Insertion of YFP at the Arabidopsis *AFL1* and *P5CS1* loci shows the potential, and potential limitations, of gene targeting for functional analysis of stress-related genes

Toshisangba Longkumer[a], Louis Grillet [a,b], Chih-Yun Chen[a,1], Hadi Putra[a,2], Wolfgang Schmidt[a], Paul E. Verslues[a,c]

a. Institute of Plant and Microbial Biology, Academia Sinica, Taipei 11528, Taiwan
b. Department of Agricultural Chemistry, National Taiwan University, Taipei 10617, Taiwan
1. Present Address: Biotechnology Center, National Chung Hsing University Taichung 40227, Taiwan
2. Present Address: Department of Plant Sciences, University of Cambridge, UK

c. Corresponding Author: paulv@gate.sinica.edu.tw

Short Title: Gene targeting at *AFL1* and *P5CS1*

Abstract

Transgenic expression of a tagged protein under control of its native promoter is the typical strategy to investigate tissue-specific accumulation and subcellular localization of a protein of interest. In some cases, difficulties in promoter identification, altered promoter function when placed in a new genomic context, and variability between transgenic lines can complicate such experiments. Crispr/CAS9-enabled homologous recombination to insert a tag in frame with an endogenous gene could circumvent these problems. Apart from work at a few loci used to test gene targeting (GT) protocols, there have been relatively few reports of GT to knock-in protein tags in *Arabidopsis thaliana*, and little data of whether expression of the knock-in allele accurately reflects that of the wild-type gene. We investigated these questions using a two-stage GT procedure where a starter line expressing hyper-accurate CAS9 (*hypaCAS9*) under control of the *Egg Cell1.1* (*EC1.1*) promoter was transformed with an mCitrine (YFP) donor sequence flanked by homology arms for homologous recombination at the N- or C-terminal ends of the stress-responsive genes *At14a-Like 1* (*AFL1*) and Δ^1*-pyrroline-5-carboxylate synthetase 1* (*P5CS1*). We observed high frequency GT to generate in frame fusions of *YFP* at either the N- or C-terminus of *AFL1* and at the N-terminus of *P5CS1*. AFL1-YFP was produced from the knock-in locus; however, the N-terminal knock-ins disrupted *AFL1* and *P5CS1* expression. These results show that the combination of *EC1.1* promoter and *hypaCAS9* along with selection of a starter line with strong *hypaCAS9* expression produced good GT frequencies, indicating that this approach can be feasible to knock-in or replace sequence at many loci. However, we also offer a cautionary tale that insertions generated by GT can interfere with gene expression for reasons which are yet unclear. Further refinement of this approach is a potentially useful tool to produce tagged proteins from their endogenous genomic loci.

Keywords: knock-in, gene targeting, abiotic stress, proline, protein localization

Introduction

Gene targeting (GT) using homologous recombination to insert or replace sequence at a specific location in the genome can be useful in many aspects of plant research (Cermak et al., 2017; Collonnier et al., 2017; Endo et al., 2016; Huang and Puchta, 2021; Huang and Puchta, 2019; Mao et al., 2019). These can include replacement of one part of the protein-coding sequence to generate a novel allele or fusion of a tag sequence to the protein coding sequence while it is still expressed from its native promoter and in its native genomic context. Since virtually all plant functions depend on correct spatial and temporal control of protein accumulation, the use of fluorescently tagged proteins to examine protein accumulation and subcellular localization is fundamental to many aspects of plant research. Use of GT to insert sequence encoding a fluorescent protein tag (or any other type of protein tag) into the native gene could have many advantages over more commonly used transgenic approaches where sequence encoding the tagged protein under control of its native promoter is inserted into unknown locations in the plant genome. While the typical native promoter transgenic approach works well for most genes, the results can be ambiguous for some genes because it is not clear which promoter fragment to use to recapitulate the endogenous gene expression pattern. Also, some promoters may not function correctly outside their native genomic context. Traditional transgenic approaches can be laborious as one needs to screen several independent transgenic lines to be sure the expression level and pattern observed is consistent and not unduly influenced by the genomic context the transgene has inserted into in any particular transgenic line. There is surprisingly little data available to test the feasibility of GT insertion of a fluorescent tag as an alternative approach to assay patterns of protein accumulation and subcellular localization. Particularly, there is little data of whether the GT procedure itself can affect expression of the targeted gene. To be most useful for molecular genetic research in Arabidopsis (or other model plants), a simple system in which a standardized set of materials and protocols can be applied for high frequency GT at many different loci is needed.

The first hurdle to high-efficiency GT is the low frequency of homologous recombination in plants and the need the generate a DNA double strand break in the desired location to induce homologous recombination-mediated repair. The limitation on generating site-specific double strand breaks has been effectively removed by use of Crispr/CAS9-mediated gene editing (Mao et al., 2019). However, most double strand breaks are repaired by Non-Homologous End Joining (NHEJ; Huang and Puchta, 2021; Shen et al., 2017). Thus,

generation of the double strand break must be coordinated with introduction of a donor DNA fragment containing sequence homologous to the region around the Crispr/CAS9-generated break site such that at least some sites are repaired by recombination rather than NHEJ. It is also critical that homologous recombination events that do occur are heritable. Heritable GT can be enhanced by specifically expressing CAS9 in egg cells or embryos (Wang et al., 2015; Wolter et al., 2018).

Recently, two types of procedures were shown to produce relatively high frequencies of heritable GT events. The first of these is a two-stage procedure using sequential transformation. In this strategy, a transgenic line expressing CAS9 (without guide RNA or with a control guide RNA unrelated to the GT sites of interest) is produced. This line is then retransformed with a construct containing the sgRNA and a donor fragment for homologous recombination (Endo et al., 2016; Miki et al., 2018). The donor DNA fragment contains a sequence to be inserted flanked by "homology arms" which flank the site targeted by the sgRNA to enable homologous recombination once a double strand break has been introduced. After the second transformation, pools of T_2 plants are screened by PCR to detect the inserted DNA fragment. Individual plants from the positive pools are then screened by PCR and positive plants are backcrossed to segregate out the CAS9 and sgRNA/homology donor constructs. Miki et al. (2018) tested three different embryo or egg cell-specific promoters to drive CAS9 expression and found that the *DD45* promoter was most effective as it gave GT efficiencies of 5-9% based on screening of T_2 seed pools. The authors concluded that the relative success of this strategy was due to the CAS9 being pre-expressed and thus ready to direct DNA cleavage at the moment when the sgRNA/homology donor construct entered the egg cell. While this strategy appears to be promising, we are not aware of subsequent studies that have used this approach in Arabidopsis or introduced further refinements.

The second strategy recently shown to promote heritable GT is a refined single stage approach where all components are introduced simultaneously by Agrobacterium transformation, but with a system to excise the donor DNA fragment from the larger T-DNA construct (Schiml et al., 2014) or with a translational enhancer to induce higher and more rapid CAS9 protein accumulation (Peng et al., 2020). In single stage GT, it was reported that use of the egg cell-specific promoter *EC1.1* could enhance heritable GT (Wolter et al., 2018). Other experiments revealed that co-expression of chromatin remodelers with the GT construct (Shaked et al., 2005), deletion of a DNA polymerase theta involved in NHEJ (van Tol et al., 2022), heterologous expression of recombinases to promote homologous

recombination (Barakate et al., 2020), use of viral replicon expression system to generate increased levels of donor DNA and CAS9 (Vu et al., 2020), or use of other nucleases such as CAS12a (Huang et al., 2021; Merker et al., 2020) could enhance GT at different test loci assayed in each study. While these studies all reported ways to enhance GT, they differed with respect to the tested loci, whether the GT test involved insertion of a new sequence (such as YFP or RFP) or just replacement of existing sequence to correct or introduce a readily assayed mutation, and the length of homology arms used in the DNA donor constructs. Thus, of the many variables affecting GT, the optimal combination of procedure and construct preparation remains unclear, as does the extent to which GT frequencies vary for different loci.

Our interest in GT stems largely from difficulties in using traditional native promoter-driven transgene constructs to study the expression and subcellular localization of some genes important for plant responses to abiotic stress. For both the stress-induced proline synthesis gene Δ^1*-pyrroline-5-carboxylate synthase 1* (*P5CS1*) (Kesari et al., 2012) and the stress-induced ABA synthesis gene *9*-cis-*epoxycarotenoid dioxygenase 3* (*NCED3)* (Kalladan et al., 2019), only partial complementation of the knockout mutant could be achieved with either *35S*- or native promoter-driven constructs. In both cases, expression of the transgene under stress was substantially less then expression of the endogenous *P5CS1* or *NCED3*, indicating that stress induction of these promoters may require the proper genomic context such that the promoter inserted into a new genomic context does not work properly. It is also possible that a longer (or shorter) promoter fragment would better recapitulate the endogenous expression; however, identifying the optimal promoter fragment may require testing of several different constructs in transgenic plants. For P5CS1, there is also very limited knowledge regarding its subcellular localization and whether its localization is altered during stress treatments that induce high levels of proline accumulation. In our previous studies of *NCED3* and *P5CS1*, even partial complementation of the mutant phenotype was sufficient to answer important biological questions (Kalladan et al., 2019; Kesari et al., 2012). However, such results where the knockout mutant cannot be fully complemented leave some doubt as to whether the subcellular localization and tissue-specific patterns of protein accumulation truly recapitulate that of the endogenous gene.

In another example of interest to our laboratory, the promoter fragment to use for transgenic expression of *At14a-Like 1* (*AFL1*) is ambiguous as it harbors a transposon (*AT3TE43765*) approximately 600 bp upstream of the transcriptional start site. Similar to

P5CS1, *AFL1* expression is upregulated by moderate severity low water potential (ψ_w) treatment, indicative of moderate drought stress, at both the transcript and protein level (Kesari et al., 2012; Kumar et al., 2015). *AFL1* encodes a plant-specific protein and there are few clues to its cellular function and subcellular localization other than observations that it may be associated with the actin cytoskeleton and vesicle trafficking components (Kumar et al., 2019; Kumar et al., 2015). Thus, AFL1 and P5CS1 are prime examples of proteins where knowledge of their tissue- and cell type-specific accumulation as well as of their subcellular localization are necessary to understand their function. However, such data has been surprisingly hard to obtain.

With this motivation, we used *AFL1* and *P5CS1* as test cases for a modified GT procedure to insert sequence encoding YFP at the 5' or 3' end of either gene. *AFL1* and *P5CS1* made good test cases for this procedure because we could use antisera recognizing AFL1 or P5CS1 to detect whether YFP-fusion proteins were successfully produced, and whether their level and stress induction matched that of the endogenous protein. For GT, we combined the two-stage procedure of Miki et al. (2018) with use of the *EC1.1* promoter which Wolter et al. (2018) found to increase GT frequency in a single-stage procedure. We also incorporated use of hyper-accurate CAS9 (hypaCAS9; Chen et al., 2017) to reduce off-target effects. By selecting a T_2 starter line having high *CAS9* expression, we obtained high (up to 20 percent) frequency of in-frame *YFP* insertion at the 3' end of *AFL1*, and found multiple homozygous GT plants in the first generation of plants screened. Recombination frequencies were lower, but still feasible at the 5' end of *AFL1* or when using T_3 generation of the starter line for GT at the 5' end of *P5CS1*. These results suggest that both the level of *CAS9* expression, which declined dramatically from the T_2 to T_3 generation of our starter line, as well as locus-specific variation in recombination frequency could influence GT success rates. While these results demonstrate the feasibility of using GT for in-frame insertion of YFP, further analysis of these lines found that insertion of YFP at the 5' end of either *AFL1* or *P5CS1* led to a dramatic decrease in mRNA accumulation and abolished protein accumulation, for reasons that are not entirely clear. In contrast, *AFL1-YFP* was correctly expressed after GT, but AFL1-YFP protein did not accumulate to the same level as endogenous AFL1, likely because the C-terminal YFP interferes with AFL1 membrane association. Thus, our results show both the potential and potential pitfalls of using GT for functional analysis of protein accumulation and subcellular localization.

Results

Generation of a starter transgenic line with high level of *hypaCAS9* expression under control of the *EC1.1* promoter

We first transformed wild type (Col-0) *Arabidopsis thaliana* with a construct for expression of *HypaCAS9* under control of the *EC1.1* promoter. To select a line where CAS9 was efficiently expressed, individual T_2 lines exhibiting a 3:1 segregation ratio were screened for *hypaCAS9* expression by QPCR. A single line with high expression (line #7; Fig. 1A) was used for second stage transformation as described below. Note that in these initial tests only a single sample from each line was used for QPCR assays as we wanted to preserve as much of the T_2 seed as possible for the second stage transformation.

One of our goals was to establish a line with high *hypaCAS9* expression which could be propagated and used repeatedly for retransformation with constructs for GT at various loci of interest. Thus, we subsequently assayed *hypaCAS9* expression in Line 1 and Line 7 in the T_3, T_4 and T_5 generations. We found that in these later generations *hypaCAS9* expression of Line 7 was no longer higher than that of line 1 (Fig. 1B). This may have contributed to lower GT efficiency we observed when using line 7 T_3 instead of T_2 as the starter plants for retransformation (see below).

High GT frequency and in-frame YFP insertion at the 3' end of *AFL1*

We first tested GT at the *AFL1* locus. A construct was prepared containing a donor DNA fragment for homologous recombination where YFP was flanked by approximately 1 kb homology arms to target recombination to the 3' end of *AFL1* to produce in frame fusion of YFP to the AFL1 C-terminus (Fig. 2A). This length of homology arm is similar to that used in previous studies (Miki et al., 2018; Shaked et al., 2005; van Tol et al., 2022; Wolter et al., 2018). The construct also encoded an sgRNA to direct DNA cleavage next to the *AFL1* stop codon (Fig. 2A). This cleavage site was not present in the donor fragment because the *AFL-YFP* fusion disrupted the sgRNA site. Thus, the sgRNA could not direct cleavage of the donor fragment or the endogenous gene after homologous recombination. This construct was then transformed into T_2 $EC1.1_{pro}$*:hypaCAS9* plants described above (line #7). Retransformed transgenic plants were selected by Basta. A first round of screening 30 T_2 pooled DNA extracts (each pool contained seedlings from one T_1 Basta-resistant plant, approximately 30 seedlings collected for each pool) identified six positive T_2 DNA pools

(Fig. 2B; Table 1), a rate that was substantially higher than those observed in a previous two stage GT procedure (Miki et al., 2018).

We then genotyped individual plants from three selected T_2 lines (Fig. 2C; Supplemental Fig. S1). In line #26 and #11, ten out of 24 plants were homozygous for the *AFL1-YFP* insertion (Fig. 2C; Table 1; Supplemental Fig. S1). This indicated that heritable homologous recombination occurred at both genomic copies of *AFL1*. Sequencing of genomic DNA from homozygous plants of line #26 confirmed that the donor fragment had been inserted intact with YFP in frame with the AFL1 coding region (Supplemental Fig. S2). Interestingly though, in line #14 amplification of the 3' end of *AFL1* (primer set 3) revealed both the expected size band for *AFL1-YFP* generated by homologous recombination as well as a larger band (Supplemental Fig. S2). In these putative "single crossover" lines, it appears that homologous recombination occurred at the 5' end of the donor construct, while at the 3' end NHEJ occurred at the end or close to the end of the 3' homology arm. This resulted in duplication of a part of the homology arm sequence, leading to the larger sized PCR band. Similarly, all of the AFL1-YFP GT lines assayed had plants where the expected size band was amplified on the 5' or 3' end (primer set 2 or 3), but there was no band amplified on the other side of the YFP insertion. In these cases, it is also possible that a repair process incorporated only part of the donor construct (or there was deletion of chromosomal DNA), such that the primer site(s) for primer set 2 or 3 were disrupted. These data indicate that a double strand break could be repaired by a combination of homologous recombination and NHEJ.

Homozygous plants from *AFL1-YFP* line #26 were backcrossed to wild type to segregate out the CAS9 and donor fragment transgenes. *AFL1-YFP* transcript and protein levels were compared to wild type at both the T_3 (before backcrossing) and T_6 (after backcross) generations. The knock-in line had slightly elevated transcript level compared to wild type in unstressed plants of both the T_3 and T_6 generation (Fig. 2D). Transcript levels of wild type and the knock-in line were similar after low ψ_w treatment. Thus, the knock-in line still had stress-induced *AFL1* gene expression, albeit that the fold-change increase was slightly less than that observed in wild type. Immunoblotting using AFL1-specific antisera (Kumar et al., 2015), demonstrated that T_3 knock-in lines produced the AFL1-YFP fusion protein (approximately 24 kD higher molecular weight than wild type AFL1), and had no detectable production of untagged AFL1 (Fig. 2E). This was further confirmed by detection of a band at the same molecular weight using YFP (GFP)-specific antisera (Fig. 2E).

However, in both the T_3 and T_6 generation, the level of AFL1-YFP protein detected was only about half the amount of untagged AFL1 detected in wild type (Fig. 2E, F and G). AFL1-YFP was detected in the hypocotyl under both control and low ψ_w treatments and was strongly induced by low ψ_w in young leaves (Fig. 2H). Higher magnification images showed that AFL1-YFP was diffusely localized in the cytoplasm (Fig. 2I) and did not form foci or actin-associated strands as we previously observed in plants expressing *35S:YFP-AFL1* (Kumar et al., 2019; Kumar et al., 2015). This suggests that the C-terminal YFP interferes with AFL1 localization, consistent with ongoing work in our laboratory indicating that the C-terminal regions of AFL1 is responsible for its membrane binding. This binding may be blocked by the C-terminal YFP fusion. Taken together, these data indicate that *AFL1* gene expression may have been slightly altered in the knock-in plants but was still responsive to stress. The reduced level of AFL1-YFP protein compared to wild type AFL1 may be due to mis-localization of AFL1-YFP and not directly related to the GT procedure. These results show that we could use GT for high-frequency insertion of YFP and production of AFL1-YFP at the native AFL1 locus.

In-frame insertion of N-terminal YFP disrupts *AFL1* expression

Simultaneously with the above-described experiments, we used a similarly designed construct (Fig. 3A) and the same *hypaCAS9*-expressing starter line for GT to insert *YFP* at the 5' end of AFL. In this case, the frequency of insertion was lower (Fig. 3B, Table 1) but still comparable to previous results (Miki et al., 2018). We identified several homozygous plants when screening individual plants from the positive T_2 line (Fig. 3C). We also confirmed that genomic DNA of homozygous knock-in plants had an in-frame insertion of YFP at the AFL1 N-terminus (Supplemental Fig. S3). However, *AFL1* transcript levels were no longer induced by low ψ_w in knock-in line (Fig. 3D) and neither YFP-AFL1 nor untagged AFL1 could be detected on immunoblots (Fig. 3E). The reason for the lack of *AFL1* transcript accumulation and YFP-AFL1 protein production in these lines is unclear.

In-frame insertion of *YFP* at the 5' end disrupts *P5CS1* expression

We also designed similar constructs for GT insertion of *YFP* at the 5' end of *P5CS1* (Fig. 4A). In this case, we transformed the donor construct into the same *hypaCAS9* expressing starter line (line 7; Fig. 1), but used T_3 seeds of the starter line rather than T_2. The T_3 seeds were used because we wanted to investigate if the same starter line could be propagated and used repeatedly for GT at additional loci, and because the T_2 seed of this line

was largely depleted by the *AFL1* experiments. Two positive T_2 seed pools were obtained (Fig. 4B; Table 1). Genotyping of individual plants showed that all plants of line #6 produced bands of the expected size at both 5' and 3' ends, and several homozygous plants were identified (Fig. 4C). However, line #13 had a higher-than-expected molecular weight band at the 5' end (primer set 3; Figure 4B and Supplemental Fig. S4). This was similar to *AFL1* C-terminal knock-in line #14 (Supplemental Fig. S1) and indicated that the 3' end of the donor construct had undergone NHEJ leading to duplication of the 5' end of *P5CS1*.

For the *P5CS1* GT line #6, sequencing of genomic DNA from homozygous plants demonstrated that *YFP* was inserted intact and in-frame with P5CS1 (Supplemental Fig. S4B). However, similar to *AFL1*, we again found that the 5'knock-in lines had greatly reduced transcript levels (Fig. 4D). We were neither able to detect YFP-P5CS1 or untagged P5CS1 protein in the knock-in line (Fig. 4E). Note that the faint bands from 95-170 kD in the knock-in line derived from a low level of cross reactivity with P5CS2. Consistent with the apparent lack of P5CS1 protein in the knock-in line, low ψ_w-induced proline accumulation was reduced (Fig. 4F) to a level that was typically observed in the *p5cs1-4* knock-out mutant (Kesari et al., 2012; Sharma and Verslues, 2010). Thus, GT insertion of *YFP* at 5' end of *P5CS1* abolished or strongly reduced *P5CS1* expression.

A similar strategy was used to attempt GT to insert *YFP* at the 3' end of *P5CS1* (Supplemental Fig. S5A). The second stage transformation was successful, as indicated by the higher molecular weight band amplified by primer set 1 (Supplemental Fig. S5B). However, none of the lines had *YFP* inserted at the *P5CS1* locus, indicated by the lack of amplification from primer sets 2 and 3 (Supplemental Fig. S5B). Whether this was due to the lower level of *hypaCAS9* expression in T_3 plants compared to T_2 of starter line 7 (Fig. 1), or whether other characteristics of this site made it more recalcitrant to homologous recombination remains unclear.

Discussion

Our goal was to test the feasibility of GT insertion of YFP at two stress-related loci where we could directly compare both the transcript and protein levels after the GT insertion of YFP to the level of transcript and untagged protein present in wild-type plants. This strategy would provide insights on whether the GT-generated lines had any disruption of gene expression or protein accumulation, and allow us to access the usefulness of these lines for future research. We also used a different promoter-CAS9 combination than tested in a

previous report of two-stage GT and performed an initial screening to identify a starter line with high expression of *hypaCAS9*. As a secondary objective, we made an initial attempt to determine whether this starter line exhibited sufficiently stable *hypaCAS9* expression to be propagated and used repeatedly for GT at multiple loci. While there has been considerable interest and attention devoted to GT lately, these are questions which, to our knowledge, have not received much attention in Arabidopsis. We obtained several positive results including high frequency GT, especially at the C-terminus of AFL1, and generation of homozygous plants at the T_2 stage. Also, AFL1-YFP fusion protein was produced after GT insertion of *YFP*. Note that the reduced protein level of AFL1-YFP was likely caused by the C-terminal YFP interfering with AFL1 function and was thus unrelated to the GT procedure. On the other hand, GT insertion of N-terminal YFP at the *AFL1* and *P5CS1* loci resulted in greatly reduced transcript levels and completely abolished protein production. The selected starter line had much lower *hypaCAS9* transcript level in the T_3 generation compared to T_2, which may have contributed to the lower GT efficiency at *P5CS1* compared to *AFL1*, although locus-specific effects cannot be ruled out. These results offer several points to consider in bringing GT closer to routine use for functional analyses of Arabidopsis genes.

There are several factors that may determine the efficiency of GT, including the promoter used for *CAS9* expression, length of the homology arms for recombination, and size of the insert between the homology arms. We used a new combination of the *EC1.1* promoter driving expression of *hypaCAS9* for two stage GT. *EC1.1* had been previously shown to be effective for GT in a single stage procedure (Wolter et al., 2018); however, it had not been tested in a two-stage procedure and had not been paired with *hypaCAS9* to reduce off-target effects. While this combination of promoter and CAS9 variant was relatively successful in our case, there is still much room for further optimization. For example, it has been reported that *LbCas12a* was effective for GT in a single stage procedure (Merker et al., 2020; Wolter and Puchta, 2019), and it could also be tested in a two stage GT procedure. For the $EC1.1_{pro}$:*hypaCAS9* combination used here, tests of additional transgenic lines could identify a line that has relatively high, but more stable *hypaCAS9* expression and that line could be propagated and used repeatedly. Such tests are currently underway in our laboratory. For the homology arms of the donor construct, we largely followed previous examples in choosing their length (around 1 kb; Miki et al., 2018; Shaked et al., 2005). Whether a different length or arrangement of homology arms could improve GT frequency is unclear and may also be locus-specific. This is particularly true as many GT studies that

have used similar length of donor DNA for recombination did not use GT to insert new sequence at the target locus, as we did here with YFP, but rather introduced a single base sequence change (or small indel) to generate an easily assayed gain or loss of function at the test loci (see for example de Pater et al., 2018; Wolter and Puchta, 2019). Thus, the coordination of insert size versus size or arrangement of the homology arms has not, to our knowledge, undergone extensive optimization. On the other hand, the fact that previous data indicate that mismatches in the homology arms can be tolerated is interesting as it implies that GT insertion of a tag such as YFP could potentially be combined with allele replacement, such as change of codon(s) to introduce an amino acid change(s).

Other strategies that have been explored to increase GT efficiency include altering the repair process to favor recombination over NHEJ (see for example Shaked et al., 2005; van Tol et al., 2022; Wolter and Puchta, 2019) or to deliberately alter chromatin structure to improve GT frequency (Shaked et al., 2005). Given our interest in using GT to study plant stress responses, these approaches would be more problematic as changes in DNA repair pathways or chromatin landscape could alter stress phenotypes (or least add ambiguity about the results). Ectopic expression of recombinases caused obvious growth and developmental phenotypes (Barakate et al., 2020) and would need considerable optimization to be useful in physiological experiments. If one must segregate the DNA repair mutation, chromatin remodeler, or recombinase away from the GT-modified gene before doing downstream experiments, this adds time and effort to the overall procedure. It also seems unneeded for many applications in Arabidopsis, given the relatively high GT frequencies we could observe even with relatively little optimization of GT procedures. That said, it is interesting to note that GT frequencies can be affected by changes in temperature or light conditions (Vu et al., 2020), and that efficient GT can be achieved by using heat shock to induce CAS9 expression (Barone et al., 2020). It could be hypothesized that stress treatments such as heat shock could also change the chromatin architecture around stress-induced genes, possibly making them more susceptible to GT. It will be of interest to check whether this is true for GT at the 3' end of *P5CS1*, which may be one such site that is relatively recalcitrant to GT.

There were two observations from our experiments that were unexpected and hard to explain. The most dramatic was the observation that insertion of *YFP* at the 5' end of either *AFL1* or *P5CS1* nearly abolished mRNA accumulation. The second was that GT frequency seemed to vary substantially for the 5' versus 3' end of *AFL1*. It has been reported that GT frequency can be dramatically affected by chromatin remodeling (Shaked et al., 2005).

Conversely, GT-mediated insertion of RFP was found to alter DNA methylation status (Lieberman-Lazarovich et al., 2013). Thus, chromatin state and DNA methylation could determine GT frequency and influence whether expression of the targeted gene after GT. It could also be expected that these effects may be greater for larger DNA insertions and for insertions at the 5' end of the gene, closer to the promoter region. Consistent with this possibility, there is a heavily methylated region around the transposon located approximately 600 bp upstream of the *AFL1* transcriptional start site that may have influenced GT frequency. Altered methylation could also explain why GT insertion at the 3' end of *AFL1* led to increased *AFL1* expression in unstressed plants (Fig. 2D). If altered methylation or chromatin state were the cause for disrupted *AFL1* expression, we could hypothesize that these factors would reset in subsequent generations after the GT, allowing gene expression to return to normal levels. However, we failed to observe any change in *AFL1* expression at three generations after GT for lines with either the 5' or 3' YFP insertion (Fig. 2D and 3D). Similarly, there was no recovery in *P5CS1* expression after backcrossing to wild type and rescreening of homozygous plants (Fig. 4D), suggesting that the block of expression may be permanent in these lines. Much GT research has used a relatively limited number of easy to assay loci. Conversely, biotechnology-oriented experiments are often focused on insertion into a limited number of "safe" intergenic sites where a whole construct can be added without disrupting any endogenous genes. Thus, we are not aware of any well-established guidelines to show which type of gene (i.e., what type of epigenetic landscape) is likely to allow for high frequency GT and which sites are likely to be more recalcitrant to GT or, conversely, which type of sites might be most likely to exhibit altered gene expression after GT.

For the GT events at the 5' ends of *AFL1* or *P5CS1*, it is also possible that the linker sequence inserted between *YFP* and *AFL1* or *P5CS1* disrupted expression. However, this is unlikely as we used the same linker sequence that is present in various Gateway cloning plant transformation vectors in which high levels of N-terminal YFP fusion protein are produced, and the linker sequence did not interfere with expression of *AFL1-YFP* after GT at the 3' end of *AFL1*.

For *P5CS1*, the low frequency of GT we observed at the 5' end and failure to find any GT events at the 3' end could be due to specific characteristic of that locus as described above, or due to the fact that the T_3 starter line plants used for transformation with the *P5CS1* donor constructs had lower *hypaCAS9* expression than the T_2 plants used for *AFL1*. Further experiments will be needed to sort out these possibilities and see whether it is the 20 percent

GT frequency at the 3' end of *AFL1* or the 0-7 percent frequency observed at *P5CS1* and *AFL1* 5' end that is more typical as this GT procedure is applied to other loci. In summary, we believe that these data and descriptions will increase the recognition that GT can be a viable strategy to study loci where traditional approaches of native promoter constructs and transgenic expression are problematic. We hope that the data presented here will help inform future experiments to overcome the remaining difficulties with this approach.

Methods

Plant materials

Wildtype *Arabidopsis thaliana* (Col-0) and other knock-in genotypes were grown on soil at 22°C in a growth room with 16 h light and 8 h dark cycle. Plate-based experiment for control and stress were carried out as described previously (Verslues, 2010). Sterilized seeds were sown on half-strength Murashige and Skoog media supplemented with 2 mM MES [2□(N-morpholino) ethanesulfonic acid] buffer (pH 5.7) and 1.75 % agar without any sugar. After stratification for 3 days at 4°C, plates were placed vertically in a growth chamber maintained at 22°C under continuous light at 90–110 μmol photons m^{-2} s^{-1}. Seven-day-old seedlings from control plates were transferred either to freshly prepared control agar plates or agar plates infused with polyethylene glycol (PEG)-8000 to impose moderate severity (-0.7 MPa) low water potential stress. Control and stress samples for gene expression or immunoblot assays were collected at 4 days after transfer.

Generation of the *EC1.1$_{pro}$:hypaCas9* transformation vector

Plant codon-optimized *HypaCAS9* expressed under the *EC1.1* promoter was generated by mutagenesis of the *EC1$_{pro}$:CAS9* sequence from the pHEE401E plasmid (Wang et al., 2015; Addgene Plasmid #71287). Two fragments were amplified by PCR with oligonucleotide primers harboring the *hypaCas9* mutations described in Chen et al. (2017). This set of PCR products overlapped at the mutation site and also overlapped with the backbone of a Gateway entry clone derived from pDONR221 including the attL1 and attL2 recombination sequences. The two PCR fragments and the vector were ligated through a Gibson Assembly reaction. The *EC1.1$_{pro}$:hypaCas9* fragment was subsequently combined with the pOLE1-OLE1-tagRFP entry clone (Shimada et al., 2010) and then then cloned into pH7m34GW destination vector (with Hygromycin resistance for plant selection) by gateway LR reaction. All oligonucleotide sequences used for construct generation are given in Supplemental Table SI

(overlapping sequences are highlighted in bold and mutations are represented by upper case letters).

Generation of the sgRNA and donor DNA transformation vectors

The 5' and 3' homology arms for targeting GT to *AFL1* and *P5CS1* were amplified from Arabidopsis Col-0 genomic DNA andYFP (mCitrine) was amplified from pEN2-mCitrine. The mCitrine CDS was initially codon-optimized for Arabidopsis using the online tool from Integrated DNA Technologies (https://sg.idtdna.com/pages/tools/codon-optimization-tool), and synthesized and cloned by AllBio company (Taiwan). Donor plasmid constructs were generated by cloning the PCR amplicon into Gateway entry pDONR vectors; the 5' arm DNA fragment in P4P1r, YFP (mCitrine) in P1P5r, the 3' arm DNA fragment in P5P2, and the U626p-tRNA-gRNA in P2rP3. These entry clones were then assembled by multi-site Gateway LR recombination into the destination binary vector pB7m34GW (with Basta resistance for plant selection). For cloning sgRNA, the target-specific crRNA sequence (20 bp) was picked manually by looking at the PAM site located close to the start codon (for GT at the 5' end of the gene) or close to the stop codon (for GT at the 3' end of the gene). Two sgRNA having the same target sequence were cloned for each gene.; initially, using primer containing 20 bp of target sequence and an additional flanking 5 bp containing BsaI sites. After BsaI digestion, this DNA was mixed with tracRNA and pre-tRNA adapters that had been excised from p1-tRNA-gRNAopti-CmR at the BsaI site, and ligated to pEN-R2-U626p-tRNA-gRNAopti1xAmp-L3 which had also been digested by BsaI (Xie et al., 2015). The sgRNA target sequence and primers used for the construct generation are given in Supplementary Table SI.

Transgenic line generation and genotyping

To generate the *hypaCas9*-expressing starter line for GT, the $EC1.1_{pro}$*:hypaCas9* construct was transformed into *Arabidopsis thaliana* (Col-0) by the Agrobacterium-mediated floral dipping method. T_1 seeds were selected on Hygromycin plates and the resistant plants transplanted to soil for generation of T_2 seeds. T_2 lines were initially selected on Hygromycin plates. For lines exhibiting 3:1 (resistant:susceptible) segregation ratios consistent with a single locus T-DNA insertion, seedlings were transplanted to soil to collect floral buds for qPCR assay of *hypaCas9* expression. To assay *hypaCas9* expression in subsequent generations, homozygous T_3 lines from lines T_2-1 and T_2-7 were identified by hygromycin

selection, and plants from the selected T_3 lines as well as subsequent T_4 and T_5 plants were grown on soil and floral buds collected for qPCR assay.

For the second stage transformation, the vector containing the sgRNA and donor DNA construct was transformed into hygromycin-resistant T_2 plants, or into homozygous T3 plants of the *hypaCas9*-expressing starter line (all experiments used starter line #7). Transformants were selected by spraying Basta herbicide solution on soil-grown plants. Seed of the T_2 generation (after retransformation) was harvested from Basta-resistant T_1 plants. Pools of around 30 T_2 seeds from an individual T_1 plant were sown on plates without any antibiotics for bulk PCR screening of seedlings. For screening of individual plants from selected lines, T_2 plants were grown on soil and leaves of individual plants harvested for genomic DNA extraction. Subsequently, T_3 plants homozygous for the GT-generated YFP insertion were backcrossed to Col-0, and lines homozygous for the YFP insertion but lacking the *EC1.1*$_{pro}$*:hypaCas9* T-DNA insertion identified by PCR screening. T_6 backcrossed seedlings were used for gene expression and immunoblot experiments.

For genotyping of plants after the second transformation, genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen). PCR was performed using standard conditions. Primer sequences are given in Supplemental Table SI.

Quantitative RT-PCR analysis of gene expression

RNA of Arabidopsis seedlings or floral buds was extracted the RNeasy Plant Mini Kit (QAIGEN) according to the manufacture's protocol. RNA amount was quantified using a Nanodrop spectrometer and reverse transcribed using SuperScript III (Invitrogen) and Oligo dT primer. RT-qPCR reactions contained 50 nM cDNA, 200 nM primers and KAPA SYBR FAST master mix (KAPA Biosystems). RT-qPCR reactions and were performed on a 7500 Fast Real-Time PCR System (Thermo Fisher Scientific). *ELF1a* was used as a reference gene to normalize transcript abundances. Primer sequences used for RT-qPCR are shown in Supplemental Table SI.

Immunoblotting

Protein sample for western blot were prepared by grinding the control and stress-treated seedlings in liquid nitrogen and suspending the homogenized samples in 200-400 µl of extraction buffer (150 mM NaCl, 50 mM Tris HCl pH 7.5, 10% glycerol, 10 mM EDTA, 10 mM DTT, 1mM NaF, 1% IGEPAL, 2X protease inhibitor, and 2X Roche phosphatase

inhibitor) according to the sample quantity (approximately 200 µl of extraction buffer per 100 µl of ground plant material). Protein was quantified using the Pierce™ 660nm Protein Assay Reagent kit (Thermo Scientific). AFL1 and P5CS1 antisera were generated in our laboratory and have been previously characterized (Bhaskara et al., 2015; Kumar et al., 2015), while anti-HSC70 and anti-GFP were obtained from Enzo and Roche, respectively. Immunoblots were performed as previously described (Kesari et al., 2012; Kumar et al., 2015; Wong et al., 2019). In brief, for SDS-PAGE, 10 µg protein was loaded per lane, resolved on 10% polyacrylamide gel, and transferred onto PVDF membrane (4°C overnight, at 30V). The membrane was incubated in 5% milk in TBS-T for at least 1 h at room temperature with shaking. The following dilutions were used for primary and secondary antibodies: anti-Hsc70 (1:5000) /secondary HRP-mouse (1:15000); anti-AFL1 or anti-P5CS1 (1:5000) /secondary HRP-rabbit (1:15000); and anti-GFP (1:5000) /secondary HRP-mouse (1:15000). For anti-GFP, blots were probed with primary antibody overnight at 4°C and for 4 hours at 4°C for the secondary antibody. All other blots were probed for 1 h at room temperature for both primary and secondary antibodies. The signal was detected using the ECL detection kit (Thermo Scientific). Band intensities were quantified using the Fiji (ImageJ) function Plot lanes.

Proline assay

Proline measurement was done using a ninhydrin assay (Bates et al., 1973) adapted to 96-well plate format (Verslues, 2010).

Confocal imaging

Control and stress-treated seedlings were stained with propidium iodide (PI) to visualize cell walls and then mounted in water or low water potential solution (–0.7 MPa mannitol solution). AFL1-YFP (excitation at 514 nm and emission at 542 nm) and PI staining were observed using a Zeiss LSM880 microscope.

Supplemental Data files

Supplemental Figures S1-S5

Supplemental Table S1: Primer and oligonucleotide sequences

Accession numbers

AFL1, AT3G28270; *P5CS1*, AT2G39800

Author Contributions

T.L. and P.E.V. conceived the research; T.L., L.G. and P.E.V. designed the experiments; L.G. supplied key materials, T.L. and P.E.V. analyzed the data; T.L. C.-Y.C. and H.P. performed the experiments; P.E.V. wrote the article with assistance from T.L.; W.S. provided material, and W.S., T.L. and L.G. edited the article.

Acknowledgements

We thank the staff of the imaging core facility of the Institute of Plant and Microbial Biology (Ji-Ying Huang and Mei-Jane Fang) for microscopy assistance and Shih-Shan Huang for laboratory assistance. This research was supported by Academia Sinica Investigator Awards to P.E.V (AS-IA-108-L04) and W.S. (AS-IA-111-L03).

Conflict of Interest Statement

The authors declare that they have no conflict of interests.

References

Barakate, A., Keir, E., Oakey, H., and Halpin, C. (2020). Stimulation of homologous recombination in plants expressing heterologous recombinases. BMC Plant Biology *20*, 336.

Barone, P., Wu, E., Lenderts, B., Anand, A., Gordon-Kamm, W., Svitashev, S., and Kumar, S. (2020). Efficient Gene Targeting in Maize Using Inducible CRISPR-Cas9 and Marker-free Donor Template. Molecular Plant *13*, 1219-1227.

Bates, L.S., Waldren, R.P., and Teare, I.D. (1973). Rapic determination of free proline for water-stress studies. Plant and Soil *39*, 205-207.

Bhaskara, G.B., Yang, T.H., and Verslues, P.E. (2015). Dynamic proline metabolism: importance and regulation in water limited environments. Frontiers in Plant Science *6*, *484*.

Cermak, T., Curtin, S.J., Gil-Humanes, J., Cegan, R., Kono, T.J.Y., Konecna, E., Belanto, J.J., Starker, C.G., Mathre, J.W., Greenstein, R.L., *et al.* (2017). A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. Plant Cell *29*, 1196-1217.

Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., and Doudna, J.A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature *550*, 407-+.

Collonnier, C., Guyon-Debast, A., Maclot, F., Mara, K., Chariot, F., and Nogue, F. (2017). Towards mastering CRISPR-induced gene knock-in in plants: Survey of key features and focus on the model *Physcomitrella patens*. Methods *121*, 103-117.

de Pater, S., Klemann, B., and Hooykaas, P.J.J. (2018). True gene-targeting events by CRISPR/Cas-induced DSB repair of the PPO locus with an ectopically integrated repair template. Scientific Reports *8*, 3338.

Endo, M., Mikami, M., and Toki, S. (2016). Biallelic Gene Targeting in Rice. Plant Physiology *170*, 667-677.

Huang, T.-K., Armstrong, B., Schindele, P., and Puchta, H. (2021). Efficient gene targeting in Nicotiana tabacum using CRISPR/SaCas9 and temperature tolerant LbCas12a. Plant Biotechnology Journal *19*, 1314-1324.

Huang, T.-K., and Puchta, H. (2021). Novel CRISPR/Cas applications in plants: from prime editing to chromosome engineering. Transgenic Research *30*, 529-549.

Huang, T.K., and Puchta, H. (2019). CRISPR/Cas-mediated gene targeting in plants: finally a turn for the better for homologous recombination. Plant Cell Reports *38*, 443-453.

Kalladan, R., Lasky, J.R., Sharma, S., Kumar, M.N., Juenger, T.E., Des Marais, D.L., and Verslues, P.E. (2019). Natural Variation in 9-*cis*-Epoxycartenoid Dioxygenase 3 and ABA Accumulation. Plant Physiology *179*, 1620-1631.

Kesari, R., Lasky, J.R., Villamor, J.G., Marais, D.L.D., Chen, Y.J.C., Liu, T.W., Lin, W., Juenger, T.E., and Verslues, P.E. (2012). Intron-mediated alternative splicing of Arabidopsis P5CS1 and its association with natural variation in proline and climate adaptation. Proceedings of the National Academy of Sciences of the United States of America *109*, 9197-9202.

Kumar, M.N., Bau, Y.C., Longkumer, T., and Verslues, P.E. (2019). Low Water Potential and At14a-Like1 (AFL1) Effects on Endocytosis and Actin Filament Organization. Plant Physiology *179*, 1594-1607.

Kumar, M.N., Hsieh, Y.F., and Verslues, P.E. (2015). At14a-Like1 participates in membrane-associated mechanisms promoting growth during drought in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America *112*, 10545-10550.

Lieberman-Lazarovich, M., Melamed-Bessudo, C., de Pater, S., and Levy, A.A. (2013). Epigenetic Alterations at Genomic Loci Modified by Gene Targeting in *Arabidopsis thaliana*. Plos One *8*, e85383.

Mao, Y.F., Botella, J.R., Liu, Y.G., and Zhu, J.K. (2019). Gene editing in plants: progress and challenges. National Science Review *6*, 421-437.

Merker, L., Schindele, P., and Puchta, H. (2020). Using CRISPR/ttLbCas12a for in planta Gene Targeting in *A. thaliana*. Current Protocols in Plant Biology *5*, e20117.

Miki, D., Zhang, W.X., Zeng, W.J., Feng, Z.Y., and Zhu, J.K. (2018). CRISPR/Cas9-mediated gene targeting in Arabidopsis using sequential transformation. Nature Communications *9*, 1967.

Peng, F., Zhang, W., Zeng, W., Zhu, J.-K., and Miki, D. (2020). Gene targeting in Arabidopsis via an all-in-one strategy that uses a translational enhancer to aid Cas9 expression. Plant Biotechnology Journal *18*, 892-894.

Schiml, S., Fauser, F., and Puchta, H. (2014). The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. Plant Journal *80*, 1139-1150.

Shaked, H., Melamed-Bessudo, C., and Levy, A.A. (2005). High-frequency gene targeting in *Arabidopsis* plants expressing the yeast *RAD54* gene. Proceedings of the National Academy of Sciences of the United States of America *102*, 12265-12269.

Sharma, S., and Verslues, P.E. (2010). Mechanisms independent of abscisic acid (ABA) or proline feedback have a predominant role in transcriptional regulation of proline metabolism during low water potential and stress recovery. Plant Cell and Environment *33*, 1838-1851.

Shen, H., Strunks, G.D., Klemann, B., Hooykaas, P.J.J., and de Pater, S. (2017). CRISPR/Cas9-Induced Double-Strand Break Repair in Arabidopsis Nonhomologous End-Joining Mutants. G3-Genes Genomes Genetics *7*, 193-202.

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

van Tol, N., van Schendel, R., Bos, A., van Kregten, M., de Pater, S., Hooykaas, P.J.J., and Tijsterman, M. (2022). Gene targeting in polymerase theta-deficient *Arabidopsis thaliana*. The Plant Journal *109*, 112-125.

Verslues, P.E. (2010). Quantification of Water Stress-Induced Osmotic Adjustment and Pro line Accumulation for Arabidopsis thaliana Molecular Genetic Studies. In Plant Stress Tolerance: Methods and Protocols, R. Sunkar, ed., pp. 301-315.

Vu, T.V., Sivankalyani, V., Kim, E.J., Doan, D.T.H., Tran, M.T., Kim, J., Sung, Y.W., Park, M., Kang, Y.J., and Kim, J.Y. (2020). Highly efficient homology-directed repair using CRISPR/Cpf1-geminiviral replicon in tomato. Plant Biotechnology Journal *18*, 2133-2143.

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

Wolter, F., Klemm, J., and Puchta, H. (2018). Efficient in planta gene targeting in Arabidopsis using egg cell-specific expression of the Cas9 nuclease of *Staphylococcus aureus*. Plant Journal *94*, 735-746.

Wolter, F., and Puchta, H. (2019). In planta gene targeting can be enhanced by the use of CRISPR/Cas12a. Plant Journal *100*, 1083-1094.

Wong, M.M., Bhaskara, G.B., Wen, T.N., Lin, W.D., Nguyen, T.T., Chong, G.L., and Verslues, P.E. (2019). Phosphoproteomics of Arabidopsis Highly ABA-Induced1 identifies AT-Hook-Like10 phosphorylation required for stress growth regulation. Proceedings of the National Academy of Sciences of the United States of America *116*, 2354-2363.

Xie, K.B., Minkenberg, B., and Yang, Y.N. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proceedings of the National Academy of Sciences of the United States of America *112*, 3570-3575.

Figure 1: *HypaCAS9* gene expression in transgenic lines.

A. Expression of *HypaCAS9* in T_2 transgenic lines. One sample from each line was assayed to give an estimate of *HypaCAS9* gene expression while still preserving other plants for retransformation. Floral buds were sampled from one plant for each line.

B. Expression of *HypaCAS9* in T_3 and later generations of plants derived from two T_2 lines with contrasting expression. Three biological replicates were assayed, each containing floral buds from one or two plants. Data are means ± S.D. (n = 3).

Figure 2: GT insertion of *YFP* at the 3' end of *AFL1*.

A. Donor Fragment design and genotyping primer positions for knock-in of C-terminal YFP at the *AFL1* locus. The *YFP* coding sequence was flanked by homology arms matching the indicated regions of *AFL1*. In the DNA sequences shown, blue indicates the AFL1 stop codon and red indicates the PAM site on the sgRNA. Note that replacement of *AFL1* endogenous sequence with the knock-in fragment disrupts the sgRNA recognition site thus preventing cutting of the donor fragment or re-cutting of endogenous *AFL1* after insertion of the donor fragment. Positions of primers used for genotyping are also shown (primer sequences and amplicon sizes can be found in Supplemental Table I).
B. Genotyping of pooled T_2 seedlings from 30 Basta resistant T_1 plants using primer set 2. Red numbers indicate T_2 seed pools with insertion of the YFP donor construct at the *AFL1* locus.
C. Genotyping of individual T_2 plants from line 26. Red numbers indicate plants homozygous for the *AFL1-YFP* knock-in. Green numbers indicate plants having putative single crossover event where homologous recombination occurred at the 3' side of the donor fragment (indicated by the expected size band amplified by primer set 3) but other putative insertion or deletion occurred on the 5' side of the donor fragment (based on lack of band amplified by primer set 2).
D. QPCR of *AFL1* in wild type compared to T_3 and T_6 homozygous plants of *AFL1-YFP* knock-in line #26. Whole seedlings were collected at four days after transfer of seven-day-old seedlings to control or stress (-0.7 MPa) treatments (C = Control, S = Stress). Three or four biological replicates were performed. Data are means ±S.D. Data sharing the same letter are not significantly difference based on ANOVA with Tukey's multiple comparison test (adjusted $p \leq 0.05$).
E. Immunoblot using antisera recognizing AFL1 (top) or YFP (bottom). In both cases, blots were stripped and re-probed with antisera recognizing HSC70 as a loading control. The AFL1 and YFP blots were prepared using the same samples. Protein was extracted from whole seedlings of wild type or T_3 homozygous plants collected four days after transfer of seven-day-old seedlings to control or stress (-0.7 MPa) treatments (C = Control, S = Stress). 10 μg of protein was loaded in each lane. The predicted molecular weight of AFL1 is 42 kD. The predicted molecular weight of the AFL1-YFP fusion protein is 66 kD.
F. Immunoblotting of T_6 seedlings after backcrossing to wild type to remove CAS9 and the recombination donor construct. Experimental conditions were as described for E.

G. Quantification of relative AFL band intensity in three biological replicates. Data sharing the same letter are not significantly different based on ANOVA with Tukey's multiple comparison test (adjusted $p \leq 0.05$).
H. AFL1-YFP protein accumulation pattern in T_6 seedlings under control or stress (four days after transfer to -0.7 MPa conditions. Scale bars indicate 500 μm.
I. Higher magnification images of AFL1-YFP in petiole cells from control or stress treated seedlings. Scale bars indicate 50 μm.

Figure 3: GT insertion of *YFP* at the 5' end of *AFL1*.

A. Donor Fragment design and genotyping primer positions for knock-in of N-terminal YFP at the *AFL1* locus. The YFP coding sequence was flanked by homology arms matching the indicated regions of *AFL1*. In the DNA sequences shown, blue indicates the AFL1 start codon and red indicates the PAM site on the sgRNA. Note that replacement of *AFL1* endogenous sequence with the knock-in fragment disrupts the sgRNA recognition site thus preventing cutting of the donor fragment or re-cutting of endogenous *AFL1* after insertion of the donor fragment. Positions of primers used for genotyping are also shown (primer sequences and amplicon sizes can be found in Supplemental Table I).
B. Genotyping of pooled T_2 seedlings from 29 Basta resistant T1 plants using primer sets 2 and 3. Red number indicates T2 seed pool from line #26 with insertion of the YFP donor construct at the *AFL1* locus.
C. Genotyping of individual T_3 plants from line 26. Red numbers indicate plants homozygous for the *AFL1-YFP* knock-in.
D. QPCR of *AFL1* in wild type compared to T3 and T6 homozygous plants of *AFL1-YFP* knock-in line #26. Whole seedlings were collected at four days after transfer of seven-day-old seedlings to control or stress (-0.7 MPa) treatments (C = Control, S = Stress). Two biological replicates were performed.
E. Immunoblot detection of AFL1. The blot was first probed with antisera recognizing both P5CS1 (as a loading control) and AFL1. Protein was extracted from whole seedlings of wild type or T6 homozygous plants collected four days after transfer of seven-day-old seedlings to control or stress (-0.7 MPa) treatments (C = Control, S = Stress). 10 μg of protein was loaded in each lane. All samples are from the same blot with intervening lanes removed for clarity. The predicted molecular weight of AFL1 is 42 kD. The predicted molecular weight of the YFP-AFL1 fusion protein is 66 kD. The predicted molecular weight of P5CS1 is 78 kD.

Figure 4: GT insertion of *YFP* at the 5' end of *P5CS1*.

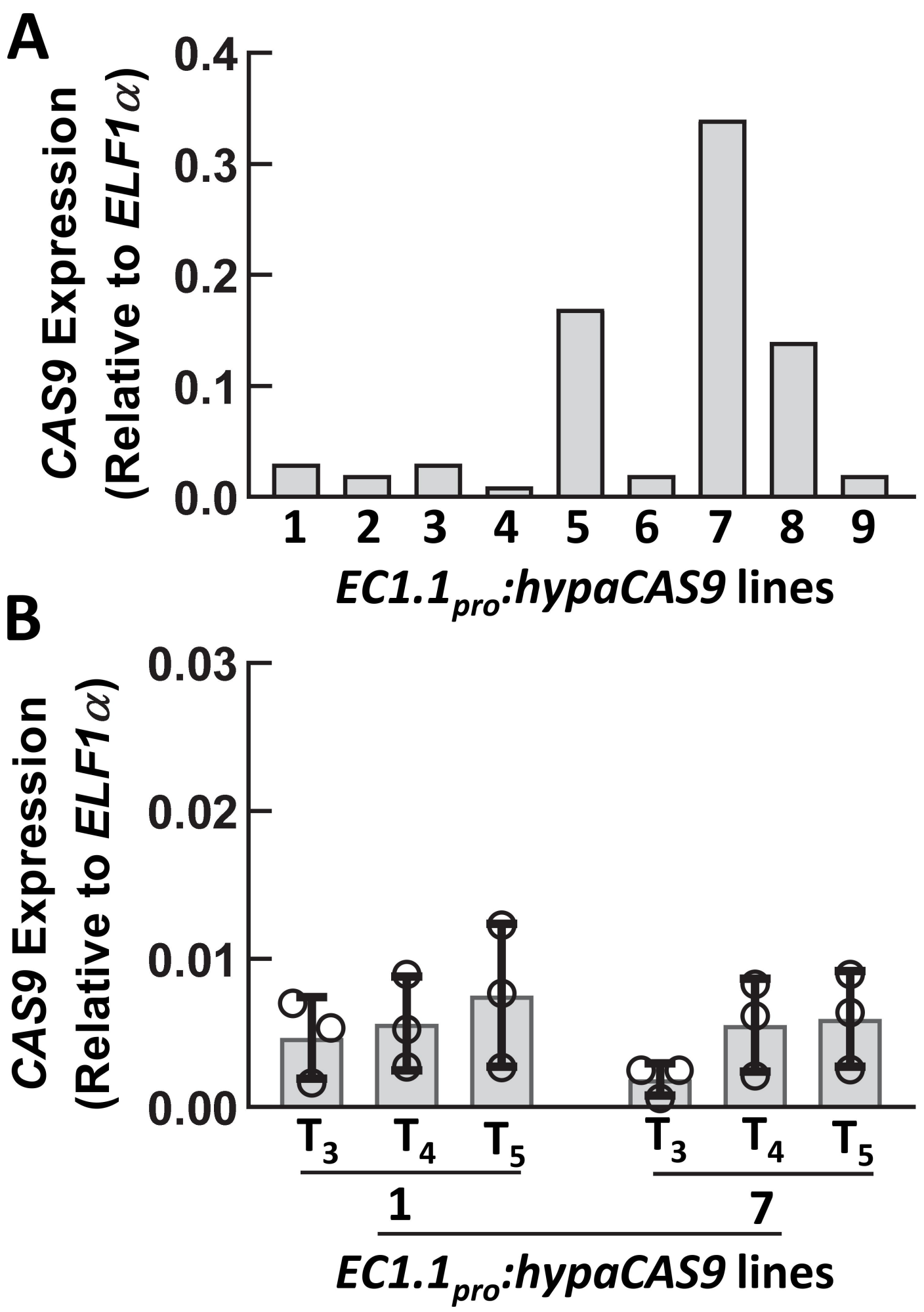
A. Donor Fragment design and genotyping primer positions for knock-in of N-terminal YFP at the *P5CS1* locus. The YFP coding sequence was flanked by homology arms matching the indicated regions of *P5CS1*. In the DNA sequences shown, blue indicates the *P5CS1* start codon and red indicates the PAM site on the sgRNA. Note that replacement of *P5CS1* endogenous sequence with the knock-in fragment disrupts the sgRNA recognition site. Positions of primers used for genotyping are also shown (primer sequences and amplicon sizes can be found in Supplemental Table I).

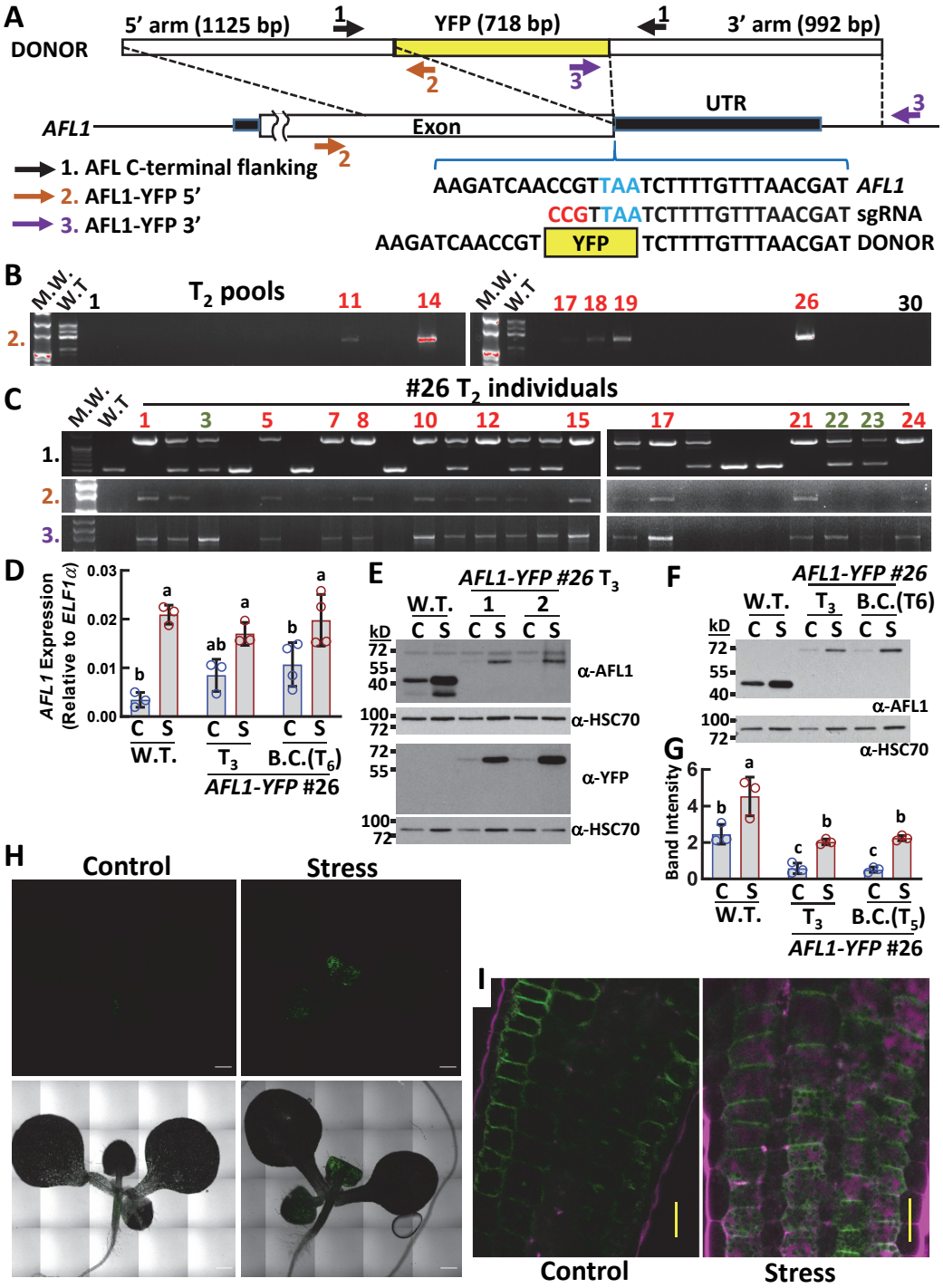
B. Genotyping of pooled T_2 seedlings from 30 Basta resistant T_1 plants using primer set 2. Red numbers indicate T2 seed pools with insertion of the YFP donor construct at the *P5CS1* locus.
C. Genotyping of individual T3 plants from line 6. Red numbers indicate plants homozygous for the *YFP-P5CS1* knock-in. Green numbers indicate plants having putative single crossover event where homologous recombination occurred at the 5' side of the donor fragment (indicated by the expected size band amplified by primer set 2) but other putative insertion or deletion occurred on the 3' side of the donor fragment (based on lack of band amplified by primer set 3).
D. QPCR of *P5CS1* in wild type compared to T_3 and T_6 homozygous plants of *YFP-P5CS1* knock-in line #6. Whole seedlings were collected at four days after transfer of seven-day-old seedlings to control or stress (-0.7 MPa) treatments (C = Control, S = Stress). Two biological replicates were performed.
E. Immunoblot using antisera recognizing P5CS1. Protein was extracted from whole seedlings of wild type or T_3 homozygous plants collected four days after transfer of seven-day-old seedlings to control or stress (-0.7 MPa) treatments (C = Control, S = Stress). 10 µg of protein was loaded in each lane. The predicted molecular weight of P5CS1 is 78 kD; however, we routinely observe that P5CS1 runs at higher than predicted molecular weight (Kesari et al., 2012). Note that the YFP-P5CS1 lanes, the faint bands seen in the 90-170 kD molecular weight range are consistent with previous observations of low level of cross reactivity of this antisera with P5CS2. The band at 57 kD is non-specific and serves as a loading control.
F. Proline accumulation in wild type and YFP-P5CS1 seedlings. Samples were collected four days after transfer of seven-day-old seedlings to control or stress (-0.7 MPa) treatments. Two or three independent experiments were performed and combined data of 3-8 samples from those experiments is shown. Error bars indicate S.D.

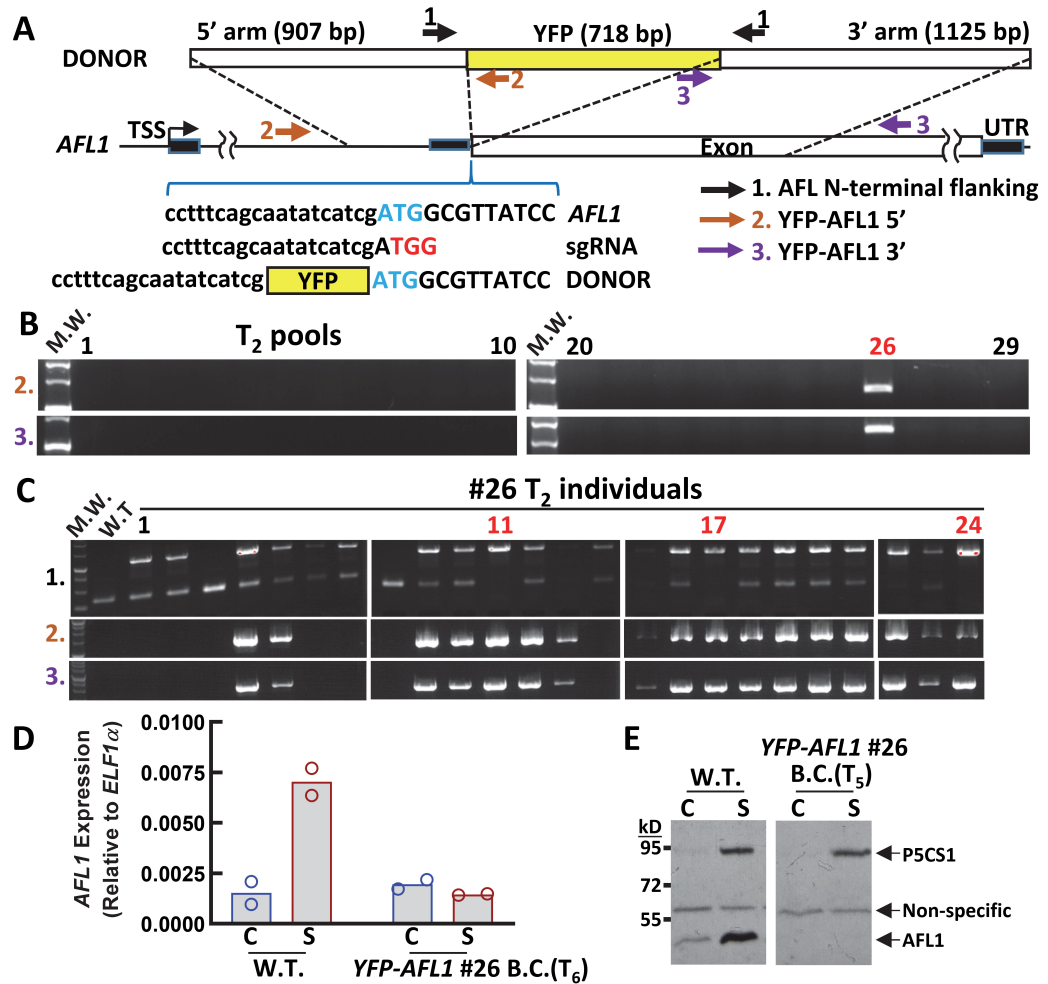
Table 1: GT genotyping results and efficiency of 5' and 3' *YFP* insertion at the *AFL1* and *P5CS1* loci.

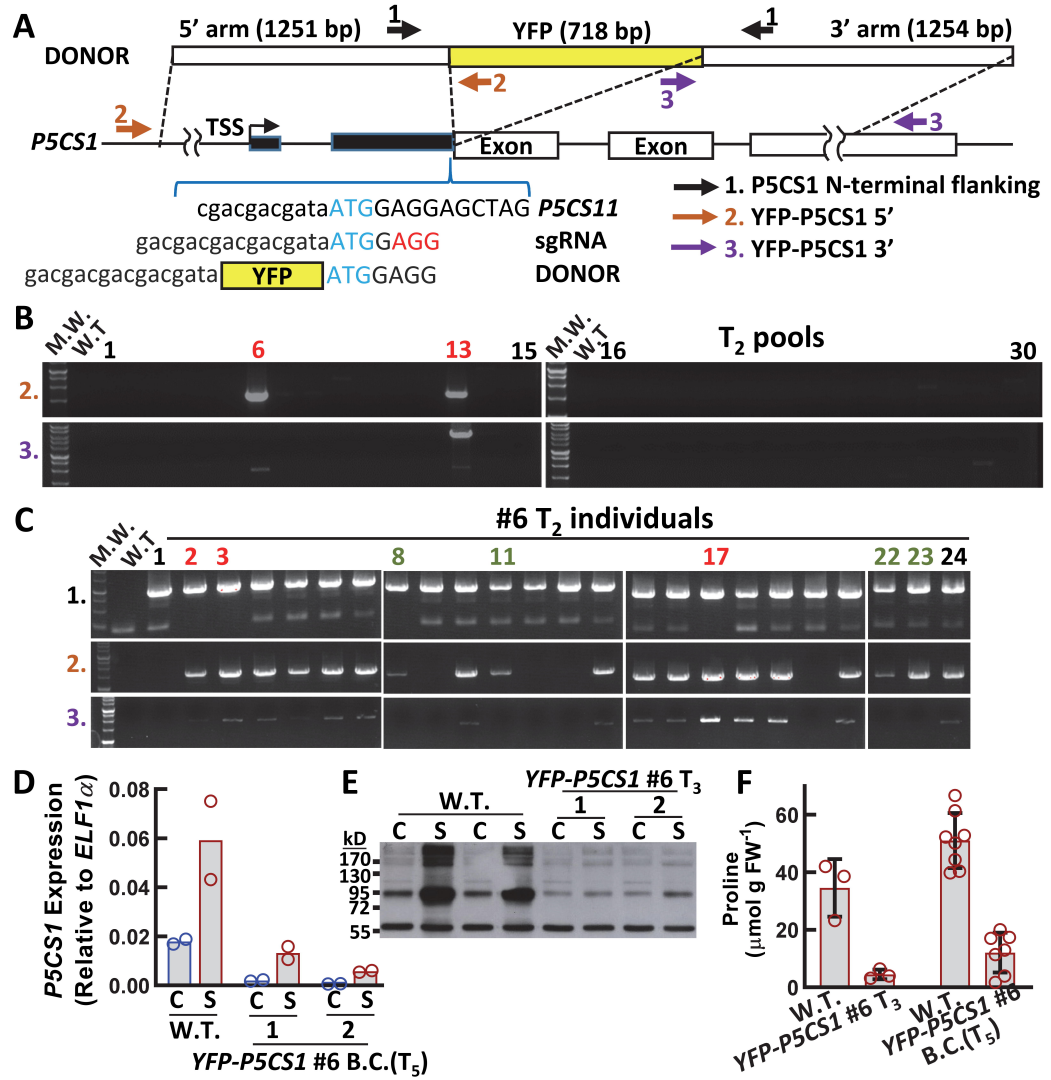
The number of homozygous, heterozygous and W.T. plants were counted based on the flanking PCR (primer set 1) and the presence bands of expected molecular weight for 5' and 3' YFP amplicons (primer sets 2 and 3). Primer positions and PCR results are shown in Figures 2, 3, and 4 as well as Supplemental Figures S1, S4 and S5. Putative single crossovers scored as plants that had amplification of the 5' or 3' YFP PCR but not the other side. Asterisks (*) indicate lines where a band of higher than expected molecular weight was observed on 3' YFP PCR indicating that the construct may have undergone NHEJ incorporating YFP and all (or a portion) of the 3' homology arm rather than homologous recombination.

	Bulk Screening of T_2		Screening Individual T_2 Plants (24 total)				
Parental (Cas9) Line generation	PCR Positive	GT Efficiency	T_2 line	Null (W.T.)	Heterozygous	Homozygous	Putative Single Crossover
	AFL1 5' (YFP-AFL1)						
Line 7, T2	1/29	3.5%	#26	2	17	5	0
	AFL1 3' (AFL1-YFP)						
Line 7, T2	6/30	20%	#11	7	7	10	7
			#14*	2	13	9	1
			#26	5	9	10	3
	P5CS1 5' (YFP-P5CS1)						
Line 7, T3	2/30	7%	#6	4	16	4	2
			#13*	4	15	5	7
	P5CS1 3' (P5CS1-YFP)						
Line 7, T3	0/40	0%	NA	NA	NA	NA	NA


A
CAS9 Expression
(Relative to ELF1α)
0.4
0.3
0.2
0.1
0.0
1
2
3
4
5
6
7
8
9
EC1.1pro:hypaCAS9 lines
B
CAS9 Expression
(Relative to ELF1α)
0.03
0.02
0.01
0.00
T3
T4
T5
T3
T4
T5
1
7
EC1.1pro:hypaCAS9 lines


A
DONOR
5' arm (1125 bp)
YFP (718 bp)
3' arm (992 bp)
AFL1
Exon
UTR
1. AFL C-terminal flanking
2. AFL1-YFP 5'
3. AFL1-YFP 3'
AAGATCAACCGTTAATCTTTTGTTTAACGAT AFL1
CCGTTAATCTTTTGTTTAACGAT sgRNA
AAGATCAACCGT YFP TCTTTTGTTTAACGAT DONOR
B
T2 pools
M.W.
W.T
C
#26 T2 individuals
D
AFL1 Expression (Relative to ELF1α)
W.T.
T3
B.C.(T6)
AFL1-YFP #26
E
AFL1-YFP #26 T3
α-AFL1
α-HSC70
α-YFP
F
AFL1-YFP #26
G
Band Intensity
H
Control
Stress
I
Control
Stress


A
5' arm (907 bp)
YFP (718 bp)
3' arm (1125 bp)
DONOR
AFL1
TSS
Exon
UTR
cctttcagcaatatcatcgATGGCGTTATCC AFL1
cctttcagcaatatcatcgATGG sgRNA
cctttcagcaatatcatcg YFP ATGGCGTTATCC DONOR
1. AFL N-terminal flanking
2. YFP-AFL1 5'
3. YFP-AFL1 3'
B
M.W.
T2 pools
1
10
20
26
29
C
#26 T2 individuals
M.W.
W.T
1
11
17
24
D
AFL1 Expression (Relative to ELF1α)
0.0100
0.0075
0.0050
0.0025
0.0000
C
S
W.T.
YFP-AFL1 #26 B.C.(T6)
E
YFP-AFL1 #26
W.T.
B.C.(T5)
kD
95
72
55
P5CS1
Non-specific
AFL1


A
5' arm (1251 bp)
YFP (718 bp)
3' arm (1254 bp)
DONOR
P5CS1
TSS
Exon
Exon
cgacgacgataATGGAGGAGCTAG P5CS11
gacgacgacgacgataATGGAGG sgRNA
gacgacgacgacgata YFP ATGGAGG DONOR
1. P5CS1 N-terminal flanking
2. YFP-P5CS1 5'
3. YFP-P5CS1 3'
B
M.W.
W.T
T2 pools
C
#6 T2 individuals
D
P5CS1 Expression (Relative to ELF1α)
W.T.
YFP-P5CS1 #6 B.C.(T5)
E
YFP-P5CS1 #6 T3
kD
170
130
95
72
55
F
Proline (μmol g FW-1)
W.T.
YFP-P5CS1 #6 T3
YFP-P5CS1 #6 B.C.(T5)