

Inverted repeat sequences involved in high-frequency T-DNA truncation in “Princettia” (*Euphorbia pulcherrima* × *Euphorbia cornastra*)

Koya Ito

Kyoto Prefectural University <https://orcid.org/0009-0001-9239-6692>

Reiko Kogishi

Kyoto Prefectural University

Sayaka Shindo

Kyoto Prefectural University

Rina Shimo

Kyoto Prefectural University

Yukiko Shimbo

Kyoto Prefectural University

Maki Ohtsubo

Kyoto Prefectural University

Keisuke Matsui

Suntory Flowers Ltd.

Kenichi Suzuki

Suntory Flowers Ltd.

Koichi Tomomatsu

Suntory Flowers Ltd.

Norihiro Ohtsubo

nohtsubo@kpu.ac.jp

Kyoto Prefectural University <https://orcid.org/0000-0002-2487-025X>

Research Article

Keywords: Agrobacterium-mediated transformation, inverted repeat, plant biotechnology, poinsettia, T-DNA deletion, T-DNA integration

Posted Date: July 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10314391/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Agrobacterium-mediated transformation is a primary method for gene transfer in plants and a crucial tool in plant biotechnology, generally leading to the integration of the entire T-DNA region into the plant genome. However, numerous plant species have documented undesirable T-DNA truncation during this integration. To address this, we investigated the sequence specificity of the truncation region using “Princettia,” an interspecific hybrid of poinsettia that exhibits this truncation at a high rate. Examination of 114 “Princettia” transformants revealed T-DNA truncation in 65 cases. These truncations occurred throughout the T-DNA, with certain regions showing higher frequencies. In a few lines, internal T-DNA sequences were missing, indicating multiple truncation events within a single T-DNA. Sequence analysis frequently identified inverted repeat (IR) sequences immediately flanking the T-DNA breakpoints. The highest truncation frequency was observed at a 113-bp IR sequence, suggesting that long IR sequences facilitate truncation, particularly those exceeding 100 bp. Additionally, partial insertions of inverted T-DNA regions were often found directly after the truncation sites, implying DNA polymerase involvement at these locations. Based on these observations, we propose a novel model where the binding of these IR sequences triggers truncation and the addition of extra sequences in their vicinity.

Introduction

Agrobacterium-mediated transformation is a major method of gene transfer in plants and an essential tool in plant biotechnology. While T-DNA is typically integrated completely from the right border (RB) to the left border (LB), truncation, leading to the deletion of portions of the T-DNA, has been documented during the steps from infection to integration. This phenomenon has been observed in various plant species, including tobacco, barley, and aspen (De Buck et al. 2000; Stahl et al. 2002; Kumar and Fladung 2002). Furthermore, large-scale truncations of several hundred or thousand base pairs have been reported in rice, grapes, and potatoes (Kim et al. 2003; Gambino et al. 2009; Yasumoto et al. 2020). Remarkably, truncations have also been observed during *Arabidopsis* transformation using the floral dip method, which bypasses tissue culture procedures like callus formation (Wei et al. 2015). These undesirable truncations reduce transformation efficiency, yet research on this issue is still limited, and the underlying causes and mechanisms are not well understood.

Agrobacterium-mediated gene transfer involves the excision of the T-DNA region, flanked by the RB and LB sequences and carrying the target gene, by VirD1/D2 proteins, and its subsequent introduction into the plant cell (De Vos and Zambryski 1989; Filichkin and Gelvin 1993; Scheiffele et al. 1995; Tzfira et al. 2004a). VirD2 protein is covalently linked to the RB of T-DNA, while VirE2 binds elsewhere on the T-DNA. Both play roles in directing the T-DNA to the nucleus and protecting it from nuclease degradation (Ward and Barnse 1988; Christie et al. 1988). Upon entering the nucleus, VirE2 dissociates from the T-DNA, whereas VirD2 remains attached to the RB until shortly before T-DNA integration into the plant genome (Dürrenberger et al. 1989; Tinland et al. 1995; Tzfira et al. 2004b). The widely accepted mechanism posits that T-DNA preferentially migrates to double-strand breaks (DSBs) in the plant genome, where it is integrated via the host's DSB repair pathways (Citovsky et al. 2007). Two integration pathways have been

identified: direct integration of double-stranded T-DNA and integration of single-stranded T-DNA followed by its conversion to a double-stranded form (Lacroix and Citovsky 2019; Thomson et al. 2023).

It has been interestingly observed that the frequency of truncation significantly rises in tobacco transformed with T-DNA containing a 200–450 bp inverted repeat (IR) sequence (Sunitha et al. 2012). While this suggested a potential IR-specific truncation mechanism, detailed sequence analysis was not conducted. We proposed using “Princettia” (*Euphorbia pulcherrima* × *Euphorbia cornea*) to investigate this further. This interspecific hybrid, in which we recently established a transformation system, shows exceptionally high T-DNA truncation rates. Therefore, we examined the sequence specificity of the truncation region in 114 “Princettia” transformants. Inverse PCR-based sequence analysis was employed to precisely determine truncation sites and analyze surrounding sequences. Based on our findings, we propose a novel model explaining the truncation mechanism. Additionally, this study reports advancements in the efficiency of poinsettia tissue culture and transformation systems.

Materials and methods

Plant materials

“Princettia” (*Euphorbia pulcherrima* × *Euphorbia cornea*) experimental lines #7801 and #7155 were provided by Suntory Global Innovation Center Limited (Kyoto, Japan) (Supplementary Fig. S1a, b, c). Aseptic plant cultures were maintained at 25°C under a 16 h light/8 h dark cycle of LED fluorescent light (40 $\mu\text{mol m}^{-2} \text{s}^{-1}$) on a modified Murashige and Skoog (MS) medium containing MS salts, 3% sucrose, 0.4% agar, 0.2% gellan gum, 0.1 mg l^{-1} BA, 1 g l^{-1} casein acid hydrolysate, 4.9 μM CuSO_4 , and 0.2% PPM™ (Plant Cell Technology, Washington D.C., USA), which was adjusted to pH 5.8. Plants were subcultured and maintained by cuttings every 60 days.

Plasmid vectors

Dr. Nobutaka Mitsuda from the Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Japan, generously provided the pBCKH-35S::*AtEIN3-SRDX* and pBCKH-35S::*AtPAP1-SRDX* constructs illustrated in Fig. 1a, b. These vectors were constructed using the binary vector pBCKH (Oshima et al. 2011).

“Princettia” transformation

Electroporation was used to introduce the vectors into *Agrobacterium tumefaciens* strains EHA105 and/or AGL0. The transformed *Agrobacterium* suspension was adjusted to an OD_{600} of 0.2 in MS liquid medium (MS with 2% sucrose) supplemented with 200 μM acetosyringone and 5 μM CuSO_4 , and subsequently used for plant infection. “Princettia” #7801 was infected using the EHA105 strain harboring pBCKH-35S::*AtEIN3-SRDX*, while both EHA105 and AGL0 strains carrying pBCKH-35S::*AtPAP1-SRDX* were used to infect lines #7801 and #7155. Stems from plants subcultured for 30 days, especially those exhibiting red stems developed in the culture medium, were selected for infection. These stems

were cut into 2 mm thick segments with a diameter of 3–4 mm and then immersed in the infection solution for 15–20 min with gentle shaking. Following immersion, excess infection solution was blotted off using sterilized paper towels (Cresia Kimtowel, Nippon Paper, Tokyo, Japan), and the stem segments were co-cultured on callus induction medium (CIM) containing MS basal with 3% sucrose, 0.5% agar, 0.05 mg l⁻¹ BA, 0.1 mg l⁻¹ NAA, 2 g l⁻¹ myo-inositol, and 5 µM CuSO₄, with Qualitative Filter Paper (ADVANTEC, Tokyo, Japan) for 3 days at 25°C in the dark. After co-cultivation, the segments were transferred to CIM containing 5 mg l⁻¹ hygromycin for selection and either 300 mg l⁻¹ timentin or 500 mg l⁻¹ carbenicillin to eliminate *Agrobacterium*. The culture medium was refreshed every two weeks. Regenerated shoots that grew to the top of the culture vessel were excised from the callus and transferred to M-225 bottles (Hakuyo Glass, Tokyo, Japan) containing shoot-inducing medium (MS, 3% sucrose, 0.5% agar, 0.05 mg l⁻¹ BA, 5 µM CuSO₄) supplemented with 5 mg l⁻¹ hygromycin.

Genomic PCR

Upper leaves (approximately 0.1 g) were detached from the transformed plants, immediately frozen in liquid nitrogen, and homogenized using a Cell Destroyer PS2000 (Bio Medical Science, Tokyo, Japan). DNA extraction was then performed using ISOPLANT $\times$ (Nippon Gene, Tokyo, Japan) according to the manufacturer's protocol. Genomic PCR was carried out using Blend Taq® -Plus- (Toyobo, Osaka, Japan) with the primers used detailed in Supplementary Table S1. A 5 µl volume of the resulting PCR product was mixed with 1 µl of 6× Loading Buffer (Takara Bio, Shiga, Japan) and subjected to electrophoresis on a 1.0% agarose gel stained with Midori Green Advance (Nippon Genetics, Tokyo, Japan).

Inverse PCR

Genomic DNA (gDNA) from lines with confirmed T-DNA truncation (via genomic PCR) was subjected to restriction enzyme digestion. For lines exhibiting truncation in the NOS region, we used *Apal*, *Ascl*, *SpeI*, *HindIII*, *EcoRI*, and *XhoI*, enzymes that do not cut between the RB and NOS regions. For lines with truncations elsewhere, we used *NcoI*, *HindIII*, *SphI*, and *PstI*, which target the 35S region. The digested DNA was then purified using the FastGene Gel/PCR Extraction Kit (Nippon Genetics) and eluted in 20 µl of the supplied elution buffer. Circularization of the resulting DNA fragments was performed with T4 DNA Ligase (New England Biolabs, Ipswich, MA, USA). The ligated DNA was precipitated with ethanol and dissolved in 10 µl of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). Inverse PCR was subsequently performed using TaKaRa Ex Premier™ DNA Polymerase (Takara Bio). Nested PCR was also carried out using the products of the inverse PCR. The primers used for this step are detailed in Supplementary Table S1. If a single band was obtained after agarose gel electrophoresis of the nested PCR products, it was purified and sequenced directly. However, if two bands were present, the PCR products were cloned into the pCRTM™-Blunt II-TOPO™ vector (Zero Blunt™ TOPO™ PCR Cloning Kit for Sequencing, Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's protocol. The plasmids were subsequently extracted using the FastGene™ PlasmidMini Kit (Nippon Genetics) and subjected to sequence analysis.

In lines lacking an internal T-DNA region, genomic PCR was conducted using gDNA and primers designed to flank the deletion. The resulting PCR products were purified and finally sent for sequence analysis.

Sequence analysis

Sanger sequencing of the purified DNA samples was conducted by AZENTA, yielding approximately 1,000 bp per sample. The resulting sequences were then analyzed using DNA Dynamo (Blue Tractor Software Ltd.), a dedicated DNA sequence analysis software, to determine the precise truncation position and examine the adjacent sequences. In addition, Unipro UGENE software (Unipro LLC.) was utilized to detect the presence of IR sequences within the T-DNA.

Results

Confirmation of T-DNA integration

Genomic PCR was used to confirm T-DNA integration in 77 transgenic “Princettia” #7801 lines carrying pBCKH-35S::*AtEIN3-SRDX* (*EIN3-SRDX* lines) and 37 lines of #7801 or #7155 carrying pBCKH-35S::*AtPAP1-SRDX* (*PAP1-SRDX* lines). PCR amplification ranges are detailed in Fig. 1a, b. The T-DNA in pBCKH-35S::*AtEIN3-SRDX* was divided into seven regions (Hyg^r, NOS, 35S, *EIN3*-a, *EIN3*-b, *EIN3*-c, SRDX-NOSTer), whereas the T-DNA in pBCKH-35S::*AtPAP1-SRDX* was divided into five regions (Hyg^r, NOS, 35S, *PAP1*, SRDX-NOSTer) (Fig. 1a, b). Our initial screening involved checking for the Hyg^r and *EIN3*-c regions in *EIN3-SRDX* lines, and the Hyg^r and *PAP1* regions in *PAP1-SRDX* lines, to verify integration of both the selectable marker and the chimeric repressor gene. If *EIN3*-c or *PAP1* were absent, we then examined the integration of the NOS and 35S regions. For *EIN3-SRDX* lines, we also checked for the *EIN3*-a and *EIN3*-b regions. Detection of *EIN3*-c or *PAP1* led to further examination for the SRDX-NOSTer region to confirm the presence of the full gene construct.

Genomic PCR analysis revealed the presence of the Hyg^r region in all 114 lines. However, *EIN3*-c was absent in 46 of the 77 *EIN3-SRDX* lines, and *PAP1* was also absent in 9 of the 37 *PAP1-SRDX* lines (Table 1; Table 2). These results suggest a high occurrence of T-DNA truncations of the LB side in these transgenic “Princettia” lines. A detailed examination of the truncated regions in these 55 lines showed that approximately half of the deletions extended to the NOS promoter region, while the other half were distributed toward the LB side, encompassing the SRDX-NOSTer region (Table 1; Table 2; see Supplementary Tables S2 and S3 for further details). To summarize, T-DNA truncation was observed in 65 of the 114 “Princettia” lines transformed with pBCKH-based vectors, primarily initiating from the LB side, with frequent termination points in NOS, *EIN3*-a, *EIN3*-c, and SRDX-NOSTer. Several lines showing truncation were selected for inverse PCR and subsequent sequence analysis. Furthermore, the #1[*EIN3*] line was included in the analysis due to the absence of the NOS region between the detected Hyg^r and 35S regions (Supplementary Table S2).

Identification of truncation positions and analysis of flanking sequences

We conducted sequence analysis on lines showing truncation within the NOS, *EIN3*-a, *EIN3*-c, and SRDX-NOSTer regions, areas where truncation ends were frequently observed, with the aim of determining the precise truncation sites and the features of the adjacent sequences. Supplementary Table S4 lists the lines used in this analysis.

Our initial analysis focused on seven lines with truncation in the NOS region: #58[EIN3], #61[EIN3], #72[EIN3], #76[EIN3], #7[PAP1], #31[PAP1], and #34[PAP1]. Inverse PCR and subsequent sequencing revealed a common truncation site in all seven lines (Fig. 2, top; Fig. 3a, b), with line #58[EIN3] also exhibiting a second distinct site (#58[EIN3]-2) (Fig. 3c). Sequence analysis of the region surrounding the truncation point identified a 113-bp IR immediately upstream of the breakpoint and near the Gene7 terminator (Fig. 2, bottom). Remarkably, the 40-bp sequence adjacent to the IR on the T-DNA's RB side was also inserted, in an inverted orientation, outside the truncation site and directly next to the IR on the LB side. This feature was shared by #61[EIN3], #76[EIN3], #7[PAP1], #31[PAP1], and #34[PAP1], where the plant genome followed the end of this inverted 40-bp sequence (Fig. 3a). Conversely, in #58[EIN3], #72[EIN3], and #58[EIN3]-2, a portion of the T-DNA sequence from the RB side was duplicated outside the truncation position (Fig. 3b, c). For these three lines, we infer that truncation also extended up to the NOS region within the duplicated T-DNA.

We then analyzed #28[EIN3] and #62[EIN3] (truncated in *EIN3*-a), and #68[EIN3] and #74[EIN3] (truncated in *EIN3*-c) (Fig. 4a). The truncation in #28[EIN3] was directly followed by the plant genome (Fig. 4b). Conversely, #62[EIN3], #68[EIN3], and #74[EIN3] showed a segment of T-DNA inserted in reverse orientation immediately after the truncation point (Fig. 4c, d, e). Furthermore, 8–10 bp IR sequences were present immediately upstream of the truncation sites in #28[EIN3], #62[EIN3], and #74[EIN3] (Fig. 5; Supplementary Fig. S2a, b). Interestingly, in #62[EIN3], a sequence adjacent to the paired IR sequence was found immediately downstream of the truncation, in an inverted orientation (Fig. 5), a pattern also observed in #61[EIN3], #76[EIN3], #7[PAP1], #31[PAP1], and #34[PAP1] (Fig. 3a). In contrast, #68[EIN3] lacked any IR sequence longer than 7 bp in the vicinity of the truncation.

Subsequently, we analyzed #44[EIN3], #48[EIN3], #71[EIN3], #12[PAP1], #20[PAP1], #27[PAP1], and #29[PAP1], all exhibiting truncations in the SRDX-NOSTer region. The precise truncation sites differed across these lines (Fig. 6a, b). At the truncation junctions, #48[EIN3], #12[PAP1], #20[PAP1], #27[PAP1], and #29[PAP1] showed plant genome-derived sequences (Fig. 6c), whereas #44[EIN3] and #71[EIN3] showed vector-derived sequences (Fig. 6d, e). #44[EIN3] contained an inverted sequence near the Hyg^r region (Fig. 6d), and #71[EIN3] had a vector backbone sequence that was originally located 1,597 bp downstream of the truncation point (Fig. 6e). Additionally, 7-bp IR sequences were identified immediately upstream of the truncation in four lines: #44[EIN3], #71[EIN3], #12[PAP1], and #27[PAP1]. Although #29[PAP1] lacked IR sequences longer than 7 bp immediately upstream, an 8-bp IR sequence was

originally present on the T-DNA directly downstream of the truncation (Fig. 7). In #48[EIN3] and #20[PAP1], no IR sequence longer than 7 bp was found in the vicinity of the truncation.

Lastly, we analyzed #1[EIN3], in which the NOS region between the Hyg^r and 35S was absent. This analysis revealed a 375 bp deletion immediately downstream of the 113-bp IR sequence within the NOS region (Supplementary Fig. S3). Furthermore, an 88-bp inverted sequence originating from the SRDX-NOSTer region, which included a 2 bp insertion, was found following the IR sequence.

To summarize, IR sequences were often found directly preceding or following the truncation positions, with the majority of truncations occurring after the longest IR sequence, which was 113 bp. Moreover, partial insertions of inverted T-DNA regions were frequently observed immediately downstream of the truncation sites.

Discussion

Our sequence analysis showed that IR sequences exceeding 7 bp are frequently found immediately flanking the truncation points. Remarkably, in the seven lines with NOS region truncations, all breakpoints coincided with the location of the 113-bp IR sequence (Fig. 2; Fig. 3a, b). These results indicate a role for IR sequences in T-DNA truncation, with a higher likelihood of truncation associated with longer IRs. The observation by Sunitha et al. (2012) of increased truncation with IR sequences longer than 200 bp supports our findings. Moreover, we often observed partial insertions of inverted T-DNA regions directly downstream of the truncation sites (Fig. 3; Fig. 4; Fig. 6). For example, in lines #61[EIN3], #76[EIN3], #7[PAP1], #31[PAP1], and #34[PAP1], the 40-bp sequence that is normally present only on the RB side was inserted in the reverse orientation after the truncation (Fig. 3a). Similarly, a portion of the inverted Hyg^r sequence was inserted immediately after the truncation in #62[EIN3] (Fig. 4c), while the Hyg^r sequence on the RB side retained its full length (Supplementary Table S2). The presence of inverted T-DNA regions immediately after the truncation points suggest these were newly synthesized sequences, potentially added by DNA polymerase at the truncation site. Given that DNA polymerase typically functions during T-DNA conversion to double-stranded DNA (Gelvin 2017), this also implies that truncation occurs before the complete integration of T-DNA into the plant genome. Furthermore, in all truncated lines, the LB side was deleted from within the T-DNA (Table 1; Table 2). This bias toward LB side truncation is likely due to the selection process favoring lines with an intact Hyg^r region on the RB side of the T-DNA. Supporting this, transformation and selection in *Chlamydomonas reinhardtii* using T-DNA with the selectable marker on the LB side resulted in 1 kbp or larger deletions on the RB side in all nine lines (Quach et al. 2023). These results collectively suggest that when T-DNA truncation occurs, only one of the regions—either the RB or LB side—is typically integrated into the plant genome.

We propose a new model for T-DNA truncation based on our findings (Fig. 8, top; Supplementary Fig. S4). This model posits that upon entry of single-stranded (ss) T-DNA into the plant nucleus and the subsequent detachment of VirE2, the T-DNA forms a stem-loop structure through the annealing of IR

sequences. At the base of this stem-loop, factors such as DNA nucleases can trigger T-DNA truncation or degradation. Subsequently, the ssT-DNA is converted to its double-stranded (ds) form and integrated into the plant genome. We hypothesize that the addition of sequences like the 40-bp fragment and Hyg^r at the truncation site occurs after the truncation event, mediated by DNA polymerase using the flanking sequence outside the IR as a template (Fig. 8, bottom). Considering that DNA polymerase θ (Pol θ) is known to convert ssT-DNA to dsT-DNA (van Kregten et al. 2016), we suggest that this enzyme may be active at the truncation point. It is established that IR sequences in single-stranded nucleic acids can form stem-loop structures through intra-strand annealing (Holters et al. 2012), which can enhance their stability (Svoboda and Di Cara 2006). Moreover, the integration of *Agrobacterium* chromosomal DNA, which lacks the border sequences necessary for T-DNA excision from binary vectors, into plant genomes has been reported (Ülker et al. 2008). This observation supports our findings that T-DNA lacking border sequences can still be integrated into the plant genome, indicating that these sequences are not strictly required for the integration process itself.

While no specific nucleases targeting foreign single-stranded DNA in plant nuclei have been reported, the enzyme responsible for T-DNA truncation might be related to DSB repair pathways, such as non-homologous end joining (NHEJ) and alternative non-homologous end joining (Alt-EJ), which are known to mediate T-DNA integration (Gelvin 2017; Lacroix and Citovsky 2019). For instance, MRE11 possesses nuclease activity against hairpin structures and is implicated in DNA repair in diverse eukaryotes (Trujillo and Sung 2001). Recent research also indicates that *Arabidopsis* MRE11 may be involved in the removal of VirD2 bound to the RB of T-DNA (Kralemann et al. 2022). Therefore, DNA repair proteins with nuclease activity, such as MRE11, are potential candidates for mediating T-DNA truncation. Additionally, Liang and Tzfira (2013) suggested that exo/endonucleases involved in plant DNA processing might degrade the unconverted 3' end of T-DNA. Considering the roles of SNM1 and COM1 in DNA end processing and NHEJ in *Arabidopsis* and rice (Johnson et al. 2011; Xu et al. 2020; Shen and Li 2022), these proteins could also be involved in T-DNA truncation.

A vector backbone sequence was found directly after the truncation point in #71[EIN3] (Fig. 6e). Integration of vector backbone sequences alongside RB or LB into plant genomes is a known phenomenon (Krizkova and Hroudá 1998; Wu et al. 2006), often attributed to errors in border sequence recognition during T-DNA excision from binary vectors (Kononov et al. 1997; De Buck et al. 2000). However, in #71[EIN3], the LB was absent, and the backbone sequence, originally located 1,597 bp downstream of the truncation, was found immediately adjacent to it. Furthermore, a 7-bp IR sequence was detected immediately upstream of the truncation. These findings suggest that the T-DNA was initially excised with the backbone sequence attached, followed by a 1,597 bp deletion encompassing the LB within the plant nucleus. In #1[EIN3], which exhibited a 375 bp deletion, the 113-bp IR sequence was located immediately upstream of the deleted region (Supplementary Fig. S3). It is probable that these deletions arose from truncation events occurring at multiple sites within the T-DNA, possibly facilitated by IR sequences, with subsequent recombination of the resulting fragments into a non-original arrangement.

Our analysis using Unipro UGENE revealed 946 and 612 IR sequences of 7–10 bp within the T-DNAs of pBCKH-35S::AtEIN3-SRDX and pBCKH-35S::AtPAP1-SRDX, respectively, with approximately 75% (706 and 451) being 7-bp IRs. This abundance of short IR sequences suggests that those found near truncation sites might be random occurrences and not necessarily causative. In fact, some lines lacked detectable IR sequences near their truncation points. These observations imply that T-DNA might undergo further excision after an initial IR-mediated truncation, or that a different mechanism than the one we proposed could be responsible. For instance, T-DNA integration can be mediated by 1–25 bp microhomologies with the plant genome (Gheysen et al. 1991; Fu et al. 2006; Kleinboelting et al. 2015). Furthermore, a 7.3 kbp deletion of T-DNA in tobacco showed significant homology between the T-DNA breakpoint and the plant genome integration site (Matsumoto et al. 1990), suggesting that truncation could also occur at homologous T-DNA/plant genome junctions. Meza et al. (2002) proposed an alternative model in which deletions occur due to T-DNA loopback through own homologous sequences. Therefore, T-DNA truncation may involve multiple pathways. A more detailed understanding necessitates considering various factors beyond IR sequences.

The pBCKH vector employed in this study was designed for the CRES-T method, a technique for repressing transcription factor activity (Oshima et al. 2011), and has been used for transformation across diverse plant species (Ishiwata et al. 2013; Kishi-Kaboshi et al. 2017; Huang et al. 2019). The 113-bp IR sequence, which may play a role in truncation, is present in the pBCKH vector and is commonly inherited by derived constructs. However, transformation of *Torenia fournieri* with pBCKH-35S::AtPAP1-SRDX rarely resulted in T-DNA truncation (Supplementary Fig. S5a, b, c), and a similar tendency was observed in *Eustoma grandiflorum* transformed with pBCKH-35S::AtEIN3-SRDX (data not shown). Moreover, to our knowledge, frequent T-DNA truncation has not been reported in plants transformed with pBCKH-based vectors. These findings suggest that the truncation observed in “Princettia” is unlikely to be attributable solely to the structure of the pBCKH vector or to a general tendency toward small-scale deletions. Rather, “Princettia” appears to have a unique predisposition to T-DNA truncation. Furthermore, we have empirically noted that frequent truncation also occurs in *Euphorbia coranstra*, the pollen parent of “Princettia”. Future research will be needed to identify the factors underlying the high frequency of T-DNA truncation in specific plant species.

Abbreviations

RB, right border of T-DNA; LB, left border of T-DNA; IR, inverted repeat; Hyg^r, hygromycin phosphotransferase; NOS, nopaline synthase; *EIN3*, *ETHYLENE INSENSITIVE3*; *PAP1*, *PRODUCTION OF ANTHOCYANIN PIGMENT 1*; SRDX, Superman Repression Domain modified ver. X; BA, 6-benzyladenine; NAA, 1-Naphthaleneacetic Acid.

Declarations

Acknowledgements

We thank Dr Yukihiisa Katsumoto, Suntory Global Innovation Center Limited (Kyoto, Japan), for providing us with the “Princettia” experimental lines #7801 and #7155. We also thank Dr. Nobutaka Mitsuda from the Bioproduction Research Institute, National Institute of AIST, Japan, for providing us with the pBCKH-35S::AtEIN3-SRDX and pBCKH-35S::AtPAP1-SRDX.

Author contribution

KI and NO conceived the study and designed the research; KM, KS and KT provided important input on the research; YS, MO and RK investigated the tissue culture and transformation conditions; RK, SS and RS produced the transgenic plants; KI and RK performed DNA extractions and PCR analysis; KI performed sequencing and analyzed the data; KI and NO wrote the manuscript; KM, KS and KT critically read and reviewed the manuscript.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interests

The authors declare no conflict of interest.

References

1. Christie PJ, Ward JE, Winans SC, Nester EW (1988) The *Agrobacterium tumefaciens virE2* gene product is a single-stranded-DNA-binding protein that associates with T-DNA. J Bacteriol 170:2659–2667. <https://doi.org/10.1128/jb.170.6.2659-2667.1988>
2. Citovsky V, Kozlovsky SV, Lacroix B, Zaltsman A, Dafny-Yelin M, Vyas S, Tovkach A, Tzfira T (2007) Biological systems of the host cell involved in *Agrobacterium* infection. Cell Microbiol 9:9–20. <https://doi.org/10.1111/j.1462-5822.2006.00830.x>
3. de Buck S, de Wilde C, van Montagu M, Depicker A (2000) T-DNA vector backbone sequences are frequently integrated into the genome of transgenic plants obtained by *Agrobacterium*-mediated transformation. Mol Breed 6:459–468. <https://doi.org/10.1023/A:1026575524345>
4. de Vos G, Zambryski P (1989) Expression of *Agrobacterium* nopaline-specific VirD1, VirD2, and VirC1 proteins and their requirement for T-strand production in *E. coli*. Mol Plant Microbe Interact 2:43–52. <https://doi.org/10.1094/mpmi-2-043>

5. Dürrenberger F, Crameri A, Hohn B, Koukolíková-Nicola Z (1989) Covalently bound VirD2 protein of *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation. *Proc Natl Acad Sci USA* 86:9154–9158. <https://doi.org/10.1073/pnas.86.23.9154>
6. Filichkin SA, Gelvin SB (1993) Formation of a putative relaxation intermediate during T-DNA processing directed by the *Agrobacterium tumefaciens* VirD1, D2 endonuclease. *Mol Microbiol* 8:915–926. <https://doi.org/10.1111/j.1365-2958.1993.tb01637.x>
7. Fu D, St Amand PC, Xiao Y, Muthukrishnan S, Liang GH (2006) Characterization of T-DNA integration in creeping bentgrass. *Plant Sci* 170:225–237. <https://doi.org/10.1016/j.plantsci.2005.08.026>
8. Gambino G, Chitarra W, Maghuly F, Laimer M, Boccacci P, Marinoni DT, Gribaudo I (2009) Characterization of T-DNA insertions in transgenic grapevines obtained by *Agrobacterium*-mediated transformation. *Mol Breed* 24:305–320. <https://doi.org/10.1007/s11032-009-9293-8>
9. Gelvin SB (2017) Integration of *Agrobacterium* T-DNA into the Plant Genome. *Annu Rev Genet* 51:195–217. <https://doi.org/10.1146/annurev-genet-120215-035320>
10. Gheysen G, Villarroel R, van Montagu M (1991) Illegitimate recombination in plants: a model for T-DNA integration. *Genes Dev* 5:287–297. <https://doi.org/10.1101/gad.5.2.287>
11. Holkers M, de Vries AAF, Gonçalves MAFV (2012) Nonspaced inverted DNA repeats are preferential targets for homology-directed gene repair in mammalian cells. *Nucleic Acids Res* 40:1984–1999. <https://doi.org/10.1093/nar/gkr976>
12. Huang PY, Zhang J, Jiang B, Chan C, Yu JH, Lu YP, Chung K, Zimmerli L (2019) NINJA-associated ERF19 negatively regulates Arabidopsis pattern-triggered immunity. *J Exp Bot* 70:1033–1047. <https://doi.org/10.1093/jxb/ery414>
13. Ishiwata A, Ozawa M, Nagasaki H, Kato M, Noda Y, Yamaguchi T, Nosaka M, Shimizu-Sato S, Nagasaki A, Maekawa M, Hirano HY, Sato Y, Two (2013) *WUSCHEL*-related homeobox genes, *narrow leaf2* and *narrow leaf3*, control leaf width in rice. *Plant Cell Physiol* 54:779–792. <https://doi.org/10.1093/pcp/pct032>
14. Johnson RA, Hellens RP, Love DR (2011) A transient assay for recombination demonstrates that *Arabidopsis* *SNM1* and *XRCC3* enhance non-homologous recombination. *Genet Mol Res* 10:2104–2132. <https://doi.org/10.4238/vol10-3gmr1347>
15. Kim SR, Lee J, Jun SH, Park S, Kang HG, Kwon S, An G (2003) Transgene structures in T-DNA-inserted rice plants. *Plant Mol Biol* 52:761–773. <https://doi.org/10.1023/a:1025093101021>
16. Kishi-Kaboshi M, Aida R, Sasaki K (2017) Generation of gene-edited *Chrysanthemum morifolium* using multicopy transgenes as targets and markers. *Plant Cell Physiol* 58:216–226. <https://doi.org/10.1093/pcp/pcw222>
17. Kleinboelting N, Huet G, Appelhagen I, Viehoveer P, Li Y, Weisshaar B (2015) The structural features of thousands of T-DNA insertion sites are consistent with a double-strand break repair-based insertion mechanism. *Mol Plant* 8:1651–1664. <https://doi.org/10.1016/j.molp.2015.08.011>
18. Kononov ME, Bassuner B, Gelvin SB (1997) Integration of T-DNA binary vector ‘backbone’ sequences into the tobacco genome: evidence for multiple complex patterns of integration. *Plant J* 11:945–

957. <https://doi.org/10.1046/j.1365-313X.1997.11050945.x>
19. Kralemann LEM, de Pater S, Shen H, Kloet SL, van Schendel R, Hooykaas PJJ, Tijsterman M (2022) Distinct mechanisms for genomic attachment of the 5' and 3' ends of *Agrobacterium* T-DNA in plants. *Nat Plants* 8:526–534. <https://doi.org/10.1038/s41477-022-01147-5>
20. Krizkova L, Hroudá M (1998) Direct repeats of T-DNA integrated in tobacco chromosome: characterization of junction regions. *Plant J* 16:673–680. <https://doi.org/10.1046/j.1365-313x.1998.00330.x>
21. Kumar S, Fladung M (2002) Transgene integration in aspen: structures of integration sites and mechanism of T-DNA integration. *Plant J* 31:543–551. <https://doi.org/10.1046/j.1365-313X.2002.01368.x>
22. Lacroix B, Citovsky V (2019) Pathways of DNA transfer to plants from *Agrobacterium tumefaciens* and related bacterial species. *Annu Rev Phytopathol* 57:231–251. <https://doi.org/10.1146/annurev-phyto-082718-100101>
23. Liang Z, Tzfira T (2013) *In vivo* formation of double-stranded T-DNA molecules by T-strand priming. *Nat Commun* 4:2253. <https://doi.org/10.1038/ncomms3253>
24. Matsumoto S, Ito Y, Hosoi T, Takahashi Y, Machida Y (1990) Integration of *Agrobacterium* T-DNA into a tobacco chromosome: Possible involvement of DNA homology between T-DNA and plant DNA. *Mol Gen Genet* 224:309–316. <https://doi.org/10.1007/BF00262423>
25. Meza TJ, Stangeland B, Mercy IS, Skårn M, Nymoen DA, Berg A, Butenko MA, Håkelién AM, Haslekås C, Meza-Zepeda LA, Aalen RB (2002) Analyses of single-copy *Arabidopsis* T-DNA-transformed lines show that the presence of vector backbone sequences, short inverted repeats and DNA methylation is not sufficient or necessary for the induction of transgene silencing. *Nucleic Acids Res* 30:4556–4566. <https://doi.org/10.1093/nar/gkf568>
26. Oshima Y, Mitsuda N, Nakata M, Nakagawa T, Nagaya S, Kato K, Ohme-Takagi M (2011) Novel vector systems to accelerate functional analysis of transcription factors using chimeric repressor gene-silencing technology (CRES-T). *Plant Biotechnol* 28:201–210. <https://doi.org/10.5511/plantbiotechnology.11.0124a>
27. Quach TN, Sato SJ, Behrens MR, Black PN, DiRusso CC, Cerutti HD, Clemente TE (2023) A facile *Agrobacterium*-mediated transformation method for the model unicellular green algae *Chlamydomonas reinhardtii*. *Vitro Cell Dev Biol Plant* 59:671–683. <https://doi.org/10.1007/s11627-023-10389-7>
28. Scheiffele P, Pansegrau W, Lanka E (1995) Initiation of *Agrobacterium tumefaciens* T-DNA processing: Purified proteins VirD1 and VirD2 catalyze site- and strand-specific cleavage of superhelical T-border DNA *in vitro*. *J Biol Chem* 270:1269–1276. <https://doi.org/10.1074/jbc.270.3.1269>
29. Shen H, Li Z (2022) DNA double-strand break repairs and their application in plant DNA integration. *Genes* 13:322. <https://doi.org/10.3390/genes13020322>

30. Stahl R, Horvath H, van Fleet J, Voetz M, von Wettstein D, Wolf N (2002) T-DNA integration into the barley genome from single and double cassette vectors. *Proc Natl Acad Sci USA* 99:2146–2151. <https://doi.org/10.1073/pnas.032645299>
31. Sunitha S, Shivaprasad PV, Sujata K, Veluthambi K (2012) High frequency of T-DNA deletions in transgenic plants transformed with intron-containing hairpin RNA genes. *Plant Mol Biol Rep* 30:158–167. <https://doi.org/10.1007/s11105-011-0327-0>
32. Svoboda P, Di Cara A (2006) Hairpin RNA: a secondary structure of primary importance. *Cell Mol Life Sci* 63:901–908. <https://doi.org/10.1007/s00018-005-5558-5>
33. Thomson G, Dickinson L, Jacob Y (2024) Genomic consequences associated with *Agrobacterium*-mediated transformation of plants. *Plant J* 117:342–363. <https://doi.org/10.1111/tpj.16496>
34. Tinland B, Schoumacher F, Gloeckler V, Bravo-Angel AM, Hohn B (1995) The *Agrobacterium tumefaciens* virulence D2 protein is responsible for precise integration of T-DNA into the plant genome. *EMBO J* 14:3585–3595. <https://doi.org/10.1002/j.1460-2075.1995.tb07364.x>
35. Trujillo KM, Sung P (2001) DNA structure-specific nuclease activities in the *Saccharomyces cerevisiae* Rad50*Mrf1 complex. *J Biol Chem* 276:35458–35464. <https://doi.org/10.1074/jbc.M105482200>
36. Tzfira T, Li J, Lacroix B, Citovsky V (2004a) *Agrobacterium* T-DNA integration: molecules and models. *Trends Genet* 20:375–383. <https://doi.org/10.1016/j.tig.2004.06.004>
37. Tzfira T, Vaidya M, Citovsky V (2004b) Involvement of targeted proteolysis in plant genetic transformation by *Agrobacterium*. *Nature* 431:87–92. <https://doi.org/10.1038/nature02857>
38. Ülker B, Li Y, Rosso MG, Logemann E, Somssich IE, Weisshaar B (2008) T-DNA-mediated transfer of *Agrobacterium tumefaciens* chromosomal DNA into plants. *Nat Biotechnol* 26:1015–1017. <https://doi.org/10.1038/nbt.1491>
39. van Kregten M, de Pater S, Romeijn R, van Schendel R, Hooykaas PJJ, Tijsterman M (2016) T-DNA integration in plants results from polymerase- θ -mediated DNA repair. *Nat Plants* 2:16164. <https://doi.org/10.1038/nplants.2016.164>
40. Ward ER, Barnes WM (1988) VirD2 protein of *Agrobacterium tumefaciens* very tightly linked to the 5' end of T-strand DNA. *Science* 242:927–930. <https://doi.org/10.1126/science.242.4880.927>
41. Wei S, Xi YZ, Song DP, Wei H, Gruber MY, Gao MJ, Parkin I, Kachatourians G, Hannoufa A (2015) Quantitative and structural analyses of T-DNA tandem repeats in transgenic *Arabidopsis* SK mutant lines. *Plant Cell Tissue Organ Cult* 123:183–192. <https://doi.org/10.1007/s11240-015-0825-0>
42. Wu H, Sparks CA, Jones HD (2006) Characterisation of T-DNA loci and vector backbone sequences in transgenic wheat produced by *Agrobacterium*-mediated transformation. *Mol Breed* 18:195–208. <https://doi.org/10.1007/s11032-006-9027-0>
43. Xu Z, Zhang J, Cheng X, Tang Y, Gong Z, Gu M, Yu H (2020) *COM1*, a factor of alternative non-homologous end joining, lagging behind the classic non-homologous end joining pathway in rice somatic cells. *Plant J* 103:140–153. <https://doi.org/10.1111/tpj.14715>

44. Yasumoto S, Sawai S, Lee HJ, Mizutani M, Saito K, Umemoto N, Muranaka T (2020) Targeted genome editing in tetraploid potato through transient TALEN expression by *Agrobacterium* infection. Plant Biotechnol 37:205–211. <https://doi.org/10.5511/plantbiotechnology.20.0525a>

Tables

Table 1 Summary of T-DNA integration results in the *EIN3-SRDX* lines

	<i>EIN3-SRDX</i> lines	Hyg ^r	NOS	35S	<i>EIN3-a</i>	<i>EIN3-b</i>	<i>EIN3-c</i>	SRDX-NOSTer
No. of integrated lines	77	77	52	49 ^a	43	39	31	26
No. of truncated lines	—	0	24 ^a	4	6	4	8	5

a: Line #1[EIN3] showed integration of the Hyg^r and 35S regions, but not the NOS region or downstream sequences of 35S. This indicates truncation within the *EIN3-a* region. Overall, 77 lines displayed Hyg^r integration, and 52 lines displayed NOS integration. When #1[EIN3] is excluded, 24 lines exhibited NOS truncation. Additionally, 52 lines showed NOS integration, four lines 35S truncation, and 49 lines 35S integration (including #1[EIN3])

Table 2 Summary of T-DNA integration results in the *PAP1-SRDX* lines

	<i>PAP1-SRDX</i> lines	Hyg ^r	NOS	35S	<i>PAP1</i>	SRDX-NOSTer
No. of integrated lines	37	37	31	30	28	23
No. of truncated lines	—	0	6	1	2	5

Figures

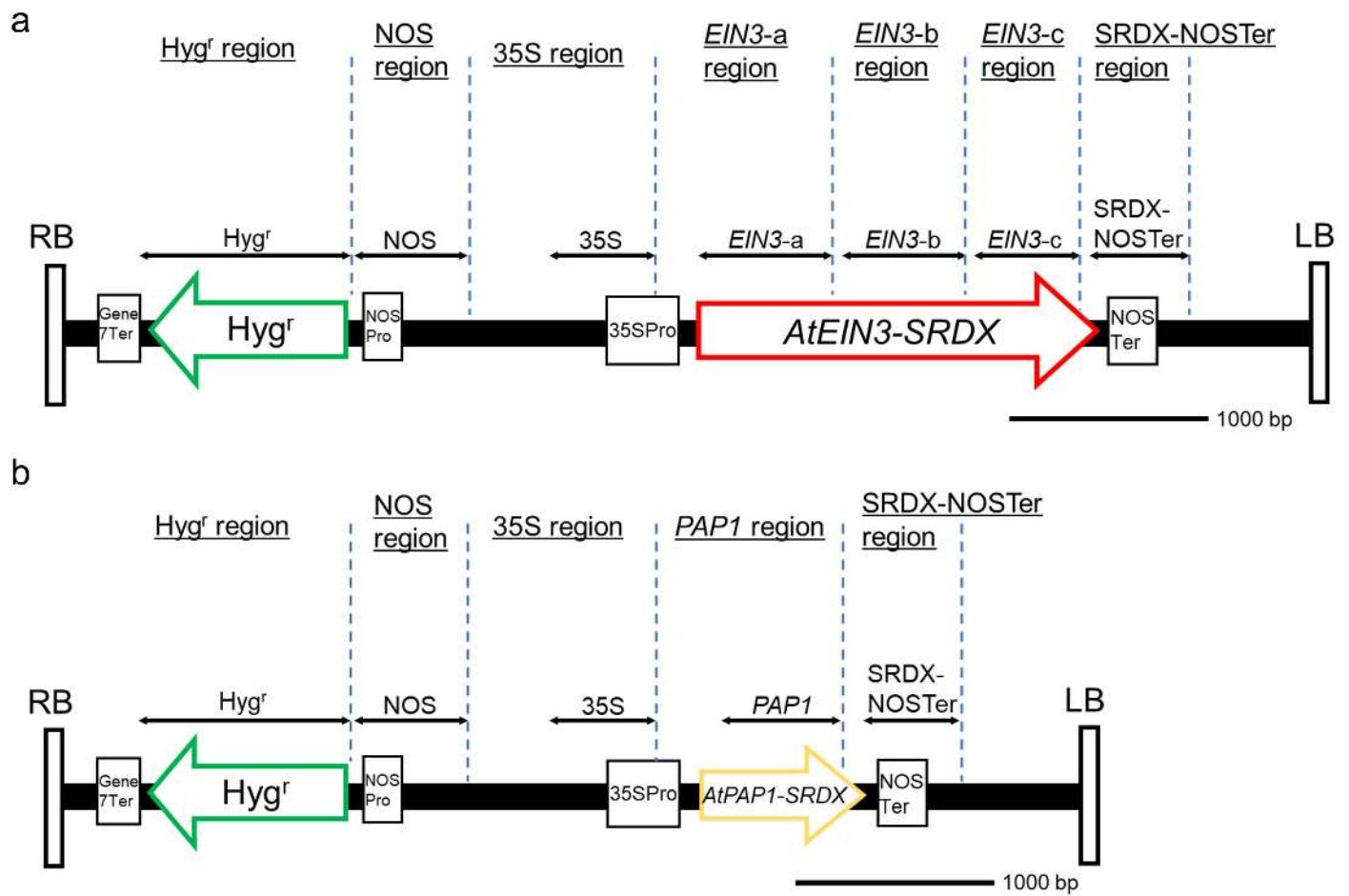


Figure 1

Schematic representation of the constructs employed in this study. Legend: RB, right border of T-DNA; Gene7Ter, octopine-type Ti plasmid gene 7 terminator; Hyg^r, hygromycin phosphotransferase (resistance); NOSPro, nopaline synthase promoter; 35SPro, Cauliflower mosaic virus 35S promoter; NOSTer, nopaline synthase terminator; LB, left border of T-DNA; *AtEIN3-SRDX* and *AtPAP1-SRDX*, chimeric repressor genes designed to convert *Arabidopsis EIN3* and *PAP1* into functional transcriptional repressors. Double-headed arrows indicate the regions amplified and presented in Tables 1 and 2. **(a)** pBCKH-35S::*AtEIN3-SRDX*. **(b)** pBCKH-35S::*AtPAP1-SRDX*

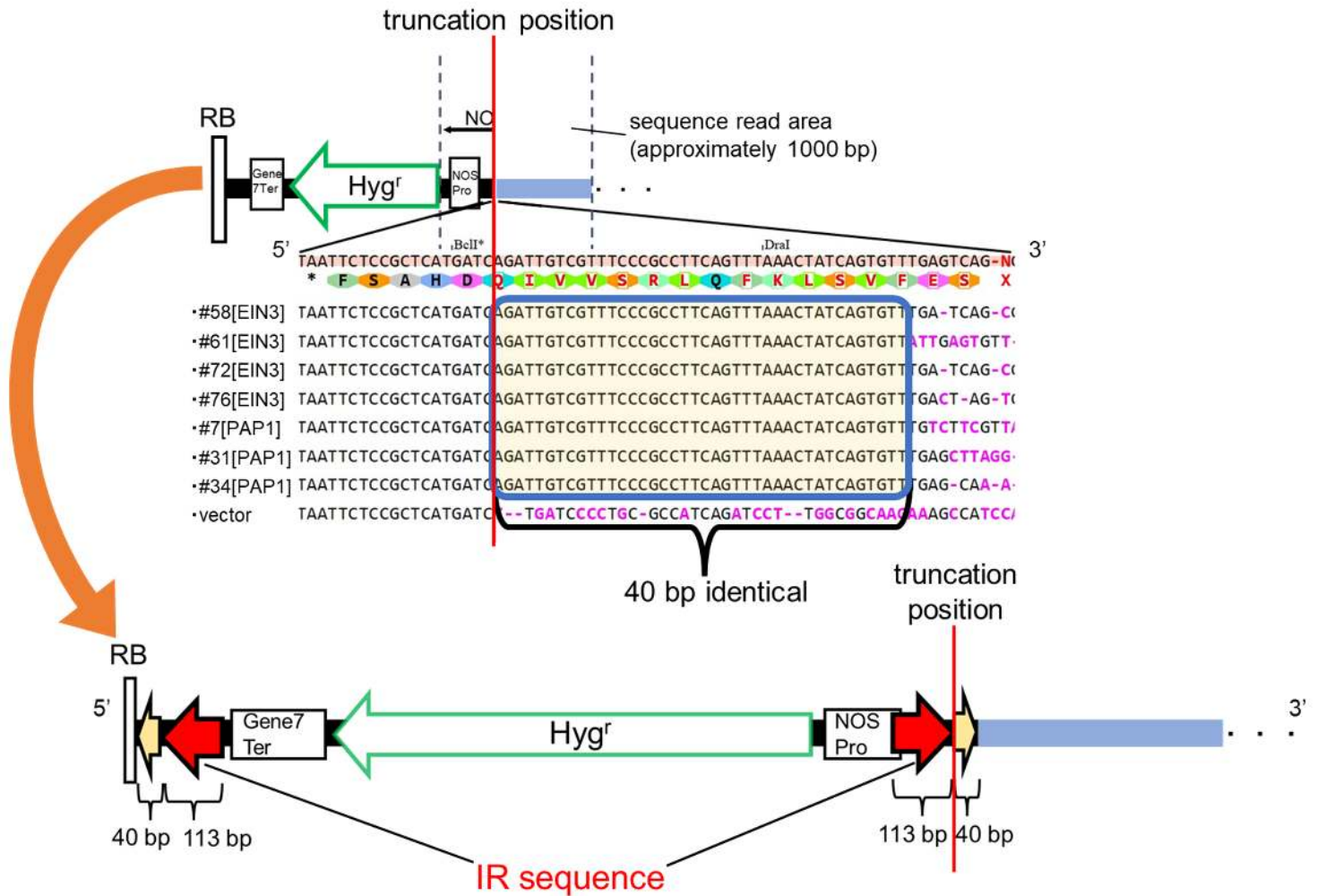


Figure 2

Schematic representation of truncation sites and surrounding sequences in seven NOS region-truncated lines: #58[EIN3], #61[EIN3], #72[EIN3], #76[EIN3], #7[PAP1], #31[PAP1], and #34[PAP1]. The top panel shows the truncation point and the immediate 5' and 3' sequences for each line. The blue-boxed region indicates a conserved 40-bp sequence immediately 3' of the truncation across all lines. The bottom panel provides a detailed view from the RB to the truncation site, highlighting a 113-bp IR sequence (red arrow) immediately 5' of the truncation and near the Gene7 terminator. Two identical 40-bp sequences (yellow arrows) are also shown: one immediately 3' of the truncation and another adjacent to the IR sequence on the RB side, oriented in opposite directions to denote their inverted relationship.

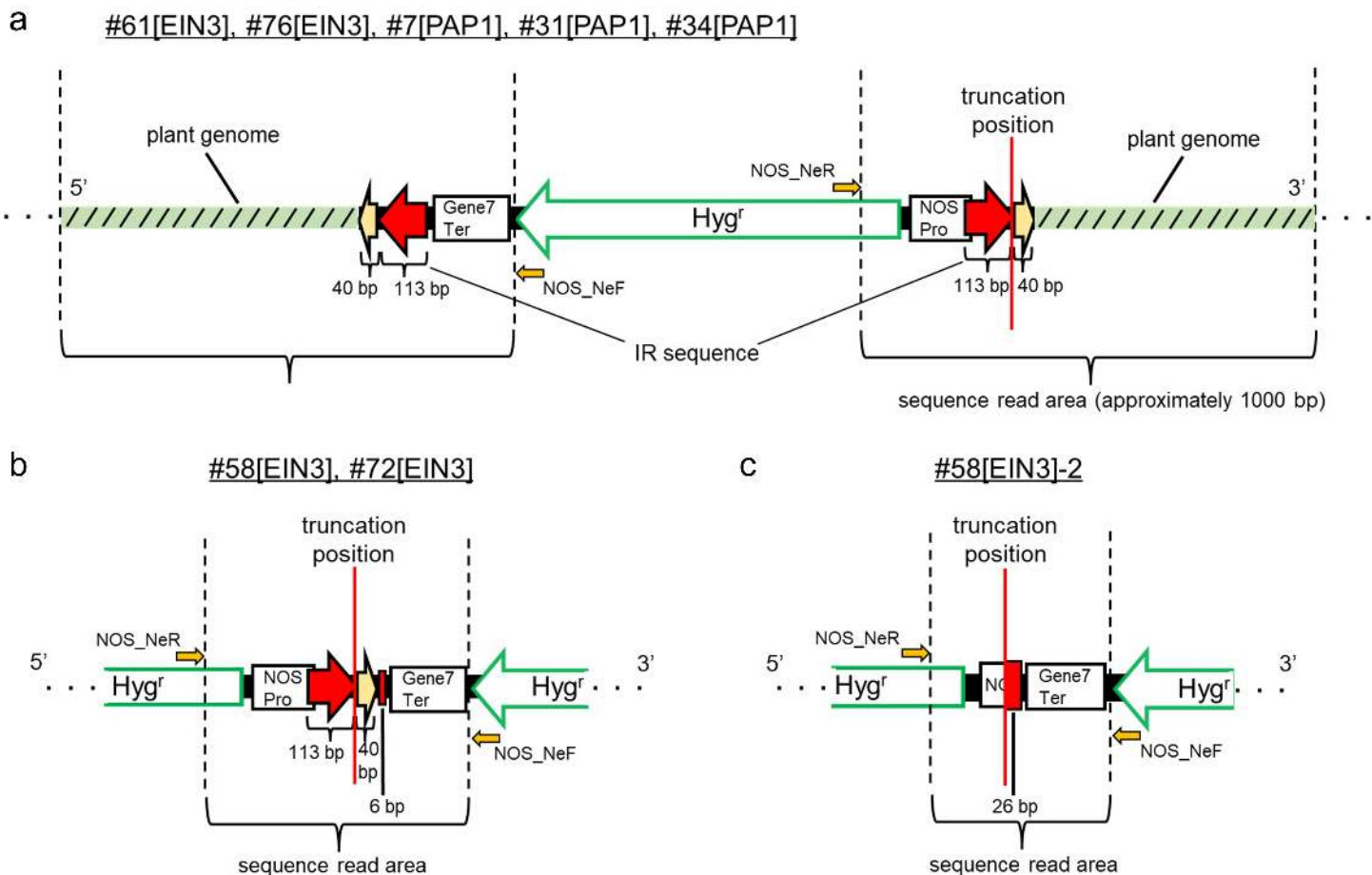


Figure 3

Sequence analysis results revealing features of NOS region truncations in seven lines: #58[EIN3], #61[EIN3], #72[EIN3], #76[EIN3], #7[PAP1], #31[PAP1], and #34[PAP1]. The primers used for sequencing were “NOS_NeF” and “NOS_NeR.” (a) In #61[EIN3], #76[EIN3], #7[PAP1], #31[PAP1], and #34[PAP1], sequence analysis showed plant genomic DNA inserted immediately 3' of the 40-bp sequence. (b) In #58[EIN3] and #72[EIN3], a segment of the T-DNA from the RB side was duplicated and inserted outside the 40 bp. (c) In #58[EIN3]-2, a segment of the T-DNA from the RB side was duplicated and inserted outside the truncation position

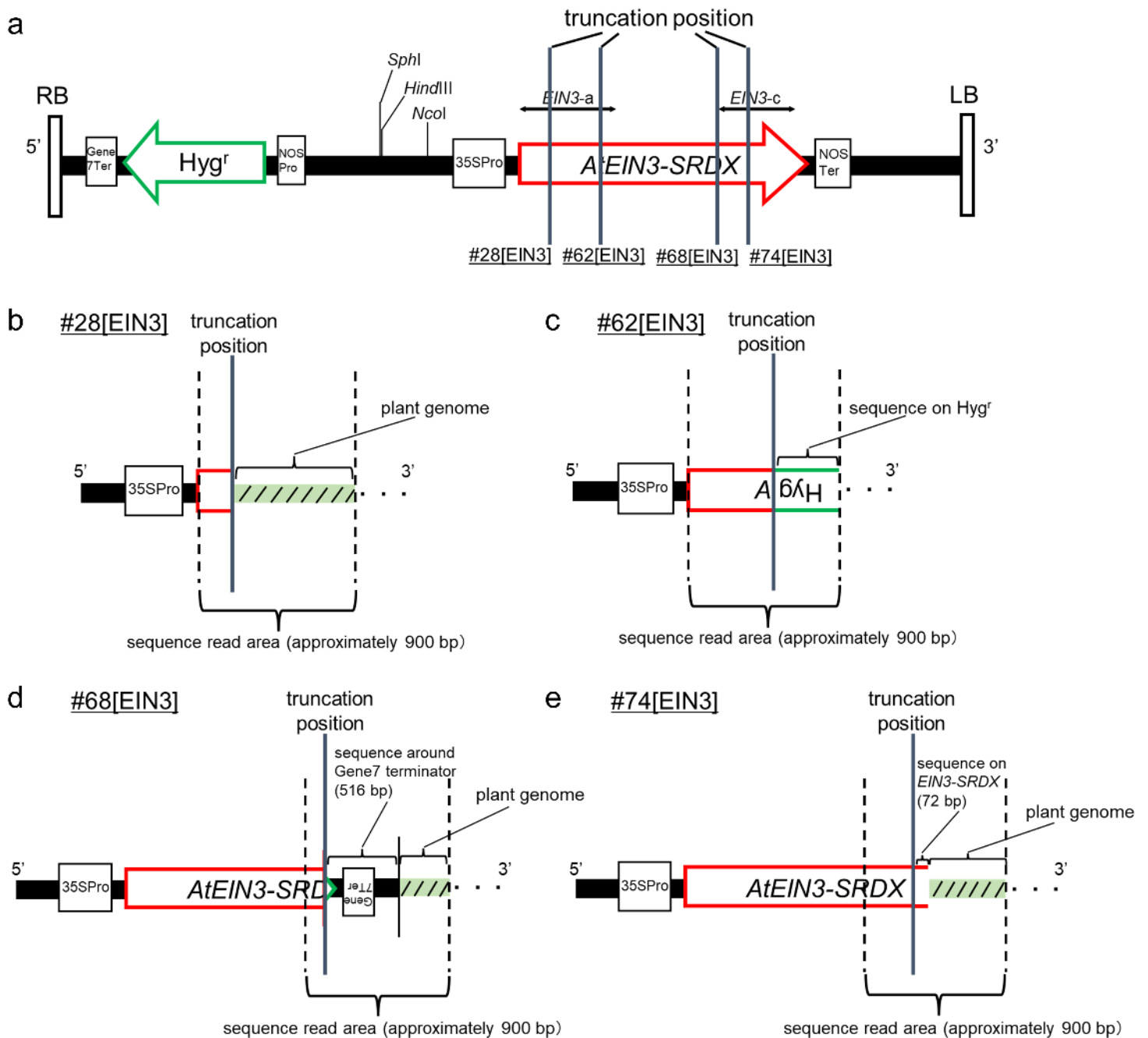


Figure 4

Sequence analysis revealing insertions at truncation sites in lines #28[EIN3], #62[EIN3], #68[EIN3], and #74[EIN3] (*EIN3*-a or *EIN3*-c region truncations). **(a)** Overview of truncation positions and the *SphI*, *HindIII*, and *NcoI* restriction sites used for digestion. **(b)** In #28[EIN3], the plant genome sequence was inserted immediately after the truncation. **(c)** In #62[EIN3], an inverted *Hyg^r* sequence was inserted immediately after the truncation. **(d)** In #68[EIN3], an inverted 516-bp sequence around the Gene7 terminator was inserted immediately after the truncation, followed by the plant genome. **(e)** In #74[EIN3], an inverted 72-bp fragment of *AtEIN3-SRDX*, originally 32 bp downstream, was inserted immediately after the truncation, followed by the plant genome

#62[EIN3]

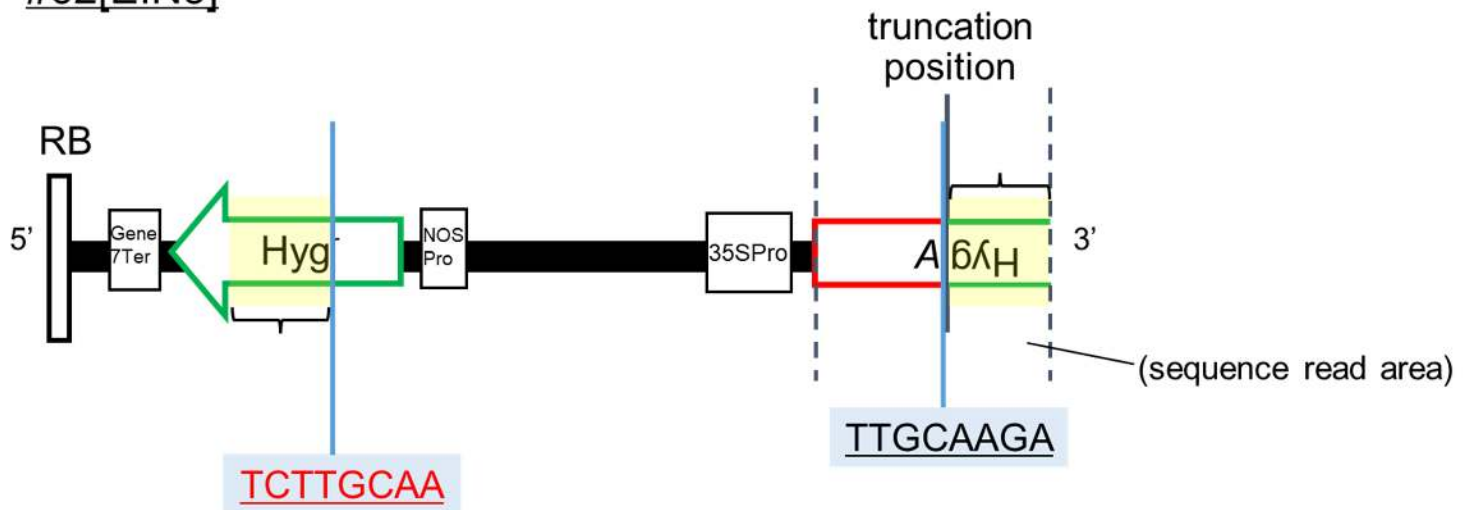


Figure 5

Schematic representation of the IR and inserted sequence identified in #62[EIN3]. An 8-bp IR sequence, TTGCAAGA, was located immediately 5' of the truncation site. A homologous IR sequence was present within the Hyg^r gene, and the sequence flanking this Hyg^r IR was inserted immediately 3' of the truncation site in an inverted orientation

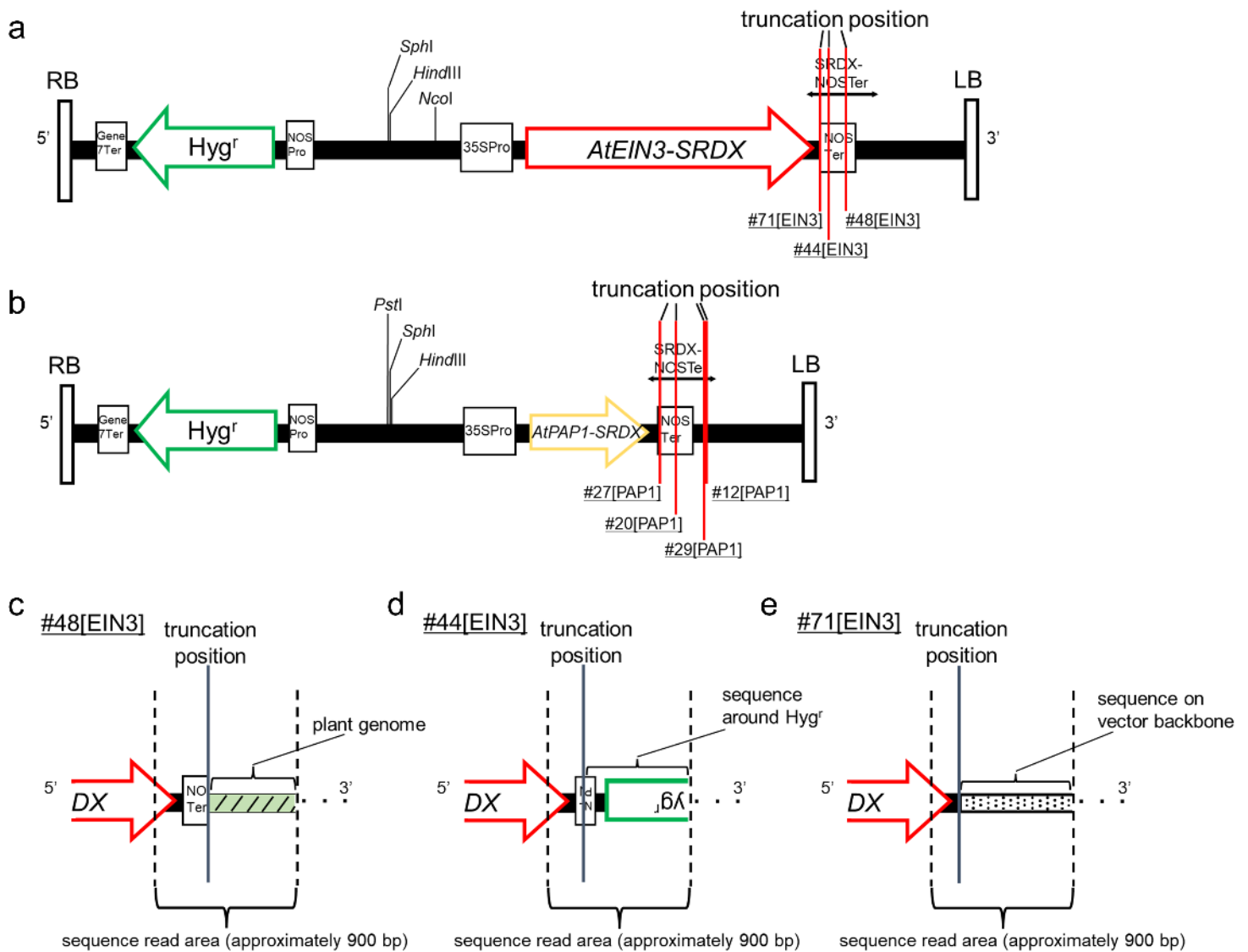


Figure 6

Sequence analysis revealing insertions at truncation sites in seven lines with SRDX-NOSTer region truncations: #44[EIN3], #48[EIN3], #71[EIN3], #12[PAP1], #20[PAP1], #27[PAP1], and #29[PAP1]. **(a)** Overview of truncation positions for #44[EIN3], #48[EIN3], and #71[EIN3], and the *SphI*, *HindIII*, and *NcoI* restriction sites used for digestion. **(b)** Overview of truncation positions for #12[PAP1], #20[PAP1], #27[PAP1], and #29[PAP1], and the *PstI*, *SphI*, and *HindIII* restriction sites used for digestion. **(c)** In #48[EIN3], the plant genome sequence was inserted immediately after the truncation. **(d)** In #44[EIN3], an inverted sequence around the *Hyg^r* region was inserted immediately after the truncation. **(e)** In #71[EIN3], a vector backbone sequence (originally 1,597 bp downstream) was inserted immediately after the truncation.

#29[PAP1]

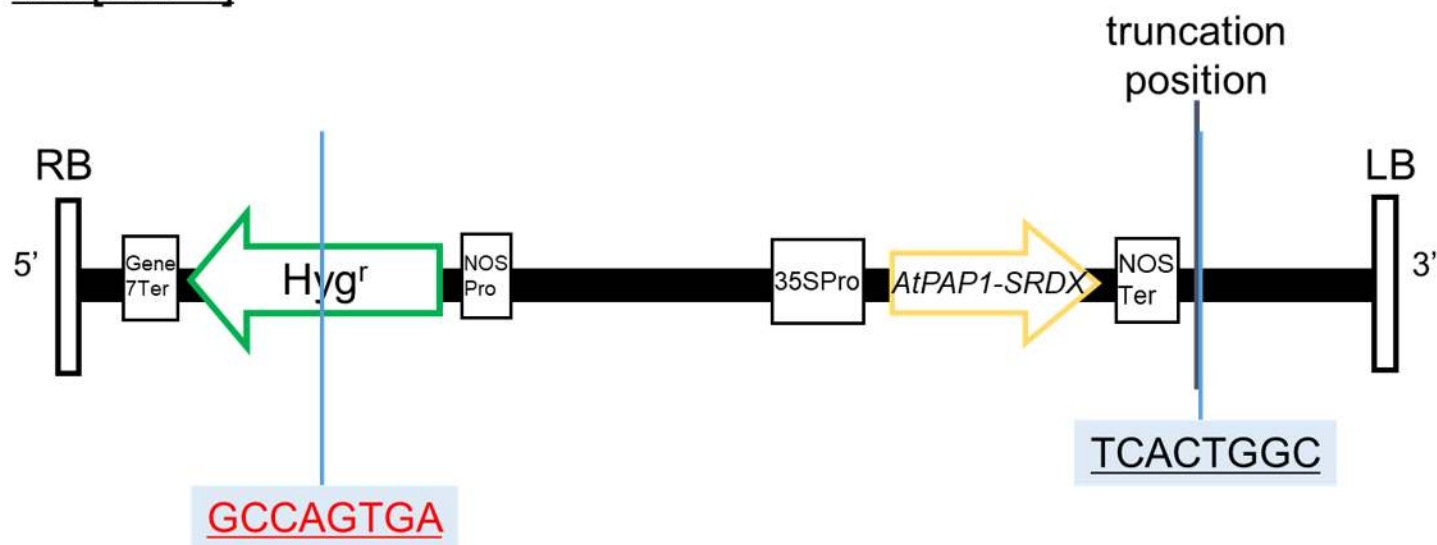


Figure 7

IR sequence in #29[PAP1]. An 8-bp IR sequence (TCACTGGC) was originally located directly after the truncation point in the T-DNA. A matching IR sequence was present within the Hyg^r

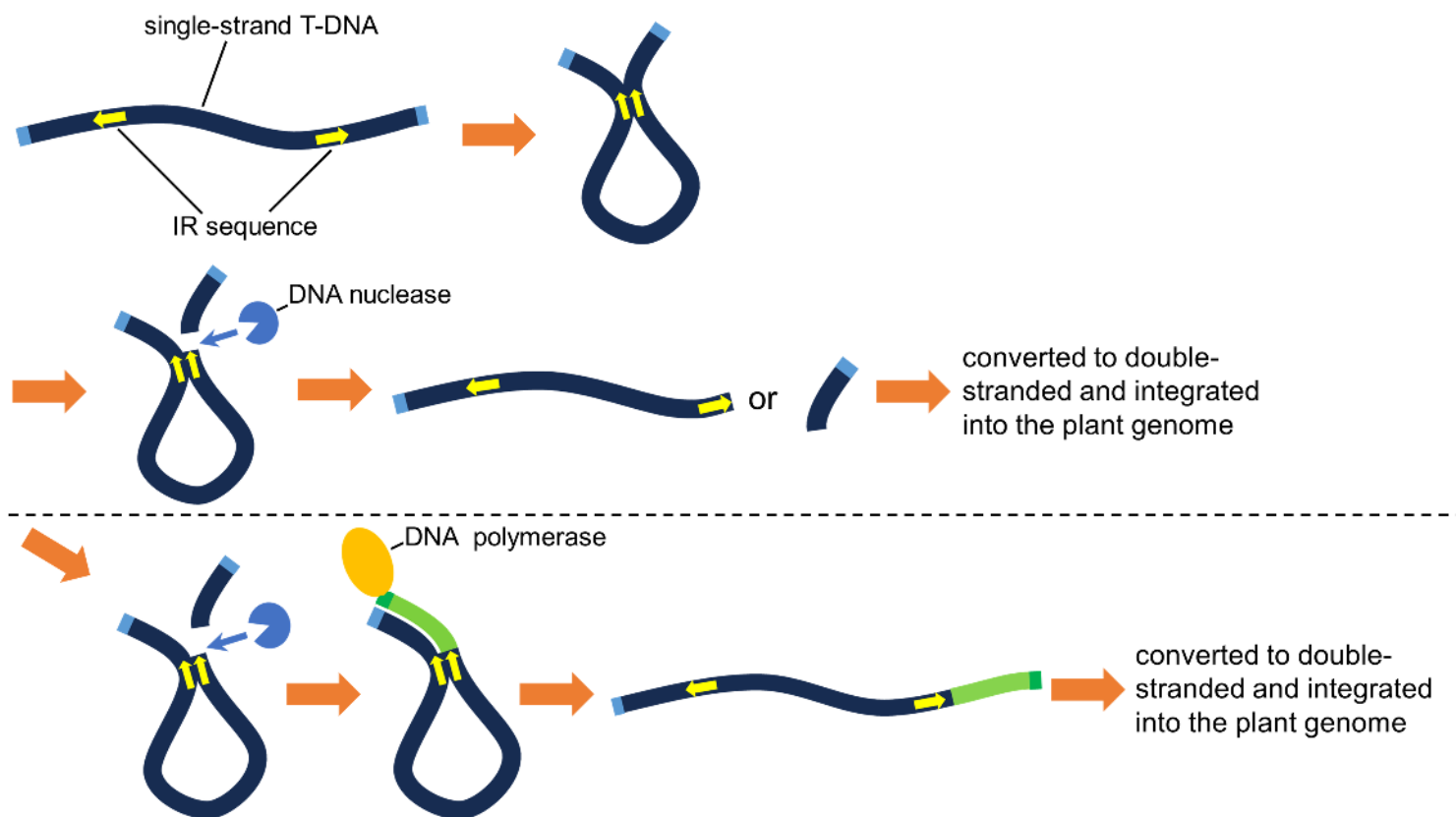


Figure 8

Model of T-DNA truncation proposed to be mediated by IR sequences. Once inside the plant nucleus, single-stranded T-DNA (ssT-DNA) folds into a stem-loop structure through the binding of its IR sequences. At the loop junction, factors like DNA nucleases can cause T-DNA truncation or degradation. Following this, the ssT-DNA (either the RB or LB strand) is converted to double-stranded DNA and integrated into the plant genome. If, after truncation, DNA polymerase uses the sequence outside the IR as a template to synthesize a complementary strand, an inverted sequence will be added directly after the truncation point

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfile1.docx](#)