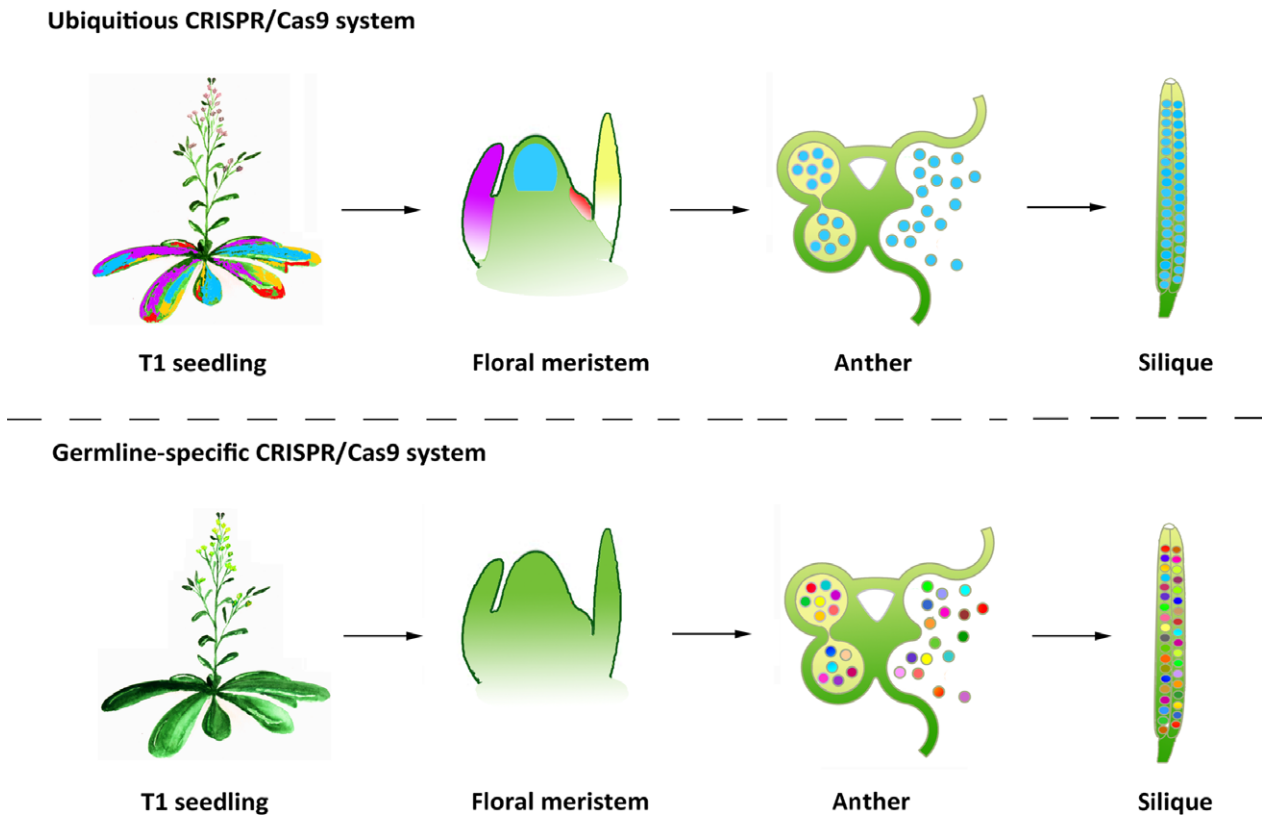


Figure 7 A model showing that the GSC system is capable of generating more diverse mutations compared to the UC system. In the ubiquitous CRISPR/Cas9 system, the T1 mutants are usually chimeras combining multiple mutation types. As most of the mutations happen in somatic cells, only a small percentage of the mutation types that occur in the progenitor cells of the gametocytes can be inherited by their progenies. Thus, the diversity of the targeted gene mutations is limited. In contrast, in the SPL germ-line-specific system, T1 mutations are induced mainly in microsporocytes, so that the NHEJ-based gene repair process is independently induced in each microspore expressing CRISPR/Cas9. Considering the large quantity of pollen cells, application of this system will strongly increase the diversity of heritable gene mutations in the T2 population.

In conclusion, we have developed a highly specific CRISPR/Cas9 system for gene modification in Arabidopsis germ-line cells. This novel system preferentially generates T2 heterozygotes with a high diversity of mutations in the targeted locus. Future application of the GSC system might facilitate the production and screening of genotypes of interest in T2 populations, especially of mutations that are lethal or cause plant sterility.

Experimental procedures

Plasmid construction

To construct the pUC vector, two DNA oligonucleotides (sgR-BsaI-F/R) (Table S6) were annealed and phosphorylated (PNK, NEB) before being inserted into the BbsI sites of the psgR-Cas9-At vector according to an online protocol (http://www.genome-engineering.org/crispr/). The EcoRI&HindIII fragment of this construct was then cloned into the corresponding site of pBIN19 to give the pUC vector.

To generate the pGSC vector, the 3.7-kb *SPL* 5′ and 2.0-kb SPL 3′ gene fragments were amplified from the Arabidopsis genome by PCR using the primer pairs SPL5′-F-XmaI/-R-BsaI and SPL3′-F-BamHI/R-KpnI, respectively (Table S6). The *SPL* 3′ DNA fragment was then inserted into pUC using BamHI and KpnI to replace the *UBQ* terminator. The resulting intermediate vector was then digested with XmaI and NcoI to remove the UBQ promoter and replace it with the SPL 5′ gene fragment to generate the pGSC vector.

For endogenous gene modification, two 23-bp DNA oligonucleotides were synthesized for each locus as listed in Table S6 and inserted into the BsaI sites of pUC and pGSC, respectively, after annealing and phosphorylation according to an online protocol (http://www.genome-engineering.org/crispr/).

The pIUC plasmids were obtained by cloning the EcoRI and HindIII fragment from each corresponding pUC plasmid into pCAMBIA1300. To generate the pGEC plasmids, the 430-bp 2 × 35s enhancer was amplified from pCAMBIA2301 using primers p35s-XmaI-F/AgeI-R (Table S6) and inserted into the XmaI site of pGSC. The XmaI/KpnI fragment of the resulting construct was then cloned into the corresponding sites of the pIUC plasmids.

To construct the pDD45-GT plasmids, a 1026-bp fragment of the *DD45* gene promoter region was amplified from the Arabidopsis genome using the primer pair DD45-F/R (Table S6). The amplified DNA fragment was digested with SalI and XhoI and ligated into the 35s-Cas9-SK plasmid to replace the position of the 35S promoter (Feng *et al.*, 2013). For target recognition, the designed pairs of GL2 targeting oligonucleotides were cloned into the BbsI sites of AtU6-26SK. The resulting sgRNA expression cassettes were excised using KpnI and SalI and ligated into pCAMBIA1300 together with the SalI/EcoRI fragment from DD45-Cas9-SK.

Similarly, the 615-bp *LAT52* gene promoter was cloned from tomato genomic DNA by PCR using the primer pair LAT52-F/R

(Table S6). The amplified DNA fragment was digested with HindIII and XhoI and ligated into the 35s-Cas9-SK plasmid. The resulting plasmid was then digested with HindIII and EcoRI and the fragment ligated into the KpnI/EcoRI sites of pCAMBIA1300 along with the HindIII/KpnI fragments of the *GL2* targeting sgRNA cassettes.

Plant growth and generation of transgenic plants

Arabidopsis plants were grown in growth rooms under long-day conditions (16-h light/8-h dark) at 22 °C. *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* Columbia-0 (Col-0) was performed using the floral dipping method as previously described (Clough and Bent, 1998). Seeds collected from the transformed plants were sterilized with 2% sodium hypochlorite for 15 min and plated on Murashige and Skoog (MS) medium containing 30 mg/L hygromycin (for pCAMBIA1300-based constructs) or 50 mg/L kanamycin (for pBIN19-based constructs) plus 50 mg/L carbenicillin to inhibit *Agrobacterium* growth. Transformants were transplanted to soil 2 weeks later.

Detection of targeted gene mutations

DNA was extracted from transgenic plants by the CTAB method (Rowland and Nguyen, 1993). Genomic regions surrounding the CRISPR target sites were amplified by PCR. Targeted gene mutations were detected by aligning the sequencing chromatograms of these PCR products with the wild-type controls.

For *AP1* and *TT4* targets, 32 T1 individual lines were analysed for each construct. Mutation detection was performed with leaf and floral tissues separately. In the T2 generation, eight seedlings from each of the 32 T1 lines were pooled together for mutation detection. Later, at least 60 T2 plants derived from eight of the 32 T1 lines were subjected to zygosity analysis by sequencing using leaf tissues. Genotypes of the T2 mutants were determined by cloning the PCR products into the pMD18T vector (Takara, Dalian, China) for DNA sequencing. Segregation ratios of these T2 gene mutations were calculated by sequencing at least 16 T3 seedlings for each T2 line.

For *GL2* targets, 36 T1 lines were analysed for each construct. Detection of targeted gene mutations was performed with leaf and floral samples separately. Twenty-four T2 progenies derived from two of the 36 T1 lines were randomly chosen for mutation frequency and zygosity analysis by sequencing.

In situ hybridization

Sections of inflorescences from the GSC transgenic plants were prepared following pretreatment and hybridization as described previously (Brewer *et al.*, 2006). The probe used for Cas9 mRNA detection was amplified by PCR using the primer pair Cas9-1F/378R (Table S6), followed by cloning into the pGEMT-easy vector (Promega, Madison, USA). Digoxigenin-labelled sense and antisense RNA probes were prepared by *in vitro* transcription (Roche, Indianapolis, USA) according to the manufacturer's protocol.

Acknowledgements

We thank members of the Zhu lab for the helpful discussion and insights with this work. We thank Dr Shu-Nong Bai, Professor from college of Life Sciences Peking University, for kindly providing the SPL gene expressional cassette. The work was supported by the Chinese Academy of Sciences. The authors declare that they have no conflicts of interest with respect to this work.

References

Alonso, J.M. and Ecker, J.R. (2006) Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. *Nat. Rev. Genet.* **7**, 524–536.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C. and Ecker, J.R. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653–657.

Bate, N., Spurr, C., Foster, G.D. and Twell, D. (1996) Maturation-specific translational enhancement mediated by the 5′-UTR of a late pollen transcript. *Plant J.* **10**, 613–623.

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.

Bortesi, L. and Fischer, R. (2014) The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* **33**, 41–52.

Brewer, P.B., Heisler, M.G., Hejatko, J., Friml, J. and Benkova, E. (2006) In situ hybridization for mRNA detection in Arabidopsis tissue sections. *Nat. Protoc.* **1**, 1462–1467.

Brooks, C., Nekrasov, V., Lippman, Z.B. and Van Eck, J. (2014) Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiol.* **166**, 1292–1297.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.

Doudna, J.A. and Charpentier, E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, **346**, 1258096.

Eady, C., Lindsey, K. and Twell, D. (1994) Differential activation and conserved vegetative cell-specific activity of a late pollen promoter in species with bicellular and tricellular pollen. *Plant J.* **5**, 543–550.

Eyal, Y., Curie, C. and McCormick, S. (1995) Pollen specificity elements reside in 30 bp of the proximal promoters of two pollen-expressed genes. *Plant Cell*, **7**, 373–384.

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 J.* **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.

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.L., Wang, Z., Zhang, Z., Zheng, R., Yang, L., Zeng, L., Liu, X. and Zhu, J.K. (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc. Natl Acad. Sci. U S A*, **111**, 4632–4637.

Forner, J., Pfeiffer, A., Langenecker, T., Manavella, P. and Lohmann, J.U. (2015) Germline-transmitted genome editing in *Arabidopsis thaliana* Using TAL-effector-nucleases. *PLoS One*, **10**, e0121056.

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. *Proc. Natl Acad. Sci. U S A*, **112**, 2275–2280.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J. and O'Connor-Giles, K.M. (2013) Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics*, **194**, 1029–1035.

Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262–1278.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R. and Joung, J.K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229.

Hyun, Y., Kim, J., Cho, S.W., Choi, Y., Kim, J.S. and Coupland, G. (2014) 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.

Ito, T., Wellmer, F., Yu, H., Das, P., Ito, N., Alves-Ferreira, M., Riechmann, J.L. and Meyerowitz, E.M. (2004) The homeotic protein AGAMOUS controls microsporogenesis by regulation of SPOROCYTELESS. *Nature*, **430**, 356–360.

Jia, H. and Wang, N. (2014) Targeted genome editing of sweet orange using Cas9/sgRNA. *PLoS One*, **9**, e93806.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B. and Weeks, D.P. (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* **41**, e188.

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.

Joung, J.K. and Sander, J.D. (2012) TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14**, 49–55.

Kim, H. and Kim, J.S. (2014) A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15**, 321–334.

Kim, Y.G., Cha, J. and Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl Acad. Sci. U S A*, **93**, 1156–1160.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.

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.

Meyerowitz, E.M. (1987) *Arabidopsis thaliana*. *Annu. Rev. Genet.* **21**, 93–111.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., Dulay, G.P., Hua, K.L., Ankoudinova, I., Cost, G.J., Urnov, F.D., Zhang, H.S., Holmes, M.C., Zhang, L., Gregory, P.D. and Rebar, E.J. (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148.

Muschietti, J., Dircks, L., Vancanneyt, G. and Mccormick, S. (1994) Lat52 protein is essential for tomato pollen development – Pollen expressing antisense Lat52 Rna hydrates and germinates abnormally and cannot achieve fertilization. *Plant J.* **6**, 321–338.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D. and Kamoun, S. (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 691–693.

Page, D.R. and Grossniklaus, U. (2002) The art and design of genetic screens: *Arabidopsis thaliana*. *Nat. Rev. Genet.* **3**, 124–136.

Rossi, L., Hohn, B. and Tinland, B. (1996) Integration of complete transferred DNA units is dependent on the activity of virulence E2 protein of *Agrobacterium tumefaciens*. *Proc. Natl Acad. Sci. U S A*, **93**, 126–130.

Rowland, L.J. and Nguyen, B. (1993) Use of polyethylene glycol for purification of DNA from leaf tissue of woody plants. *Biotechniques*, **14**, 734–736.

Sanders, P.R., Winter, J.A., Barnason, A.R., Rogers, S.G. and Fraley, R.T. (1987) Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. *Nucleic Acids Res.* **15**, 1543–1558.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.L. and Gao, C. (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 686–688.

Steffen, J.G., Kang, I.H., Macfarlane, J. and Drews, G.N. (2007) Identification of genes expressed in the Arabidopsis female gametophyte. *Plant J.* **51**, 281–292.

Symington, L.S. and Gautier, J. (2011) Double-strand break end resection and repair pathway choice. *Annu. Rev. Genet.* **45**, 247–271.

Upadhyay, S.K., Kumar, J., Alok, A. and Tuli, R. (2013) RNA-guided genome editing for target gene mutations in wheat. *G3 (Bethesda)*, **3**, 2233–2238.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646.

Waki, T., Miyashima, S., Nakanishi, M., Ikeda, Y., Hashimoto, T. and Nakajima, K. (2013) A GAL4-based targeted activation tagging system in *Arabidopsis thaliana*. *Plant J.* **73**, 357–367.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F. and Jaenisch, R. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, **153**, 910–918.

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 Biol.* **16**, 144.

Waterhouse, P.M., Graham, H.W. and Wang, M.B. (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl Acad. Sci. U S A*, **95**, 13959–13964.

Weigel, D., Ahn, J.H., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Malancharuvil, E.J., Neff, M.M., Nguyen, J.T., Sato, S., Wang, Z.Y., Xia, Y., Dixon, R.A., Harrison, M.J., Lamb, C.J., Yanofsky, M.F. and Chory, J. (2000) Activation tagging in Arabidopsis. *Plant Physiol.* **122**, 1003–1013.

Xie, K. and Yang, Y. (2013) RNA-guided genome editing in plants using a CRISPR-Cas system. *Mol. Plant*, **6**, 1975–1983.

Xing, H.L., Dong, L., Wang, Z.P., Zhang, H.Y., Han, C.Y., Liu, B., Wang, X.C. and Chen, Q.J. (2014) A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* **14**, 327.

Yang, W.C., Ye, D., Xu, J. and Sundaresan, V. (1999) The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev.* **13**, 2108–2117.

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 J.* **19**, 249–257.

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y., Yang, L., Xu, N. and Zhu, J.K. (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* **12**, 797–807.

Zheng, X., Deng, W., Luo, K., Duan, H., Chen, Y., McAvoy, R., Song, S., Pei, Y. and Li, Y. (2007) The cauliflower mosaic virus (CaMV) 35S promoter sequence alters the level and patterns of activity of adjacent tissue- and organ-specific gene promoters. *Plant Cell Rep.* **26**, 1195–1203.

Supporting information

Additional Supporting information may be found in the online version of this article:

Figure S1 A germ-line-specific Cas9 system designed for gene modification in Arabidopsis.

Figure S2 Sequencing chromatograms of T2 mutants generated by the GSC and UC CRISPR/Cas9 systems at the AP1 locus.

Figure S3 Sequencing chromatograms of T2 mutants generated by the GSC and UC CRISPR/Cas9 systems at the TT4 locus.

Figure S4 *In situ* hybridization showing the Cas9 expression pattern directed by the SPL cassette in germ-line cells.

Figure S5 Modified UC and GSC CRISPR/Cas9 systems for gene modification in Arabidopsis.

Figure S6 Design of two gamete-specific gene modification systems.

Table S1 Gene-editing efficiency of the UC and GSC CRISPR/Cas9 systems in T1 and T2 generations.

Table S2 Zygosity of T2 plants derived from the UC, GSC, IUC and GEC CRISPR/Cas9 systems.

Table S3 Summary of the T2 mutation types resulted by UC and GSC system at the AP1-24 and AP1-178 loci.

Table S4 Gene-editing efficiency of the IUC and GEC systems compared to UC and GSC systems in T1 and T2 generations.

Table S5 Zygosity of T2 plants derived from the DD45 and LAT52 CRISPR/Cas9 systems.

Table S6 Primers used in this study.

Data S1 Sequences of the sgRNA and Cas9 expression cassettes.