The Mechanism of T-DNA Integration: Some Major Unresolved Questions

Kamy Singer

Contents

Abstract The mechanism of T-DNA integration into plant genomes during *Agrobacterium*-mediated genetic transformation is still not understood. As genetic transformation of plants via *Agrobacterium* has become a routine practice among plant biologists, understanding T-DNA integration remains important for several reasons. First, T-DNA is the final step in one of the unique cases of inter-kingdom horizontal gene transfer in nature. Second, understanding T-DNA integration is important for biotechnological applications. For example, better knowledge of this process may help develop methods to transform species that are currently not susceptible to *Agrobacterium*-mediated transformation. In addition, regulatory agencies usually require "clean" and "precise" transgenic insertion events, whereas

K. Singer
Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA

K. Singer (✉)
Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA
e-mail: kamys@umich.edu

transgenic insertions are commonly complex unpredictable structures. Furthermore, whereas T-DNA integration under natural conditions occurs randomly, technology to direct T-DNA to specific sites in the genome is highly desired. A better understanding of T-DNA integration may help develop methods to achieve more desirable results. Finally, gene targeting methods that require a foreign DNA template for precise DNA modifications in plants often utilize *Agrobacterium* to deliver the DNA template. Better understanding of the fate of T-DNA in the plant nucleus may help utilize T-DNA for more efficient gene targeting. For introducing gene targeting reagents, efficient delivery of T-DNA without ectopic integration would be useful. The following review summarizes current knowledge related to T-DNA integration. Five major open questions related to T-DNA integration are being presented. Finally, different models for T-DNA integration are being discussed, and a revised model is proposed.

1   Introduction

Agrobacterium tumefaciens is a soil-borne bacterium well known for its unique ability of inter-kingdom horizontal gene transfer. In nature, this plant pathogen causes the disease crown gall (Smith and Townsend 1907). The disease is characterized by galls appearing at the plant's root, stem, or crown area. These galls are tumor growths that form as a result of a transfer of a region of the *Agrobacterium* tumor-inducing (Ti) plasmid into a plant cell and stable integration of this DNA into the plant genome (Zaenen et al. 1974; Chilton et al. 1977; Fig. 1). The transferred DNA (T-DNA) of the Ti plasmid carries genes that cause uncontrolled cell divisions by modifying the plant's hormonal balance, whereas other genes encode proteins involved in the production of opines, compounds that are utilized by the *Agrobacterium* colonies surrounding the galls.
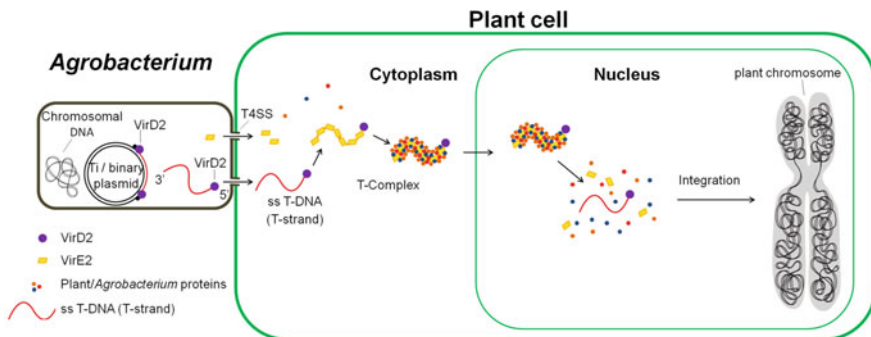

Fig. 1 Schematic illustration of a model describing DNA transfer from *Agrobacterium* to the plant cell (see relevant book chapters for further details). Adapted from Singer (2013) doctoral dissertation

Agrobacterium has been extensively studied because of its role as a "natural genetic engineer." Moreover, *Agrobacterium* has been harnessed by humans as a gene vector to genetically engineer plants. *Agrobacterium*-mediated transformation was the first method to generate transgenic plants (Barton et al. 1983; Zambryski et al. 1983). Three decades later, this bacterium is still a key player in many of the plant molecular genetics techniques used in agricultural biotechnologies (reviewed in Shiboleth and Tzfira 2012; Altpeter et al. 2016).

Whereas early events during *Agrobacterium* infection are relatively well studied, the final events in the plant nucleus are relatively less understood. Notably, the mechanism behind T-DNA integration into the plant genome is still unknown. Consequently, investigators often adopt different models to explain T-DNA integration. For instance, the process of T-DNA integration resulting in chromosome truncation has been explained in two different ways recently. In Teo et al. (2011), T-DNA is a single-stranded (ss) molecule during integration, whereas in Nelson et al. (2011) T-DNA is a double-stranded (ds) molecule. Moreover, according to the model in Teo et al. (2011), bacterial VirD2 is involved in integration, whereas according to Nelson et al. (2011), integration is mediated by plant host proteins. Thus, the literature may present conflicting models for T-DNA integration which are based on three decades of research evidence leading to somewhat conflicting conclusions. This review describes the various open questions that contribute to current models of T-DNA integration.

1.1 The Transfer of T-DNA

T-DNA integration requires the transfer of T-DNA from *Agrobacterium* into the plant cells. T-DNA transfer is a process related to bacterial conjugation (for review, see Lessl and Lanka 1994; Christie et al. 2005). Transfer begins inside the bacterium when T-DNA is separated from its parent plasmid, a Ti plasmid in natural strains or a binary plasmid in laboratory strains. The separation of T-DNA is initiated when a protein complex of VirD1, a helicase, and VirD2, an endonuclease, attaches to the left border (LB) and right border (RB) of a T-DNA region (Durrenberger et al. 1989; Scheiffele et al. 1995; Relic et al. 1998) (Fig. 2). The LB and RB are 25 base pairs (bp) of imperfect direct repeats (Yadav et al. 1982). VirD2 nicks the lower DNA strand between the third and fourth nucleotides of each of these repeats (Fig. 2) (Yanofsky et al. 1986; Wang et al. 1987). Consequently, a single-stranded (ss) T-DNA, termed the T-strand, is generated from the parent plasmid (Albright et al. 1987). A single VirD2 protein remains covalently attached to the 5′ end of the ss T-DNA (the RB side) (Herrera-Estrella et al. 1988; Ward and Barnes 1988; Young and Nester 1988; Vogel and Das 1992) (Fig. 2).

VirD2 protein pilots the ss T-DNA from its 5′ end through the *Agrobacterium* type IV protein secretion (T4S) system into the plant cytoplasm (Vergunst et al. 2005; van Kregten et al. 2009) (Fig. 1). In the plant cytoplasm, before entering the nucleus, the transported ss T-DNA is thought to be coated by multiple VirE2 ss
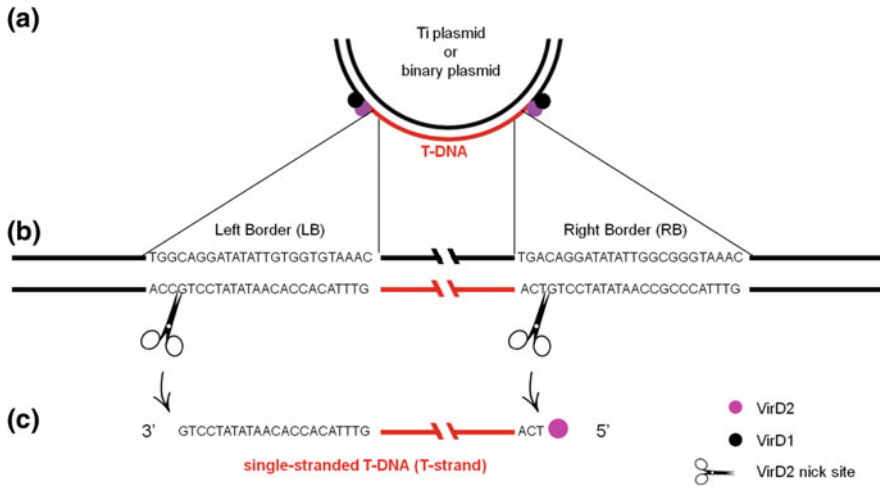
Fig. 2 Schematic illustration of T-DNA processing in *Agrobacterium*. **a** Only part of the plasmid (Ti or binary) is illustrated in the figure. The T-DNA region in the plasmid is marked in red (the lower DNA strand is processed and transferred). VirD1 (marked in black circle) and VirD2 (marked in purple circle) bind the left border (LB) and right border (RB) of the T-DNA region. **b** VirD2 nicks between the third and fourth nucleotide of each border (the 25 base pair DNA sequence of the LB and RB is illustrated; the nicking site is indicated by the scissors). **c** After the single-stranded T-DNA is separated from the parent plasmid, VirD2 protein remains attached to the 5′ end of the T-DNA (the RB of the T-DNA). Adapted from Singer (2013) doctoral dissertation

DNA binding proteins (Citovsky et al. 1988; Das 1988; Abu-Arish et al. 2004), which are secreted into the plant cell from *Agrobacterium* independent of T-DNA (Otten et al. 1984; Citovsky et al. 1992; Binns et al. 1995; Li and Pan 2017). In addition, it is believed that other bacterial and plant proteins interact with ViE2 and VirD2 in the formation of a "T-complex" (for reviews, see Lacroix et al. 2006; Gelvin 2010). The proposed role of the T-complex is to protect the ss T-DNA from degradation (Durrenberger et al. 1989; Tinland et al. 1995; Rossi et al. 1996) and to facilitate its transport into the nucleus. There are a number of facilitators of T-complex nuclear transport. VirD2 facilitates nuclear transport by its nuclear localization signal (NLS) domain (Herrera-Estrella et al. 1990; Shurvinton et al. 1992; Howard et al. 1992; Tinland et al. 1992; Ziemienowicz et al. 2001; van Kregten et al. 2009). Similarly, VirE2 may facilitate T-complex transport via its two NLS domains (Citovsky et al. 1992, 1994; Zupan et al. 1996; Ziemienowicz et al. 2001). However, a number of groups have reported that VirE2 remains mostly in the cytoplasm (Bhattacharjee et al. 2008; Grange et al. 2008; Lee et al. 2008; Sakalis et al. 2014; Shi et al. 2014). Another proposed facilitator of nuclear transport of the T-complex is the VirE2-interacting protein 1 (VIP1), a plant transcription factor that enters the nucleus upon activation of the defense response (Tzfira et al. 2001; Li et al. 2005a; Djamei et al. 2007; Pitzschke et al. 2009; Wang et al. 2017). VirE3, which binds VirE2 in the plant cytoplasm, may substitute for VIP1 to facilitate nuclear transport (Lacroix et al. 2005). Recently, however,

Shi et al. (2014) concluded that VIP1 is not important for *Agrobacterium*-mediated transformation. Finally, nuclear transport of the T-complex may be facilitated by additional host nuclear transporters that interact with T-complex components (Ballas and Citovsky 1997; Lacroix et al. 2005; Bako et al. 2003; Bhattacharjee et al. 2008).

1.2 Stable and Transient T-DNA in the Plant Cell

In the plant nucleus, T-complex proteins must be stripped off the ss T-DNA before integration. It has been hypothesized that stripping off these proteins from the T-complex is mediated by VirF and the host proteasomal degradation machinery (Schrammeijer et al. 2001; Tzfira et al. 2004a, b; Zaltsman et al. 2013). It is known that several T-DNA molecules can enter the plant cell simultaneously (Virts and Gelvin 1985). Whereas the number is unknown and likely varies under different conditions, it has been shown that the percentage of T-DNA molecules in the nucleus that eventually integrate into the plant genome is relatively low (Narasimhulu et al. 1996; Maximova et al. 1998; De Buck et al. 2000; Ghedira et al. 2013). Integration of T-DNA into the plant genome results in stable genetic transformation, whereas T-DNA that does not integrate into the plant genome results in transient genetic transformation.

Stable genetic transformation by *Agrobacterium*-mediated genetic transformation is the preferred method used by plant biologists to generate transgenic plants. Commonly, the desired DNA sequence is cloned between T-DNA borders. Consequently, after T-DNA integrates into the plant genome, a transgenic plant is produced. The *Agrobacterium* strains used for biotechnological applications are themselves genetically modified. This modification includes removal of the natural tumor-inducing genes from T-DNA so that the strains become "disarmed." However, the ability of disarmed strains to transfer a modified T-DNA is unaffected because the only elements on the T-DNA that are necessary for T-DNA transfer are the T-DNA left border (LB) and right border (RB) (Fig. 2) (Hoekema et al. 1983; Ream et al. 1983; Wang et al. 1984). Moreover, in order to make genetic engineering simpler, T-DNA is often placed on a smaller binary plasmid instead of the natural Ti plasmid because the former is easier to work with and can replicate in *E. coli* as well as in *Agrobacterium* (Hoekema et al. 1983; for review, see Tzfira and Citovsky 2006). In addition to introducing desired genes into the plant genome, T-DNA has also been instrumental for the creation of large mutant and enhancer trap libraries because integration occurs randomly in the plant genome. These T-DNA insertion collections have been especially important for studies of *Arabidopsis* and rice (Sessions et al. 2002; Sallaud et al. 2004; O'Malley and Ecker 2010).

Transient expression of foreign genes is another method often used for manipulating plant genomes. Gene expression of those T-DNA molecules lasts for a few days before the genes are silenced (Johansen and Carrington 2001). Manipulating plant genomes can be achieved by transient expression of engineered nucleases

such as meganucleases, ZFNs, TALENs, and Cas9. These nucleases have been used to target specific genomic sites and create double-strand breaks (DSBs) required for gene editing (for review, see Kumar and Jain 2015; Yin et al. 2017). In addition to its use for genomic modifications, *Agrobacterium*-mediated transient expression is an important investigative tool for plant biologists. For example, transient expression is commonly used to investigate cellular localization of proteins or to produce and isolate proteins *in planta* (Sparkes et al. 2006). Recently, transient expression by *Agrobacterium*-mediated transformation has been applied commercially using plants as factories for products such as vaccines and antibodies (for review, see Ko et al. 2009; Komarova et al. 2010).

In nature, T-DNA integration occurs during *Agrobacterium* infection of certain dicotyledonous plants and gymnosperms. However, under laboratory conditions, scientists have harnessed *Agrobacterium* to transform an increasing variety of plants, including monocotyledonous plant species. *Agrobacterium*-mediated transformation has also been successfully applied for transformation of non-plant eukaryotes (for review, see Soltani et al. 2010), such as yeast (Bundock et al. 1995; Bundock and Hooykaas 1996; Rolloos et al. 2014, 2015; Ohmine et al. 2016) and other fungi (de Groot et al. 1998; Korn et al. 2015), as well as for human cells (Kunik et al. 2001). Whereas studying T-DNA integration in non-plant organisms may contribute to understanding T-DNA integration in plants, T-DNA integration in non-plant organisms may involve mechanisms and enzymatic pathways that differ from that of T-DNA integration into plants. Therefore, the following review focuses on T-DNA integration in plants.

2   The Mechanism of T-DNA Integration

The evidence for much of our understanding of T-DNA integration has been facilitated by post-integration sequence analysis of T-DNA/plant genome junctions. This approach has been important for the development of the early models of T-DNA integration because it revealed the general patterns of T-DNA insertions (Mayerhofer et al. 1991; Gheysen et al. 1991). One of the earliest observations was that T-DNA integrates at random locations in the genome (Chyi et al. 1986; Gheysen et al. 1987). This topic is further discussed under the section "*Which is the genomic site prerequisite for T-DNA integration*".

DNA sequencing of the junctions between the integrated T-DNA and the surrounding plant genome also revealed that no homology, or only a few homologous nucleotides (nt) at the junction point, existed (Mayerhofer et al. 1991; Gheysen et al. 1991). Therefore, it was evident that homologous recombination is normally not involved in T-DNA integration, and the terms "illegitimate" recombination (IR) and nonhomologous recombination (NHR) have been used to describe T-DNA integration in plants (Mayerhofer et al. 1991; Gheysen et al. 1991; Bleuyard et al. 2006). More recently, the nonhomologous end-joining (NHEJ) DNA repair pathway of plants is frequently mentioned as the likely pathway responsible for T-DNA

integration. The NHEJ pathway is typically associated with a DNA repair pathway which is responsible for end joining between double-stranded DNA ends such as those present at genomic double-strand breaks (DSBs). Therefore, the NHEJ pathway may not describe well a model involving a single-stranded T-DNA intermediate. "*Does a T-DNA integrate into the plant genome as a single- or a double-stranded intermediate*" is discussed later on. It should be noted that the NHEJ repair pathway is usually associated with key enzymatic components such as the Ku70/Ku80 heterodimer (Critchlow and Jackson 1998). Nevertheless, recent studies have revealed the existence of additional DSBs repair pathways, often described as alternative NHEJ (A-NHEJ) or microhomology-mediated end joining (MMEJ), employing different enzymatic pathways and mechanisms (for review, see McVey and Lee 2008; Bleuyard et al. 2006). Recently, van Kregten et al. (2016) showed that DNA polymerase theta (pol θ) has an important role in T-DNA integration. This topic is further discussed under the section "*What are the bacterial and plant factors involved in T-DNA integration?*"

T-DNA integration is neither a "precise" nor a "clean" process (e.g., Kwok et al. 1985; Spielmann and Simpson 1986). It is not precise because T-DNA seldom preserves its two borders after integration in plants. It is not clean because insertions often include other DNA sequences from *Agrobacterium*. Commonly, the extra DNA sequences are derived from the parent plasmid (Ti or binary; Martineau et al. 1994; Kononov et al. 1997) but may also include DNA from unknown and known sources such as *Agrobacterium* chromosomal DNA (Ulker et al. 2008; Kleinboelting et al. 2015) and plant DNA (Kleinboelting et al. 2015). In addition, it is common for insertion sites to include two or more T-DNA molecules adjacent to each other (Cluster et al. 1996; Krizkova and Hrouda 1998; De Buck et al. 2000). The integration patterns differ under different experimental conditions and plant species (Grevelding et al. 1993; De Buck et al. 2009). In addition, transformed plants may contain more than a single T-DNA insertion site; each insertion may contain a single copy of T-DNA, or a cluster of T-DNA copies (e.g., Alonso et al. 2003; Rosso et al. 2003). Finally, major chromosomal aberrations may result from T-DNA integration (Nacry et al. 1998; Tax and Vernon 2001; Clark and Krysan 2010). The topic of "*Why and how do complex T-DNA insertions form?*" is integral to the question of T-DNA integration.

Finally, T-DNA integration models that include depictions of T-DNA vary in different publications. Often T-DNA is depicted as a straight line. In addition, the timing at which T-DNA ends interact with the target genome varies between the different models (for recent reviews of T-DNA integration, see Ziemienowicz et al. 2010; Windels et al. 2010; Magori and Citovsky 2011; Gelvin 2017). This topic is further discussed under the section "*What is the spatial and temporal arrangement of T-DNA during integration?*"

The following review attempts to explain the bigger question of T-DNA integration by presenting smaller questions. This has been done for the purpose of discussion. However, it should be emphasized that T-DNA integration may be mediated by different pathways under different conditions, and perhaps simultaneously. It is most likely a complex and multistep process. As such, the different

questions related to T-DNA integration are interrelated and may have more than a single answer. Correspondingly, whereas several major models have been proposed (reviewed in Tzfira et al. 2004a, b), when adopting different assumptions for different open questions, more models are possible.

3 The Major Unresolved Questions Related to the Mechanism of T-DNA Integration

3.1 Does T-DNA Integrate into the Plant Genome as a Single- or a Double-Stranded Intermediate?

T-DNA enters the plant nucleus as a single-stranded (ss) DNA but it is ultimately a double-stranded (ds) DNA when it becomes part of the host genome. However, without being able to visualize the integration process as it occurs, it is difficult to determine the timing of conversion from ss T-DNA to ds T-DNA. Mayerhofer et al. (1991) and Gheysen et al. (1991) discussed this question when proposing models for T-DNA integration via a mechanism of illegitimate recombination. According to the proposed ds T-DNA integration model, conversion from ss to ds T-DNA occurs extrachromosomally. Therefore, when T-DNA begins integration into the plant's genome, it is already a ds T-DNA intermediate (Mayerhofer et al. 1991). On the other hand, according to the proposed ss T-DNA integration model, the integration process begins with an ss T-DNA intermediate and the conversion to ds T-DNA happens during integration (Mayerhofer et al. 1991; Gheysen et al. 1991).

The ss T-DNA integration model was refined by Tinland et al. (1995) and became widely accepted soon thereafter (Tinland et al. 1995; Tinland 1996). According to this model, integration begins when the LB side of the ss T-DNA (the 3′ end) anneals to homologous sequences in the plant DNA, possibly by invading A-T-rich regions of melted chromosomal DNA (Brunaud et al. 2002). This annealing through homology may not include parts of the sequences at the 3′ distal end of the LB side, resulting in the loss of some of the 3′ side of the T-DNA due to exonuclease degradation. Next, the RB side of the ss T-DNA (the 5′ end) ligates to the 3′ end of the plant DNA. Unlike the 3′ end of the LB side, the RB is protected from exonuclease degradation by VirD2. VirD2 may also be involved in ligation of the 5′ ss T-DNA end to a 3′ end of the plant DNA. Several observations from different early studies support the ss T-DNA integration model: (a) T-DNA enters the nucleus as ss DNA molecule. Moreover, extrachromosomal recombination assays suggested that the T-DNA derivatives inside the plant nucleus are mainly ss T-DNA molecules (Tinland et al. 1994; Yusibov et al. 1994); (b) When ss DNA was introduced into plant protoplasts, the integration rate was comparable to (Furner et al. 1989) or higher than (Rodenburg et al. 1989) that of ds DNA; (c) The deletions of T-DNA post-integration are usually more severe at the LB side in comparison to the RB side (e.g., Tinland 1996; Kumar and Fladung 2002; Kim

et al. 2003; Zhang et al. 2008); d) The junctions between the T-DNA LB side and plant DNA after integration have been shown to contain higher microhomology levels compared to junctions involving the RB side of the T-DNA (e.g., Matsumoto et al. 1990; Tinland et al. 1995; Brunaud et al. 2002; Kim et al. 2003; Zhu et al. 2006; Thomas and Jones 2007); (e) There is evidence, although inconclusive, that VirD2 is involved in T-DNA integration (Pansegrau et al. 1993; Scheiffele et al. 1995; Mysore et al. 1998).

It should be noted that the ss T-DNA integration model could, in principle, apply to a T-DNA with ss DNA overhangs and a ds DNA internal body (Gheysen et al. 1991), although usually an ss T-DNA intermediate has been assumed (Gheysen et al. 1991; Tinland 1996; Brunaud et al. 2002; Meza et al. 2002). Moreover, the role of VirD2 in integration is inconclusive, and therefore the possible T-DNA intermediates depicted in Fig. 3 are shown with and without VirD2 attached to the 5′ end of the RB (Fig. 3).

The T-DNA integration model involving a ds intermediate is supported by evidence that T-DNA integration is linked to the repair of genomic DNA double-strand breaks (DSBs). The first evidence came from the observation that, when genomic DSBs are induced in protoplasts by X-ray irradiation, integration of foreign plasmid DNA is in enhanced (Kohler et al. 1989). Salomon and Puchta (1998) showed that when genomic site-specific DSBs are induced by *Agrobacterium*-mediated transient expression of the homing endonuclease I-*Sce*I, DSBs are often repaired with a T-DNA captured within the repaired break.
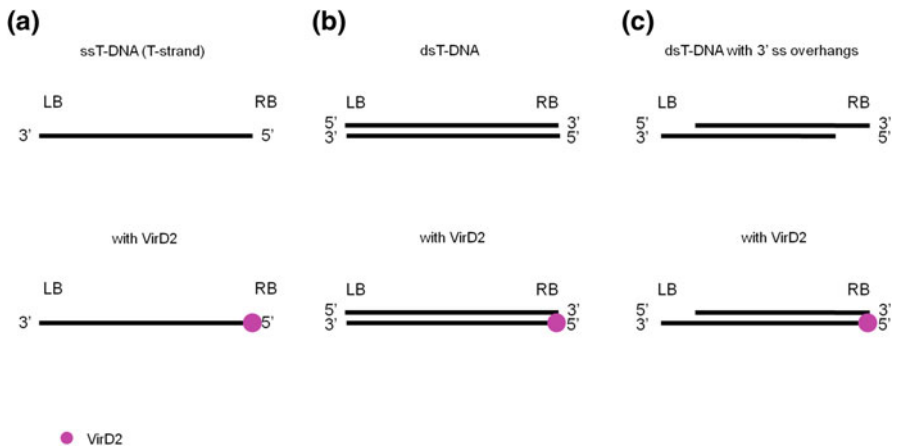

Fig. 3 Possible configurations of the T-DNA integration intermediate. **a** Single-stranded (ss) T-DNA (also termed T-strand). The 5′ end is always the RB side, while the 3′ end is always the LB side. Illustration underneath demonstrates VirD2 (in purple) attached to the 5′ end. **b** Double-stranded (ds) T-DNA with blunt ends. Illustration underneath demonstrates VirD2 (in purple) attached to the 5′ end of the RB side. **c** Double-stranded (ds) T-DNA internal body with 3′ single-stranded overhangs. Illustration underneath demonstrates VirD2 (in purple) attached to the 5′ end of the RB side and protecting the end from possible resection. Adapted from Singer (2013) doctoral dissertation

Moreover, early studies involving the sequencing of the junctions between T-DNA and plant DNA (e.g., Gheysen et al. 1991; Mayerhofer et al. 1991; Ohba et al. 1995; Takano et al. 1997) revealed patterns similar to those found in later studies of the mechanisms of DNA DSBs repair in plants (Gorbunova and Levy 1997; Salomon and Puchta 1998). The patterns of DSB repair in plants exhibited the characteristics of illegitimate/nonhomologous recombination. These included DNA deletions close to the breaks, repeated sequences or DNA from an unknown source ("filler" DNA), and little or no homology between DNA sequences forming the junctions. The notion that T-DNA integrates at genomic DSBs favors a model of a ds T-DNA as an intermediate because repair of DSBs involves end joining between two ds DNA ends. Moreover, Tzfira et al. (2003) and Chilton and Que (2003) provided evidence that T-DNAs captured at genomic DSBs were already ds intermediates prior to integration.

Evidence supporting the ds T-DNA model is also derived from the common formation of complex T-DNA insertions, in particular, complex insertions that include two T-DNAs ligated at their LB–LB sides or RB–RB sides without any microhomology within the ligated junction. The reason that this arrangement is difficult to explain via an ss model is that direct LB–LB end joining ("tail-to-tail" ligation) or RB–RB end joining ("head-to-head" ligation) cannot occur directly between the transferred ss T-DNA because there are always 3′ ends at the LB side and 5′ ends at the RB side (Gheysen et al. 1991; Mayerhofer et al. 1991; De Neve et al. 1997). In addition, extrachromosomal double-stranded circular T-DNA structures (T-circles) from *Agrobacterium*-infected plants have been isolated (Singer et al. 2012). By analyzing the DNA sequences of the extrachromosomal structures, it was found that the DNA junctions within the structures show the characteristic patterns of repaired DSBs. Importantly, it was possible to study the complete structure of the molecules (a feat more difficult to achieve in a genomic background when complex DNA repeats are involved). The structures included configurations such as multiple T-DNA copies arranged adjacent to each other or binary vector fragments attached to T-DNA sequences. Such structures are common post-T-DNA integration in transgenic plants. For example, according to different reports (e.g., Castle et al. 1993; Rios et al. 2002; De Buck et al. 2009), the integration of T-DNAs in clusters of two or more copies can account for about 50% of the integration events, and about 30–70% of events include sequences from the T-DNA parent binary plasmid (e.g., Martineau et al. 1994; Kononov et al. 1997; Nicolia et al. 2017). Therefore, these same structures were captured as ds DNA molecules before integration supports a notion that most T-DNA molecules integrate as ds T-DNA intermediates.

If T-DNA integrates as a ds T-DNA intermediate, then an important question is what mechanism accounts for the synthesis of the complementary strand. Liang and Tzfira (2013) showed that oligonucleotides can efficiently interact with the ss T-DNA and convert ss T-DNA to ds T-DNA molecules. Whereas the mechanism of this conversion is still unknown, it is has been shown that introduction of ss DNA into protoplasts using either electroporation or polyethylene glycol resulted in rapid synthesis of the complementary strand (Rodenburg et al. 1989; Furner et al. 1989). Therefore, this process can be mediated entirely by the plant DNA repair

machinery. There is also evidence for the existence of extrachromosomal ds T-DNA molecules after *Agrobacterium* infection. The first piece of evidence is the rapid and broad transient expression of T-DNA genes in infected leaves (Janssen and Gardner 1990), no matter if the transferred ss T-DNA is the coding or non-coding strand (Narasimhulu et al. 1996). In addition, experiments involving homologous recombination between extrachromosomal T-DNA constructs delivered as noncomplementary strands suggested that at least one of the ss T-DNA constructs must have been converted to ds T-DNA prior to recombination (Offringa et al. 1990). Recently, Dafny-Yelin et al. (2015) showed that blocking ss T-DNA to ds T-DNA conversion reduced T-DNA gene expression. Therefore, although there is no question that extrachromosomal ds T-DNAs exist in plants immediately after *Agrobacterium* infection, the question remains whether they are the only, or the predominant, intermediates in the integration process.

Identification of plant components that are important for T-DNA integration can provide more clues regarding the form, or the predominant form, of the T-DNA intermediate during integration. For example, evidence that Ku70/80 heterodimer is important for integration may support the ds T-DNA theory because Ku70/80 is involved in nonhomologous end joining (NHEJ) between ds DNA ends (Critchlow and Jackson 1998). However, studies to identify the plant components important for T-DNA integration are still ongoing (discussed below).

3.2 What Are the Bacterial and Plant Factors Involved in T-DNA Integration?

The major approaches to identifying the proteins involved in the process of *Agrobacterium*-mediated transformation are forward and reverse genetics. These approaches have led to the identification of bacterial and plant factors involved in the transformation process and its last step of T-DNA integration. Experimental assays can distinguish between a block in T-DNA transfer, a step prior to actual integration, from a block in T-DNA integration. The principle allowing this distinction is that mutants blocked in stable T-DNA integration but not T-DNA transfer will be able to transiently express genes in plant cells, but not generate stable transgenic plants or plant calli. Therefore, many of the proteins involved in T-DNA integration, including the specific protein domains important for this process, have been identified by this principle coupled with protein localization and protein–protein interaction studies. In addition, large-scale screens in *Arabidopsis* have been conducted to identify host proteins involved in *Agrobacterium* transformation (Zhu et al. 2003; Anand et al. 2007a; Gelvin 2010). When considering the commonly used experimental methods to identify genes involved in stable T-DNA integration, a possible scenario should be noted. If T-DNA randomly integrates into and mutates a gene leading to increased gene silencing, fewer stable transgenic event would be recovered through selection if the selection gene has

been silenced. Thus, a mutant may show reduced stable transformation, whereas stable T-DNA integration occurs at same or even higher rate as wild-type plants. This was demonstrated by Park et al. (2015) who analyzed T-DNA integration biochemically. Therefore, T-DNA integration does not necessarily equate to stable transformation.

Several lines of evidence suggest that plant factors mostly, if not entirely, mediate the process of T-DNA integration. First, there are not many *Agrobacterium* candidate proteins that can be involved in the process because T-DNA itself does not encode proteins that are required for T-DNA integration and only a few Vir proteins are known to be transported into the plant nucleus. Second, DNA sequencing of T-DNA/plant DNA junctions suggests that integration occurs through the same pathways responsible for DNA end-joining repair by the host plant cell (i.e., illegitimate/nonhomologous recombination). Another support for the notion that the host cell is responsible for T-DNA integration comes from results of *Agrobacterium*-mediated transformation of yeast (Bundock et al. 1995). In yeast, T-DNA can integrate via homologous recombination, a major pathway of DSBs repair that is used by this organism to repair DSBs, if sufficient homology between T-DNA and yeast sequences exists. Third, foreign DNA can be introduced into plant cells by other methods that do not include *Agrobacterium*, such as electroporation, polyethylene glycol, and particle bombardment transformation. By these methods, the introduced DNA integrates through illegitimate/nonhomologous recombination into the genome, demonstrating that the plant's own DNA repair machinery can potentially accomplish the task of T-DNA integration without the assistance of foreign genes (for review, see Somers and Makarevitch 2004). As ongoing studies are improving our understanding of the mechanisms and pathways behind DNA DSB repair in plants, we anticipate a better understanding of how plant factors facilitate T-DNA integration.

DNA end joining during DSBs repair is less understood in plants in comparison to yeast or mammalian cells. However, it is known that the major pathway of DNA DSBs repair in plants is the nonhomologous end joining (NHEJ) pathway, which is the major pathway for DSB repair in higher eukaryotes. The NHEJ pathway includes the key heterodimer Ku70/Ku80 that binds double-stranded DNA ends formed by the DSBs (Critchlow and Jackson 1998). Several studies investigated the role of Ku80, both in repairing DNA DSBs in plants and in T-DNA integration. Friesner and Britt (2003) reported that Ku80-deficient plants are more sensitive to DSB inducing gamma radiation and are reduced in T-DNA integration rates. The results of Friesner and Britt (2003) supported the involvement of the NHEJ repair pathway in T-DNA integration. Li et al. (2005b) further demonstrated that Ku80 is important for T-DNA integration. First, overexpression of Ku80 in plants enhanced T-DNA integration, whereas Ku80-deficient plants were deficient in T-DNA integration. In addition, Ku80 interacted with ds T-DNA *in planta*, as demonstrated by immunoprecipitation experiments. Both of these studies were done in *Arabidopsis* plants, and results by Jia et al. (2012) and Mestiri et al. (2014) also support the notion that Ku proteins are involved stable T-DNA integration in *Arabidopsis*. Moreover, in rice, knockdown of the Ku70/80 heterodimer also confirmed reduced

stable transformation rates (Nishizawa-Yokoi et al. 2012). On the other hand, contradictory results have been presented by other research groups. Gallego et al. (2003) found that whereas Ku80 has a role in NHEJ in *Arabidopsis* plants, a Ku80-deficient plant was not deficient in T-DNA integration. Park et al. (2015) examined a set of the NHEJ mutant genes in *Arabidopsis*, including *Ku80* and *Ku70*, and determined that deficiency in NHEJ proteins increased the rate of T-DNA integration. According to the authors of that study, the contradictory results can be explained by increased random DNA DSBs in the plant genome that results from deficiency in NHEJ proteins. This results in T-DNA having more available target sites for integration. Therefore, T-DNA integration rate could be affected either way from a deficiency in NHEJ factors: whereas it may be enhanced from increased availability of genomic DSBs as a result of deficiency in NHEJ factors, the integration rate may also be reduced because of reduced ligation ability of the T-DNA into DSBs.

Similar conflicting results have been obtained for Ligase IV, another key component of the NHEJ pathway. Whereas the importance of Ligase IV for NHEJ DNA repair has been demonstrated in plants (Friesner and Britt 2003; van Attikum et al. 2003), Ligase IV has been shown to be both dispensable (van Attikum et al. 2003; Park et al. 2015) and involved but nonessential (Friesner and Britt 2003; Nishizawa-Yokoi et al. 2012) for T-DNA integration. Other components of the NHEJ pathway have been shown to be either required (Jia et al. 2012) or dispensable (Park et al. 2015; Vaghchhipawala et al. 2012; Mestiri et at. 2014) for T-DNA integration (reviewed in Saika et al. 2014).

Alternative NHEJ pathways, such as the microhomology-mediated end joining (MMEJ), have received increasing attention in recent years (reviewed in Wang and Xu 2017). Mestiri et al. (2014) showed that mutations in several alternative NHEJ pathway genes reduced T-DNA integration. In addition, a quadruple *Arabidopsis* mutant disabling several end-joining pathways, including NHEJ, was severely compromised in *Agrobacterium*-mediated transformation. However, that T-DNA integration still occurred suggests additional pathways. On the other hand, Park et al. (2015) found that T-DNA integration was not reduced, but increased, in a *parp1* mutant. Finally, disabling another MMEJ component, polymerase theta (pol θ), completely eliminated T-DNA integration according to van Kregten et al. (2016). However, Gelvin's group found that these same mutants can still be transformed at ∼20% of wild-type levels (personal communication). Furthermore, a recent analysis of filler DNA at T-DNA junctions provided more support for an MMEJ mechanism acting at the LB end of the T-DNA (van Kregten et al. 2016).

The discrepancies of some of the results of these studies may be the result of different experimental conditions when measuring transient and stable T-DNA transformation, as well as conflating stable transformation with T-DNA integration. More interestingly, these discrepancies may point to other alternative pathways that are active under different conditions, such as tissue type and developmental stage.

Other plant proteins that have been identified as being involved or as affecting T-DNA integration include proteins that are involved in the chromatin structure or proteins that direct the T-DNA to the chromatin (for review, see Magori and

Citovsky 2011). In particular, evidence suggests that histones play an important role in T-DNA integration. Several studies demonstrated that plants deficient in different histones were reduced in T-DNA integration rate, whereas overexpression of histones resulted in increased stable transformation and T-DNA integration (Mysore et al. 2000; Yi et al. 2002, 2006; Anand et al. 2007b; Iwakawa et al. 2017). In addition, a domain in VIP1 has been shown to be important for the interaction with histone proteins and for stable transformation (Li et al. 2005a). Finally, VIP2 a transcriptional regulator influencing histone mRNA levels has been shown to be important for T-DNA integration (Anand et al. 2007b).

The bacterial factors that are potential candidates to be involved in T-DNA integration are limited to those that are secreted into the plant cells. They include VirE2, VirE3, VirF, VirD5, and VirD2 (Vergunst et al. 2000; Schrammeijer et al. 2003; Vergunst et al. 2005). Indirect evidence for involvement of Vir proteins in T-DNA integration can be derived from the patterns of T-DNA integration. If comparing T-DNA integration to integration of foreign DNA delivered by other non-*Agrobacterium* methods, T-DNA integration is usually much more efficient. Also, although T-DNA integration can result in complex insertions, these are usually considered more "simple" and precise compared to insertions produced via other methods (Hu et al. 2003; Makarevitch et al. 2003; Travella et al. 2005). These observations may suggest that T-DNA integration, in contrast to DNA delivered by other methods, uses another bacterial factor or factors in addition to the host DNA repair machinery to facilitate integration. However, it is also possible that this may be merely a result of more efficient nuclear localization due to VirD2 piloting the T-DNA, or the protection of T-DNA from degradation through VirE2 coating.

The bacterial candidate important for T-DNA integration that has been studied most extensively is VirD2, because it is transferred into the nucleus while attached to the 5′ end of the single-stranded T-DNA (T-strand). The earliest support for VirD2 involvement in integration was provided in an *in vitro* assay showing that VirD2 has an ability to rejoin ends from the cutting reaction (Pansegrau et al. 1993). Therefore, it has been suggested that the 5′ end of a T-DNA is ligated to the plant DNA via VirD2. Potentially supporting the notion that VirD2 has a ligase-like activity *in planta*, Tinland et al. (1995) reported a VirD2 mutant (R129G) that resulted in reduced precision of the RB side after T-DNA integration. However, this mutation did not reduce the efficiency of T-DNA integration, suggesting that the loss of precision may be only due to VirD2's role in protecting the 5′ end. Moreover, a different in vitro study rejected a general ligation activity of VirD2 (Ziemienowicz et al. 2000).

A better understanding of the potential role of VirD2 in T-DNA integration required investigating the different VirD2 domains. Whereas the N-terminal region of VirD2 contains a relaxase domain that is important for border nicking in *Agrobacterium* (Ward and Barnes 1988), the C-terminal domain contains three regions: a DUF domain, a bipartite NLS, and an omega (Ω) domain. The role of the DUF domain has been shown to be delivery of the ss T-DNA through the T4S system (van Kregten et al. 2009), whereas the role of the bipartite NLS domain is in nuclear transport (Howard et al. 1992; Shurvinton et al. 1992; Tinland et al. 1992,

1995; Rossi et al. 1993; Bravo-Angel et al. 1998; Mysore et al. 1998; van Kregten et al. 2009). The Ω domain has been shown to be important for tumorigenesis (Shurvinton et al. 1992; Bravo-Angel et al. 1998); however, its involvement in T-DNA integration is undetermined. Several reports have shown that a deletion or substitution mutation at the Ω domain reduced T-DNA integration to about 1–4% of the wild-type T-DNA rate (Shurvinton et al. 1992; Narasimhulu et al. 1996; Mysore et al. 1998), whereas the T-DNA transfer rate is reduced to only 20–30% of the wild-type rate (Narasimhulu et al. 1996; Bravo-Angel et al. 1998; Mysore et al. 1998). However, Bravo-Angel et al. (1998) and van Kregten et al. (2009) concluded that the Ω domain has no role in integration. Moreover, inducible expression of VirD2 in plants reduced the transformation efficiency (Hwang et al. 2006). Therefore, it is still controversial if VirD2 or any of the other bacterial Vir proteins have a direct role in T-DNA integration.

Recently, Zhang et al. (2017) showed that in yeast, VirD5 localizes to the centromeres/kinetochores in the nucleus and causes chromosome instability. The authors also showed that VirD5 inhibited cell growth in yeast and also in plants. Therefore, whether VirD5 is involved somehow in T-DNA integration is an interesting question.

## 3.3 What Is the Genomic Site Prerequisite for T-DNA Integration?

Large-scale analysis of T-DNA insertions has shown that insertions are distributed randomly among the plant chromosomes (e.g., Alonso et al. 2003; Sallaud et al. 2004). At the chromosome and gene level, there may be a distribution bias, although this is controversial. It has been suggested that T-DNA integrates preferably at genomic regions that are actively transcribed because T-DNA insertions are generally found more frequently at 5′ and 3′ regions of genes, but less frequent at regions closer to the centromeres and telomeres (Brunaud et al. 2002; Szabados et al. 2002; Alonso et al. 2003; Chen et al. 2003; An et al. 2003; Sallaud et al. 2004; Li et al. 2006; Zhang et al. 2007). A plausible explanation is that during transcription genomic DNA is more "open" and therefore more accessible to incoming T-DNA molecules. Indeed, it has been shown that T-DNA integration sites are preferably found in A-T-rich regions that have a relatively lower DNA duplex stability (e.g., Brunaud et al. 2002; Chen et al. 2003). In addition, a component of the T-complex may interact with host factors, such as a TATA-binding protein, that are involved in gene transcription (Bako et al. 2003). This way, they can guide the T-DNA to actively transcribed regions. Active regions may also be more prone to DNA damage, such as DSBs, and this may create "hot spots" for DNA repair factors and T-DNA integration.

On the other hand, results from previous large-scale studies may have been biased by the experimental method that relied on marker-based selection and

regeneration of plants. If not selected, T-DNA integration events may be excluded from a studied collection. In most studies, the analysis is based on selection via the T-DNA's own marker gene, such as antibiotic or herbicide resistance. However, if T-DNA integrates but the marker gene is not expressed, plants will not survive selection and therefore will not be included in the studied collection. In this regard, Francis and Spiker (2005) showed that in about 30% of transformed plants T-DNA genes are not transcribed. Furthermore, it has been shown that when not applying selection to detect T-DNA insertions, they are distributed randomly and are equally represented in centromeric and telomeric regions (Francis and Spiker 2005; Kim et al. 2007; Shilo et al. 2017). In addition, T-DNA integration can be mutagenic and therefore can disrupt genes that are essential for selection and recovery of plants. However, this likely occurs in a relatively small number of cases.

Different events have been proposed to stimulate the integration of T-DNA into specific sites in the plant DNA. The events include single-strand DNA nicks in the plant DNA, a relaxed duplex DNA forces that allow "invasion" of a T-DNA to the plant DNA, and genomic double-strand breaks (DSBs). Early models by Gheysen et al. (1991) and Mayerhofer et al. (1991) suggested that a nick in the plant DNA is first generated (Fig. 4a). This nick is later converted, via 5′ to 3′ exonuclease activity, into a gap (the "single-strand gap-repair" model). The LB and RB sides of a single-stranded T-DNA can anneal to the plant DNA at this gap through microhomologies and initiate integration. Revision of this model postulated that instead of annealing to DNA within a gap, the LB side invades and anneals to regions of microhomology at the plant DNA. This may happen more often at A-T-rich regions due to lower duplex stability (Tinland et al. 1995; Brunaud et al. 2002) (Fig. 4b). Recently, a link between T-DNA integration and genomic DSBs has become increasingly accepted (for review, see Magori and Citovsky 2011). It has been suggested that genomic DSBs are the prerequisite for T-DNA integration (Fig. 4c). The breaks may be spontaneous and may occur randomly in the genome under natural conditions. Extrachromosomal T-DNA molecules may be directed to DSBs, likely guided by host DNA repair proteins, and possibly also *Agrobacterium* proteins of the T-complex. Direct support for this model is that T-DNA can be directed to integrate into artificiality induced genomic DSBs (Salomon and Puchta 1998; Tzfira et al. 2003; Chilton and Que 2003; Zhang et al. 2018). Muller et al. (2007) reported that T-DNA insertions are found more frequently near or at palindromic sequences in the plant genome. This observation supports integration of T-DNA into genomic DSBs because palindromic regions are often found at sites of DSBs repair in plants (Muller et al. 1999) and, therefore, may be more susceptible to breaks due to their secondary structure. In addition, induction of genomic DSBs by irradiation increases integration of foreign DNA into plant genome (Kohler et al. 1989). Another possibility is that under natural conditions *Agrobacterium* can induce DSBs in order to facilitate T-DNA integration, as microbial pathogens have been shown to trigger host DNA DSBs in plants (Song and Bent, 2014). However, currently there is no evidence for such an activity induced by any of the *Agrobacterium* virulence factors.

Fig. 4 Possible genomic
pre-conditions for T-DNA
integration. **a** Nick (later
expanded into a gab).
b Relaxed duplex DNA
region. **c** Double-strand break
(DSB). Adapted from Singer
(2013) doctoral dissertation
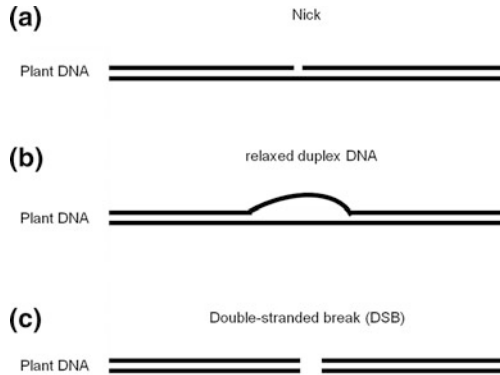
3.4 What Is the Spatial/Temporal Arrangement of T-DNA During Integration?

The T-DNA region of natural *Agrobacterium* strains is 5-25 kbp in length (Barker et al. 1983; Suzuki et al. 2000). The size range of engineered T-DNA constructs used in laboratory strains may be similar, but also T-DNA constructs of up to 150 kbp can be successfully transferred and integrated into the plant genome (Hamilton et al. 1996). On the other hand, T-DNA can integrate into the plant genome with remarkably minimal damage to the plant DNA. For example, Meza et al. (2002), Windels et al. (2003), and Kleinboelting et al. (2015) reported that in the T-DNA collections they chose to analyze, most sequenced integration events (sites in which both LB and RB T-DNA junctions with the genomic DNA had been sequenced) showed a deletion of 100 bp or less of plant genomic DNA bordering the integration site. This result raises the question how T-DNA is spatially arranged during the process of integration, considering that T-DNA is a large molecule in comparison to the small integration site (Fig. 5a). The current popular models do not provide an explanation for this question. The temporal mode of T-DNA integration, on the other hand, has been discussed in early models for T-DNA integration. The model proposed by Tinland (1996) suggests that the LB side of T-DNA interacts with the plant DNA first, following which the RB side attaches to plant DNA (Fig. 5b). That the LB is the initiator of integration is based on the observations that T-DNA insertions share higher degree of microhomology at the T-DNA LB side with the plant DNA junctions, in comparison to the RB side (Tinland et al. 1995; Brunaud et al. 2002; Kim et al. 2003; Zhu et al. 2006; Thomas and Jones 2007). However, several studies reported similar frequencies of microhomologies at both ends (Meza et al. 2002; Windels et al. 2003; Forsbach et al. 2003; Kleinboelting et al. 2015). Therefore, Meza et al. (2002) proposed that in some cases the RB side is the first side to initiate integration into the plant DNA (Fig. 5c).
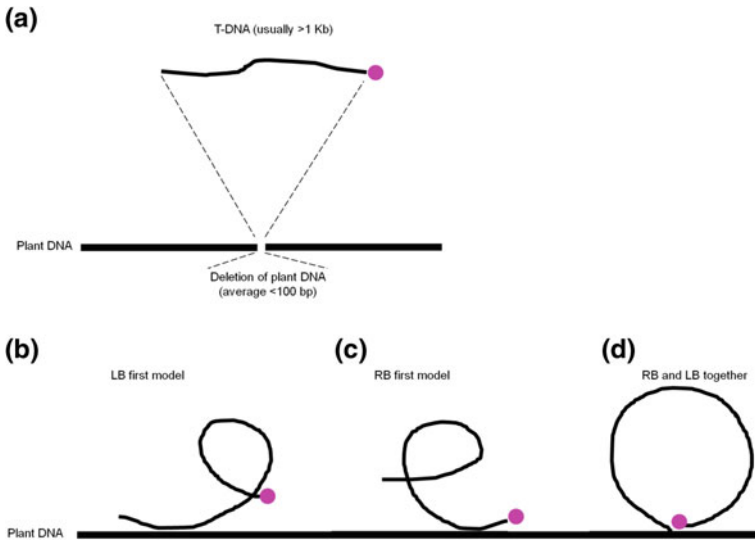
Fig. 5 Possible spatial arrangements of T-DNA during integration. **a** T-DNA, at a size of a few kbp DNA, integrates in most cases without causing major deletions at the target genomic site. **b** LB first model suggests that the LB anneals first via microhomology. **c** RB first model suggests that the RB anneals first via microhomology. **d** LB and RB are in close proximity during integration. Adapted from Singer (2013) doctoral dissertation

Interestingly, the Muller et al. (2007) analysis of T-DNA/plant junctions revealed that T-DNA integration also involves microhomologies in inverted orientation. Based on this finding, these authors proposed that the LB and RB sides of a T-DNA strand anneal to plant DNA simultaneously via microhomologies and that T-DNA ends are in close proximity during integration. However, the model of Muller et al. (2007) does not explain how and when the two ends of a T-DNA are brought into close proximity. The discovery of T-DNA circles (T-circles) provides a possible explanation (Singer et al. 2012), as circular double-stranded T-DNA in plants contains LB and RB sequences ligated extrachromosomally. These results suggest that the LB and RB ends are recognized by plant DNA repair factors and that these factors pull the two ends of a T-DNA toward each other before integration (Fig. 5c). Therefore, it is possible that double-stranded T-DNA approaches the plant genome with the LB and RB sides already in close proximity (Figs. 5d and 6).

3.5 Why and How Do Complex T-DNA Insertions Form?

As mentioned above, T-DNA integration is not a "precise" or "clean" process. Early studies indicated that T-DNA integration can often result in complex T-DNA insertions (Ooms et al. 1982; Kwok et al. 1985; Spielmann and Simpson 1986;
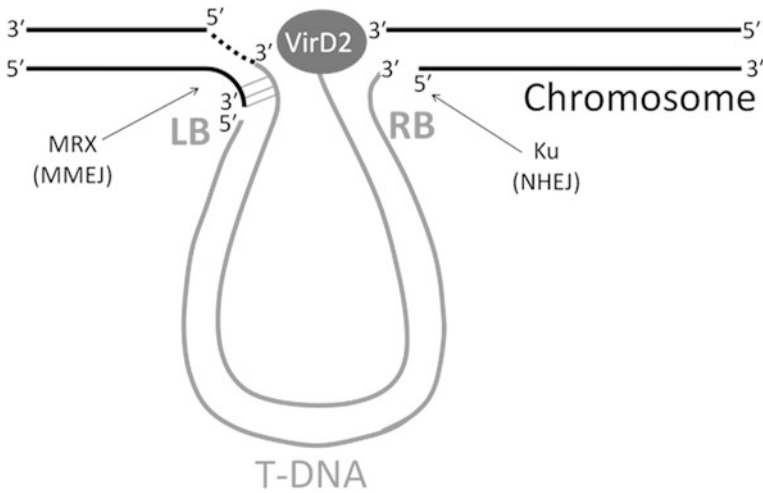
Fig. 6 A proposed model for T-DNA integration. Schematic illustration of double-stranded T-DNA (gray lines) and double-stranded plant DNA (black lines) during T-DNA integration into the plant genome. T-DNA is arranged in a looped mode in which the left border (LB) end and the right border (RB) end are in close proximity. The LB end has a 3′ single-stranded overhang that aligns via short microhomologies to the plant 3′ overhang through the Mre11, Rad50, and Xrs2 (MRX) complex of the MMEJ repair pathway. The dashed line represents a region of template-dependent DNA synthesis of the complementary strand. The RB end, with VirD2 covalently attached to the 5′ end, aligns to the plant DNA through the Ku70/80 mediated NHEJ DNA repair pathway. Illustration adapted from Singer (2013) Doctoral dissertation

Gheysen et al. 1987; Grevelding et al. 1993; Ohba et al. 1995). A complex T-DNA insertion can include, in addition to T-DNA, DNA sequences from various sources. The DNA sequences can be derived from the *Agrobacterium* binary or Ti plasmid (Martineau et al. 1994; Kononov et al. 1997) or even from bacterial chromosomal DNA (Ulker et al. 2008; Kleinboelting et al. 2015). In addition, plant DNA at the site of integration may be re-arranged and include duplications of plant DNA sequences that were not part of the original pre-integration genomic site (Gheysen et al. 1987; Takano et al. 1997; Windels et al. 2003; Kleinboelting et al. 2015). Also, several copies of T-DNA, or parts of the T-DNA sequence, are often clustered together at the integration site (Jorgensen et al. 1987; De Neve et al. 1997). In some cases, the additional DNA sequence that is found at the insertion site does not have any homology to a known DNA sequence. This kind of DNA is termed "filler" DNA, a term that is also used to describe additional DNA at DSB repair sites. The term "filler" is also used to describe additional DNA sequences that share homology with known DNA, such as DNA that is homologous to plant or *Agrobacterium* DNA (Gheysen et al. 1987; Gorbunova and Levy 1997; Windels et al. 2003; Kleinboelting et al. 2015).

The formation or appearance of the different DNA sequences that accompany complex T-DNA insertions can be explained in several ways. Therefore, each

complex insertion can be explained by a different mechanism or by a combination of mechanisms. Nevertheless, it is possible to distinguish between two general types of potential sources for the DNA that is found in complex T-DNA insertions. The first type includes DNA fragments that are present in the nucleus at the time of integration, which may also be described as "free-floating" DNA fragments. The free DNA fragments can ligate with T-DNA prior to or during integration and form complex insertions. The second type includes DNA that is synthesized during the process of DNA repair in the plant nucleus. During DNA repair and ligation, synthesis of DNA can occur using random DNA sequences as templates. This process is also known as synthesis-dependent strand annealing (SDSA). It involves a single-stranded DNA strand invading random DNA sequence *in cis* (DNA from the same molecule) or *in trans* (DNA from a different molecule), using it as a template, and often switching between different templates. Filler DNA also characterizes ds DNA end joining in higher eukaryotes (Gorbunova and Levy 1997; Salomon and Puchta 1998). It is difficult to determine whether a specific DNA sequence in a complex insertion is the result of ligation between free existing DNA fragments or the result of DNA synthesis. However, as discussed below, in many cases it is possible to surmise the origin of the DNA sequence from the sequence identity, length, and overall arrangement in the complex structure.

Ligation between free extrachromosomal DNA fragments is likely when the DNA sequence can be traced to *Agrobacterium* chromosomal DNA, pTi, or binary plasmid DNA sequences. In many instances, T-DNA is transferred together with the backbone of the parent plasmid, termed a "read-through" transfer, due to incorrect processing of the T-DNA borders in the *Agrobacterium* (Kononov et al. 1997; Wenck et al. 1997). However, non-read through *Agrobacterium* DNA often found in complex T-DNA insertions may be transferred from *Agrobacterium* independently and ligated to T-DNA molecules in plants before integration, or alternatively, transferred from *Agrobacterium* already linked to T-DNA. Clusters of two or more T-DNAs probably result from T-DNA molecules that were transferred independently, ligated into the plant nucleus, and then integrated. De Neve et al. (1997) provided compelling evidence supporting this notion by transforming plants simultaneously with different *Agrobacterium* strains that contained different T-DNA constructs. The authors showed that the two types of T-DNAs can integrate adjacent to each other. Similarly, Singer et al. (2012) isolated extrachromosomal T-DNA structures composed of T-DNA originating from two different *Agrobacterium* strains.

Synthesis-dependent strand annealing (SDSA) is likely the mechanism that accounts for other regions of DNA at the junctions between end-joined DNA fragments. This mechanism can sometimes generate a patchwork of short sequences resulting from consecutive template switches (Gorbunova and Levy 1997; Salomon and Puchta 1998; van Kregten et al. 2016). These sequences can be identical to those of T-DNA or plant DNA; therefore, it is a matter of debate whether a specific DNA fragment is a broken fragment of molecule patched together with another DNA or a new synthesis product. The recent discovery that DNA polymerase θ (pol θ) is involved in T-DNA integration (van Kregten et al. 2016) supports the latter, as pol θ is associated with microhomology annealing and low-fidelity DNA synthesis

(Wang and Xu 2017). The shorter the sequence and the more it is "scrambled", the more likely it is a synthesis product.

T-DNA insertions, where T-DNA copies are arranged adjacent to each other in clusters, may also result from T-DNA replication after transfer. In that case, the replicated T-DNA copy integrates adjacent to its template, as proposed by Van Lijsebettens et al. (1986) and Jorgensen et al. (1987) based on analyzing structures of integration events that include adjacent T-DNAs. In some cases, a pair of adjacent T-DNAs shared the same truncation point at their ends. Therefore, a truncated T-DNA replicated to produce another identical copy with the same truncation. T-DNA replication has been supported by statistical analysis of co-transformation and integration of different T-DNAs at the same locus (De Buck et al. 2009).

Truncations of T-DNA ends, especially at the LB side, are another common pattern of T-DNA integration. There may be several reasons for T-DNA insertion having more severe truncations at the LB side. First, T-DNA is transferred from its RB side piloted by VirD2; therefore, the LB side may be more prone to incorrect processing or breaks during the transfer process. VirD2 attached to the 5′ end of the T-DNA may protect the RB side from exonuclease activity, whereas the LB is exposed to such activity. Second, in the plant nucleus, some of the LB side of T-DNA may be lost during synthesis of a complementary strand (Liang and Tzfira 2013). Synthesis of the complementary strand cannot start from the end of the LB, because the LB side is the 3′ end, whereas synthesis is from the 5′ to 3′ end and requires priming. Third, the LB side may be lost in the process of integration when the single-stranded LB anneals through microhomologies to the plant genome (or another T-DNA). Microhomology usually resides in a region internal to the LB end; in that case, the remaining LB side that is not annealed may be degraded and lost (Tinland 1996).

4   A Proposed Model

A T-DNA that is transferred as a linear DNA molecule may circularize via end joining between its LB and RB ends, thus generating a T-circle with head-to-tail end joining (Singer et al. 2012). Throughout different experiments, the majority of the detected T-circles were cases of simple end joining between the LB and RB sides of a single T-DNA, with some small deletions or additions of DNA (Singer and Gelvin, unpublished data). On the other hand, in some cases, T-circles were multimers comprised of several T-DNA molecules or other complex structures. Interestingly, when two T-DNAs are involved in an end-joining event, T-DNA ends preferably ligate tail-to-tail and head-to-head (unpublished data). These results suggest that when the LB and RB sides of a T-DNA are brought into close proximity, a process that is likely mediated by the plant DNA repair pathways, the LB and RB ends are not favorably ligated to each other. This condition can favor T-DNA integration if the T-DNA is situated next to a plant DNA double-stranded

break, because the T-DNA ends may prefer to ligate to the plant DNA instead of ligating to themselves. A T-DNA LB–RB ligation event may not occur if the RB side is a blunt end that preferably utilizes an NHEJ DNA repair pathway, whereas the LB side is a 3′ overhang that preferably utilizes MMEJ DNA repair pathway.

It is unlikely that circularized (and ligated) T-DNAs are intermediates of T-DNA integration because T-DNAs after integration generally maintain the original linear left and right borders, whereas integration of a circular molecule will inevitably result in T-DNA with circularly permuted, random borders. On the other hand, an integration model involving a linear T-DNA in which the two ends are positioned at the opposite poles of molecule is also not likely, because precise and efficient T-DNA integration may require having the two ends of the T-DNA in close proximity.

Therefore, in the model presented T-DNA is proposed to integrate as a double-stranded DNA molecule that is spatially arranged in a looped form (Fig. 6). A looped configuration in which T-DNA ends are in close proximity can explain how T-DNA is often inserted into the genome without the target genomic sites suffering major deletions (Meza et al. 2002; Windels et al. 2003). The exposed T-DNA ends are likely brought together during the initial stage of the repair process by the DNA repair factors coating the ends. VirD2 may also be involved in bringing the T-DNA ends together, as purified VirD2 in vitro has been shown to catalyze end-joining reactions with single-stranded T-border DNA (Pansegrau et al. 1993). These factors may also facilitate the targeting of T-DNA ends to chromosomal sites where DNA repair occurs, such as sites of random genomic DSBs. Integration into these sites occurs when the ends of a T-DNA do not end join to each other to generate a T-circle, but instead end join with the chromosomal DNA (Fig. 6). It should be noted that the proposed model is simplified and does not explain other different outcomes of integration. For example, the frequent formation of filler DNA can be explained by synthesis activity of pol θ (van Kregten et al. 2016). Integration of other more complex structures can occur similarly following their formation extrachromosomally.

Whereas the proposed model suggests a spatial arrangement of T-DNA during integration, it also speculates that the two T-DNA ends utilize different DNA repair pathways for integration into the plant genome. The involvement of different repair pathways in T-DNA integration can explain the conflicting evidence regarding the importance of some key components of DNA repair pathways. It can also explain the tendency of T-DNA ends to generate LB–LB and RB–RB junctions. However, testing this model will require further biochemical and genetic experiments.

5   Conclusions

Agrobacterium tumefaciens remains the main vector used by plant biologists to genetically transform plants. However, there are still many questions to be answered in order to understand the mechanism of T-DNA integration. Because

most of the questions presented in this review are interrelated, understanding T-DNA integration will require different experimental approaches to answer the different questions. In particular, because T-DNA integration most likely relies mostly on plant host factors, a further understanding of pathways of DNA repair in plants is important for improving the understanding of T-DNA integration.

References

Abu-Arish A, Frenkiel-Krispin D, Fricke T et al (2004) Three-dimensional reconstruction of *Agrobacterium* VirE2 protein with single-stranded DNA. J Biol Chem 279:25359–25363

Albright LM, Yanofsky MF, Leroux B et al (1987) Processing of the T-DNA of *Agrobacterium tumefaciens* generates border nicks and linear, single-stranded T-DNA. J Bacteriol 16:1046–1055

Alonso JM, Stepanova AN, Leisse TJ et al (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301:653–657

Altpeter F, Springer NM, Bartley LE et al (2016) Advancing crop transformation in the era of genome editing. Plant Cell 2:1510–1520

An S, Park S, Jeong DH et al (2003) Generation and analysis of end sequence database for T-DNA tagging lines in rice. Plant Physiol 133:2040–2047

Anand A, Krichevsky A, Schornack S et al (2007a) *Arabidopsis* VIRE2 INTERACTING PROTEIN2 is required for *Agrobacterium* T-DNA integration in plants. Plant Cell 19:1695–1708

Anand A, Vaghchhipawala Z, Ryu CM et al (2007b) Identification and characterization of plant genes involved in *Agrobacterium*-mediated plant transformation by virus-induced gene silencing. Mol Plant Microbe Interact 20:41–52

Bako L, Umeda M, Tiburcio AF et al (2003) The VirD2 pilot protein of *Agrobacterium*-transferred DNA interacts with the TATA box-binding protein and a nuclear protein kinase in plants. Proc Natl Acad Sci U S A 100:10108–10113

Ballas N, Citovsky V (1997) Nuclear localization signal binding protein from *Arabidopsis* mediates nuclear import of *Agrobacterium* VirD2 protein. Proc Natl Acad Sci U S A 94:10723–10728

Barker RF, Idler KB, Thompson DV et al (1983) Nucleotide sequence of the T-DNA region from the *Agrobacterium tumefaciens* octopine Ti plasmid pTi15955. Plant Mol Biol 2:335–350

Barton KA, Binns AN, Matzke AJ et al (1983) Regeneration of intact tobacco plants containing full length copies of genetically engineered T-DNA, and transmission of T-DNA to R1 progeny. Cell 32:1033–1043

Bhattacharjee S, Lee LY, Oltmanns H et al (2008) IMPa-4, an *Arabidopsis* importin alpha isoform, is preferentially involved in *Agrobacterium*-mediated plant transformation. Plant Cell 20:2661–2680

Binns AN, Beaupre CE, Dale EM (1995) Inhibition of VirB-mediated transfer of diverse substrates from *Agrobacterium tumefaciens* by the IncQ plasmid RSF1010. J Bacteriol 177:4890–4899

Bleuyard JY, Gallego ME, White CI (2006) Recent advances in understanding of the DNA double-strand break repair machinery of plants. DNA Repair 5:1–12

Bravo-Angel AM, Hohn B, Tinland B (1998) The omega sequence of VirD2 is important but not essential for efficient transfer of T-DNA by *Agrobacterium tumefaciens*. Mol Plant-Microbe Interact 11:57–63

Brunaud V, Balzergue S, Dubreucq B et al (2002) T-DNA integration into the *Arabidopsis* genome depends on sequences of pre-insertion sites. EMBO Rep 3:1152–1157

Bundock P, den Dulk-Ras A, Beijersbergen A et al (1995) Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to Saccharomyces cerevisiae. EMBO J 14:3206–3214

Bundock P, Hooykaas PJ (1996) Integration of *Agrobacterium tumefaciens* T-DNA in the *Saccharomyces cerevisiae* genome by illegitimate recombination. Proc Natl Acad Sci U S A 93:15272–15275

Castle LA, Errampalli D, Atherton TL et al (1993) Genetic and molecular characterization of embryonic mutants identified following seed transformation in *Arabidopsis*. Mol Gen Genet 241:504–514

Chen S, Jin W, Wang M et al (2003) Distribution and characterization of over 1000 T-DNA tags in rice genome. Plant J 36:105–113

Chilton MD, Drummond M, Merlo D et al (1977) Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. Cell 11:263–271

Chilton MD, Que Q (2003) Targeted integration of T-DNA into the tobacco genome at double-stranded breaks: new insights on the mechanism of T-DNA integration. Plant Physiol 133:956–965

Christie PJ, Atmakuri K, Krishnamoorthy V et al (2005) Biogenesis, architecture, and function of bacterial type IV secretion systems. Annu Rev Microbiol 59:451–485

Chyi YS, Jorgensen RA, Goldstein D et al (1986) Locations and stability of *Agrobacterium*-mediated T-DNA insertions in the *Lycopersicon* genome. Mol Gen Genet 204:64–69

Citovsky V, DE Vos G, Zambryski P (1988) Single-stranded DNA binding protein encoded by the *virE* locus of *Agrobacterium tumefaciens*. Science 240:501–504

Citovsky V, Warnick D, Zambryski P (1994) Nuclear import of *Agrobacterium* VirD2 and VirE2 proteins in maize and tobacco. Proc Natl Acad Sci U S A 91:3210–3214

Citovsky V, Zupan J, Warnick D et al (1992) Nuclear localization of *Agrobacterium* VirE2 protein in plant cells. Science 256:1802–1805

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

Cluster PD, O'Dell M, Metzlaff M et al (1996) Details of T-DNA structural organization from a transgenic *Petunia* population exhibiting co-suppression. Plant Mol Biol 32:1197–1203

Critchlow SE, Jackson SP (1998) DNA end-joining: from yeast to man. Trends Biochem Sci 23:394–398

Dafny-Yelin M, Levy A, Dafny R et al (2015) Blocking single-stranded transferred DNA conversion to double-stranded intermediates by overexpression of yeast DNA REPLICATION FACTOR A. Plant Physiol 167:153–163

Das A (1988) *Agrobacterium tumefaciens virE* operon encodes a single-stranded DNA-binding protein. Proc Natl Acad Sci U S A 85:2909–2913

De Buck S, De Wilde C, Van Montagu M et al (2000) Determination of the T-DNA transfer and the T-DNA integration frequencies upon cocultivation of *Arabidopsis thaliana* root explants. Mol Plant-Microbe Interact 13:658–665

De Buck S, Podevin N, Nolf J et al (2009) The T-DNA integration pattern in *Arabidopsis* transformants is highly determined by the transformed target cell. Plant J 60:134–145

de Groot MJ, Bundock P, Hooykaas PJ et al (1998) *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. Nat Biotechnol 16:839–842

De Neve M, De Buck S, Jacobs A et al (1997) T-DNA integration patterns in co-transformed plant cells suggest that T-DNA repeats originate from co-integration of separate T-DNAs. Plant J 11:15–29

Djamei A, Pitzschke A, Nakagami H et al (2007) Trojan horse strategy in *Agrobacterium* transformation: abusing MAPK defense signaling. Science 318:453–456

Durrenberger F, Crameri A, Hohn B et al (1989) Covalently bound VirD2 protein of *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation. Proc Natl Acad Sci U S A 86:9154–9158

Forsbach A, Schubert D, Lechtenberg B et al (2003) A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome. Plant Mol Biol 52:161–176

Francis KE, Spiker S (2005) Identification of *Arabidopsis thaliana* transformants without selection reveals a high occurrence of silenced T-DNA integrations. Plant J 41:464–477

Friesner J, Britt AB (2003) Ku80- and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. Plant J 34:427–440

Furner IJ, Higgins ES, Berrington AW (1989) Single-stranded DNA transforms plant protoplasts. Mol Gen Genet 220:65–68

Gallego ME, Bleuyard JY, Daoudal-Cotterell S et al (2003) Ku80 plays a role in non-homologous recombination but is not required for T-DNA integration in *Arabidopsis*. Plant J 35:557–565

Gelvin SB (2010) Plant proteins involved in *Agrobacterium*-mediated genetic transformation. Annu Rev Phytopathol 48:45–68

Gelvin SB (2017) Integration of *Agrobacterium* T-DNA into the plant genome. Annu Rev Genet 51:195–217

Ghedira R, De Buck S, Van Ex F et al (2013) T-DNA transfer and T-DNA integration efficiencies upon *Arabidopsis thaliana* root explant cocultivation and floral dip transformation. Planta 238:1025–1037

Gheysen G, Montagu MV, Zambryski P (1987) Integration of *Agrobacterium tumefaciens* transfer DNA (T-DNA) involves rearrangements of target plant DNA sequences. Proc Natl Acad Sci U S A 84:6169–6173

Gheysen G, Villarroel R, Van Montagu M (1991) Illegitimate recombination in plants: a model for T-DNA integration. Genes Dev 5:287–297

Gorbunova V, Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Res 25:4650–4657

Grange W, Duckely M, Husale S et al (2008) VirE2: a unique ssDNA-compacting molecular machine. PLoS Biol 6(2):e44

Grevelding C, Fantes V, Kemper E et al (1993) Single-copy T-DNA insertions in *Arabidopsis* are the predominant form of integration in root-derived transgenics, whereas multiple insertions are found in leaf discs. Plant Mol Biol 23:847–860

Hamilton CM, Frary A, Lewis C et al (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. Proc Natl Acad Sci U S A 93:9975–9979

Herrera-Estrella A, Chen ZM, Van Montagu M et al (1988) VirD proteins of *Agrobacterium tumefaciens* are required for the formation of a covalent DNA-protein complex at the 5' terminus of T-strand molecules. EMBO J 7:4055–4062

Herrera-Estrella A, Van Montagu M, Wang K (1990) A bacterial peptide acting as a plant nuclear targeting signal: the amino-terminal portion of *Agrobacterium* VirD2 protein directs a beta-galactosidase fusion protein into tobacco nuclei. Proc Natl Acad Sci U S A 87:9534–9537

Hoekema A, Hirsch PR, Hooykaas PJJ et al (1983) A binary plant vector strategy based on separation of vir- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. Nature 303:179–180

Howard EA, Zupan JR, Citovsky V et al (1992) The VirD2 protein of *A. tumefaciens* contains a C-terminal bipartite nuclear localization signal: implications for nuclear uptake of DNA in plant cells. Cell 68:109–118

Hu T, Metz S, Chay C et al (2003) *Agrobacterium*-mediated large-scale transformation of wheat (*Triticum aestivum* L.) using glyphosate selection. Plant Cell Rep 21:1010–1019

Hwang HH, Mysore KS, Gelvin SB (2006) Transgenic *Arabidopsis* plants expressing *Agrobacterium tumefaciens* VirD2 protein are less susceptible to *Agrobacterium* transformation. Mol Plant Pathol 7:473–484

Iwakawa H, Carter BC, Bishop BC, Ogas J, Gelvin SB (2017) Perturbation of H3K27me3-associated epigenetic processes increases *Agrobacterium*-mediated transformation. Mol Plant Microbe Interact 30:35–44

Janssen BJ, Gardner RC (1990) Localized transient expression of GUS in leaf discs following cocultivation with *Agrobacterium*. Plant Mol Biol 14:61–72

Jia Q, Bundock P, Hooykaas PJJ et al (2012) *Agrobacterium tumefaciens* T-DNA integration and gene targeting in *Arabidopsis thaliana* non-homologous end-joining mutants. J Bot 2012:1–13

Johansen LK, Carrington JC (2001) Silencing on the spot. Induction and suppression of RNA silencing in the *Agrobacterium*-mediated transient expression system. Plant Physiol 126:930–938

Jorgensen R, Snyder C, Jones JDG (1987) T-DNA is organized predominantly in inverted repeat structures in plants transformed with *Agrobacterium tumefaciens* C58 derivatives. Mol Gen Genet 207:471–477

Kim SI, Veena, Gelvin SB (2007) Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the *Arabidopsis* genome generated under non-selective conditions. Plant J 51:779–791

Kim SR, Lee J, Jun SH et al (2003) Transgene structures in T-DNA-inserted rice plants. Plant Mol Biol 52:761–773

Kleinboelting N, Huep G, Appelhagen I et al (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

Ko K, Brodzik R, Steplewski Z (2009) Production of antibodies in plants: approaches and perspectives. Curr Top Microbiol Immunol 332:55–78

Köhler F, Cardon G, Pöhlman M et al (1989) Enhancement of transformation rates in higher plants by low-dose irradiation: Are DNA repair systems involved in the incorporation of exogenous DNA into the plant genome? Plant Mol Biol 12:189–199

Komarova TV, Baschieri S, Donini M et al (2010) Transient expression systems for plant-derived biopharmaceuticals. Expert Rev Vaccines 9:859–876

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

Korn M, Schmidpeter J, Dahl M et al (2015) A genetic screen for pathogenicity genes in the hemibiotrophic fungus *Colletotrichum higginsianum* identifies the plasma membrane proton ump Pma2 required for host penetration. PLoS ONE 10:e0125960

Krizkova L, Hrouda M (1998) Direct repeats of T-DNA integrated in tobacco chromosome: characterization of junction regions. Plant J 16:673–680

Kumar S, Fladung M (2002) Transgene integration in aspen: structures of integration sites and mechanism of T-DNA integration. Plant J 31:543–551

Kumar V, Jain M (2015) The CRISPR-Cas system for plant genome editing: advances and opportunities. J Exp Bot 66:47–57

Kunik T, Tzfira T, Kapulnik Y et al (2001) Genetic transformation of HeLa cells by *Agrobacterium*. Proc Natl Acad Sci U S A 98:1871–1876

Kwok WW, Nester EW, Gordon MP (1985) Unusual plasmid DNA organization in an octopine crown gall tumor. Nucleic Acids Res 13:459–471

Lacroix B, Li J, Tzfira T et al (2006) Will you let me use your nucleus? How *Agrobacterium* gets its T-DNA expressed in the host plant cell. Can J Physiol Pharmacol 84:333–345

Lacroix B, Vaidya M, Tzfira T et al (2005) The VirE3 protein of *Agrobacterium* mimics a host cell function required for plant genetic transformation. EMBO J 24:428–437

Lee L-Y, Fang MJ, Kuang LY et al (2008) Vectors for multi-color bimolecular fluorescence complementation to investigate protein-protein interactions in living plant cells. Plant Methods 4:24–4811-4-24

Lessl M, Lanka E (1994) Common mechanisms in bacterial conjugation and Ti-mediated T-DNA transfer to plant cells. Cell 77:321–324

Li J, Krichevsky A, Vaidya M et al (2005a) Uncoupling of the functions of the *Arabidopsis* VIP1 protein in transient and stable plant genetic transformation by *Agrobacterium*. Proc Natl Acad Sci U S A 102:5733–5738

Li J, Vaidya M, White C et al (2005b) Involvement of KU80 in T-DNA integration in plant cells. Proc Natl Acad Sci U S A 102:19231–19236

Li X, Pan SQ (2017) *Agrobacterium* delivers VirE2 protein into host cells via clathrin-mediated endocytosis. Sci Adv 3:e1601528

Li Y, Rosso MG, Ülker B et al (2006) Analysis of T-DNA insertion site distribution patterns in *Arabidopsis thaliana* reveals special features of genes without insertions. Genomics 87:645–652

Liang Z, Tzfira T (2013) In vivo formation of double-stranded T-DNA molecules by T-strand priming. Nat Commun 4:2253

Magori S, Citovsky V (2011) Epigenetic control of *Agrobacterium* T-DNA integration. Biochim Biophys Acta 1809:388–394

Makarevitch I, Svitashev SK, Somers DA (2003) Complete sequence analysis of transgene loci from plants transformed via microprojectile bombardment. Plant Mol Biol 52:421–432

Martineau B, Voelker TA, Sanders RA (1994) On defining T-DNA. Plant Cell 6:1032–1033

Matsumoto S, Ito Y, Hosoi T et al (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

Maximova SN, Dandekar AM, Guiltinan MJ (1998) Investigation of *Agrobacterium*-mediated transformation of apple using green fluorescent protein: high transient expression and low stable transformation suggest that factors other than T-DNA transfer are rate-limiting. Plant Mol Biol 37:549–559

Mayerhofer R, Koncz-Kalman Z, Nawrath C et al (1991) T-DNA integration: a mode of illegitimate recombination in plants. EMBO J 10:697–704

McVey M, Lee SE (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet 24:529–538

Mestiri I, Norre F, Gallego ME et al (2014) Multiple host-cell recombination pathways act in *Agrobacterium*-mediated transformation of plant cells. Plant J 77:511–520

Meza TJ, Stangeland B, Mercy IS et al (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

Muller AE, Atkinson RG, Sandoval RB et al (2007) Microhomologies between T-DNA ends and target sites often occur in inverted orientation and may be responsible for the high frequency of T-DNA-associated inversions. Plant Cell Rep 26:617–630

Muller AE, Kamisugi Y, Gruneberg R et al (1999) Palindromic sequences and A+T-rich DNA elements promote illegitimate recombination in *Nicotiana tabacum*. J Mol Biol 291:29–46

Mysore KS, Bassuner B, Deng X et al (1998) Role of the *Agrobacterium tumefaciens* VirD2 protein in T-DNA transfer and integration. Mol Plant-Microbe Interact 11:668–683

Mysore KS, Nam J, Gelvin SB (2000) An *Arabidopsis* histone H2A mutant is deficient in *Agrobacterium* T-DNA integration. Proc Natl Acad Sci U S A 97:948–953

Nacry P, Camilleri C, Courtial B et al (1998) Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. Genetics 149:641–650

Narasimhulu SB, Deng XB, Sarria R et al (1996) Early transcription of *Agrobacterium* T-DNA genes in tobacco and maize. Plant Cell 8:873–886

Nelson AD, Lamb JC, Kobrossly PS et al (2011) Parameters affecting telomere-mediated chromosomal truncation in *Arabidopsis*. Plant Cell 23:2263–2272

Nicolia A, Ferradini N, Veronesi F et al (2017) An insight into T-DNA integration events in *Medicago sativa*. Int J Mol Sci 18. https://doi.org/10.3390/ijms18091951

Nishizawa-Yokoi A, Nonaka S, Saika H et al (2012) Suppression of Ku70/80 or Lig4 leads to decreased stable transformation and enhanced homologous recombination in rice. New Phytol 196:1048–1059

Offringa R, de Groot MJ, Haagsman HJ et al (1990) Extrachromosomal homologous recombination and gene targeting in plant cells after *Agrobacterium* mediated transformation. EMBO J 9:3077–3084

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

Ohmine Y, Satoh Y, Kiyokawa K et al (2016) DNA repair genes RAD52 and SRS2, a cell wall synthesis regulator gene SMI1, and the membrane sterol synthesis scaffold gene ERG28 are important in efficient *Agrobacterium*-mediated yeast transformation with chromosomal T-DNA. BMC Microbiol 16:58-016-0672-0

O'Malley RC, Ecker JR (2010) Linking genotype to phenotype using the *Arabidopsis* unimutant collection. Plant J 61:928–940

Ooms G, Bakker A, Molendijk L et al (1982) T-DNA organization in homogeneous and heterogeneous octopine-type crown gall tissues of *Nicotiana tabacum*. Cell 30:589–597

Otten L, De Greve H, Leemans J, Hain R, Hooykaas P, Schell J (1984) Restoration of virulence of *Vir* region mutants of *Agrobacterium tumefaciens* strain B6S3 by coinfection with normal and mutant *Agrobacterium* strains. Mol Gen Genet 195:159–163

Pansegrau W, Schoumacher F, Hohn B et al (1993) Site-specific cleavage and joining of single-stranded DNA by VirD2 protein of *Agrobacterium tumefaciens* Ti plasmids: analogy to bacterial conjugation. Proc Natl Acad Sci U S A 90:11538–11542

Park SY, Vaghchhipawala Z, Vasudevan B et al (2015) *Agrobacterium* T-DNA integration into the plant genome can occur without the activity of key non-homologous end-joining proteins. Plant J 81:934–946

Pitzschke A, Djamei A, Teige M et al (2009) VIP1 response elements mediate mitogen-activated protein kinase 3-induced stress gene expression. Proc Natl Acad Sci U S A 106(43):18414–18419

Ream LW, Gordon MP, Nester EW (1983) Multiple mutations in the T region of the *Agrobacterium tumefaciens* tumor-inducing plasmid. Proc Natl Acad Sci U S A 80:1660–1664

Relic B, Andjelkovic M, Rossi L et al (1998) Interaction of the DNA modifying proteins VirD1 and VirD2 of *Agrobacterium tumefaciens*: analysis by subcellular localization in mammalian cells. Proc Natl Acad Sci U S A 95:9105–9110

Rios G, Lossow A, Hertel B et al (2002) Rapid identification of *Arabidopsis* insertion mutants by non-radioactive detection of T-DNA tagged genes. Plant J 32:243–253

Rodenburg KW, de Groot MJ, Schilperoort RA et al (1989) Single-stranded DNA used as an efficient new vehicle for transformation of plant protoplasts. Plant Mol Biol 13:711–719

Rolloos M, Dohmen MH, Hooykaas PJ et al (2014) Involvement of Rad52 in T-DNA circle formation during *Agrobacterium tumefaciens*-mediated transformation of *Saccharomyces cerevisiae*. Mol Microbiol 91:1240–1251

Rolloos M, Hooykaas PJ, van der Zaal BJ (2015) Enhanced targeted integration mediated by translocated I-*Sce*I during the *Agrobacterium* mediated transformation of yeast. Sci Rep 5:8345

Rossi L, Hohn B, Tinland B (1993) The VirD2 protein of *Agrobacterium tumefaciens* carries nuclear localization signals important for transfer of T-DNA to plant. Mol Gen Genet 239:345–353

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

Rosso MG, Li Y, Strizhov N et al (2003) An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol Biol 53:247–259

Saika H, Nishizawa-Yokoi A, Toki S (2014) The non-homologous end-joining pathway is involved in stable transformation in rice. Front Plant Sci 5:560

Sakalis PA, van Heusden GP, Hooykaas PJ (2014) Visualization of VirE2 protein translocation by the *Agrobacterium* type IV secretion system into host cells. Microbiologyopen 3(1):104–117

Sallaud C, Gay C, Larmande P et al (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics. Plant J 39:450–464

Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO J 17:6086–6095

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

Schrammeijer B, den Dulk-Ras A, Vergunst AC et al (2003) Analysis of Vir protein translocation from *Agrobacterium tumefaciens* using *Saccharomyces cerevisiae* as a model: evidence for transport of a novel effector protein VirE3. Nucleic Acids Res 31:860–868

Schrammeijer B, Risseeuw E, Pansegrau W et al (2001) Interaction of the virulence protein VirF of *Agrobacterium tumefaciens* with plant homologs of the yeast Skp1 protein. Curr Biol 11:258–262

Sessions A, Burke E, Presting G et al (2002) A high-throughput *Arabidopsis* reverse genetics system. Plant Cell 14:2985–2994

Shi Y, Lee LY, Gelvin SB (2014) Is VIP1 important for *Agrobacterium*-mediated transformation? Plant J 79:848–860

Shiboleth Y, Tzfira T (2012) *Agrobacterium*-mediated plant genetic transformation. In: Hasegawa PM (ed) Altman A. Plant Biotechnology and Agriculture Academic Press, San Diego, pp 99–116

Shilo S, Tripathi P, Melamed-Bessudo C et al (2017) T-DNA-genome junctions form early after infection and are influenced by the chromatin state of the host genome. PLoS Genet 13: e1006875

Shurvinton CE, Hodges L, Ream W (1992) A nuclear localization signal and the C-terminal omega sequence in the *Agrobacterium tumefaciens* VirD2 endonuclease are important for tumor formation. Proc Natl Acad Sci U S A 89:11837–11841

Singer K (2013) T-DNA integration during *Agrobacterium*-mediated transformation of plants (doctoral dissertation). Retrieved from: https://deepbluelibumichedu/handle/202742/99891? show=full University of Michigan, Horace H. Rackham School of Graduate Studies

Singer K, Shiboleth YM, Li J et al (2012) Formation of complex extrachromosomal T-DNA structures in *Agrobacterium tumefaciens*-infected plants. Plant Physiol 160:511–522

Smith EF, Townsend CO (1907) A plant-tumor of bacterial origin. Science 25:671–673

Soltani J, van Heusden GPH, Hooykaas PJJ (2010) *Agrobacterium*-mediated transformation of non-plant organisms. In: Tzfira T, Citovsky V (eds) *Agrobacterium*: from biology to biotechnology. Springer, New York, pp 650–675

Somers DA, Makarevitch I (2004) Transgene integration in plants: poking or patching holes in promiscuous genomes? Curr Opin Biotechnol 15:126–131

Song J, Bent AF (2014) Microbial pathogens trigger host DNA double-strand breaks whose abundance is reduced by plant defense responses. PLoS Pathog 10(4):e1004030

Sparkes IA, Runions J, Kearns A et al (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc 1:2019–2025

Spielmann A, Simpson R (1986) T-DNA structure in transgenic tobacco plants with multiple independent integration sites. Mol Gen Genet 205:34–41

Suzuki K, Hattori Y, Uraji M et al (2000) Complete nucleotide sequence of a plant tumor-inducing Ti plasmid. Gene 242:331–336

Szabados L, Kovacs I, Oberschall A et al (2002) Distribution of 1000 sequenced T-DNA tags in the *Arabidopsis* genome. Plant J 32:233–242

Takano M, Egawa H, Ikeda JE et al (1997) The structures of integration sites in transgenic rice. Plant J 11:353–361

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

Teo CH, Ma L, Kapusi E et al (2011) Induction of telomere-mediated chromosomal truncation and stability of truncated chromosomes in *Arabidopsis thalian*a. Plant J 68:28–39

Thomas CM, Jones JD (2007) Molecular analysis of *Agrobacterium* T-DNA integration in tomato reveals a role for left border sequence homology in most integration events. Mol Genet Genom 278:411–420

Tinland B (1996) The integration of T-DNA into plant genomes. Trends Plant Sci 1:178–184

Tinland B, Hohn B, Puchta H (1994) *Agrobacterium tumefaciens* transfers single-stranded transferred DNA (T-DNA) into the plant cell nucleus. Proc Natl Acad Sci U S A 91:8000–8004

Tinland B, Koukolikova-Nicola Z, Hall MN et al (1992) The T-DNA-linked VirD2 protein contains two distinct functional nuclear localization signals. Proc Natl Acad Sci U S A 89:7442–7446

Tinland B, Schoumacher F, Gloeckler V et al (1995) The *Agrobacterium tumefaciens* virulence D2 protein is responsible for precise integration of T-DNA into the plant genome. EMBO J 14:3585–3595

Travella S, Ross SM, Harden J et al (2005) A comparison of transgenic barley lines produced by particle bombardment and *Agrobacterium*-mediated techniques. Plant Cell Rep 23:780–789

Tzfira T, Citovsky V (2006) *Agrobacterium*-mediated genetic transformation of plants: biology and biotechnology. Curr Opin Biotechnol 17:147–154

Tzfira T, Frankman LR, Vaidya M et al (2003) Site-specific integration of *Agrobacterium tumefaciens* T-DNA via double-stranded intermediates. Plant Physiol 133:1011–1023

Tzfira T, Li J, Lacroix B et al (2004a) *Agrobacterium* T-DNA integration: molecules and models. Trends Genet 20:375–383

Tzfira T, Vaidya M, Citovsky V (2001) VIP1, an *Arabidopsis* protein that interacts with *Agrobacterium* VirE2, is involved in VirE2 nuclear import and *Agrobacterium* infectivity. EMBO J 20:3596–3607

Tzfira T, Vaidya M, Citovsky V (2004b) Involvement of targeted proteolysis in plant genetic transformation by *Agrobacterium*. Nature 431:87–92

Ulker B, Li Y, Rosso MG et al (2008) T-DNA-mediated transfer of *Agrobacterium tumefaciens* chromosomal DNA into plants. Nat Biotechnol 26:1015–1017

Vaghchhipawala ZE, Vasudevan B, Lee S et al (2012) *Agrobacterium* may delay plant nonhomologous end-joining DNA repair via XRCC4 to favor T-DNA integration. Plant Cell 24:4110–4123

van Attikum H, Bundock P, Overmeer RM et al (2003) The *Arabidopsis* AtLIG4 gene is required for the repair of DNA damage, but not for the integration of *Agrobacterium* T-DNA. Nucleic Acids Res 31:4247–4255

van Kregten M, de Pater S, Romeijn R et al (2016) T-DNA integration in plants results from polymerase-theta-mediated DNA repair. Nat Plants 2:16164

van Kregten M, Lindhout BI, Hooykaas PJ et al (2009) *Agrobacterium*-mediated T-DNA transfer and integration by minimal VirD2 consisting of the relaxase domain and a type IV secretion system translocation signal. Mol Plant-Microbe Interact 22:1356–1365

Van Lijsebettens M, Inze D, Schell J et al (1986) Transformed cell clones as a tool to study T-DNA integration mediated by *Agrobacterium tumefaciens*. J Mol Biol 188:129–145

Vergunst AC, Schrammeijer B, den Dulk-Ras A et al (2000) VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. Science 290:979–982

Vergunst AC, van Lier MC, den Dulk-Ras A et al (2005) Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium*. Proc Natl Acad Sci U S A 102:832–837

Virts EL, Gelvin SB (1985) Analysis of transfer of tumor-inducing plasmids from *Agrobacterium tumefaciens* to *Petunia* protoplasts. J Bacteriol 162:1030–1038

Vogel AM, Das A (1992) Mutational analysis of *Agrobacterium tumefaciens* VirD2: tyrosine 29 is essential for endonuclease activity. J Bacteriol 174:303–308

Wang H, Xu X (2017) Microhomology-mediated end joining: new players join the team. Cell Biosci 7:6-017-0136-8. eCollection 2017

Wang K, Herrera-Estrella L, Van Montagu M et al (1984) Right 25 bp terminus sequence of the nopaline T-DNA is essential for and determines direction of DNA transfer from *Agrobacterium* to the plant genome. Cell 38:455–462

Wang K, Stachel SE, Timmerman B et al (1987) Site-specific nick in the T-DNA border sequence as a result of *Agrobacterium* vir gene expression. Science 235:587–591

Wang L, Lacroix B, Guo J et al (2017) The *Agrobacterium* VirE2 effector interacts with multiple members of the *Arabidopsis* VIP1 protein family. Mol Plant Pathol 19:1172–1183

Ward ER, Barnes WM (1988) VirD2 protein of *Agrobacterium tumefaciens* is very tightly linked to the 5′ end of T-strand DNA. Science 242:927–930

Wenck A, Czako M, Kanevski I et al (1997) Frequent collinear long transfer of DNA inclusive of the whole binary vector during *Agrobacterium*-mediated transformation. Plant Mol Biol 34:913–922

Windels P, De Buck S, Depicker A (2010) Patterns of T-DNA integration into the host genome. In: Tzfira T, Citovsky V (eds) *Agrobacterium*: from biology to biotechnology. Springer, New York, NY, pp 441–481

Windels P, De Buck S, Van Bockstaele E et al (2003) T-DNA integration in *Arabidopsis* chromosomes. Presence and origin of filler DNA sequences. Plant Physiol 133:2061–2068

Yadav NS, Vanderleyden J, Bennett DR et al (1982) Short direct repeats flank the T-DNA on a nopaline Ti plasmid. Proc Natl Acad Sci U S A 79:6322–6326

Yanofsky MF, Porter SG, Young C et al (1986) The *virD* operon of *Agrobacterium tumefaciens* encodes a site-specific endonuclease. Cell 47:471–477

Yi H, Mysore KS, Gelvin SB (2002) Expression of the *Arabidopsis* histone H2A-1 gene correlates with susceptibility to *Agrobacterium* transformation. Plant J 32:285–298

Yi H, Sardesai N, Fujinuma T et al (2006) Constitutive expression exposes functional redundancy between the *Arabidopsis* histone H2A gene HTA1 and other H2A gene family members. Plant Cell 18:1575–1589

Yin K, Gao C, Qiu JL (2017) Progress and prospects in plant genome editing. Nat Plants 3:17107

Young C, Nester EW (1988) Association of the VirD2 protein with the 5′ end of T strands in *Agrobacterium tumefaciens*. J Bacteriol 170:3367–3374

Yusibov VM, Steck TR, Gupta V et al (1994) Association of single-stranded transferred DNA from *Agrobacterium tumefaciens* with tobacco cells. Proc Natl Acad Sci U S A 91(8):2994–2998

Zaenen I, van Larebeke H, Teuchy H et al (1974) Supercoiled circular DNA in crown-gall inducing *Agrobacterium* strains. J Mol Biol 86:109–127

Zaltsman A, Lacroix B, Gafni Y et al (2013) Disassembly of synthetic *Agrobacterium* T-DNA-protein complexes via the host SCF(VBF) ubiquitin-ligase complex pathway. Proc Natl Acad Sci U S A 110:169–174

Zambryski P, Joos H, Genetello C et al (1983) Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. EMBO J 2:2143–2150

Zhang J, Cai L, Cheng J et al (2008) Transgene integration and organization in cotton (*Gossypium hirsutum* L.) genome. Transgenic Res 17:293–306. https://doi.org/10.1007/s11248-007-9101-3

Zhang J, Guo D, Chang Y et al (2007) Non-random distribution of T-DNA insertions at various levels of the genome hierarchy as revealed by analyzing 13 804 T-DNA flanking sequences from an enhancer-trap mutant library. Plant J 49:947–959

Zhang Q, Xing HL, Wang ZP et al (2018) Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in Arabidopsis and its prevention. Plant Mol Biol 96(4–5):445–456

Zhang X, van Heusden GPH, Hooykaas PJJ (2017) Virulence protein VirD5 of *Agrobacterium tumefaciens* binds to kinetochores in host cells via an interaction with Spt4. Proc Natl Acad Sci U S A 114:10238–10243

Zhu QH, Ramm K, Eamens AL et al (2006) Transgene structures suggest that multiple mechanisms are involved in T-DNA integration in plants. Plant Sci 171:308–322

Zhu Y, Nam J, Humara JM et al (2003) Identification of *Arabidopsis* rat mutants. Plant Physiol 132:494–505

Ziemienowicz A, Merkle T, Schoumacher F et al (2001) Import of *Agrobacterium* T-DNA into plant nuclei: two distinct functions of VirD2 and VirE2 proteins. Plant Cell 13:369–383

Ziemienowicz A, Tinland B, Bryant J et al (2000) Plant enzymes but not *Agrobacterium* VirD2 mediate T-DNA ligation in vitro. Mol Cell Biol 20:6317–6322

Ziemienowicz A, Tzfira T, Hohn B (2010) Mechanisms of T-DNA Integration. In: Tzfira T, Citovsky V (eds) *Agrobacterium*: From Biology to Biotechnology Springer, New York, NY, p 395–440

Zupan JR, Citovsky V, Zambryski P (1996) *Agrobacterium* VirE2 protein mediates nuclear uptake of single-stranded DNA in plant cells. Proc Natl Acad Sci U S A 93:2392–2397