Journal of Experimental Botany, Vol. 75, No. 11 pp. 3248–3258, 2024
https://doi.org/10.1093/jxb/erae114 Advance Access Publication 13 March 2024

REVIEW PAPER

A word of caution: T-DNA-associated mutagenesis in plant reproduction research

Karel Raabe[1,2,*], Limin Sun[3], Cédric Schindfessel[3], David Honys[1], and Danny Geelen[3,*]

[1] Laboratory of Pollen Biology, Institute of Experimental Botany of the Czech Academy of Sciences, Rozvojová 263, 165 02 Prague 6, Czech Republic
[2] Department of Experimental Plant Biology, Faculty of Science, Charles University, Viničná 5, 128 44 Prague 2, Czech Republic
[3] Horticell, Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000 Ghent, Belgium

* Correspondence: raabe@ueb.cas.cz or Danny.Geelen@UGent.be

Received 15 January 2024; Editorial decision 7 March 2024; Accepted 12 March 2024

Editor: Aline Probst, Université Clermont Auvergne, France

Abstract

T-DNA transformation is prevalent in Arabidopsis research and has expanded to a broad range of crops and model plants. While major progress has been made in optimizing the *Agrobacterium*-mediated transformation process for various species, a variety of pitfalls associated with the T-DNA insertion may lead to the misinterpretation of T-DNA mutant analysis. Indeed, secondary mutagenesis either on the integration site or elsewhere in the genome, together with epigenetic interactions between T-DNA inserts or frequent genomic rearrangements, can be tricky to differentiate from the effect of the knockout of the gene of interest. These are mainly the case for genomic rearrangements that become balanced in filial generations without consequential phenotypical defects, which may be confusing particularly for studies that aim to investigate fertility and gametogenesis. As a cautionary note to the plant research community studying gametogenesis, we here report an overview of the consequences of T-DNA-induced secondary mutagenesis with emphasis on the genomic imbalance on gametogenesis. Additionally, we present a simple guideline to evaluate the T-DNA-mutagenized transgenic lines to decrease the risk of faulty analysis with minimal experimental effort.

Keywords: Arabidopsis, chromosome, chromosome rearrangement, gametophyte, meiosis, reproduction, T-DNA, translocation.

Introduction

Agrobacterium tumefaciens is a Gram-negative soil bacterium that transfers DNA sequences (T-DNA) into plants and integrates it into the plant nuclear genome (Gelvin, 2021). *Agrobacterium*-mediated transformation is the most widely used method for transgenesis in plant research and biotechnological applications. Large collections of *Arabidopsis thaliana* T-DNA insertion lines have been generated to characterize genes in reverse genetics studies as well as in forward genetic screens (Alonso and Ecker, 2006). Currently, publicly available T-DNA collections contain ~350 000 T-DNA insertion lines which provide a vast pool of possible knockout mutants for reverse genetic research (O'Malley *et al.*, 2015; Pucker *et al.*, 2021). Similar collections

also exist in crop species, for example in rice (An *et al.*, 2005). In addition to these collections, *Agrobacterium*-mediated T-DNA insertion is used for the creation of custom-made stable transgenic lines carrying various types of constructs for modulation of gene and protein expression or site-directed mutagenesis including the CRISPR (clustered regularly interspaced palindromic repeats) technology. The mechanism by which a T-DNA is inserted into the host genomic sequence remains only partially understood and involves a double strand break (DSB) as a starting point for T-DNA integration into the host genome and ligation by the plant DSB repair machinery (Gelvin, 2021).

A T-DNA line preferentially contains a single, complete T-DNA inserted into the plant genome. However, *Agrobacterium*-mediated transformation often leads to multiple secondary or extraneous mutations (Fig. 1). T-DNA-associated mutagenesis encompasses various changes, including insertions and/or deletions of the plant DNA or T-DNA segment at the integration site, as well as more complex integration of multiple copies of T-DNA at a single or multiple loci in the genome (Gelvin, 2021; Thomson *et al.*, 2024). Additionally, genomic changes may also include epigenetic modifications. T-DNA integration often triggers chromosomal rearrangements such as translocations, inversions, or large segment deletions/duplications (Hu *et al.*, 2017; Thomson *et al.*, 2024). The incidence of such rearrangements increases with the number of T-DNA insertions (Clark and Krysan, 2010; O'Malley *et al.*, 2015).

Reverse genetics, aiming at functional characterization of genes, involves evaluating the phenotype of T-DNA insertion lines or CRISPR/Cas9 (CRISPR-associated protein 9) generated mutants. At the initial stage of characterization, cryptic mutations or genomic rearrangements may go unnoticed, and a reduced fertility phenotype may be inadvertently linked with the disruption of the gene of interest. While there are reports of cryptic mutations in lines from T-DNA collections, it is less well documented for CRISPR/Cas9-generated mutations. Considering the popularity of CRISPR/Cas9 mutagenesis, the incidence of translocation upon transformation is likely to be an obstacle as with conventional T-DNA insertion mutagenesis. Indeed, Cas9 constructs are usually transformed using *Agrobacterium* transformation whereby the Cas9-mediated cleavage mechanism intrinsically leads to additional DSBs. These Cas9-induced DSBs could be specifically used to induce a targeted reciprocal translocation in plant species with possible applications for chromosome engineering of crops (Beying *et al.*, 2020).

With this report, we aim to raise awareness among plant scientists and describe the genetic and phenotypic consequences of T-DNA-associated mutagenesis, and thoroughly discuss the most problematic chromosome rearrangements for reproduction research in T-DNA transgenic lines—translocations. A simple guideline and brief method description is proposed to help in avoiding misinterpretation of sterility or other phenotypes in T-DNA-associated mutagenesis.

Common characteristics of T-DNA-induced mutagenesis

Structure and detection of a single T-DNA insertion

The T-DNA transferred from *Agrobacterium* Ti-plasmid into plant cells is marked at the borders by imperfect direct repeats, designated the left border (LB) and right border (RB). Between the LB and RB, a custom DNA sequence can be inserted such as a selective marker for the isolation of transgenic lines (Nath Radhamony *et al.*, 2005). T-DNA primers are used to characterize the insertion flanking plant genome sequences. With a flanking sequence tag (FST), the T-DNA insertion is mapped onto the genome and located at a specific locus (O'Malley *et al.*, 2015). In publicly available collections, the T-DNA insertion lines are indexed using the LB flanking sequences and their distribution covering the entire genome (Wu *et al.*, 2015).

The success of the Arabidopsis T-DNA insertion libraries has boosted the development of T-DNA mutant collections of crops. Also, these T-DNA mutants are being genotyped using a pair of locus-specific primers encompassing the insertion site and T-DNA insertion-specific primers. However, it is common practice that lines are first screened for a specific phenotype and only the most promising candidates are then studied in more detail. Although this approach is time efficient, the presence of cryptic mutations is potentially overlooked, and insertion lines with multiple mutations might be selected for further studies.

The problem with mutants carrying additional mutations is easily avoided when two independent T-DNA lines are available showing similar phenotypes or by performing genetic complementation tests. However, not every gene is sufficiently covered by the T-DNA libraries, and complementation tests require time-consuming cloning and transformation. Regarding CRISPR/Cas9 mutagenesis, the additional mutagenesis induced by the Cas9 T-DNA should be avoided by segregating or crossing out the Cas9 and guide RNA (gRNA) T-DNA from the mutant background before detailed phenotype analysis.

Integration site-associated T-DNA-induced mutagenesis

T-DNA integration has been shown to frequently induce additional insertions, deletions, and duplications of the plant genomic DNA at the site of integration (Fig. 1) (Kleinboelting *et al.*, 2015; Wu *et al.*, 2015; Schouten *et al.*, 2017). These deletions are usually smaller than 30 bp, but can span several hundreds of base pairs affecting more than one gene (Kleinboelting *et al.*, 2015). Apart from the sequence between the LB and RB, the additionally inserted DNA originates from the Ti-plasmid

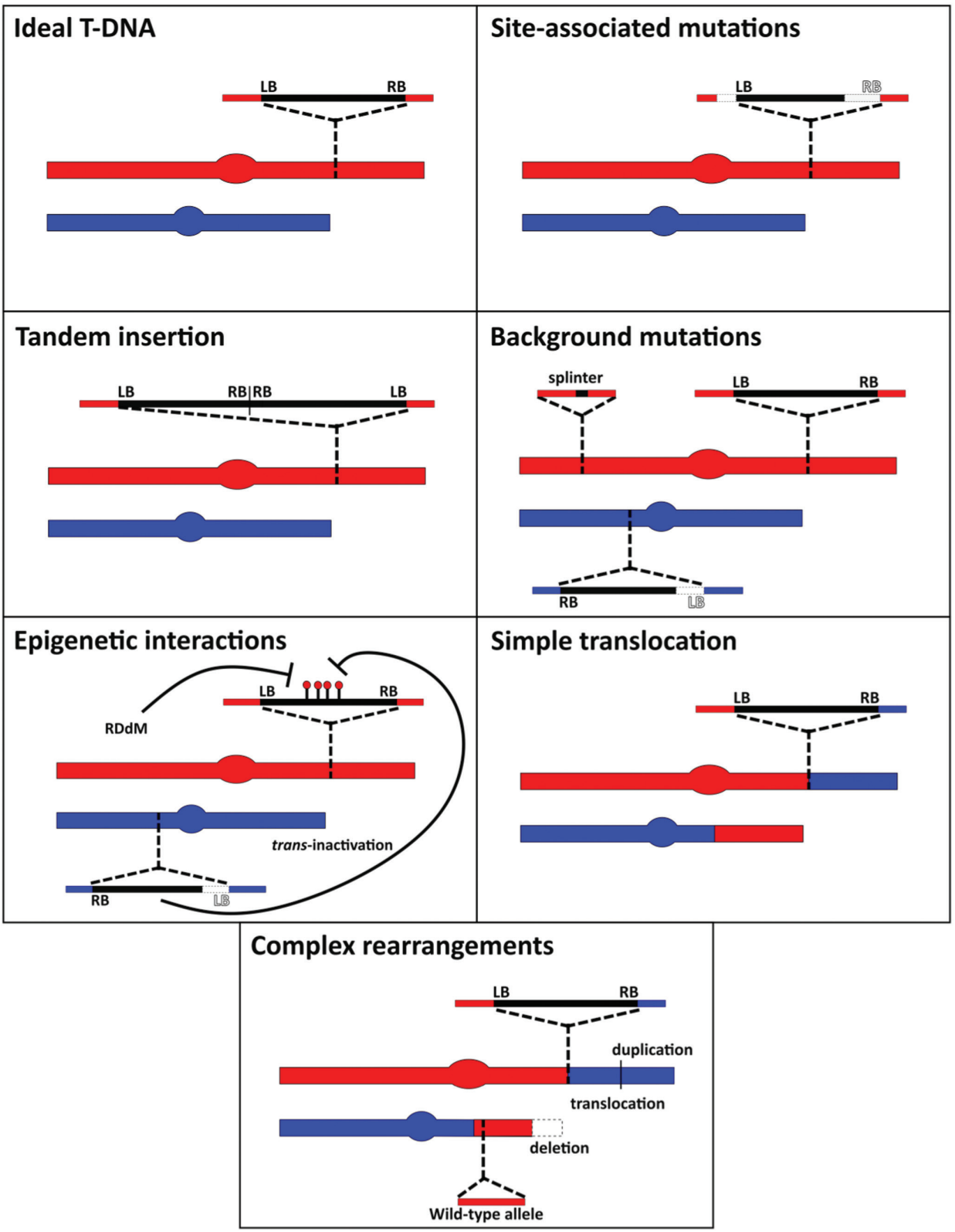

Fig. 1. Overview of the T-DNA-associated mutagenesis. The ideal T-DNA insertion is shown as a simple scheme of two metacentric chromosomes (red and blue) and T-DNA insertion in one of the chromosomes with complete left border (LB) and right border (RB) sequences. Anything other than the ideal T-DNA insertion could lead to misinterpretations of the phenotypical changes in the analysed mutant line. The site-associated mutation, here simplified into a plant genome deletion next to the LB site and truncation of the T-DNA sequence at the RB site, might cause trouble in PCR genotyping. Similarly,

a tandem insertion, here shown in RB–RB configuration, could lead to a failed PCR. Background mutations, here depicted as splinter DNA and additional truncated T-DNA in the second chromosome, can lead to a phenotype unrelated to the inspected T-DNA, particularly difficult to detect when the additional insertion lacks sequences that are used in the initial insertion mapping (LB site). Robust segregation and complementation analysis prevent the faulty genotype–phenotype associations. Epigenetic interactions, here shown as the silencing of the T-DNA by RNA-directed DNA methylation (RdDM) or *trans*-inactivation by a homologous background T-DNA, can change the epigenetic landscape of the insertion site, including the T-DNA. This could lead to the loss of the T-DNA selection marker or in phenotype loss in intronic insertions. Frequently occurring chromosome translocations, here shown as the reciprocal exchange of chromosome arms between the red and blue chromosomes at the insertion site, can lead to severe gametophytic phenotypes due to a meiotic imbalance. More complex rearrangements, here shown as a translocation combined with deletion and duplication of large genomic segments, have further potential to complicate the T-DNA analysis. For detailed description of the T-DNA-associated mutagenesis and their detection, see the respective sections in the main text.

or binary vector backbone, the *Agrobacterium* chromosome, the plant genome, or the inserted sequence does not show homology and is then referred to as 'filler' DNA (Wu *et al.*, 2015; Singer, 2018). The T-DNA fragment is also often incompletely integrated, with truncated or rearranged borders (Kleinboelting *et al.*, 2015; Wu *et al.*, 2015; Sun *et al.*, 2019). The most frequent erroneous insertions consist of two T-DNA copies, commonly ligated in LB–LB or RB–RB configuration or multiple copies in adjacent clusters (De Buck *et al.*, 1999; Singer, 2018).

Alterations at the T-DNA target site are usually discovered rather quickly, as common genotyping focuses on the genomic locus and the gene of interest. It is good practice, however, to examine both T-DNA–genomic junctions of the inserted T-DNA not only by PCR but also by sequencing the PCR products for both the LB and RB. Failure to obtain a PCR product or failure to map the product on the LB or RB might indicate that the T-DNA insertion site is rearranged, and it is reasonable to pursue sequencing the genome of the inspected mutant.

T-DNA-induced mutagenesis outside the inspected locus

A common phenomenon is the integration of multiple T-DNA fragments throughout the genome, which is estimated to be on average 1.5 T-DNA inserts per transgenic Arabidopsis line (O'Malley et al., 2015). The number of insertions in seed databases may be underestimated because borders are truncated or the selection marker is missing (Kleinboelting *et al.*, 2015). Additionally, the FST prediction may fail in the case of integration into paralogous or repetitive sequences (Huep *et al.*, 2014). Furthermore, small inserts, dozens of base pairs long originating from the T-DNA named 'splinter' DNA, have been reported (Fig. 1) (Schouten *et al.*, 2017). Such cryptic untagged mutations can be responsible for, or contribute to, the plant phenotype studied. Additional T-DNA insertions as well as cryptic mutations can be detected and removed by outcrosses with wild-type plants, and detailed segregation analysis of the allele and phenotype can be performed

T-DNA-associated epigenetic changes

In addition to local sequence rearrangements, T-DNA integration can also influence the epigenetic landscape surrounding the insertion site (Fig. 1). The loss of chromatin marks and alteration in levels of euchromatin and heterochromatin histone marks, including *de novo* H3K27me3, have been reported (Jupe *et al.*, 2019), and T-DNA sequences can be methylated, leading to the silencing of the selectable marker (Jupe *et al.*, 2019). T-DNA insertions can also supress other T-DNA copies introduced during transformation or after crossing with another T-DNA mutant, so-called *trans*-inactivation (Daxinger *et al.*, 2008). The *trans*-inactivation depends on the sequence homology between both interacting T-DNAs and can, for instance, be induced by the presence of a cauliflower mosaic virus (CaMV) 35S promoter present in both insertions (Daxinger *et al.*, 2008). Additionally, a cryptic untagged T-DNA in the genome can silence a tagged T-DNA. An intronic T-DNA that affects the splicing of a gene can lose the initially observed phenotype upon interaction with another T-DNA, known as T-DNA suppression (Xue *et al.*, 2012; Gao and Zhao, 2013; Sandhu *et al.*, 2013; Osabe *et al.*, 2017; Jiang *et al.*, 2019). This inactivation depends on heterochromatinization of the intronic T-DNA sequences which restores the splicing of the gene (Osabe *et al.*, 2017), leading to partial or complete loss of the phenotype, lasting over several generations even after the interacting T-DNA is crossed out (Jiang *et al.*, 2019).

The introduction of multiple T-DNAs into a single line should be carefully designed to minimize the sequence homology between the inserts, especially in complementation studies. When studying double mutants, the combination of T-DNA insertions derived from different libraries (e.g. SALK and SAIL) may help in avoiding T-DNA *trans*-inactivation. It is also good practice to screen the crossed lines using PCR genotyping rather than by antibiotic selection. The annotation of T-DNA sequences used in large collections is summarized here (Ulker *et al.*, 2008).

T-DNA-induced genomic rearrangements

T-DNA integration may trigger genomic rearrangements that include translocations, duplications, deletions, inversions, or combinations thereof (Fig. 1). While these rearrangements often lead to gametophytic lethality, a genetic structure may arise in homozygous T-DNA lines that do not display the lethal phenotype (Fig. 2). How these genotypes may emerge is discussed in the next section. In the description, we use a simple reciprocal translocation line and its typical phenotypic consequences. This ideal example might not be what a researcher encounters, given

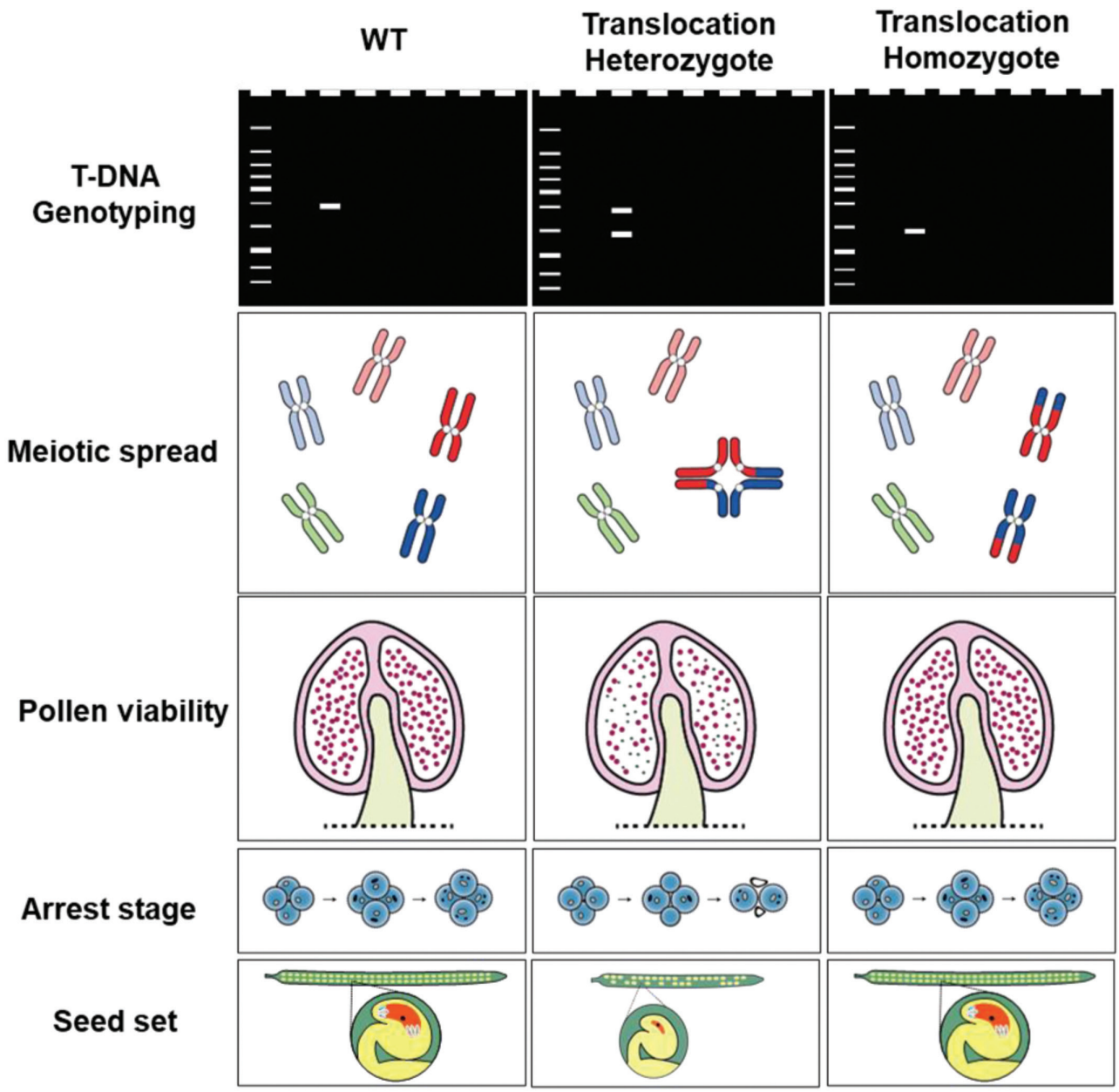

Fig. 2. Schematic overview of the simple T-DNA translocation phenotype in *Arabidopsis thaliana*. The matrix shows comparison of different phenotypical traits between wild-type plants and plants carrying a T-DNA-induced chromosome translocation in the heterozygous or homozygous state. Typical PCR genotyping agarose electrophoresis band pattern of the T-DNA insertion is shown, together with chromosome arrangements during the meiotic prophase; pollen viability analysis by Alexander staining (pink indicates viable pollen, grey indicates unviable pollen). The bottom two lines show the pollen development arrest stage and the seed set defect with the indicated ovule development arrest stage. The generalized scenario is depicted, where the defects are observed in ~50% of gametes. However, the severity and penetrance of the gamete defect can vary between translocation lines.

that genomic rearrangements can be complex or in combination with other types of mutagenesis described earlier, and that the phenotypical effects might show reduced penetrance.

T-DNA-induced translocations

Introduction of T-DNA in the plant genome can induce the exchange of two DNA segments between two non-homologous chromosomes, a reciprocal translocation (Fig. 1) (Gelvin, 2021), and has been reported many times in Arabidopsis T-DNA mutants containing genomic rearrangements (Castle *et al.*, 1993; Nacry *et al.*, 1998; Tax and Vernon, 2001; Forsbach *et al.*, 2003; Lafleuriel *et al.*, 2004; Yuen *et al.*, 2005; Curtis *et al.*, 2009; Ruprecht *et al.*, 2014; Schouten *et al.*, 2017; Min *et al.*, 2020; Alling and Galindo-Trigo, 2023), and in T-DNA mutants of *Solanum lycopersum* and *Oryza sativa* (Ohba *et al.*, 1995; Takano

et al., 1997). Surprisingly, the frequency of translocations in Arabidopsis T-DNA collections varies ~10–20% (Clark and Krysan, 2010; Pucker *et al.*, 2021), albeit this is likely to be an underestimation because of limitations in detection (Clark and Krysan, 2010).

Meiotic imbalance due to reciprocal chromosome translocation

The consequences of the frequently occurring chromosome translocation are typically observed in the heterozygous state of the T-DNA (Fig. 2) (Clark and Krysan, 2010). This is caused by a set of gametophytic developmental defects that are a result of unbalanced gametes produced by erratic meiotic divisions (Curtis *et al.*, 2009). During meiosis, in wild-type plants and in homozygous translocation lines, regular bivalents are formed by crossover between homologous regions (Zickler and Kleckner, 2015). However, in the translocation heterozygotes, translocated segments form crossovers with the homologous parts present in a non-homologous chromosome, creating an imbalance in the distribution of the chromosomes over the gametes (Curtis *et al.*, 2009). Gametes with chromosome imbalance arrest at an early phase of development and eventually abort. It has been hypothesized that the arrest in developmental progress is either caused by the loss of essential genes in the unbalanced gametes (Clark and Krysan, 2010) or induced by genomic stress and perturbation of pre-mitotic control mechanisms (Ruprecht *et al.*, 2014). Balanced gametes containing the translocation with a full chromosome complement are viable and capable of passing the translocation to the next generation.

Gametophytic defects in a translocation line

During male reproduction, microspores are released from tetrads whereby the balanced gametes progress to develop into mature pollen (Hafidh and Honys, 2021), and the unbalanced gametes arrest before, or shortly after, pollen mitosis I (PMI) (Curtis *et al.*, 2009). The arrested microspores collapse and form a raisin-like structure devoid of nucleus and cytoplasm (Fig. 2) (Clark and Krysan, 2010). The timing of the arrest varies slightly, being mostly at the unicellular stage (Curtis *et al.*, 2009) or the early bicellular stage (Min *et al.*, 2020). In translocation heterozygotes, the frequency of unbalanced aborted pollen reaches up to a maximum of 50%.

Female meiosis produces four meiocytes of which three degenerate and the remaining one develops into a functional embryo sac inside the ovule (Skinner and Sundaresan, 2018). In a translocation heterozygote, up to 50% of the ovules stop developing before the first mitotic division, which leads to reduced seed set and shorter siliques (Fig. 2) (Curtis *et al.*, 2009). As in male gametogenesis, some gametes undergo recombination events that do not lead to a loss of essential genetic material and develop normally (Ray *et al.*, 1997). Gamete abortion is a typical consequence of chromosomal rearrangement. However, gamete abortion is not fully penetrant in all translocation lines, and abortion can be caused by T-DNA disrupting an essential gene (Clark and Krysan, 2010).

T-DNA characterization in the translocation line

Translocations are present in T-DNA lines obtained from publicly available collections (such as SALK, SAIL, Gabi-Kat, etc.) as well as in in-house-generated transgenic lines. In T-DNA-induced translocation lines, the allelic state of the T-DNA insertion allele usually reflects the allelic state of the translocation as well (Fig. 2). Notably, the translocation can affect the T-DNA insertion in such a way that the LB and RB are separated on different chromosomes and segregate independently (Clark and Krysan, 2010). The absence of a border PCR product may also occur due to the truncated T-DNA and is hence not sufficient evidence for a translocation event (O'Malley *et al.*, 2015).

T-DNA transmission and non-Mendelian segregation after crosses

As we described earlier, in a translocation heterozygote about half of the viable gametes contain the chromosome translocation and pass it on to the next generation. Consequently, the T-DNA insertion allele is transmitted through pollen and ovules at the same frequency as the translocation, having no reduction in the transmission through the male and the female germline. However, a non-Mendelian segregation can occur after crossing multiple T-DNA lines carrying insertions at different chromosomes (Curtis *et al.*, 2009). In such cases, a cross of two lines, one being a translocation homozygous without phenotypical differences, leads to an F_1 double heterozygote now showing the gametophytic phenotype caused by the translocation. Subsequently, due to the rearranged chromosomes, the analysis of the selfed F_2 can lead to a non-Mendelian segregation of the two traits (Curtis *et al.*, 2009).

Complex genomic rearrangements

Some T-DNA-induced chromosome rearrangements are complex, and include multiple translocations, duplications, and inversions (Fig. 1) (Hu *et al.*, 2017). A reported example is line *seb19* that contains a translocation, inversion, and duplication on chromosomes 4 and 5, producing progeny with two distinct phenotypes (Hu *et al.*, 2017). *seb19*-L plants showed the typical translocation phenotype that includes reduced seed set, aborted pollen, and regular gametic transmission, while *seb19*-S plants showed a delayed sporophytic development and smaller plant size (Hu *et al.*, 2017). Genome sequencing of Seb19-L revealed that these plants are heterozygous for the genomic rearrangement and the Seb19-S plants are homozygous. A 180 kb deletion in *seb19*-S was responsible for the vegetative growth phenotype (Hu *et al.*, 2017).

When the rearranged genome includes a duplication, it may contain more than two alleles of the characterized gene of interest. In such cases, the PCR genotyping does not reflect the allelic status of the inspected gene of interest (Tax and Vernon, 2001). In the example described in Tax and Vernon (2001), plants that are homozygous for the translocation are PCR genotyped as heterozygous for the T-DNA insertion, but do not show the gametophytic phenotype. Also, the T-DNA allele does not segregate in the next generation of the uniform population of plants without phenotype. Upon outcrossing to the wild type, PCR genotyping of the T-DNA identifies a uniform population of heterozygous plants as well; however, in this case, all plants display the gametophytic phenotype. In the F_2, the T-DNA insertion genotype does not fully link with the gametophytic phenotype as the PCR genotyping scores translocation heterozygotes and translocation homozygotes as T-DNA heterozygotes.

As translocation mutant phenotypes are not always fully penetrant, the above outcome can be misinterpreted. The phenotype should be complemented with a functional target gene. However, in the case of simple or complex translocation T-DNA lines, the complementation test can result in the selection of plants that contain the mutated allele and the complementing allele, and are homozygous for the translocation. These plants display no mutant phenotype and hence are seemingly complemented and could lead to a faulty association of the phenotype with the T-DNA locus. In cases where the inspected allele shows reduced penetrance and the phenotype resembles the translocation-associated phenotype, the genome should be sequenced in such a line.

Methodological approach to detecting chromosome translocations

With every T-DNA line, the possibility of extraneous mutations should be evaluated. Investigating T-DNA lines with or without a reproductive phenotype should be investigated used the methods described here. A practical guideline for the detection of and dealing with T-DNA-associated mutagenesis using methods listed here is presented in Fig. 3.

T-DNA mapping by PCR

Typically, the LB genome–T-DNA junction is used to map the genomic location of a T-DNA insertion and presented in the FST database (O'Malley *et al.*, 2015). The RB is frequently lost during T-DNA integration; however, the PCR-based genotyping using insertion- and locus-specific primers should be performed on a segregating population to assess both the LB and RB genome–T-DNA junction position and whether the junctions segregate together (O'Malley *et al.*, 2015). In combination with sequencing of the PCR products, it reveals potential mutations proximal to the T-DNA insertion site. When one of the borders cannot be mapped to the insertion site, a truncated border, multiple T-DNA inserts, a simple translocation, or other complex mutations is likely to be present, and sequencing of the genome is recommended in that case.

Reciprocal backcrosses

Reciprocal crosses between the wild type and the T-DNA line are used to check for T-DNA transmission bias through the male and female gametes. T-DNA lines carrying a translocation causing a gametophytic lethal phenotype do not show a transmission bias. Outcross analysis can also clear the T-DNA line of potential multiple insertions or cryptic untagged mutations. Although the presence of the latter cannot formally be checked in this manner, an outcross analysis is essential to check co-segregation of the T-DNA with the phenotype of interest.

Genetic complementation

Complementation of the observed phenotype with a construct carrying the wild-type allele of the mutated locus is a strong indicator that the phenotype is indeed caused by the intended T-DNA. However, complementation using T-DNA-mediated insertion of the complementing gene can cause additional mutations or *trans*-silencing of the T-DNA, requiring the analysis of multiple independent transformation events.

Gametophytic defects

Typical for translocation mutants is the reduced fertility of heterozygous lines that commonly correlates with fruit size. In Arabidopsis, this is readily observed as a decreased silique length. Dissection of the silique under a stereomicroscope shows a reduced number of seeds and empty seed slots. In other plant species, in addition to fruit size, one may also check the number of seeds per fruit. The stage of the gametophyte arrest can be assessed by clearing in Hoyer's solution in combination with differential interference contrast (DIC) microscopy (Weigel and Glazebrook, 2002).

Pollen grain viability is typically analysed by bright field microscopy after Alexander´s staining (Alexander, 1969). Observation of a large number of aborted pollen warrants further investigation. Young flower buds are then dissected and stained with DAPI (Vergne *et al.*, 1987) to assess microspore development.

A subset of publicly available T-DNA lines (for example about half of the SAIL lines) are in the *quartet* (*qrt*) mutant background, in which the microspores do not separate post-meiosis and are released in tetrads. In translocation heterozygotes in the *qrt* background, the gametophytic defects are manifested with a pollen abortion distribution corresponding to 4:0, 2:2, and 0:4 (normal:collapsed) in the tetrad (Curtis *et al.*, 2009).

Inspection of all genotype classes (heterozygous, homozygous, and wild-type segregants for the T-DNA) is recommended, as

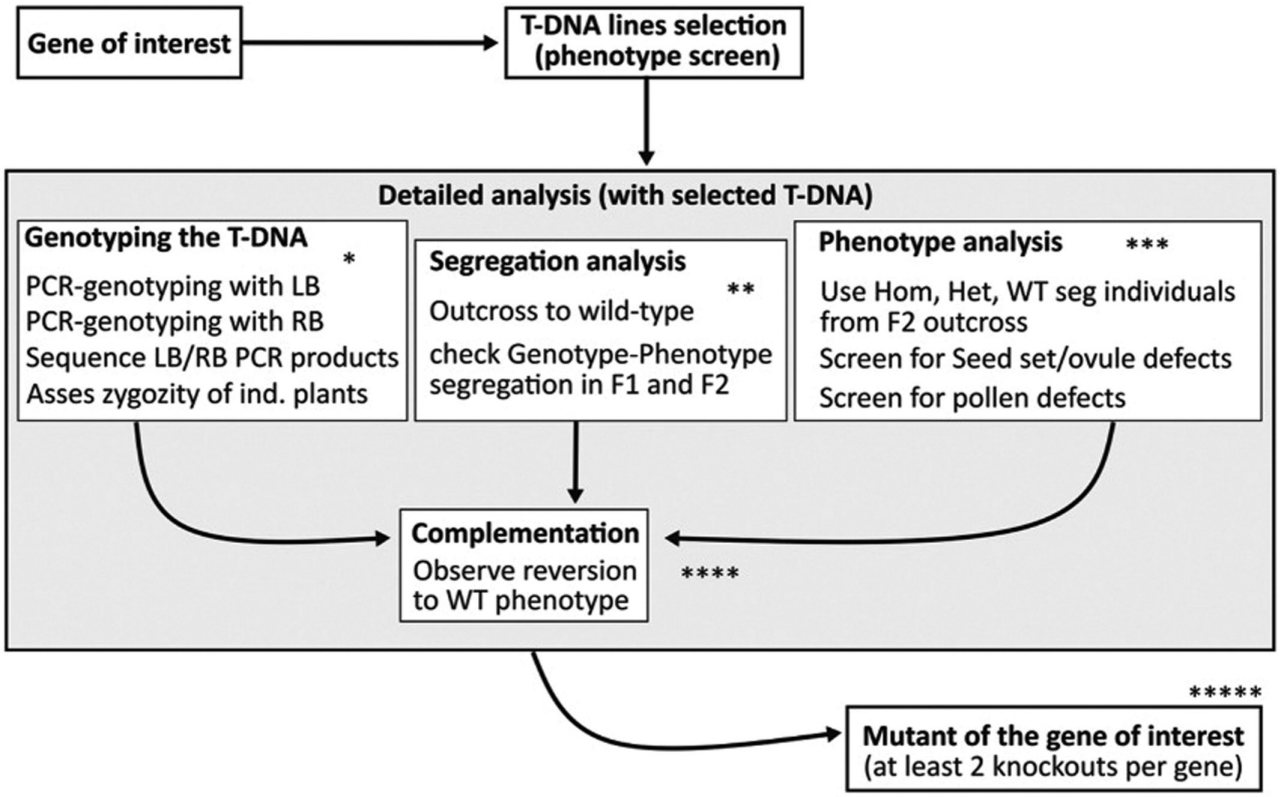

Fig. 3. Minimalistic guideline for novel T-DNA insertion lines. The most effective approach to T-DNA lines in reverse genetics is to order and investigate all available lines with an insertion mapped to the inspected locus. A typical first step is to select for mutants with an observable phenotype. Having such lines selected, a more detailed analysis of the mutant allele should be routinely performed. Here we present a minimalistic guideline to detect most if not all T-DNA-associated mutations that might lead to a misinterpretation of the phenotype being linked to your gene of interest. When encountering any results that suggest additional mutations, whole-genome sequencing should be performed to fully characterize the line.
*PCR-genotype multiple individuals grown from the seed stock (good practice is ~20 plants) to assess their zygosity and map the insertion. It is important to genotype and sequence both junctions from the LB and RB site and confirm that both borders map to your gene. The FST predictions could differ from your sequencing results. Any additional mutations at the insertion site (truncated T-DNA, plant genome deletion, or filler DNA) can be detected by this approach. Failure in obtaining a PCR product from the LB or RB junction might hint that there is some additional mutagenesis that needs further inspection (truncated, tandem duplications, or translocation). Consider testing more genomic or insertion-specific primers when the initial pair fails. Include PCRs with LB and RB primers and no genomic primer to see if the product is detected, increasing the possibility of tandem insertions. With continuous trouble, genome sequencing is recommended to fully characterize the locus.
**Perform an outcross of homozygous and/or heterozygous T-DNA mutants with the wild-type (WT) genotype to segregate any background mutations. Screening the genotype–phenotype link in the F_1 and F_2 outcross generations should reveal both dominant and recessive mutations. Carefully check whether there is no segregation of the T-DNA insertion and the phenotype. It is recommended to continue in the downstream analyses with the cleaned F_2 outcross lines. Analysing the progeny from a cross between a heterozygous plant and the WT will reveal the allele transmission rates.
***Plants of all zygotic states of the T-DNA (homozygous, heterozygous, and segregating WT) should be checked for gametophytic defects to exclude the possibility of translocations or other genomic imbalances that can arise in T-DNA lines. It is important to note that gametophytic defects are not exclusive to translocations and could truly be an effect of the disrupted gene. In that case, the homozygous plants usually display a more severe phenotype than heterozygous plants; the allele has reduced transmission and/or does not affect both gametophytes (depending, of course, on the nature of the disrupted gene). However, observing shorter siliques with empty seeds slots (up to 50%) together with a defect in both pollen (early abortion in up to 50%) and ovules could hint at translocation. The possibility of translocation increases when this phenotypical difference is observed in plants heterozygous but not homozygous for the T-DNA insertion. Importantly, in translocation lines, the transmission of the T-DNA allele through both the male and female gametes is not affected. When such a phenotype is observed, a quick check of meiotic spreads with FISH analysis can confirm aberrant chromosome interactions.
****Genetic complementation with a functional construct that rescues the mutant phenotype should confirm that the phenotypical differences were an effect of the disruption of the gene of interest. Complemented lines should not display the phenotype differences described in the T-DNA mutant line. Note that complementing a phenotype caused by translocation (translocation heterozygotes) could lead to selection of translocation homozygotes (no phenotype) that would be falsely interpreted as successfully complemented. Also, introducing the complementing expression cassette (a new T-DNA) in the mutant background can lead to novel mutagenesis; therefore, it is recommended to screen multiple independent transgenic lines. Additionally, it is best to avoid using sequences homologous to the T-DNA in the complementation transgene to avoid *trans*-inactivation.
*****With a line where you have excluded all possible T-DNA-associated mutations, you can proceed with further phenotype analysis. It is good practice to have at least two knockout T-DNA lines analysed per gene of interest. In situations where only one mutant allele is available, whole-genome sequencing should be done to exclude additional mutagenesis. Be aware of the *trans*-inactivation of other T-DNAs (mutant collections or custom-made expression cassettes) that you might plan to introduce into your T-DNA mutant background and try to avoid sequence homology that might cause unwanted silencing.

the defects in translocation lines only occur in heterozygotes. Around 50% abortion of both male and female gametes, in T-DNA heterozygous plants only, in combination with an arrest around the first mitosis, hints at a translocation-dependent phenotype.

Meiotic chromosome spreading

The meiotic recombination abnormalities typical for a translocation event can be observed histologically by means of meiotic chromosome spreads (Ross *et al*., 1997). During prophase I of meiosis, homologous chromosomes align and recombine. At the diakinesis and metaphase I stages, so-called bivalents or pairs of interacting homologous chromosomes are visualized by DAPI staining (Mercier *et al*., 2015; Zickler and Kleckner, 2015). The number of bivalents equals the haploid chromosome number of a species (n=5 in the case of Arabidopsis). The crossovers are crucial to the formation of bivalents and normally form only between homologous chromosomes. In translocation lines, the homology between homologous pairs of chromosomes is disrupted (Fig. 2). In the heterozygous state, crossovers will form between non-homologous chromosomes. Histologically this is manifested by the formation of so-called tetravalents (two interacting bivalents). The frequency at which tetravalents occur depends on the size of the translocated chromosome region. In the case of a translocation, the ectopic interaction that leads to a tetravalent always occurs between the same chromosomes. Therefore, it might be informative to assess chromosome identity by using fluorescent *in situ* hybridization (FISH) probes on top of the meiotic spreads (Curtis *et al*., 2009). In the homozygous state, translocated chromosomes can form unique pairs, giving rise to only bivalents which will segregate normally.

Genetic mapping and genome sequencing

Although the aforementioned phenotypical analyses point to the presence of a chromosome translocation, unambiguous identification of a chromosome rearrangement is obtained by genetic and sequence analyses. A simple genetic approach for the mapping of the T-DNA insertion is based on polymorphism markers (Clark and Krysan, 2010; Ruprecht *et al*., 2014). Polymorphism markers are identified by PCR in a segregating population using, for example, Arabidopsis accession Columbia-0 (containing the T-DNA and the putative translocation) and the outcrossing parent accession Landsberg *erecta*. This approach is based on the genetic linkage of the T-DNA and a set of genetic markers with known genomic location. In a regular transgenic line, the T-DNA is genetically linked to a single genomic position, corresponding to the position of the predicted insertion site. In a translocation line, linkage will be observed between the T-DNA and the predicted insertion site and additional markers on a second chromosome close to the translocation breakpoint. This approach has been shown to identify a translocation even if it does not confer a phenotype (Clark and Krysan, 2010). Consequently, finding a translocation based on genetic mapping in a plant that has a gametophytic phenotype does not necessarily mean the translocation is causal for the phenotype, and multiple lines of evidence are needed for linking the locus to the gametophytic lethal phenotype.

Highly complex translocation lines and sequence-level mutations induced by T-DNA insertion may require a more comprehensive investigation that can be achieved using long read sequencing technologies (Amarasinghe *et al*., 2020) such as PacBio (Rhoads and Au, 2015) and Oxford Nanopore sequencing (ONT) (Wang *et al*., 2021). This technology has been used successfully to investigate genomic rearrangements in T-DNA lines in Arabidopsis (Pucker *et al*., 2021) and birch (Gang *et al*., 2019).

Practical guideline for novel T-DNA insertion lines

The starting point of the characterization of T-DNA mutants is to order all available lines with insertions mapped to the gene of interest. Then a first round of selection is usually to screen for lines with a phenotype in the inspected traits. For lines with an interesting phenotype, a more detailed analysis of the mutant allele should be routinely performed. In Fig. 3 we present a guideline with a minimalistic approach to detect most T-DNA-associated mutagenesis that might lead to misinterpretations of the phenotype being linked to your gene of interest. The current best practice to avoid this misinterpretation is to include at least two alleles of the gene of interest with similar phenotypical differences and/or complement the mutant. On top of that, with whole-genome ONT sequencing becoming more affordable, the rising trend is to sequence all mutant lines of interest and evaluate the genomic assembly. However, as it might not be an option for all researchers for various reasons, the minimalistic guidelines should be followed with all T-DNA lines.

Conclusion

Agrobacterium-mediated transformation of plants is a powerful tool in plant biotechnology that is used with ease in a wide range of plant species. However, the erroneous integration of T-DNA warrants careful examination of every transgenic line under study. The recent boom in CRISPR/Cas9 mutagenesis and the fact that the Cas9-mediated cleavage mechanism intrinsically involves DSB formation calls for vigilance for the risk of chromosome rearrangement that can confound the phenotype–genotype link. Since the relatively high frequency of T-DNA-induced mutagenesis, including chromosomal rearrangements, is often overlooked, this report may help to eliminate the futile investigation of such mutants especially in a context of plant reproduction and gametogenesis studies.

Acknowledgements

KR is grateful to his supervisor Dr Christos Michailidis for encouragement and guidance, and to his mentor Dr Lukáš Fischer for helpful comments on parts considering epigenetics.

Author contributions

KR, LS, and CS: writing—original draft; KR, LS, CS, and DG: conceptualization; KR and LS: visualization; DH and DG: writing—review and editing, funding acquisition.

Conflict of interest

No conflict of interest declared.

Funding

KR and DH acknowledge support from the project TowArds Next GENeration Crops, reg. no. CZ.02.01.01/00/22_008/0004581 of the ERDF Programme Johannes Amos Comenius and the project 'Grant Schemes at CU' (reg. no. CZ.02.2.69/0.0/0.0/19_073/0016935). KR was funded by CSF/GACR grant no. 21-15856S and DH received funding from CSF/GACR grant no. 23-07000S. CS received funding from the Research Foundation – Flanders (FWO; grant no. 11G7421N). LS was granted funding by the China Scholarship Council (CSC; grant no. 201806350263).

References

Alexander MP. 1969. Differential staining of aborted and nonaborted pollen. Stain Technology **44**, 117–122.

Alling R, Galindo-Trigo S. 2023. Reproductive defects in the abscission mutant *ida-2* are caused by T-DNA-induced genomic rearrangements. Plant Physiology **193**, 2292–2297.

Alonso JM, Ecker JR. 2006. Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. Nature Reviews. Genetics **7**, 524–536.

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biology **21**, 30.

An G, Lee S, Kim S-H, Kim S-R. 2005. Molecular genetics using T-DNA in rice. Plant and Cell Physiology **46**, 14–22.

Beying N, Schmidt C, Pacher M, Houben A, Puchta H. 2020. CRISPR–Cas9-mediated induction of heritable chromosomal translocations in *Arabidopsis*. Nature Plants **6**, 638–645.

Castle LA, Errampalli D, Atherton TL, Franzmann LH, Yoon ES, Meinke DW. 1993. Genetic and molecular characterization of embryonic mutants identified following seed transformation in *Arabidopsis*. Molecular & General Genetics **241**, 504–514.

Clark KA, Krysan PJ. 2010. Chromosomal translocations are a common phenomenon in *Arabidopsis thaliana* T-DNA insertion lines. The Plant Journal **64**, 990–1001.

Curtis MJ, Belcram K, Bollmann SR, Tominey CM, Hoffman PD, Mercier R, Hays JB. 2009. Reciprocal chromosome translocation associated with TDNA-insertion mutation in *Arabidopsis*: genetic and cytological analyses of consequences for gametophyte development and for construction of doubly mutant lines. Planta **229**, 731–745.

Daxinger L, Hunter B, Sheikh M, Jauvion V, Gasciolli V, Vaucheret H, Matzke M, Furner I. 2008. Unexpected silencing effects from T-DNA tags in *Arabidopsis*. Trends in Plant Science **13**, 4–6.

De Buck S, Jacobs A, Van Montagu M, Depicker A. 1999. The DNA sequences of T-DNA junctions suggest that complex T-DNA loci are formed by a recombination process resembling T-DNA integration. The Plant Journal **20**, 295–304.

Forsbach A, Schubert D, Lechtenberg B, Gils M, Schmidt R. 2003. A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome. Plant Molecular Biology **52**, 161–176.

Gang H, Liu G, Zhang M, Zhao Y, Jiang J, Chen S. 2019. Comprehensive characterization of T-DNA integration induced chromosomal rearrangement in a birch T-DNA mutant. BMC Genomics **20**, 311.

Gao Y, Zhao Y. 2013. Epigenetic suppression of T-DNA insertion mutants in *Arabidopsis*. Molecular Plant **6**, 539–545.

Gelvin SB. 2021. Plant DNA repair and *Agrobacterium* T-DNA integration. International Journal of Molecular Sciences **22**, 8458.

Hafidh S, Honys D. 2021. Reproduction multitasking: the male gametophyte. Annual Review of Plant Biology **72**, 581–614.

Hu Y, Chen Z, Zhuang C, Huang J. 2017. Cascade of chromosomal rearrangements caused by a heterogeneous T-DNA integration supports the double-stranded break repair model for T-DNA integration. The Plant Journal **90**, 954–965.

Huep G, Kleinboelting N, Weisshaar B. 2014. An easy-to-use primer design tool to address paralogous loci and T-DNA insertion sites in the genome of *Arabidopsis thaliana*. Plant Methods **10**, 28.

Jiang W, Li Z, Yao X, Zheng B, Shen W-H, Dong A. 2019. *jaw-1D*: a gain-of-function mutation responsive to paramutation-like induction of epigenetic silencing. Journal of Experimental Botany **70**, 459–468.

Jupe F, Rivkin AC, Michael TP, *et al*. 2019. The complex architecture and epigenomic impact of plant T-DNA insertions. PLoS Genetics **15**, e1007819.

Kleinboelting N, Huep G, Appelhagen I, Viehoever 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. Molecular Plant **8**, 1651–1664.

Lafleuriel J, Degroote F, Depeiges A, Picard G. 2004. A reciprocal translocation, induced by a canonical integration of a single T-DNA, interrupts the HMG-I/Y *Arabidopsis thaliana* gene. Plant Physiology and Biochemistry **42**, 171–179.

Mercier R, Mézard C, Jenczewski E, Macaisne N, Grelon M. 2015. The molecular biology of meiosis in plants. Annual Review of Plant Biology **66**, 297–327.

Min Y, Frost JM, Choi Y. 2020. Gametophytic abortion in heterozygotes but not in homozygotes: implied chromosome rearrangement during T-DNA insertion at the ASF1 locus in *Arabidopsis*. Molecules and Cells **43**, 448–458.

Nacry P, Camilleri C, Courtial B, Caboche M, Bouchez D. 1998. Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. Genetics **149**, 641–650.

Nath Radhamony R, Mohan Prasad A, Srinivasan R. 2005. T-DNA insertional mutagenesis in *Arabidopsis*: a tool for functional genomics. Electronic Journal of Biotechnology **8**.

Ohba T, Yoshioka Y, Machida C, Machida Y. 1995. DNA rearrangement associated with the integration of T-DNA in tobacco: an example for multiple duplications of DNA around the integration target. The Plant Journal **7**, 157–164.

O'Malley RC, Barragan CC, Ecker JR. 2015. A user's guide to the *Arabidopsis* T-DNA insertion mutant collections. Methods in Molecular Biology **1284**, 323–342.

Osabe K, Harukawa Y, Miura S, Saze H. 2017. Epigenetic regulation of intronic transgenes in *Arabidopsis*. Scientific Reports **7**, 45166.

Pucker B, Kleinbölting N, Weisshaar B. 2021. Large scale genomic rearrangements in selected *Arabidopsis thaliana* T-DNA lines are caused by T-DNA insertion mutagenesis. BMC Genomics **22**, 599.

Ray SM, Park SS, Ray A. 1997. Pollen tube guidance by the female gametophyte. Development **124**, 2489–2498.

Rhoads A, Au KF. 2015. PacBio sequencing and its applications. Genomics, Proteomics & Bioinformatics **13**, 278–289.

Ross KJ, Fransz P, Armstrong SJ, Vizir I, Mulligan B, Franklin FC, Jones GH. 1997. Cytological characterization of four meiotic mutants of *Arabidopsis* isolated from T-DNA-transformed lines. Chromosome Research **5**, 551–559.

Ruprecht C, Carroll A, Persson S. 2014. T-DNA-induced chromosomal translocations in *feronia* and *anxur2* mutants reveal implications for the mechanism of collapsed pollen due to chromosomal rearrangements. Molecular Plant **7**, 1591–1594.

Sandhu KS, Koirala PS, Neff MM. 2013. The *ben1-1* brassinosteroid-catabolism mutation is unstable due to epigenetic modifications of the intronic T-DNA insertion. G3 **3**, 1587–1595.

Schouten HJ, Vande Geest H, Papadimitriou S, Bemer M, Schaart JG, Smulders MJM, Perez GS, Schijlen E. 2017. Re-sequencing transgenic plants revealed rearrangements at T-DNA inserts, and integration of a short T-DNA fragment, but no increase of small mutations elsewhere. Plant Cell Reports **36**, 493–504.

Singer K. 2018. The mechanism of T-DNA integration: some major unresolved questions. Current Topics in Microbiology and Immunology **418**, 287–317.

Skinner DJ, Sundaresan V. 2018. Recent advances in understanding female gametophyte development. F1000Research **7**, F1000 Faculty Rev-804.

Sun L, Ge Y, Sparks JA, Robinson ZT, Cheng X, Wen J, Blancaflor EB. 2019. TDNAscan: a software to identify complete and truncated T-DNA insertions. Frontiers in Genetics **10**, 685.

Takano M, Egawa H, Ikeda JE, Wakasa K. 1997. The structures of integration sites in transgenic rice. The Plant Journal **11**, 353–361.

Tax FE, Vernon DM. 2001. T-DNA-associated duplication/translocations in *Arabidopsis*. Implications for mutant analysis and functional genomics. Plant Physiology **126**, 1527–1538.

Thomson G, Dickinson L, Jacob Y. 2024. Genomic consequences associated with *Agrobacterium*-mediated transformation of plants. The Plant Journal **117**, 342–363.

Ulker B, Peiter E, Dixon DP, Moffat C, Capper R, Bouché N, Edwards R, Sanders D, Knight H, Knight MR. 2008. Getting the most out of publicly available T-DNA insertion lines. The Plant Journal **56**, 665–677.

Vergne P, Delvallee I, Dumas C. 1987. Rapid assessment of microspore and pollen development stage in wheat and maize using DAPI and membrane permeabilization. Stain Technology **62**, 299–304.

Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. 2021. Nanopore sequencing technology, bioinformatics and applications. Nature Biotechnology **39**, 1348–1365.

Weigel D, Glazebrook J. 2002. Arabidopsis: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Wu L, Di D-W, Zhang D, Song B, Luo P, Guo G-Q. 2015. Frequent problems and their resolutions by using thermal asymmetric interlaced PCR (TAIL-PCR) to clone genes in *Arabidopsis* T-DNA tagged mutants. Biotechnology, Biotechnological Equipment **29**, 260–267.

Xue W, Ruprecht C, Street N, Hematy K, Chang C, Frommer WB, Persson S, Niittylä T. 2012. Paramutation-like interaction of T-DNA loci in *Arabidopsis*. PLoS One **7**, e51651.

Yuen CYL, Sedbrook JC, Perrin RM, Carroll KL, Masson PH. 2005. Loss-of-function mutations of ROOT HAIR DEFECTIVE3 suppress root waving, skewing, and epidermal cell file rotation in *Arabidopsis*. Plant Physiology **138**, 701–714.

Zickler D, Kleckner N. 2015. Recombination, pairing, and synapsis of homologs during meiosis. Cold Spring Harbor Perspectives in Biology **7**, a016626.