




CRISPR RNA-guided integrase enables high-efficiency targeted genome engineering in *Agrobacterium tumefaciens*

Ephraim Aliu^{1,2,3} , Keunsub Lee^{1,2}  and Kan Wang^{1,2,*} 

¹Department of Agronomy, Iowa State University, Ames, Iowa, USA

²Crop Bioengineering Center, Iowa State University, Ames, Iowa, USA

³Interdepartmental Plant Biology Major, Iowa State University, Ames, Iowa, USA

Received 19 April 2022;

revised 3 June 2022;

accepted 8 June 2022.

*Correspondence (Tel: 1-515-294-4429;

Fax: 515-294-3163;

email kanwang@iastate.edu)

Keywords: Cre-*loxP* recombination, Large DNA fragment deletion, Transposon, Targeted DNA insertion, T-DNA.

Summary

Agrobacterium tumefaciens, the causal agent of plant crown gall disease, has been widely used to genetically transform many plant species. The inter-kingdom gene transfer capability made *Agrobacterium* an essential tool and model system to study the mechanism of exporting and integrating a segment of bacterial DNA into the plant genome. However, many biological processes such as *Agrobacterium*-host recognition and interaction are still elusive. To accelerate the understanding of this important plant pathogen and further improve its capacity in plant genetic engineering, we adopted a CRISPR RNA-guided integrase system for Transposable Elements by Guide RNA-Assisted TargEting (INTEGRATE) can efficiently generate DNA insertions to enable targeted gene knockouts. In addition, in conjunction with Cre-*loxP* recombination system, we achieved precise deletions of large DNA fragments. This work provides new genetic engineering strategies for *Agrobacterium* species and their gene functional analyses.

Introduction

Plant genetic transformation is critical for fundamental biological research and modern crop improvement. One widely adopted transformation technology is the *Agrobacterium*-mediated method. Naturally, the gram-negative soil bacterium *Agrobacterium tumefaciens* exists as an opportunistic pathogen in plants, and this niche is due to the presence of a large tumour-inducing (Ti) plasmid. During pathogenesis, *A. tumefaciens* transfers a segment of its DNA (transfer DNA or T-DNA) carrying plant growth hormone genes into the host plant's genome (Banta and Montenegro, 2008; Gelvin, 2021). The overproduction of the phytohormones often leads to neoplastic growth on the infection sites, termed 'crown gall' disease (Păcurar *et al.*, 2011). Furthermore, the T-DNA harbours a set of genes required to produce opines, which are used as energy sources for specific *Agrobacterium* strains. For the T-DNA delivery, *Agrobacterium* senses environmental signals in the form of low pH, simple sugars, and phenolic exudates from wounded plants through the aid of virulence (*vir*) genes encoded on the chromosome and Ti plasmid (Gelvin, 2003, 2017). Due to their ability to transfer T-DNAs, *A. tumefaciens* have been modified over the years for plant transformation. The modifications included 'disarming' the Ti plasmid by deleting the natural T-DNA that carries 'plant tumour' genes (Hood *et al.*, 1986; Koncz and Schell, 1986; Ooms *et al.*, 1981), and utilizing a synthetic T-DNA(s) from a binary plasmid (Komari *et al.*, 2006). Other modifications include the generation of 'disarmed' strains such as EHA101 and EHA105 from highly virulent variants (Hood *et al.*, 1986; Hood *et al.*, 1993; Sciaky *et al.*, 1978); generation of the 'super-

binary vector' (Komari *et al.*, 2006) and 'helper plasmid' (Anand *et al.*, 2018) that carries an additional copy of essential *vir* genes; and inactivating the *recA* recombinase gene to reduce plasmid recombination (Lazo *et al.*, 1991). More recently, auxotrophic *Agrobacterium* strains have been generated to minimize post-infection bacterial overgrowth during the plant transformation processes (Aliu *et al.*, 2020; Anand *et al.*, 2018; Ranch *et al.*, 2012).

While the engineered *Agrobacterium* strains and binary vectors are popular for plant genome engineering, they do not work equally well on all plants. For example, strains and vectors that can be used for transforming maize may not work efficiently for yam plants. In addition, the genetic and molecular basis for *Agrobacterium* host range determination remains unknown (Gelvin, 2003, 2021). Moreover, even though the T-DNA can be successfully transferred and integrated into the plant genome, the information and knowledge of its integration mechanisms are still lacking. As a result, the transformation of many plant species, especially economically important crops, has been challenging (Alteter *et al.*, 2016). With the availability of the whole genome sequences of *Agrobacterium tumefaciens* C58 (Goodner *et al.*, 2001; Wood *et al.*, 2001), an opportunity is presented to further explore its biology through gene editing and genome engineering. The traditional molecular approach for gene knockin or knockout in *Agrobacterium* requires homologous recombination (HR) or transposon insertional mutagenesis (Morton and Fuqua, 2012). However, these methods have a combined drawback of being laborious, time-consuming, and inefficient; as culturable and transformable bacteria often require screening many colonies to ascertain true mutants (Aliu *et al.*, 2020; Tamzil *et al.*, 2021).

The advent of CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated protein) technology has made it easier to perform genome editing. Rodrigues *et al.* reported the generation of loss-of-function mutations of *recA* gene in a popular disarmed *Agrobacterium* strain EHA105 using CRISPR-mediated base editing (Rodrigues *et al.*, 2021). Recently, RNA-guided DNA integration systems have been developed for genome engineering in prokaryotes. These integration systems have been utilized to engineer bacterial cells with new functions via the stable integration of genes and pathways into the genome (Klompe *et al.*, 2019; Makarova *et al.*, 2018; Strecker *et al.*, 2019). Especially, the **IN**sertion of **T**ransposable **E**lements by **Guide RNA**–**A**ssisted **T**argEting (INTEGRATE) system utilizing a Type 1-F CRISPR-Cas system from *Vibrio cholerae* showed highly precise and efficient bacterial genome engineering capability (Klompe *et al.*, 2019). INTEGRATE system consists of TniQ-Cascade complex for RNA-guided DNA targeting and transposase proteins TnsA, TnsB, and TnsC for cargo DNA transposition (Klompe *et al.*, 2019). Target site is determined by a dinucleotide 5'-CC-3' protospacer adjacent motif (PAM) and ~32-nt protospacer sequence, and DNA integration occurs about 48–52 bp downstream from the protospacer (Klompe *et al.*, 2019). Insertion DNA can be cloned into a mini transposon cargo between the left and right end sequences (L-end and R-end), and the cargo DNA size is negatively correlated with the targeted DNA insertion frequency (Klompe *et al.*, 2019; Vo *et al.*, 2021). INTEGRATE vectors stably replicate within the bacterial cells and can be easily removed after targeted DNA integration using a negative selectable marker such as *sacB* gene (Gay *et al.*, 1983; Steinmetz *et al.*, 1983) on the vector backbone. INTEGRATE does not require target site homology as in HR or the use of selectable markers and has been reported to achieve an insertion efficiency of ~100% in gram-negative bacteria, including *Escherichia coli*, *Klebsiella oxytoca*, and *Pseudomonas putida* for single gene targeting (Vo *et al.*, 2021). Because INTEGRATE system does not generate double-strand DNA breaks (DSB), there is little to no off-target effect (Vo *et al.*, 2021). In addition, although the target sequence (PAM + protospacer) is not abolished by the cargo DNA insertion, the presence of the cargo provides 'immunity' preventing repetitive insertions at the same target site allowing precise single-copy insertion events (Vo *et al.*, 2021). Furthermore, when combined with the *Cre-loxP* recombination system, small and large precise genomic deletions can be achieved (Vo *et al.*, 2021). However, as with other gene-editing systems, determining how to strategically port the INTEGRATE system into other bacteria of scientific relevance like *A. tumefaciens* poses a significant challenge due to cross-species incompatibilities and host-specific factors (Wannier *et al.*, 2020).

In this paper, we report optimization and successful implementation of the INTEGRATE system in *A. tumefaciens*. We demonstrate that single and multiplex integration can be achieved in *Agrobacterium* with high efficiency. In addition, we show that insertional mutants resulting from INTEGRATE are stable, exhibit the expected phenotypes, and have similar plant transformation ability compared to the wild-type (WT) when evaluated by transient gene expression. Furthermore, we demonstrate that by combining INTEGRATE with the *Cre-loxP* recombination systems, we can precisely and effectively make large T-DNA deletions of 25 kb and 42.6 kb in the WT Ti plasmids pTiC58 and pTiBo542, respectively. This work provides a simple and efficacious genome editing tool for *Agrobacterium* to

accelerate gene functional analysis and explore many unanswered questions in *Agrobacterium* biology.

Results

Optimized multiplex DNA integration in *Agrobacterium*

To implement the INTEGRATE system for *Agrobacterium* genome engineering, we tested two different plasmid backbones with broad host range origin of replication (ORI) pBBR1 and pVS1, respectively (Figure 1a), using the *Agrobacterium* strains listed in Table S1. As a cargo, we cloned a red fluorescent protein marker gene *mCherry* (Shaner *et al.*, 2004) driven by a constitutive J23107 promoter (Iverson *et al.*, 2016) between the two transposon ends (L- and R-end) (Figures 1b, 2a,b). For easy curing of the plasmids after targeted DNA insertion, we cloned a negative selection marker gene *sacB* (Figure S1a, b), which confers sucrose sensitivity by converting sucrose into toxic levan (Gay *et al.*, 1983; Steinmetz *et al.*, 1983).

To test the effects of plasmid ORIs and the feasibility of multiplexed DNA insertions in *Agrobacterium*, we selected three target genes, *virG* (transcriptional activator of *vir* genes), *apg1* (bacteriophytochrome protein), and *flaA* (flagella associated protein), based on their non-lethality and known null mutant phenotypes (Chesnokova *et al.*, 1997; Gelvin, 2003; Merritt *et al.*, 2007; Xue *et al.*, 2021). Because the presence of the INTEGRATE vectors (harbouring a constitutive promoter driving *mCherry* reporter) conferred red fluorescence even in the absence of targeted insertion, we used PCR screening to identify insertion mutants, as insertion of the *mCherry* cassette in the genome generates larger amplicons (>1.5 kb) compared to the WT (~0.5 kb, Figure 1b; Table S2). To estimate the targeted DNA insertion frequencies, we screened a total of 144 individual colonies per target gene (48 colonies x 3 independent experiments) and classified them into three groups: WT (no DNA insertion); Targeted insertion (targeted DNA insertion without WT band); Mixed insertion (targeted DNA insertion and WT band). As shown in Figure 1c, the INTEGRATE vector with pVS1 ORI performed much better than the pBBR1 ORI vector at all three target sites: Targeted insertion frequencies without WT band were 3.5–22.2% and 20.1–75.0% for pBBR1 and pVS1 ORIs, respectively (Figure 1c; paired *t*-test, *t* = 3.0, *df* = 8, *P* < 0.05). Differences in total targeted insertion frequencies including the colonies with Mixed insertion were even greater with 8.3–42.4% for pBBR1 ORI and 45.1–98.6% for pVS1 ORI, respectively (Figure 1c; paired *t*-test, *t* = 5.6, *df* = 8, *P* < 0.001). Pure insertion mutants were readily obtained from the Mixed insertion colonies by re-plating after serial dilution as described in the methods. The three tested gRNAs exhibited varying levels of efficiency ranging from 0.7% (gRNA1 with pBBR1 ORI) to 75% (gRNA3 with pVS1 ORI; Figure 1c). Interestingly, gRNA2 exhibited smaller difference in Targeted insertion frequency (22.2% vs. 32.6% = 1.5-fold) compared to the other two gRNAs (gRNA1, 3.5% vs. 20.1% = 5.7-fold; gRNA3, 12.5% vs. 75.0% = 6.0-fold). Consistent with the previous report (Vo *et al.*, 2021), we also observed that gRNA efficiencies were dependent on the relative position of the spacer within the crRNA array: i.e., the first spacer (gRNA3, *flaA*) exhibited highest efficiency followed by the second (gRNA2, *apg1*) and the third (gRNA1, *virG*) spacers (Figure 1c; Figure S1). Importantly, with pVS1 ORI, colonies with targeted insertions were readily obtained with an average of 42% at all three target sites (triple insertion mutants) highlighting the

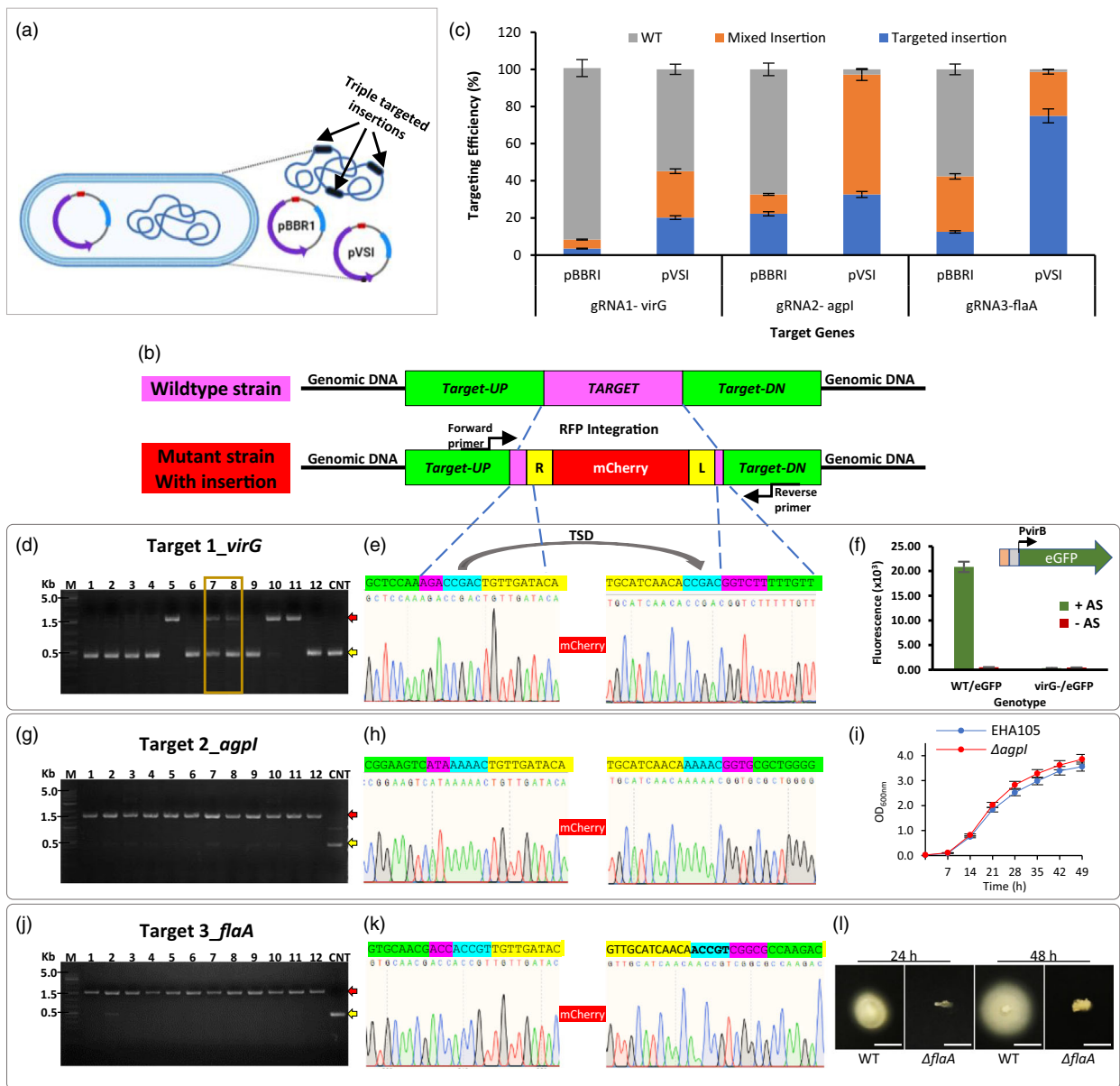


Figure 1 Optimized multiplex DNA integration in *Agrobacterium*. (a) Schematic of INTEGRATE-mediated targeted DNA insertions in *Agrobacterium* with distinct vector backbones, pBBR1 and pVS1. (b) Schematics of mCherry insertion at a target site. Target-UP/DN, upstream and downstream sequence of a target site, respectively; forward and reverse primers for screening targeted insertion. (c) Targeted DNA integration frequency (Targeting efficiency) at three target sites estimated by PCR analysis. Targeting efficiency (%) was estimated by screening a total of 144 single colonies per target gene with three biological replicates (48 × 3 = 144). Data are shown as mean ± s.d. for n = 3 biological replicates. (d,g,j) Verification of targeted DNA insertion by PCR screening. Red arrowheads indicate representative triple targeted insertion mutant; yellow arrowheads indicate WT control band; gold square indicates mixed insertion mutant. (e,h,k) Chromatograms show the Sanger sequencing results confirming the targeted mCherry insertion for the target sites of *virG* (e), *agpl* (h), and *flaA* (k). TSD, target site duplication (blue coloured region). A larger band (~1.5 kb) indicates targeted mCherry insertion, whereas a smaller band (~0.5 kb) indicates no DNA integration (WT). Lanes 1–12, screened single colonies; CNT, wild-type (WT) *Agrobacterium* EHA105 control. (f,i,l) Phenotypic analysis of the triple insertion mutant (EHA105VAF). The mutant strain exhibited expected mutant phenotypes: lack of *vir* gene induction for $\Delta virG$ (f), faster growth rate at 25 °C for $\Delta agpl$ (i), and compromised motility for $\Delta flaA$ (l). Scale bars in (l) are 1 cm.

efficacy of our INTEGRATE system for targeted DNA insertions in *Agrobacterium*.

To confirm the targeted DNA insertion into the target sites, PCR products were purified and subjected to Sanger sequencing. Targeted mCherry insertion was confirmed from all three target sites (Figures 1d,g,j), and the integration site was 47–54 bp downstream from the protospacer with either forward (T-

RL) or reverse cargo orientation (T-LR), which is consistent with the previous report (Vo *et al.*, 2021) (Table S3). Sanger sequencing also revealed that all targeted insertion events were accompanied by a 5-bp target site duplication (TSD) at the L- or R-end (Figures 1e,h,k), a well-known feature generated by transposition of Tn7-like elements (Peters *et al.*, 2017), further confirming that targeted DNA insertion generated insertional

mutations. We then cured the INTEGRATE vectors of the triple insertional mutant strain (EHA105VAF) deficient of *virG*, *agp1*, and *flaA* by plating the cells on a solid YEP medium amended with 5% sucrose. Emerging single colonies with antibiotics sensitivity were further tested for targeted insertion at all three target sites by PCR and Sanger sequencing before being used for phenotyping analysis.

For the *virG* null mutant phenotype, we used a *vir* gene induction assay. *Agrobacterium* utilizes VirA/G two-component system to regulate *vir* gene expression (Gelvin, 2003, 2017). A well-known *Agrobacterium* *vir* gene inducer is acetosyringone (AS), which is a phenolic exudate from wounded plants (Ashby *et al.*, 1987) and is commonly used for *Agrobacterium*-mediated plant transformation. AS has the ability to interact with

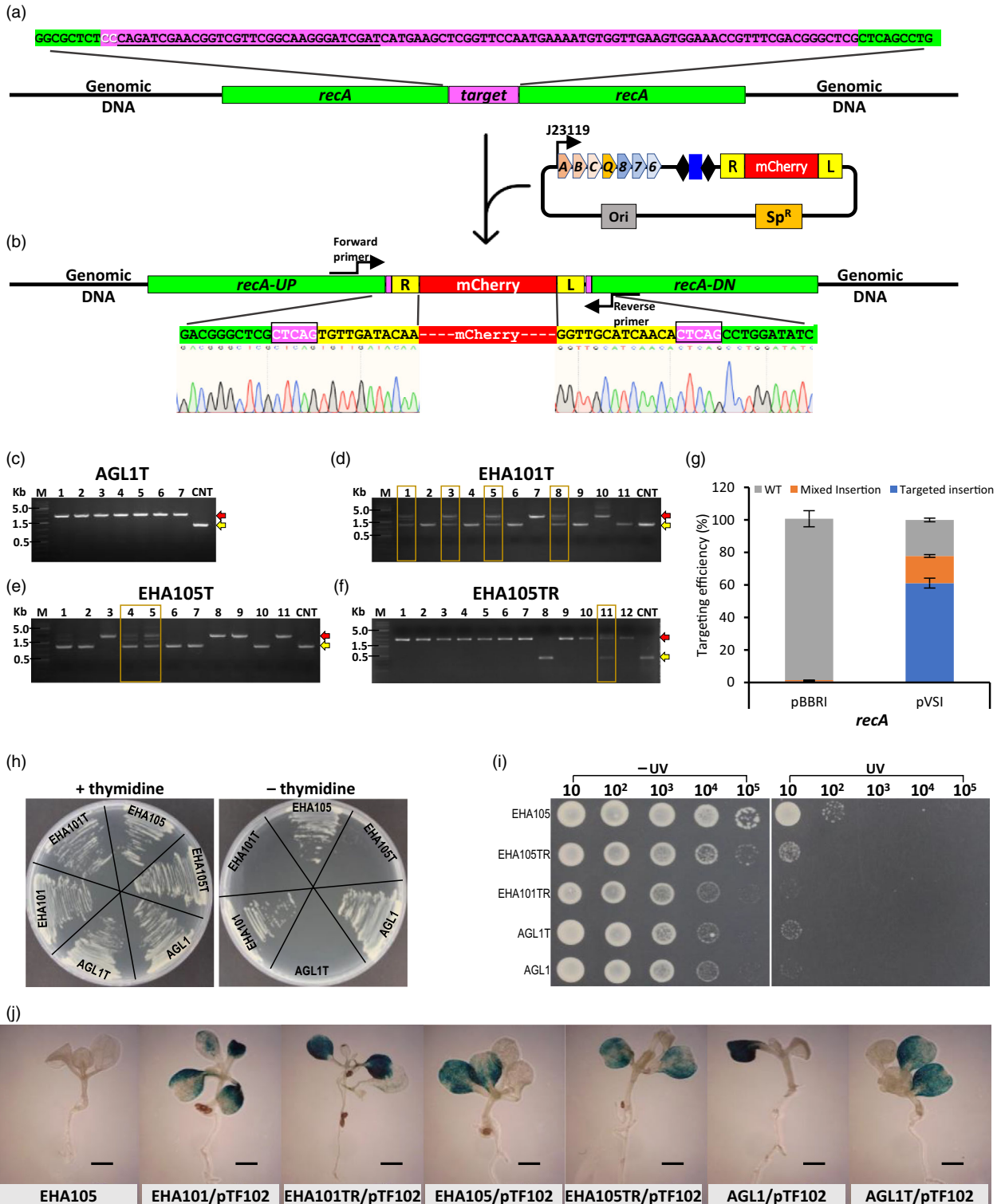


Figure 2 Utilization of INTEGRATE to generate useful *Agrobacterium* strains for plant transformation. (a and b) Schematics of the *recA* genomic locus before and after targeted mCherry insertion. The pVS1-based INTEGRATE construct is used for targeted insertion. (b) Chromatogram data confirm the targeted mCherry insertion at the *recA* genomic locus. The 5-bp TSD (CTCAG) is highlighted with pink colour. PCR screening of targeted insertion at *thyA* of AGL1 (c), EHA101 (d), EHA105 (e), and at *recA* of EHA105Thy- (f) strains. Red arrowheads indicate representative targeted insertion mutant; yellow arrowheads indicate WT control; and gold squares show mixed insertion mutant. (g) PCR-based quantification of integration efficiency at *recA* locus using two different origin of replications, pBBR1 and pVS1. Targeting efficiency (%) was estimated by screening a total of 144 single colonies with three biological replicates (48 x 3 = 148). Data are shown as mean \pm s.d. for n = 3 biological replicates. (h) Thymidine-dependent growth of auxotrophic *Agrobacterium* strains shown via streaked agar plates of *Agrobacterium* cells grown with/without 50 mg/L of thymidine. (i) Ultraviolet (UV) radiation assay for *recA*-deficient *A. tumefaciens* strains: EHA105TR, EHA101TR, and AGL1T. EHA105 (WT) and AGL1 (WT) served as negative and positive controls, respectively. (j) Transient GUS expression assay using *Arabidopsis efr-1* seedlings to test the T-DNA delivery capacity of *Agrobacterium* insertion mutants. All strains were transformed with a binary vector pTF102 with a *gus* reporter gene for AGROBEST assay. Wild-type EHA105 strain without the binary vector served as a negative control. Scale bar = 2 mm.

transmembrane VirA protein, which in turn phosphorylates VirG. Once phosphorylated, VirG acts as a transcriptional activator for other *vir* genes (Winans *et al.*, 1988). We generated a reporter plasmid, pEA183 (Table 1; Figure S2b) that has a green fluorescent protein (EGFP) gene under a VirG-inducible *virB* promoter (PvirB) as described in the methods. The reporter pEA183 was introduced into the wild type EHA105 and the triple mutant strain EHA105VAF. To monitor the *virG* activity in these strains, we measured the green fluorescence with and without induction by AS. As shown in Figure 1f, the control strain (WT/eGFP) showed induction of the EGFP expression, whereas only a basal level of EGFP expression was observed in the EHA105VAF mutant strain (*virG*-eGFP), indicating that the triple mutant strain indeed lacks the ability to activate *vir* gene expression.

We then tested if the triple mutant strain showed *agp1* null mutant phenotype via a temperature-sensitive growth assay. *Agp1/2* are two phytochrome photoreceptor genes encoded in *A. tumefaciens* C58 genome which are responsible for sensing lights in blue, red, and far-red wavelengths (Lamparter *et al.*, 2021). Previous studies in *Agrobacterium* have shown that *Agp1/2* can inhibit cell propagation under nutrient deprivation (Xue *et al.*, 2021), as the single and double knockout mutants exhibited faster growth and reached higher cell densities than

the WT strain during an expanded growth period. Because the histidine kinase domain of *Agp1* is temperature-dependent, exhibiting lower phosphorylation activity as temperature increases above 25 °C (Njimonu and Lamparter, 2011), we grew the mutant and WT strains at 25 °C and monitored cell density over the course of 49 h. Consistent with the previous report (Xue *et al.*, 2021), we observed that the *agp1*-deficient strain grew faster than the WT and the difference became evident after 21 h (Figure 1i; Student *t*-test, $t = 3.6$, $df = 5$, $P < 0.05$), suggesting growth inhibition by the *Agp1* in the WT strain under nutrient deprivation conditions.

Lastly, we tested the *flaA* mutant phenotype using the swimming motility assay. Bacteria depend on external appendages, including pili and flagella, for movement through their environment. Swimming motility in *Agrobacterium* is mediated by flagella, which is encoded by four discrete flagellin gene homologues (*flaABCD*), and mutants of the flagellin genes have been reported nonmotile and attenuated for virulence (Chesnokova *et al.*, 1997; Merritt *et al.*, 2007). We assessed the swimming motility of the triple mutant strain on 0.25% swim agar plates (Figure 1j). Compared to the WT strain, the triple mutant strain demonstrated irregular colony shape and reduced motility, consistent with the *flaA* knockout mutant phenotype

Table 1 *Agrobacterium* strains and plasmids used and produced in this study

Item	Description	Reference
Strains		
AGL1T	<i>thyA</i> mCherry insertion mutant derived from AGL1	This study; Figure S3
Bo542 Δ T-DNA	Disarmed Bo542 strain	This study; Figure 3
C58 Δ T-DNA	Disarmed C58 strain	This study; Figure 3
EHA101TR	<i>recA</i> mCherry insertion mutant derived from EHA101Thy-	This study; Figure 2i
EHA105TR	<i>recA</i> mCherry insertion mutant derived from EHA105Thy-	This study; Figure 2i
EHA105VAF	<i>virG</i> , <i>agp1</i> and <i>flaA</i> mCherry insertion mutant derived from EHA105	This study; Figure 1
Plasmids		
pEA106	Inducible PvirB-mCherry construct	This study; Addgene #187872
pEA183	Inducible PvirB-eGFP construct	This study; Figure S2b; Addgene #187873
pEA186	Empty pVS1-based INTEGRATE vector with mCherry cargo; dual <i>Bsa</i> I-sites for new spacer cloning; <i>sacB</i> gene for plasmid curing; spectinomycin resistant.	This study; Addgene #187874
pKL2310	Empty pVS1-based INTEGRATE vector with <i>loxP</i> cargo; cargo can be replaced by <i>Pst</i> I and <i>Xho</i> I digestion; dual <i>Bsa</i> I-sites for new spacer cloning; spectinomycin resistant.	This study; Addgene #187875
pKL2315	Cre recombinase expression vector; <i>sacB</i> gene for plasmid curing; pVS1 ORI; kanamycin resistant.	This study; Figure S4c; Addgene #187876

observed in previous reports (Chesnokova *et al.*, 1997; Deakin *et al.*, 1999). Together, these results confirm that our improved INTEGRATE system with pVS1 ORI is an efficient tool for multiplex targeted mutagenesis. The system can accelerate gene functional analyses in *Agrobacterium* and related bacteria.

Utilization of INTEGRATE to engineer *Agrobacterium* strains

Using the improved INTEGRATE system with pVS1 ORI (Figures 2a,b), we sought to engineer *Agrobacterium* strains to enhance plant transformation. As an essential biotechnology tool for plant gene functional analyses and crop trait improvements, the easy removal of *Agrobacterium* cells after the co-cultivation period is one of the highly desired traits. Auxotrophic *Agrobacterium* strains have been proven to be useful for plant transformation (Anand *et al.*, 2018; Ranch *et al.*, 2012). We previously generated thymidine auxotrophic strains EHA101Thy- and EHA105Thy- using a suicide vector and homologous recombination approach (Aliu *et al.*, 2020). However, this traditional mutagenesis approach suffered a low efficiency in generating thymidine knockout mutants (Aliu *et al.*, 2020). Additionally, the recombination-based mutagenesis cannot be used for *Agrobacterium* strains with suppressed functions in homologous recombination. For example, AGL1 strain is one of the widely used public strains for plant transformation due to insertional mutagenesis of the *recA* gene (Lazo *et al.*, 1991), which encodes a recombinase enzyme essential for homologous recombination (Chen *et al.*, 2008). The AGL1 strain is known to have greater plasmid stability when the binary vector carries DNA sequences with homologies.

Our previous attempt to generate thymidine auxotrophic AGL1 strain using the traditional recombination mutagenesis approach was not successful, because the *recA* mutation suppressed targeted *thyA* gene deletion via homologous recombination (Aliu *et al.*, 2020). Using our improved INTEGRATE system, we attempted to generate thymidine auxotrophic AGL1 strain (AGL1T) as well as both thymidine auxotrophic/*recA*-deficient (TR) EHA101 and EHA105 strains (Figure 2). The mCherry cargo used in the multiplex DNA insertions described above was also used here for single gene insertions (Figures 2a,b). A 32-bp spacer targeting *recA* (*atu1874*) or *thyA* (*atu2047*) were cloned into the crRNA repeats as described in the methods. The resulting constructs (pEA189 and pEA190, Table S1) were introduced into EHA101, EHA105, and AGL1 strains via electroporation. After re-plating (see Methods), emerging single colonies were screened by PCR to identify single colonies with targeted insertion (Figures 2c-f), and Sanger sequencing analysis verified the precise insertion of the mCherry cargo into the target site (Figure 2b). As shown in Figure 2c, single colonies with targeted insertion were readily obtained for all target sites (*thyA* and *recA*) and strains (AGL1, EHA101, and EHA105). In addition, consistent with the multiplexing experiments, INTEGRATE vector with pVS1 ORI showed a much higher insertion frequency than the pBBR1 vector for *recA* gene (61.1% vs. 0.7%, Figure 2g), further demonstrating much improved targeted DNA insertion efficiency using the pVS1 ORI.

The resulting *thyA* mutant strains were tested for their thymidine dependence. Consistent with our previous observation (Aliu *et al.*, 2020), all thymidine auxotrophic strains (AGL1T, EHA101T, and EHA105T) only grew in the presence of 50 mg L⁻¹ thymidine in contrast to the corresponding wild-type strain (Figure 2h), thus confirming successful knockout of the *thyA* gene by insertional mutagenesis. To monitor the stability of the

insertional mutants generated by the INTEGRATE system, bacterial plate cultures were streaked every two days for 20 days. Colonies formed only in the presence of supplemented thymidine with a constitutive expression of red fluorescent protein, indicating that these strains did not lose the inserted mCherry cargo. In addition, Sanger sequencing analysis on the colonies from the 8th-streaked plate showed that the target site in the AGL1T strain did not have additional sequence changes (Figure S3).

The *recA* gene encodes a protein that is essential for DNA repair (Chau *et al.*, 2008). To test the *recA* insertion mutants (EHA101TR and EHA105TR), we subjected the *Agrobacterium* strains to UV-B irradiation to observe their UV susceptibility. Consistent with the previous report (Chau *et al.*, 2008), EHA101TR, EHA105TR, AGL1, and AGL1T exhibited increased sensitivity to UV irradiation compared to the wild-type EHA105 strain (Figure 2i), confirming that these strains are indeed *recA*-deficient mutants.

Lastly, we tested if these strains could deliver T-DNAs into plant cells as well as their WT strains using a transient transgene expression assay, AGROBEST (Wu *et al.*, 2014) (Figure 2j). A binary vector containing a GUS reporter gene, pTF102 (Frame *et al.*, 2002), was introduced into the mutant and corresponding WT strains via electroporation. These strains were then used for infecting *Arabidopsis thaliana* T-DNA insertion mutant *efr-1* (SALK 044334) (see Methods). Transient GUS staining results showed that there was no noticeable difference among the mutants and their parental strains as all tested strains showed strong GUS expression in the Arabidopsis seedlings (Figure 2j). These results indicate that our auxotrophic and *recA*-deficient *Agrobacterium* strains (EHA101TR, EHA105TR, and AGL1T) generated by the INTEGRATE system are stable and suitable for plant transformation. Furthermore, their engineered traits would be advantageous for controlling *Agrobacterium* overgrowth or handling large plasmids with repetitive sequences.

Large DNA fragment deletions

In conjunction with Cre-*loxP* recombination system, we utilized INTEGRATE system to demonstrate an efficient protocol to disarm two natural Ti plasmids (pTiBo542 and pTiC58) by deleting the T-DNAs. The *loxP* site was inserted into the mini cargo. Two spacer sequences targeting the flanking regions of the T-DNAs of the two pTi plasmids were cloned into the crRNA array in the INTEGRATE vector with pVS1 backbone resulting in pKL2308 (for pTiBo542) and pKL2313 (for pTiC58) (Figures S4a,b; Table S1).

As illustrated in Figure 3a, a large fragment deletion (42.6 kb) was conducted in two steps. The first step was to insert the *loxP* sites into the two target sites. Because of the improved efficiency with the pVS1 ORI, homogeneous *Agrobacterium* colonies with simultaneous insertions at both target sites were obtained from the Bo542 cells. Four out of 36 colonies carried homogeneous targeted insertion at both target sites (Table S4). However, targeted insertion efficiency was lower in the C58 cells as only 11% of screened colonies (3/28) had *loxP* insertion at either target site, but none had simultaneous insertions at both targets (Table S1). An extra re-plating step was performed using a colony with *loxP* insertion at target 2 to obtain homogenous colonies with insertions at both target sites with a 70% frequency (14/20, Table S4). Sanger sequencing analysis was used to identify deletion-ready colonies, which have *loxP* insertions at both target sites with the same orientation. Three out of four Bo542/2308 colonies with targeted insertions had forward cargo orientation (T-RL) at both target sites (Table S5), thus they were eligible for

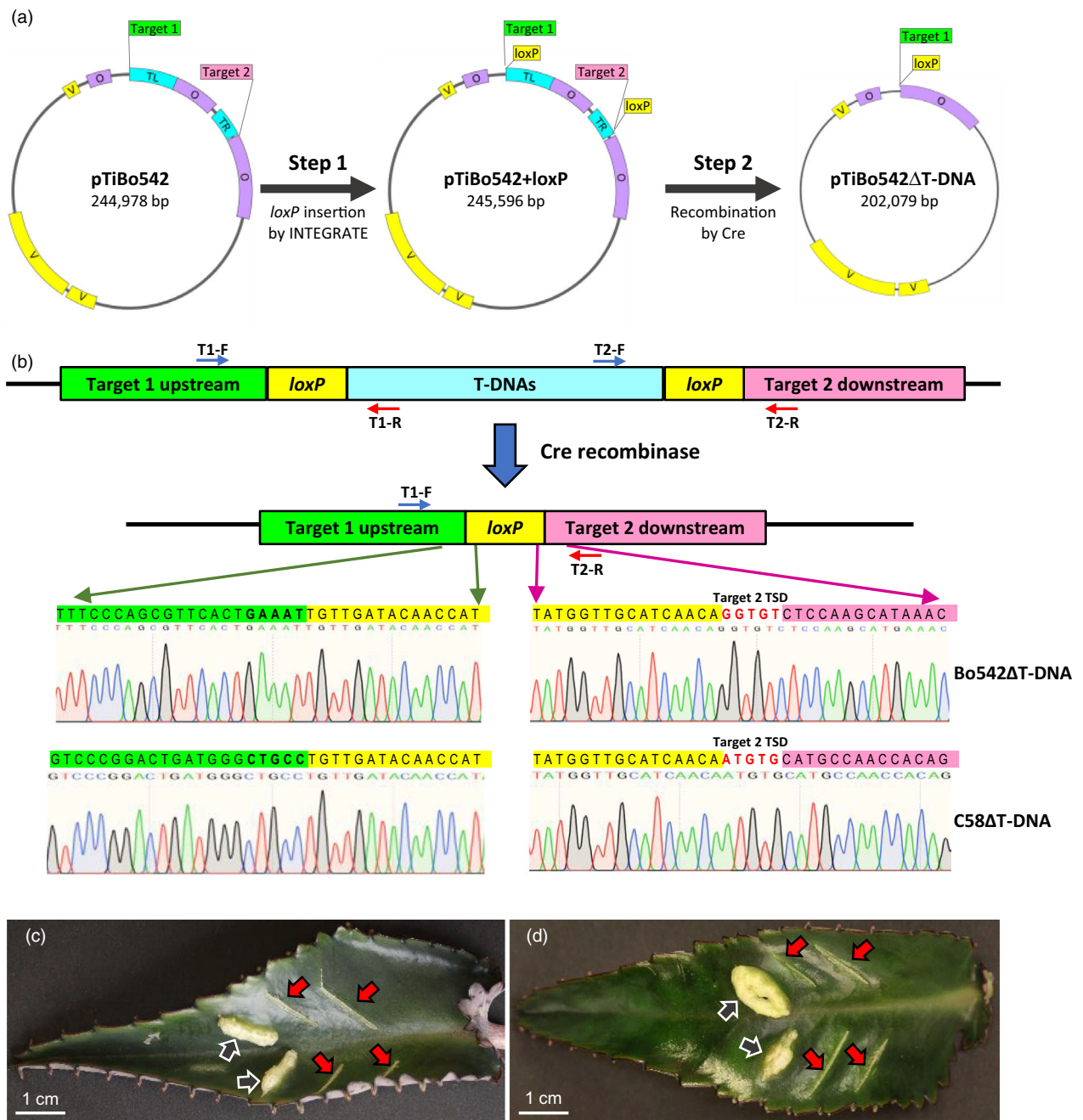


Figure 3 Large DNA fragment deletion in conjunction with Cre-loxP recombination system. (a) Schematic diagrams of two-step deletion of natural T-DNAs from pTiBo542. O, opine metabolism genes; TL, left T-DNA; TR, right T-DNA; V, virulence genes. (b) Illustration of loxP insertions into two target sites flanking the T-DNAs in pTiBo542 and pTiC58. The chromatograms show the Sanger sequencing results for the strains Bo542ΔT-DNA (top) and C58ΔT-DNA (bottom) strains after recombination. TSD, target site duplication. T1, target 1; T2, target 2; F, forward primer; R, reverse primer. (c and d) Tumorigenicity assay of Bo542ΔT-DNA (c) and C58ΔT-DNA (d) strains on Kalanchoe leaves. Grey arrows, wild-type strain; red arrows, disarmed strains. Scale bars in (c and d) are 1 cm.

Cre recombinase-mediated deletion. Likewise, 8 out of the 14 C58/2313 colonies with loxP insertions at both target sites were analysed by Sanger sequencing, and all of them had the same loxP orientation (T-RL) at both targets (Table S5).

The second step was to induce recombination between the two loxP sites using Cre recombinase. Competent cells were prepared from the Bo542/2308–3 and C58/2313–9 colonies. A plasmid expressing Cre recombinase (pKL2315, Table 1, Figure S4c) was introduced via electroporation. Recombination

between the two loxP sites was screened by PCR using the forward primer from target 1 (Table S2) and reverse primer from target 2 (Table S2). Recombination was detected from the first-generation colonies in both Bo542/2308–3 and C58/2313–9 strains but the efficiency was lower in Bo542, presumably due to the larger distance between the two loxP sites (42.6 kb) compared to C58 (25 kb) (Table S4). An extra replating step in Bo542/2308–3 ensured obtaining homogeneous colonies without remaining target 2 sites (Table S4), and Sanger sequencing

analysis confirmed targeted T-DNA deletion by recombination between the two *loxP* sites as shown in Figure 3b. The signature 5-bp TSD from the *loxP* cargo insertion at the target 2 remained after recombination. Because the Cre recombinase vector (pKL2315) carries a negative selection marker *sacB* gene, it was easily cured by spreading *Agrobacterium* cells on a solid YEP medium amended with 5% sucrose. The resulting colonies were susceptible to kanamycin, indicating that the *Agrobacterium* Δ T-DNA mutant strains were free of pKL2315.

To verify the deletion of the natural T-DNAs, the mutants (Bo542 Δ T-DNA and C58 Δ T-DNA) and their WT strains were inoculated onto Kalanchoe leaves for tumorigenicity assay. Three weeks after inoculation, only the WT strains incited tumours indicating that these disarmed strains indeed lost the oncogenes encoded on the T-DNAs (Figures 3c,d). These results demonstrate that INTEGRATE/Cre-*loxP* recombination system makes a powerful tool to generate precise large fragment deletions in *Agrobacterium* cells.

Discussion

We successfully implemented *V. cholerae* INTEGRATE system as an efficient genome engineering tool for *Agrobacterium*. A set of new vectors were developed with different origin of replication (pBBR1 and pVS1), transposon cargos (mCherry and *loxP*), and a negative selectable marker gene (*sacB*) to optimize the INTEGRATE system for targeted DNA integration and large fragment deletion in *Agrobacterium* (Table 1). These vectors can be easily modified to clone new DNA cargo or spacers using the existing restriction enzyme sites. For example, pEA186 and pKL2310 are empty (no guide RNA) pVS1-based INTEGRATE vectors with *mCherry* and *loxP* cargo, respectively. They can be used for targeted insertion in *Agrobacterium* strains. In addition, *Agrobacterium* strains generated from this study (Table 1) can be useful for plant genetic transformation to control post co-cultivation overgrowth or to enhance plasmid DNA stability.

Targeted insertional mutagenesis is a powerful tool to study gene functions. We showed the efficacy of this approach by targeting 5 different genes: *thyA*, *recA*, *virG*, *flaA*, and *agp1* located on the *A. tumefaciens* chromosome and Ti plasmid, respectively. In particular, the *thyA* and *recA* genes were selected to generate useful *Agrobacterium* strains for plant transformation. Auxotrophic strains are increasingly useful not only for more controlled growth during plant transformation but also for easing the biosafety issues associated with the genome engineering tools such as CRISPR/Cas systems. Because the auxotrophic strains cannot survive outside of the laboratory conditions, using these strains to deliver genome-editing reagents can decrease the risk of environmental release. In addition, the *recA* mutant strains can be useful when handling large plasmid DNAs with repetitive sequences or constructs with same promoters driving multiple genes, which often recombine within bacterial cells exhibiting plasmid instability problems. Prior to this study, three *thyA*-deficient mutant strains have been reported including EHA101Thy- (Aliu *et al.*, 2020), EHA105Thy- (Aliu *et al.*, 2020), and LBA4404Thy- (Ranch *et al.*, 2012). No successful attempt has been reported for generating *thyA* mutant using the homologous recombination approach from AGL1 strain, which has a deficiency in DNA repair and recombination due to the insertional mutation in the *recA* gene (Farrand *et al.*, 1989; Lazo *et al.*, 1991). Recently, Tamzil *et al.* reported the generation of threonine auxotrophic AGL1 strain using Tn5 transposon (Tamzil

et al., 2021). Unlike RNA-guided INTEGRATE system demonstrated here, the random insertional mutagenesis approach requires screening a large number of mutants as Tamzil *et al.* obtained 20 putative auxotrophs from 3315 mutants, demonstrating the cumbersome procedure of the traditional insertional mutagenesis approach (Tamzil *et al.*, 2021). The *recA* mutant strains such as AGL1 are desired for plant transformation due to the increased plasmid stability within *Agrobacterium* cells, but AGL1 has been the only public *recA* mutant strain for almost 30 years until Rodrigues *et al.* generated *recA*-deficient EHA105 strain using a cytosine base editor (Rodrigues *et al.*, 2021). Using the INTEGRATE system, we readily generated both single and double knockout mutants lacking *thyA* and *recA* genes for the popular *Agrobacterium* strains EHA101, EHA105, and AGL1 strains (Table 1).

With further harnessing, INTEGRATE system can be scaled-up for genome-wide gene functional analyses. As demonstrated in other CRISPR-Cas systems using large libraries of guide RNAs (Koike-Yusa *et al.*, 2014; Sanjana, 2017; Sharon *et al.*, 2018), INTEGRATE system could be utilized for high-throughput insertional mutagenesis assays to identify unknown *Agrobacterium* genes or genetic elements that play an important role for plant transformation or host range determination. Another useful aspect of the INTEGRATE system is the capability to precisely delete large DNA fragments in conjunction with the Cre-*loxP* recombination system. A precise large genomic DNA fragment deletion approach could be suitable for synthetic biology approaches to modify a specific pathway by deleting an entire operon or introducing a set of genes up to 10 kb (Vo *et al.*, 2021) to build new traits. Furthermore, as demonstrated here, novel Ti plasmids can be readily disarmed and utilized for plant transformation purposes. This can be especially useful for exploring novel *Agrobacterium* strains for wider host ranges, as diverse *Agrobacterium* strains have been sequenced. Comparative genomic and transcriptomic studies can provide new tools for understanding *Agrobacterium*-plant interactions (Otten, 2021; Weisberg *et al.*, 2020).

Like other bacterial genome editing systems, INTEGRATE also possesses certain drawbacks. First, due to end sequence-specific recognition by the INTEGRATE machinery, this system cannot be leveraged for applications involving point mutations and scar-free insertions (Klompe *et al.*, 2019; Vo *et al.*, 2021). Second, guide RNA design tool is not yet available for targeted insertions other than the 5'-CC-3' PAM preference (Klompe *et al.*, 2019), thus designing efficient guide RNAs would require testing multiple gRNAs for successful targeted DNA insertions. This study observed that certain guide RNAs have higher efficiency than others. For instance, the two guide RNAs used for pTiBo542 were more efficient than those used for the pTiC58 (Table S4). Guide RNA design tools can be developed in the future when more *in vitro* and *in vivo* data are accumulated, as were the case with previous CRISPR/Cas systems. Third, insertional mutagenesis may not be used for multiple rounds of transformation as there is a chance of remobilization of the previous cargo when the same INTEGRATE system is used in the subsequent round of engineering (Vo *et al.*, 2021). This can be avoided by deploying additional RNA-guided DNA insertion systems such as the ShCAST that utilizes a type V-K CRISPR effector (Strecker *et al.*, 2019). As an emerging tool for bacterial genome engineering, RNA-guided DNA insertion systems utilizing transposon-associated CRISPR systems such as INTEGRATE and ShCAST would be further tuned for robust and

precise DNA insertion-mediated bacterial genome engineering applications.

In summary, we successfully implemented and applied the RNA-guided DNA integration system, INTEGRATE, for insertional mutagenesis approaches as well as precise large DNA fragment deletion for *Agrobacterium* genome engineering. INTEGRATE systems with highly efficient insertional mutagenesis and multiplexing capability can accelerate gene functional analyses and the generation of useful bacterial strains for synthetic biology or plant genetic transformation.

Materials and methods

Bacterial strains and growth conditions

Agrobacterium tumefaciens strains, plasmids, and primer sequences used in this study are listed in Tables S1 and S2, respectively. Three *A. tumefaciens* strains EHA101 (Hood et al., 1986), EHA105 (Hood et al., 1993), and AGL1 (Lazo et al., 1991) were used to generate thymidine auxotrophs via insertional gene mutagenesis using the INTEGRATE system. EHA101Thy- and EHA105Thy- previously derived from EHA101 and EHA105, respectively, by homologous recombination-mediated *thyA* knockout (Aliu et al., 2020), were used to generate *recA* mutants, while EHA105 was used for INTEGRATE multiplexing studies. All strains have the *A. tumefaciens* C58 chromosomal background.

Vector construction

All *V. cholerae* INTEGRATE plasmid constructs listed in Table S1 were generated from the pSPIN vectors pSL1142 and pSL1765, which were gifts from Dr. Samuel H. Sternberg (Addgene plasmid # 160730 and # 160734), using a combination of polymerase chain reaction (PCR), restriction enzyme digestion, ligation, and Gibson assembly (Gibson et al., 2009). All PCR fragments for cloning were generated using the Phusion high-fidelity DNA polymerase (ThermoFisher Scientific, MA, USA) or Q5 DNA Polymerase (New England Biolabs, MA, USA). crRNAs (for targeting genomic sites were designed with 32-nt spacers) and all spacer sequences used for this study are available in Table S2.

Construction of the single and multiplex targeting constructs

We first made pBBR1-based INTEGRATE vector pEA179 (Figure S1a) using pSL1142 and pSL1765 (Table S1). pVS1-based INTEGRATE vector pEA191 (Figure S1b) was made by transferring the INTEGRATE components from pSL1765 to the vector pTF505 (Lee et al., 2013). To utilize the *BsaI*-mediated spacer cloning in the INTEGRATE vectors, an endogenous *BsaI* restriction site within the pTF505 backbone (Figure S2a) was cured by site-directed mutagenesis using the primers (primer IDs 29 and 30) listed in Table S2. As a cargo, the mCherry cassette containing J23107 promoter and *rrnB* gene terminator (J23107-mCherry-TrnB) was PCR amplified with overlapping sequences for Gibson assembly with pSL1142 (or pSL1765) digested with *XhoI* and *PstI*. For single targeted insertion, each oligonucleotide pair for a new spacer were annealed and ligated with the *BsaI*-digested vector backbone. Triple-spacer arrays for multiplexing were cloned by combining two or three oligonucleotide duplexes with compatible sticky ends in a single ligation reaction. crRNAs were designed with 32-nt spacers targeting genomic sites with a 5'-CC-3' PAM. For easy curing of the INTEGRATE plasmids after targeted DNA insertion, a negative selection marker gene *sacB* was inserted via

Gibson assembly by utilizing a unique *KpnI* site (Figures S1a,b). Cloning reactions were used to transform DH5 α *E. coli* competent cells, and plasmids were extracted using QIAprep Spin Miniprep kit (Qiagen, Hilden, Germany) and confirmed by Sanger sequencing. Transformed cells were cultured in liquid Terrific Broth (TB) (Tartof and Hobbs, 1987) media amended with 50 mg L⁻¹ of kanamycin (pBBR1 backbone) or 100 mg L⁻¹ spectinomycin (pVS1 backbone).

Transposition assays in *Agrobacterium*

All transposition assays for *A. tumefaciens* were performed by introducing the INTEGRATE vectors via electroporation. Forty μ L of electrocompetent cells were electroporated with 50–100 ng of plasmid DNA, in 2 mm electroporation cuvettes using a Bio-Rad Gene Pulser (Bio-Rad, CA, USA). Cells were recovered in 0.5 mL of SOC (or LB) media for 2 h in a shaking incubator at 28 °C and plated on yeast extract peptone (YEP) agar media with appropriate antibiotics and incubated at 28 °C for 48 h to yield the parent colonies.

Isolation of clonally integrated *A. tumefaciens* colonies

Based on previous report that single colonies might be genetically heterogeneous when integration occurs simultaneously with colony expansion (Vo et al., 2021), isolation of colonies was preceded by a 're-plating' step where the parent colonies were scraped and pooled together, then resuspended in YEP broth, and plated at an appropriate dilution to obtain a fresh set of colonies. *A. tumefaciens* cells transformed with INTEGRATE vectors were scraped from YEP agar plates and resuspended in liquid YEP. After 24 h growth in a shaking incubator at 28 °C with 200 rpm, cell density was measured using a spectrophotometer at 600 nm. About 50 μ L of cell suspension after 10⁶-fold dilution prepared from the overnight culture was plated on YEP agar with appropriate antibiotics and incubated at 28 °C for 48 h to yield the first-generation colonies that were screened by PCR using suitable primers (Table S2) to detect the integration of the transposon into the genome of *A. tumefaciens*.

PCR analysis of transposition

For detection of genomic integration, two genome-specific primers flanking the anticipated integration region were used, and to determine the transposon orientation (T-RL or T-LR), one genome-specific and one cargo-specific primers were used. Thermocycling conditions were as follows: initial denaturation at 98 °C for 30 s, followed by 30 cycles of 10 s at 98 °C, 30 s at 63 °C, 30 s at 72 °C, with a final extension of at 72 °C for 5 min. PCR amplicons were resolved by electrophoresis on 1–1.3% agarose gels stained with Red Safe (Bulldog Bio, NH, USA). Colonies yielding the expected PCR products were cleaned up using the QIAquick PCR purification kit (Qiagen) and further subjected to Sanger sequencing to verify the targeted insertion site.

Thymidine-dependent growth of auxotrophs

Seed cultures of thymidine auxotrophs generated by homologous recombination (HR) (Aliu et al., 2020) and INTEGRATE (EHA101T, EHA101TR, EHA105T, EHA105TR, and AGL1T) and their prototrophs were inoculated in 5 mL of YEP medium (supplemented with 50 mg L⁻¹ of thymidine for the auxotrophs) in 50 mL falcon tubes and grown in a shaking incubator for 15 h at 28 °C with 200 rpm. A batch culture was prepared by transferring a calculated amount of overnight culture to a 50 mL of YEP

medium in a 250 mL flask to a cell density of OD₆₀₀ of 0.02. Batch cultures were grown in a shaking incubator (28 °C, 200 rpm) and 0.5 mL of culture was sampled every 4 h for 24 h to measure cell density using a spectrophotometer at 600 nm.

Evaluation of T-DNA delivery capability

Evaluation of T-DNA delivery capability of thymidine auxotrophic strains was done by AGROBEST assay (Wu *et al.*, 2014). The *efr-1* (SALK 044334) T-DNA insertional mutant *Arabidopsis thaliana* line was obtained from the Arabidopsis Biological Resource Center (Columbus, Ohio). For surface sterilization of seeds, 50% bleach (3% sodium hypochlorite) and 0.1% SDS solution was used for about 200–400 *efr-1* seeds in a 2 mL tube. After incubating for 15 min on a rotary mixer, seeds were rinsed four times with sterile water and 1 mL of half strength Murashige and Skoog medium (½ MS) containing 0.5% sucrose was added to the tube before being transferred to a 60 mm petri dish using a pipette and a wide-bore tip. The *efr-1* seeds were grown for 7 days in a growth chamber at 22 °C under a 16 h/8 h light/dark cycle after 48 h cold treatment at 4 °C for uniform seed germination. The binary vector pTF102 (Frame *et al.*, 2002) harbouring a GUS reporter gene (β-glucuronidase) driven by a cauliflower mosaic virus 35S promoter was transformed by electroporation into the thymidine auxotrophs and their corresponding WT strains serving as positive controls. For negative control, EHA105 strain without pTF102 was used. The strains were grown for 16 to 20 h in 5 mL of YEP medium supplemented with 50 mg L⁻¹ thymidine (for thymidine auxotrophs) and appropriate antibiotics (50 mg L⁻¹ kanamycin and 100 mg L⁻¹ spectinomycin for EHA101T; 100 mg L⁻¹ spectinomycin for EHA105T and AGL1T) in 50 mL tubes at 28 °C with 200 rpm. *Agrobacterium* cells were pelleted by centrifugation prior to infection and re-suspended in AB induction medium (Gelvin, 2006) to a density of OD₆₀₀ of 0.04. Ten to fifteen *efr-1* seedlings were transferred to each well of a 12-well plate using sterile inoculation loops and 0.5 mL of ½ MS medium supplemented with 50 mg L⁻¹ thymidine was pipetted into each well followed by the addition of 0.5 mL of *Agrobacterium* cell suspensions. The 12-well plates were sealed with 3 M micropore tape and incubated in a growth chamber for two days at 22 °C under a 16 h/8 h light/dark cycle. *Agrobacterium* cells were removed by pipetting after a two-day co-cultivation, and 1 mL of fresh ½ MS medium amended with 100 mg L⁻¹ cefotaxime and 100 mg L⁻¹ timentin were added. Arabidopsis seedlings were grown in the growth chamber for two more days before GUS staining. With slight modifications, GUS staining was conducted as previously described (Cervera, 2005) to visualize transient transgene expression. For GUS staining, the remaining MS medium was removed before adding 1 mL of GUS staining solution to the *efr-1* seedlings. After overnight incubation at 37 °C, GUS staining solution was removed by pipetting and 1 mL of 75% ethanol was added to remove chlorophyll from the *efr-1* seedlings and incubated at room temperature for 12 to 16 h. Images were taken on a white background to aid T-DNA delivery comparison between the auxotrophs and their corresponding prototrophs.

Virulence gene induction assay of *virG* mutant strain

virG mutant strain and its corresponding WT were transformed with a reporter plasmid with VirG-inducible PvirB promoter driving a green fluorescent protein gene (EGFP). The

Agrobacterium strains were grown for 20 h in 5 mL of YEP medium supplemented with 100 mg L⁻¹ spectinomycin in 50 mL tubes at 28 °C with 200 rpm. *Agrobacterium* cells were then pelleted by centrifugation and re-suspended in AB induction medium (Gelvin, 2006) to a density of OD₆₀₀ of ~0.2 with or without 200 μM AS, and grown at 25 °C with 150 rpm for 21 h. EGFP induction was quantified by a fluorometer using a SYBR filter (492 nm for excitation, 516 nm for emission).

Ultraviolet (UV) light sensitivity assay of *recA* mutants

UV-B light was used to test the EHA101 and EHA105 *recA* mutant strains. An overnight culture of each strain was prepared by inoculating a single colony into 50 mL conical tube containing 5 mL YEP medium and incubated at 28 °C with 200 rpm for 20 h. Cells were then diluted at a 1:100 ratio in a fresh YEP medium and grown to a cell density of OD₆₀₀ of 0.7–0.9. A serially diluted bacterial suspension (10–10⁶-fold) was prepared in a six-well plate and exposed to UV-B irradiation for 2 h in the dark using a 6 W handheld UV lamp (Model UVM; Analytik Jena US LLC, Upland, CA, USA) placed at 20 cm from the top of the six-well plate. After 2 h, 4 μL of bacterial suspensions were pipetted on YEP agar plates, which were covered with aluminium foil and incubated at 28 °C for 48 h. As a control, *Agrobacterium* cell suspensions from a duplicate six-well plate without UV irradiation were also pipetted on YEP agar plates and incubated for 48 h. The *recA*-deficient *Agrobacterium* AGL1 strain was used as a positive control, while EHA105 WT strain was used as a negative control.

Temperature sensitivity assay of *agp1* mutant

Seed cultures of EHA105 WT and the triple mutant EHA105VAF were inoculated in 5 mL of LB medium in a 50 mL falcon tube and grown in a shaking incubator for 15 h at 28 °C with 200 rpm. A batch culture was prepared by transferring a calculated volume of overnight culture to a 50 mL of LB medium in a 250 mL flask to a cell density of 0.02 at 600 nm. Batch cultures were grown in a shaking incubator (25 °C and 110 rpm) and 0.5 mL of culture was sampled every 7 h for 49 h to measure cell density using a spectrophotometer.

Flagellar motility assay

Swimming phenotypes of the EHA105 WT and the triple mutant EHA105VAF strains were assessed on YEP swim plates containing 0.25% Bacto agar (Xue *et al.*, 2021). Single colonies were inoculated in 5 mL of liquid YEP medium in 50 mL conical tubes and the overnight cultures were then diluted to a density of OD₆₀₀ of 0.4. Inoculation on swim plates was done by pipetting 1 μL of diluted culture onto the agar at the centre of the plate. Plates were incubated at 28 °C and images were captured at 24 h and 48 h to observe the phenotypic difference.

Cre-*loxP*-mediated large DNA fragment deletion experiments

Cre-*loxP*-mediated targeted DNA deletion was performed with two-step procedures. The first step was to insert the *loxP* sites into the two target sites using the INTEGRATE system. The *loxP* site was inserted into the cargo and two crRNAs targeting the flanking sequences of the T-DNAs of pTiBo542 and pTiC58 were cloned into the CRISPR array of the INTEGRATE vector with pVS1 backbone (pKL2310) using the oligonucleotides listed in the Table S2. *A. tumefaciens* strain Bo542 and C58 were transformed with the pKL2308 and pKL2313, respectively, by

electroporation as described above and the resulting parental colonies were pooled, diluted, and spread on fresh solid YEP medium (re-plated). After two days, resulting colonies were screened by PCR using the target-specific primer pairs listed in Table S2. PCR was performed using the Q5 2X master mix (New England Biolabs, MA, USA) and the thermocycling conditions defined above. For samples with only targeted insertion event (i.e., larger bands about 800 bp), Sanger sequencing analysis was followed to confirm the insertion of *loxP* sites and their orientation. The second step was to introduce another pVS1-based vector expressing Cre recombinase (pKL2315; Figure S4c) into the Bo542 or C58 cells that have two *loxP* site insertions with same orientation (T-RL). Because the Cre recombinase vector carries a different antibiotic resistance gene (kanamycin), the second step eliminates the INTEGRATE vector (spectinomycin) from *Agrobacterium* cells. PCR screening using the forward primer of target 1 and reverse primer of target 2 was carried out to identify *Agrobacterium* colonies with 42.6 kb or 25 kb deletions in Bo542 or C58 strains, respectively. Control PCRs using the target-specific primers were also performed to detect the presence of undeleted T-DNAs. The re-plating step was implemented if the control PCRs were positive, indicating that the colonies were heterogeneous. Sanger sequencing analysis was used to validate the precise recombination and deletion of the T-DNAs.

Tumorigenesis assay on Kalanchoe leaves

Tumorigenesis assay on Kalanchoe leaves was used to confirm the deletion of oncogenes encoded on the T-DNAs. *Agrobacterium* strains were grown for 20 h in 5 mL liquid YEP medium in 50 mL conical tubes in a shaking incubator at 28 °C with 200 rpm. Cells were collected by centrifugation and resuspended in AB induction medium (Gelvin, 2006). After adjusting the cell density to OD₆₀₀ of 0.2, *Agrobacterium* cells were grown under *vir* gene induction condition with 200 µM AS for 20 h at 25 °C with 200 rpm. Six to eight weeks old Kalanchoe leaves were inoculated as described previously (Guo et al., 2007). Hypodermic needle was first used to wound the leaves, then T-DNA deletion mutants and their WT strain were inoculated onto the same leaves. Three microliters of bacterial inoculum were pipetted onto each wound and tumour formation was monitored after 2–3 weeks.

Statistical analysis

Differences in targeted DNA insertion frequencies were compared using two-tailed paired sample *t*-test (Hsu and Lachenbruch, 2014).

Acknowledgements

We thank Samuel H. Sternberg from Columbia University for the INTEGRATE vector and for discussion at the initial stage of this work. This work was supported partially by National Science Foundation Plant Genome Research Program Grant IOS-1725122 and IOS-1917138 to K.W. and K.L., by the Iowa State University Interdepartmental Plant Biology Major fellowship to E.A., by the seed grant fund from Crop Bioengineering Center of Iowa State University to K.L., and by the USDA NIFA Hatch project #IOW04714, by State of Iowa funds. K.W.'s contribution to this work is partially supported by (while serving at) the National Science Foundation. Open access funding provided by the Iowa State University Library.

Author contributions

K.W. conceived the research idea, K.L. and E.A. designed and performed experiments. E.A. performed single and multiplex insertion assays and analysed loss of function mutants. K.L. performed programmed genomic deletion assays and analysed deletion mutants. K.L. and E.A. analysed the data. K.L., E.A. and K.W. wrote the paper.

Competing interests

The authors declare no competing interest.

Data availability

All study data are included in the article and/or SI Appendix.

References

- Aliu, E., Azanu, M.K., Wang, K. and Lee, K. (2020). *Generation of thymidine auxotrophic agrobacterium tumefaciens* strains for plant transformation. *bioRxiv* preprint doi: <https://doi.org/10.1101/2020.08.21.261941>.
- Alteter, F., Springer, N.M., Bartley, L.E., Blechl, A.E., Brutnell, T.P., Citovsky, V., Conrad, L.J. et al. (2016) Advancing crop transformation in the era of genome editing. *Plant Cell*, **28**, 1510–1520.
- Anand, A., Bass, S.H., Wu, E., Wang, N., McBride, K.E., Annaluru, N., Miller, M. et al. (2018) An improved ternary vector system for *agrobacterium*-mediated rapid maize transformation. *Plant Mol. Biol.* **97**, 187–200.
- Ashby, A.M., Watson, M.D. and Shaw, C.H. (1987) A Ti-plasmid determined function is responsible for chemotaxis of *agrobacterium tumefaciens* towards the plant wound product acteosyringone. *FEMS Microbiol. Lett.* **41**, 189–192.
- Banta, L.M. and Montenegro, M. (2008) *Agrobacterium* and plant biotechnology. In *Agrobacterium: From Biology to Biotechnology* (Tzfira, T. and Citovsky, V., eds), pp. 73–147. New York, NY: Springer Science+Business Media.
- Cervera, M. (2005) Histochemical and fluorometric assays for uidA (GUS) gene detection. In *Transgenic Plants: Methods and Protocols* (Peña, L., ed), pp. 203–213. Totowa, NJ: Humana Press Inc.
- Chau, A., Giang, K., Leung, M. and Tam, N. (2008) Role of RecA in the protection of DNA damage by UV-A in *Escherichia coli*. *J. Exp. Microbiol. Immuno.* **12**, 39–44.
- Chen, Z., Yang, H. and Pavletich, N. (2008) Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature*, **453**, 489–494.
- Chesnokova, O., Coutinho, J.B., Khan, I.H., Mikhail, M.S. and Kado, C.I. (1997) Characterization of flagella genes of *agrobacterium tumefaciens*, and the effect of a bald strain on virulence. *Mol. Microbiol.* **23**, 579–590.
- Deakin, W.J., Parker, V.E., Wright, E.L., Ashcroft, K.J., Loake, G.J. and Shaw, C.H. (1999) *Agrobacterium tumefaciens* possesses a fourth flagellin gene located in a large gene cluster concerned with flagellar structure, assembly and motility. *Microbiology*, **145**, 1397–1407.
- Farrand, S.K., Omorchoe, S.P. and McCutchan, J. (1989) Construction of an *agrobacterium tumefaciens* C58 recA mutant. *J. Bacteriol.* **171**, 5314–5321.
- Frame, B.R., Shou, H., Chikwamba, R.K., Zhang, Z., Xiang, C., Fonger, T.M., Pegg, S.E.K. et al. (2002) *Agrobacterium tumefaciens*-mediated transformation of maize embryos using a standard binary vector system. *Plant Physiol.* **129**, 13–22.
- Gay, P., Le Coq, D., Steinmetz, M., Ferrari, E. and Hoch, J.A. (1983) Cloning structural gene *sacB*, which codes for exoenzyme levansucrase of *Bacillus subtilis*: expression of the gene in *Escherichia coli*. *J. Bacteriol.* **153**, 1424–1431.
- Gelvin, S.B. (2003) *Agrobacterium*-mediated plant transformation: the biology behind the “gene-jockeying” tool. *Microbiol. Mol. Biol. Rev.* **67**, 16–37.
- Gelvin, S.B. (2006) *Agrobacterium* virulence gene induction. In *Agrobacterium Protocols*, 2nd edn (Wang, K., ed), pp. 77–84. Totowa, NJ: Humana Press.

- Gelvin, S.B. (2017) Integration of *agrobacterium* T-DNA into the plant genome. *Annu. Rev. Genet.* **51**, 195–217.
- Gelvin, S.B. (2021) Plant DNA repair and *agrobacterium* T-DNA integration. *Int. J. Mol. Sci.* **22**, 8458.
- Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., III and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343–345.
- Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Quorllo, B., Goldman, B.S. *et al.* (2001) Genome sequence of the plant pathogen and biotechnology agent *agrobacterium tumefaciens* C58. *Science*, **294**, 2323–2328.
- Guo, M., Hou, Q.M., Hew, C.L. and Pan, S.Q. (2007) *Agrobacterium* VirD2-binding protein is involved in tumorigenesis and redundantly encoded in conjugative transfer gene clusters. *Mol. Plant-Microbe Interact.* **20**, 1201–1212.
- Hood, E.E., Gelvin, S.B., Melchers, L.S. and Hoekema, A. (1993) New *agrobacterium* helper plasmids for gene transfer to plants. *Transgenic Res.* **2**, 208–218.
- Hood, E.E., Helmer, G.L., Fraley, R.T. and Chilton, M.D. (1986) The hypervirulence of *agrobacterium tumefaciens* A281 is encoded in a region of pTiBo542 outside of T-DNA. *J. Bacteriol.* **168**, 1291–1301.
- Hsu, H. and Lachenbruch, P.A. (2014). *Paired t test. Wiley stats ref: statistics reference online.*
- Iverson, S.V., Haddock, T.L., Beal, J. and Densmore, D.M. (2016) CIDAR MoClo: improved MoClo assembly standard and new *E. coli* part library enable rapid combinatorial Design for Synthetic and Traditional Biology. *ACS Synth. Biol.* **5**, 99–103.
- Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S. and Sternberg, S.H. (2019) Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*, **571**, 219–225.
- Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, M.D.C. and Yusa, K. (2014) Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273.
- Komari, T., Takakura, Y., Ueki, J., Kato, N., Ishida, Y. and Hiei, Y. (2006) Binary vectors and super-binary vectors. In *Agrobacterium Protocols*, 2nd edn (Wang, K., ed), pp. 15–41. Totowa, NJ: Humana Press.
- Koncz, C. and Schell, J. (1986) The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimeric genes carried by a novel type of *agrobacterium* binary vector. *Mol. Gen. Genet.* **204**, 383–396.
- Lamparter, T., Xue, P., Elkurdi, A., Kaeser, G., Sauthof, L., Scheerer, P. and Krauß, N. (2021) Phytochromes in *agrobacterium fabrum*. *Front. Plant Sci.* **12**, 642801.
- Lazo, G.R., Stein, P.A. and Ludwig, R.A.A. (1991) DNA transformation-competent Arabidopsis genomic library in *agrobacterium*. *Biotechnology*, **9**, 963–967.
- Lee, K., Huang, X., Yang, C., Lee, D., Ho, V., Nobuta, K., Fan, J.B. *et al.* (2013) A genome-wide survey of highly expressed non-coding RNAs and biological validation of selected candidates in *agrobacterium tumefaciens*. *PLoS One*, **8**, e70720.
- Makarova, K.S., Wolf, Y.I. and Koonin, E.V. (2018) Classification and nomenclature of CRISPR-Cas systems. *The CRISPR Journal*, **1**, 5–336.
- Merritt, P.M., Danhorn, T. and Fuqua, C. (2007) Motility and chemotaxis in *agrobacterium tumefaciens* surface attachment and biofilm formation. *J. Bacteriol.*, **189**, 8005–8014.
- Morton, E.R. and Fuqua, C. (2012) Genetic manipulation of *agrobacterium*. *Curr. Protoc. Microbiol.* **25**, 3D.2.1–3D.2.15.
- Njimona, I. and Lamparter, T. (2011) Temperature effects on *agrobacterium* phytochrome Agp1. *PLoS One*, **6**, e25977.
- Ooms, G., Hooykaas, P.J.J., Moolenaar, G. and Schilperoort, R.A. (1981) Crown gall plant tumors of abnormal morphology, induced by *agrobacterium tumefaciens* carrying mutated octopine Ti plasmids; analysis of T-DNA functions. *Gene*, **14**, 33–50.
- Otten, L. (2021) T-DNA regions from 350 *agrobacterium* genomes: maps and phylogeny. *Plant Mol. Biol.* **106**, 239–258.
- Păcurar, D.I., Thordal-Christensen, H., Păcurar, M.L., Pamfil, D., Botez, C. and Bellini, C. (2011) *Agrobacterium tumefaciens*: from crown gall tumors to genetic transformation. *Physiol. Mol. Plant Pathol.* **76**, 76–81.
- Peters, J.E., Makarova, K.S., Shmakov, S. and Koonin, E.V. (2017) Recruitment of CRISPR-Cas systems by Tn7-like transposons. *PNAS*, **114**, E7358–E7366.
- Ranch, J.P., Liebergesell, M., Garnaat, C.W. and Huffman, G.A. (2012). *Auxotrophic Agrobacterium* for plant transformation and methods thereof. U.S. Patent No. 8,334,429. Washington, DC: U.S. Patent and Trademark Office.
- Rodrigues, S.D., Karimi, M., Impens, L., van Lerberge, E., Coussens, G., Aesaert, S., Rombaut, D. *et al.* (2021) Efficient CRISPR-mediated base editing in *agrobacterium* spp. *PNAS*, **118**, 1–8.
- Sanjana, N.E. (2017) Genome-scale CRISPR pooled screens. *Anal. Biochem.* **532**, 95–99.
- Sciaky, D., Montoya, A.L. and Chilton, M.D. (1978) Fingerprints of *agrobacterium* Ti plasmids. *Plasmid*, **1**, 238–253.
- Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E. and Tsien, R.Y. (2004) Improved monomeric red, orange, and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572.
- Sharon, E., Chen, S.A.A., Khosla, N.M., Smith, J.D., Pritchard, J.K. and Fraser, H.B. (2018) Functional genetic variants revealed by massively parallel precise genome editing. *Cell*, **175**, 544–557.
- Steinmetz, M., Le Coq, D., Djemia, H.B. and Gay, P. (1983) Genetic analysis of *sacB*, the structural gene of a secreted enzyme, levansucrase of *Bacillus subtilis* Marburg. *Mol. Gen. Genet.* **191**, 138–144.
- Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J.L., Makarova, K.S., Koonin, E.V. and Zhang, F. (2019) RNA-guided DNA insertion with CRISPR-associated transposases. *Science*, **365**, 48–53.
- Tamzil, M.S., Alfiko, Y., Mubarak, A.F., Purwantomo, S., Suwanto, A. and Budiarti, S. (2021) Development of auxotrophic *agrobacterium tumefaciens* AGL1 by Tn5 transposon for Rice (*Oryza sativa* L.) transformation. *Biotechnol. Bioproc. E.* **26**, 641–649.
- Tartof, K.D. and Hobbs, C.A. (1987) Improved media for growing plasmid and cosmid clones. *Bethesda Res. Lab. Focus*, **9**, 12–16.
- Vo, P.L.H., Ronda, C., Klompe, S.E., Chen, E.E., Acree, C., Wang, H.H. and Sternberg, S.H. (2021) CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nature Biotech.* **39**, 480–489.
- Wannier, T.M., Nyerges, A., Kuchwara, H.M., Czikkely, M., Balogh, D., Filsinger, G.T., Borders, N.C. *et al.* (2020) Improved bacterial recombineering by parallelized protein discovery. *PNAS*, **117**, 13689–13698.
- Weisberg, A.J., Davis, E.W., II, Tabima, J., Belcher, M.S., Miller, M., Kuo, C.H., Loper, J.E. *et al.* (2020) Unexpected conservation and global transmission of *agrobacterium* virulence plasmids. *Science*, **368**, eaba5256.
- Winans, S.C., Kerstetter, R.A. and Nester, E.W. (1988) Transcriptional regulation of the *virA* and *virG* genes of *agrobacterium tumefaciens*. *J. Bacteriol.* **170**, 4047–4054.
- Wood, D.W., Setubal, J.C., Kaul, R., Monks, D.E., Kitajima, J.P., Okura, V.K., Zhou, Y. *et al.* (2001) The genome of the natural genetic engineer *agrobacterium tumefaciens* C58. *Science*, **294**, 2317–2323.
- Wu, H., Liu, K.H., Wang, Y.C., Wu, J.F., Chiu, W.L., Chen, C.Y., Wu, S.H. *et al.* (2014) AGROBEST: an efficient *agrobacterium*-mediated transient expression method for versatile gene function analyses in Arabidopsis seedlings. *Plant Methods*, **10**, 1–16.
- Xue, P., Bai, Y., Rottwinkel, G., Averbukh, E., Ma, Y., Roeder, T., Scheerer, P. *et al.* (2021) Phytochrome mediated responses in *agrobacterium fabrum*: growth, motility and plant infection. *Curr. Microbiol.* **78**, 2708–2719.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1–S5 Supplementary Tables.
Figure S1–S4 Supplementary Figures.