Review TRENDS in Genetics Vol.20 No.8 August 2004

Full text provided by www.sciencedirect.com
SCIENCE DIRECT

Agrobacterium T-DNA integration: molecules and models

Tzvi Tzfira, Jianxiong Li, Benoît Lacroix and Vitaly Citovsky

Department of Biochemistry and Cell Biology, State University of New York, Stony Brook, NY 11794-5215, USA

Genetic transformation mediated by *Agrobacterium* involves the transfer of a DNA molecule (T-DNA) from the bacterium to the eukaryotic host cell, and its integration into the host genome. Whereas extensive work has revealed the biological mechanisms governing the production, *Agrobacterium*-to-plant cell transport and nuclear import of the *Agrobacterium* T-DNA, the integration step remains largely unexplored, although several different T-DNA integration mechanisms have been suggested. Recent genetic and functional studies have revealed the importance of host proteins involved in DNA repair and maintenance for T-DNA integration. In this article, we review our understanding of the specific function of these proteins and propose a detailed model for integration.

Agrobacterium is well known for its natural capability of trans-kingdom DNA transfer [1]. Although used mainly for plant genetic engineering [2], *Agrobacterium* can transform virtually any living cell, from other prokaryotes [3] to yeast [4] and fungi [5,6] to human cells [7]. The molecular basis for the ability of *Agrobacterium* to genetically transform its hosts has therefore been the subject of numerous studies over the past several decades [8–12].

In nature, *Agrobacterium* causes crown-gall disease, a neoplastic growth that results from the transfer of a transferred DNA (T-DNA) segment from the bacterial tumor-inducing (Ti) plasmid to the host cell, its integration into the host genome and the expression of its encoded genes [8–12]. The molecular machinery needed for T-DNA generation and transport into the host cell comprises proteins that are encoded by the bacterial chromosomal virulence (*chv*) genes and encoded by the T-plasmid virulence (*vir*) genes [8–12]. A great deal is known about the biological functions of most of the *Agrobacterium* Vir proteins, and about their function in T-DNA production and processing inside the bacterial cells. Less is known, however, about the mechanism by which the T-DNA travels into the host cell, and we have only recently begun exploring the function of host proteins in the transformation process, especially in its last two stages: T-DNA nuclear import and integration [2,12].

The genetic transformation process

The Ti plasmid carries two components needed for genetic transformation: the *vir* and T-DNA regions (Figure 1a). Although the *vir* region encodes most of the bacterial proteins necessary for processing, transport and nuclear import of the T-DNA, the T-DNA carries no specific targeting signal nor does it encode any transport or integration functions. In fact, the T-DNA borders, which comprise two 25-bp direct repeats (Figure 1b), are the only required *cis* elements that delineate and thereby determine the T-DNA region on the Ti plasmid [13], enabling replacement of the native T-DNA sequences between the borders with DNA of interest for genetic engineering [2].

Plant genetic transformation initiates with the induction of the *Agrobacterium vir* region by specific host signals, usually small phenolic and certain monosaccharide molecules [14]. The induced VirD1 and VirD2 proteins act together as a site-specific nuclease [15], which cuts the bottom strand at the T-DNA borders (Figure 1b) and releases a single-stranded (ss) T-DNA molecule (T-strand). The T-strand (Figure 1c), together with several Vir proteins, is then exported to the host cell through a channel that is formed by the VirB and VirD4 proteins [16]. Once inside the host cell cytoplasm, the T-strand presumably exists as a nucleoprotein complex (T-complex) with a single VirD2 molecule covalently attached to its 5′-end, and numerous VirE2 molecules covering its entire length. The T-complex is imported into the nucleus of the host cell with the assistance of VirD2, VirE2 and their cellular interactors [10,17]. VirD2 and VirE2 might also be involved in the final stage of the transformation, T-DNA integration [18–21], although it is generally recognized that plant factors, and not just the bacterial Vir proteins, have a crucial role in this process [8,22–25]. Although T-DNA integration is the crucial step of the transformation process, our understanding of its mechanism is still poor. In the following sections, we present the current knowledge of the function of bacterial and plant proteins in T-DNA integration and we discuss possible mechanisms for this process.

From junction sequences to illegitimate recombination models

Early studies concentrated on modeling the T-DNA integration mechanism by analyzing the junctions and structures of the integrated T-DNA (Figure 1d–f). This showed that T-DNA integration is not site-specific, and that integrated T-DNAs appeared to be distributed

Corresponding author: Tzvi Tzfira (ttzfira@ms.cc.sunysb.edu).

 0168-9525/$ - see front matter © 2004 Elsevier Ltd. All rights reserved. doi:10.1016/j.tig.2004.06.004

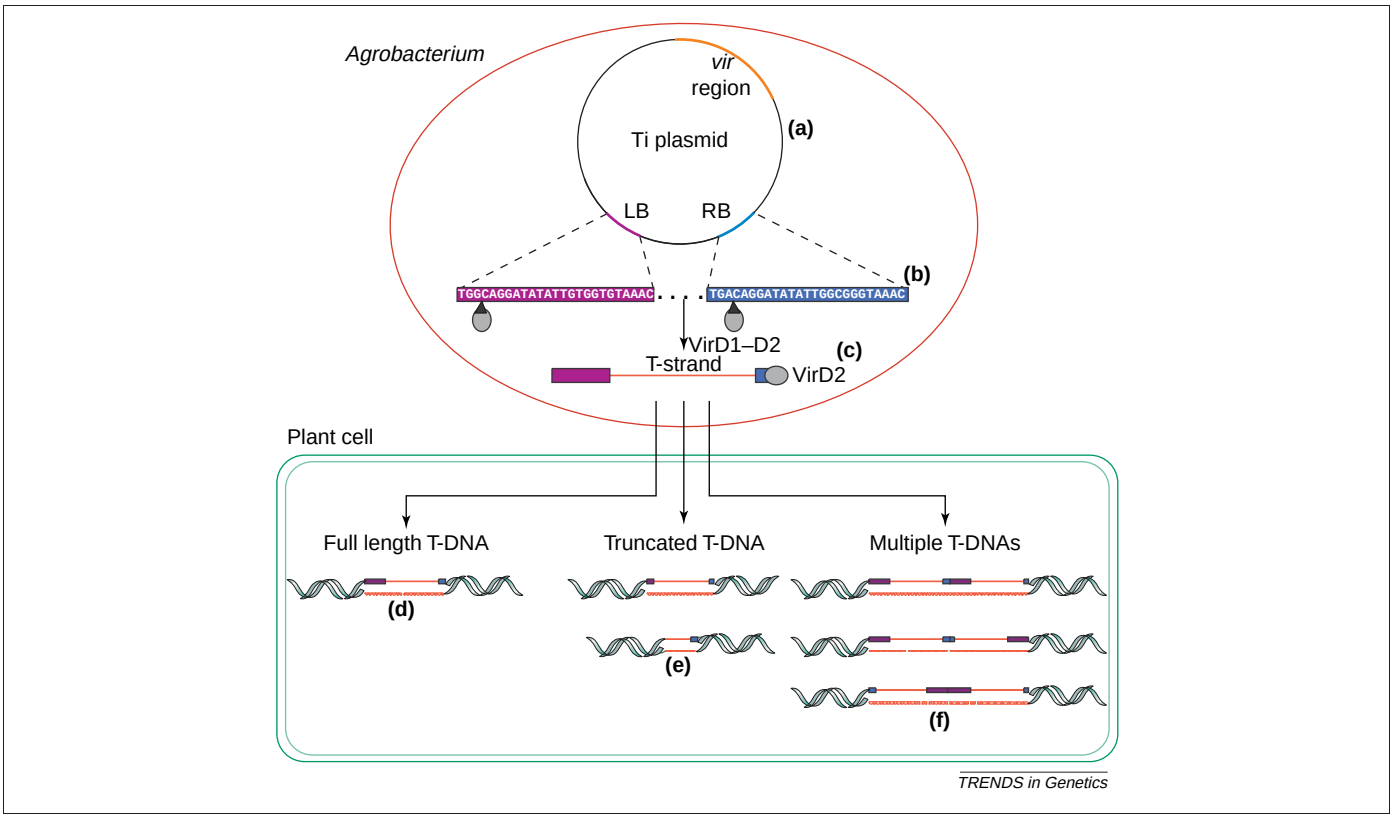

Figure 1. Schematic structure of the *Agrobacterium* Ti-plasmid and mobile and integrated T-DNA molecules. **(a)** The Ti plasmid contains the T-DNA region, which is defined by its right border (RB, in blue) and left border (LB, in purple), *in trans* to the virulence (*vir*) region (orange). **(b)** The T-DNA borders are 25-bp repeats and serve as targets for the VirD2–VirD1 endonuclease complex, which nicks between the third and the fourth nucleotides of each of the border sequences [15,71,72]. **(c)** The mobile copy of the T-DNA is released as single-stranded (ss) DNA molecule (T-strand) with a single VirD2 molecule attached to its 5′-end. The residual sequences of the T-DNA borders, which no longer have biological function, are often used as reference points for T-DNA orientation and integrity in the plant cell. T-DNA typically integrates into the host genome **(d)** as a single full-length or **(e)** truncated molecule in addition to **(f)** multiple molecules ligated to each other in various orientations.

randomly throughout the plant genome, although later studies [26,27] suggested that intergenic regions are more susceptible to T-DNA integration (discussed in the following sections). Two pioneering studies, performed on 13 T-DNA insertions in *Arabidopsis thaliana* plants and two T-DNA insertions in tobacco plants [28,29], reported that T-DNA integration can be accompanied by small deletions in the plant DNA and the invading T-DNA, and that microhomologies could be observed between the T-DNA ends and the pre-insertion sites. Moreover, deletions in the T-DNA, when they occurred, were more severe at the T-DNA 3′-end compared with its 5′-end [28,29], and longer microhomologous regions were usually observed near the 3′-end of the T-DNA. The observed microhomologies between T-DNA and the pre-insertion site led to the suggestion that both plant DNA and T-DNA end sequences have an important role in T-DNA integration. Two possible models for T-DNA integration were originally proposed [28]: the double-strand-break repair (DSBR; Figure 2a) and single-strand-gap repair (SSGR; Figure 2b) integration models. In the DSBR integration model (Figure 2a), a double-stranded (ds) break (DSB) in the target DNA is a prerequisite for T-DNA integration. The unwound or exonuclease-processed ends of the T-DNA are then annealed to the target DNA and, following ss-overhang removal by endo- and/or exonucleases, the ends are repaired and ligated. According to the SSGR integration model (Figure 2b), a single nick, later converted to a gap by a 5′ → 3′-endonuclease, is required for the initiation of T-DNA integration. The T-strand ends then partially anneal to the target DNA and the T-strand overhangs are trimmed. Following ligation of the T-strand to the target DNA, a nick is introduced in the second strand of the target DNA and extended to a gap by exonucleases. The synthesis of a strand that is complementary to the T-strand and ligation of the 3′-end of this newly synthesized complementary strand to the target DNA finalize the integration. Although the DSBR integration model requires that the T-DNA be converted to a ds form before its integration into the DSB, the SSGR integration model favors T-DNA integration as a ss molecule. Because earlier research had shown that the transformation frequency is higher for ssDNA than for dsDNA [30], it was suggested that ss T-DNA represents the preferred substrate for integration [28,29]. Supported by the fact that T-DNA is transferred into the host cell as a ss molecule, SSGR became the preferred model for T-DNA integration, and it was further suggested that VirD2 is involved directly in the integration process by recognizing nicks in the plant DNA and by assisting in the T-strand ligation to the target DNA.

VirD2 and the microhomology-based T-strand integration model

The SSGR integration model suggests a unique dual function for VirD2: it not only acts as a component of a

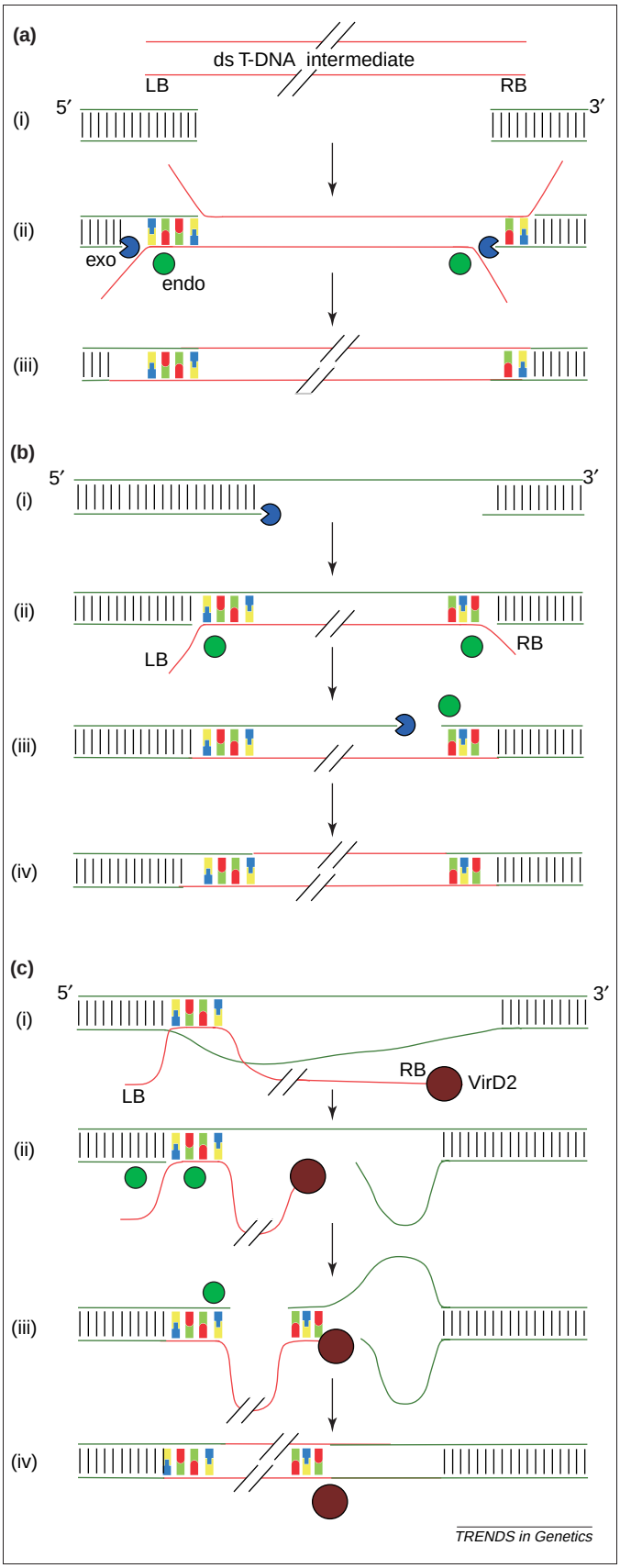

Figure 2. Double-strand-break repair (DSBR), single-strand-gap repair (SSGR) and microhomology-dependent models for *Agrobacterium* T-DNA integration. **(a)** In the DSBR model, T-DNA integration initiates with the production of a double-stranded (ds) break (DSB) in the target DNA and the conversion of the T-strand to a dsDNA form (step i). The ds ends of both the T-DNA and the DSB in the target DNA can be either unwound or processed by exonucleases. The resulting single-stranded (ss) T-DNA ends are then annealed to the ss ends of the DSB in areas of microhomology (indicated by the multicolored vertical bars), ss overhangs are removed by exonucleases (exo) and/or endonucleases (endo) (step ii) and the ends are repaired and ligated (step iii). **(b)** In the SSGR model, T-DNA integration initiates with the production of a nick, which is then converted to an ss gap by endonucleases (step i). Both 5′-end and 3′-end T-DNA regions anneal to microhomology areas within the ss portion of the target DNA gap, and the T-strand overhangs are trimmed (step ii). Following ligation of the T-strand to the target DNA, a nick in the upper strand of the target DNA is produced by endo activity and extended to a gap by exo (step iii). The gap in the upper strand of the target DNA is repaired, using the inserted T-strand as a template (step iv). **(c)** In the microhomology-dependent model, T-DNA integration initiates by annealing of the T-strand 3′-end, or its adjacent sequences, to an area of microhomology in the target DNA (step i). The 3′-end overhang of the T-strand is trimmed, and the bottom strand of the target DNA is nicked (step ii). Annealing of the 5′-end of the T-strand to the target DNA is promoted by additional microhomologies and is transiently stabilized by putative plant proteins, while the top strand of the target DNA is nicked and partly removed (step iii). Finally, complementation of the T-strand to dsDNA completes the integration. VirD2, which can assist ligation of the 5′-end of the T-strand to the target DNA, is released (step iv). Parts (a) and (b) are adapted with permission from Ref. [28], and part (c) is adapted with permission from Ref. [32]. Abbreviations: LB, left border; RB, right border.

site-specific endonuclease in the bacterial cell (Figure 1b) but also as a DNA ligase in the host cell. Indeed, VirD2 has been shown to possess DNA ligase-like activity, because it is able to re-join the products of its nuclease activity *in vivo* [20], and was thus proposed to be involved in T-DNA integration *in planta* [18]. Similar to many site-specific recombinases from several bacteriophages [31], VirD2 contains a conserved H-R-Y integrase motif, which most probably resides within its proposed active site. R-to-G mutations in this H-R-Y motif affect the integration process, rendering it less precise albeit no less efficient [18]. The loss of integration precision, shown by the loss of part of the 5′-end of the T-DNA sequence, and the unaffected nuclear import capabilities of this mutated VirD2 further supported the function of VirD2 during the integration step itself. Additional analyses of T-DNA insertion sites [28,29] revealed microhomologies between T-DNA borders and pre-insertion sites, prompting the introduction of a revised T-DNA model (Figure 2c) [32]. According to this model, T-DNA integration initiates by microhomology-dependent annealing of the T-strand 3′-end, or its adjacent sequences, to the host DNA that has been unwound partially by an as yet unknown mechanism. The 3′-end overhang of the T-strand is trimmed, and the bottom strand of the target DNA is nicked. Then the 5′-end of the T-strand anneals to the target DNA, while the top strand of the target DNA is nicked and partly removed. Complementation of the T-strand to dsDNA completes the integration.

The revised ss T-DNA integration model assumes that T-DNA integration relies on microhomologies and on the ligase function of VirD2 within the plant cell [32]. However, if VirD2 acts directly as a ligase, the R-to-G mutation in its active site should render it completely inactive for T-DNA integration, and should affect not only the precision of T-DNA integration but also the overall transformation efficiency. The limited effect of this mutation suggests that other motifs, and/or other factors, are involved in the T-DNA integration and ligation steps. Indeed, another motif within VirD2 has been shown to be important for the infectivity of *Agrobacterium* [33]. This C-terminal motif, termed omega, is not involved directly in VirD2 endonuclease activity or T-DNA nuclear targeting,

Table 1. Yeast and *Arabidopsis* DNA repair genes and their function in T-DNA integration[a]

Host gene	Cellular function	Yeast		Plants		Refs
		HR	NHR	*In planta* transformation assay	Root tumor assay	
Ku70–AtKU70	Binds and stabilizes dsDNA ends	Not required	Required	ND[a]	ND	[55,56]
Ku80–AtKU80	Binds and stabilizes dsDNA ends	ND	ND	Susceptible and/or reduced susceptibility	Resistant[b]	[22,25]
Lig4–AtLIG4	ATP-dependent DNA ligase	Not required	Required	Susceptible and/or reduced susceptibility	Susceptible	[22,55,56,62]
Sir4	Interacts with Ku70; forms a complex with Sir3–Sir2; possible involvement in histone acetylation	ND	Required	Unknown homolog		[55,56]
Rad50	Rad50, Mre11 and Xrs2 form a complex that has probable exonuclease activity, which functions in NHR and HR	Not required	Required	ND	ND	[55,56]
Mre11–AtMRE11		Not required	Required	ND	ND	[55,56]
Xrs2		Not required	Required	Unknown homolog		[55,56]
Rad51–AtRAD5	ATP-dependent homologous function in DNA pairing	Required	Not required	ND	Reduced susceptibility (*rad5* mutant)	[55–57]
Rad52	SsDNA binding and annealing	Required	Not required	Unknown homolog		[55,56]
AtH2A	Core histone; DNA packaging	Unknown homolog		Susceptible	Resistant	[24]

[a]Abbreviations: HR, homologous recombination; NHR, non-homologous recombination; ND, not determined; ss, single stranded.
[b]J. Li *et al.*, unpublished.

although replacing four of its amino acids with two serine residues resulted in a less virulent *Agrobacterium* strain [33]. Further studies of this domain attributed the reduction in transformation efficiency mainly to a reduction in stable transformation and not a reduction in T-DNA nuclear import [21,34]. It was thus shown that VirD2 activity in T-DNA integration is not limited to its putative function as a DNA ligase through its conserved H-R-Y integrase motif but most likely consists of several functions, determined by different motifs.

Whether VirD2 is indeed a DNA ligase remains controversial. Although earlier research suggested that VirD2 can ligate ssDNA [20], a recent study indicated that VirD2 is not a DNA ligase [35]. In this study, ligation of ssDNA to various target oligonucleotides was assayed *in vitro*. A ligation–integration reaction was promoted by plant extracts or prokaryotic DNA ligase, whereas purified VirD2 failed to facilitate this reaction. It was thus concluded that a plant ligase, and not VirD2, is the functional enzyme during T-DNA integration. However, *in vivo*, VirD2 and plant proteins might still function together: VirD2 could recruit plant ligase to the integration site or plant factors might be required for VirD2 to act as a ligase. Interestingly, an *Arabidopsis* mutant, in which *AtLIG4*, a DNA ligase-encoding gene, was knocked out (Table 1), was not impaired in T-DNA integration [23], suggesting that other plant cell ligases (e.g. DNA ligase 1) function during T-DNA integration.

Complex T-DNA integration patterns and integration of ds T-DNA

The revised ss T-DNA integration model proposed by Tinland *et al.* [32] can easily explain the integration of single-copy T-DNA, even without postulating VirD2 function as a DNA ligase. However, it does not provide a mechanism for the production of complex T-DNA inserts (Figure 1f), which are composed of two or more T-DNA copies arranged in the same (direct) or in reverse (indirect) orientation relative to each other, with or without filler DNA between them (e.g. [36–38]). It has been suggested that T-DNA rearrangements occur before integration [37,38]. Using two binary plasmids (carried by two different Agrobacteria) containing two different selectable markers on their T-DNAs, doubly-transformed *Arabidopsis* plants were obtained in which both T-DNAs, each derived from a separate *Agrobacterium*, were found to integrate at the same location on the plant chromosome [39]. In some cases, sequence analysis of these T-DNA insertions revealed a precise fusion between two right border ends ('heads'), whereas the left border junctions ('tails') often exhibited imprecise fusion and/or contained filler DNA [36–38]. Because the T-DNA molecules can not recombine head-to-head at their right borders when they are in ssDNA form but can do so when in ds form [37,38], it was suggested that T-strands must be converted to ds molecules before their integration. A model for T-DNA integration via ds intermediates was thus proposed [38], describing the insertion of multiple T-DNAs into the same chromosomal site (Figure 3). This model assumes a simple ligation mechanism for head-to-head (right border-to-right border) orientation of two T-DNAs (Figure 3a), and an illegitimate recombination mechanism for head-to-tail (right border-to-left border; Figure 3b) or tail-to-tail (left border-to-left border; Figure 3c) orientations. Illegitimate recombination is also assumed to mediate the integration of both the head and the tail of the T-DNA molecule (or of several conjoined T-DNA molecules) into the host target DNA (Figure 3d,e). According to this model, the VirD2 molecule remains attached to the 5′-end (corresponding to the right border or the 'head') of the T-strand, even after its conversion to a ds form. Potentially, VirD2 molecules that are attached to two T-DNAs can bind to each other and bring the T-DNAs together in a head-to-head orientation before their ligation by host factors, explaining the

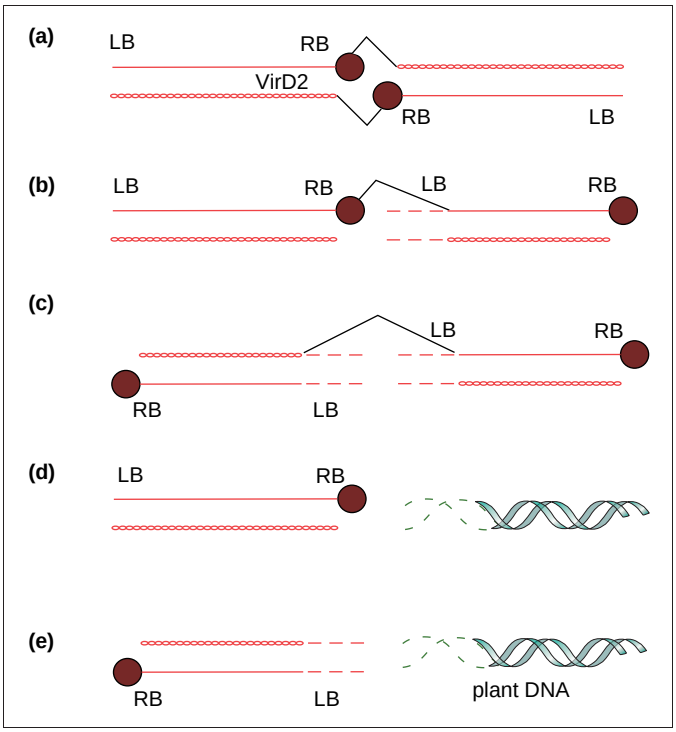

Figure 3. A model for co-integration of two individual T-DNA molecules into the same target DNA site via double-stranded (ds) intermediates. The integration of two T-DNAs in different orientations relative to each other probably occurs by different mechanisms, specific for each orientation. **(a)** Head-to-head fusion of two T-DNAs occurs via simple ligation of two ds T-DNA intermediates, leading to the precise joining of the molecules at their T-DNA right borders (RB). Non-homologous (illegitimate) recombination is the preferred mechanism for the production of **(b)** head-to-tail or **(c)** tail-to-tail orientations, leading to imperfect joining between the T-DNA LB and RB, which can be accompanied by insertion of filler DNA (broken lines). The integration of the conjoint ds T-DNA intermediates into the plant genome also occurs via illegitimate recombination, **(d)** at the right or **(e)** left T-DNA borders, with or without the insertion of filler DNA (reproduced with permission from Ref. [37]).

relatively lower frequency of tail-to-tail integrations, as compared with head-to-head or head-to-tail integrations of the T-DNA molecules [39]. The ability of VirD2 molecules to interact with each other [40], as well as with a cellular DNA-binding protein [41], further supports the role of VirD2 in recruiting individual T-DNAs to ligation before integration.

DSBs represent a major pathway for T-DNA integration

The model of T-DNA integration via ds intermediates (Figure 3) does not explain the origin of filler DNA, which is sometimes found between repeating copies of co-integrated T-DNA molecules (e.g. [28,29,42,43]). This filler DNA is also observed during the integration of transfected plasmids into the plant genome. For example, DNA inserts between the integrated plasmid and the tobacco genome were reported to contain mostly scrambled sequences that were derived from the plasmid or tobacco genomes [44]. The authors suggested that DSBs are involved in foreign DNA integration, in formation of deletions and in insertion of filler DNA at the junctions. Indeed, recently it has been shown that induction of DSBs in the tobacco DNA by transgenic expression of a rare-cutting restriction enzyme results in frequent incorporation of T-DNA into these breaks [45]. In addition, the repair of DSBs was sometimes associated with the insertion of other genomic sequences into the break, which could explain the presence of filler DNA between multiple co-integrated copies of T-DNA [37,38,43]. It was thus proposed that DSBs might have a significant role in T-DNA integration and that the number of naturally-occurring DSBs in the host genome might represent the limiting factor for T-DNA integration [45]. Nevertheless, the possibility of ss T-DNA integration into DSBs by a synthesis-dependent annealing (SDA) mechanism [45] was not ruled out. The second option of a ds T-DNA intermediate integrating into a DSB by either two consecutive single-strand-annealing (SSA) reactions or by a single SSA reaction followed by a simple ligation step was also proposed [45].

In a comparative analysis of the relationship between DSB repair and T-DNA integration in *Arabidopsis*, T-DNA insertions were found to be associated with much smaller deletions [43] than those typical for DSB repair [46,47]. It was also found that ~40% of all T-DNA junctions harbored filler DNA [43], which is usually not introduced during DSB repair in *Arabidopsis* [46], and that the composition of this filler DNA differed from that reported for DSBs repair in other plant species (e.g. tobacco and maize) [43]. These observations led the authors to suggest that although DSB repair and T-DNA integration are probably related they still possess some distinct characteristics, and that these differences might depend on the plant species that differ in their DSB repair mechanisms [46,47].

In a modified version of the DSB induction system, the recognition site for a rare-cutting restriction enzyme was placed not only in the target plant DNA but also in the invading T-DNA [48,49]. Induction of DSBs in the host DNA by the restriction enzyme resulted, in numerous cases, in the integration of truncated T-DNA molecules, which had been digested previously, *in vivo*, by the same restriction enzyme [48,49]. The precise ligation of such truncated T-DNA molecules into the restriction site within the target genome provided the first direct evidence that T-DNA molecules are, at least partially, converted to a ds form before integration is completed. In addition, by avoiding direct selection for T-DNA integration into DSBs, these breaks are most probably the preferred sites for integration [48]. X-ray irradiation, which is known to cause DSBs [50], enhances transgene integration [51] and further supports the role of DSBs in T-DNA integration. Thus, T-DNA integration into DSBs might represent a default mode of T-DNA entry into the host genome [49]. In this scenario, naturally occurring DSBs and/or DSB-associated DNA-repair activity might act as molecular 'bait' for incoming T-DNA molecules [48]. Such a mechanism might help to explain the well-known phenomenon of integration of multiple T-DNA molecules, even of those that originated from different *Agrobacterium* cells (discussed previously), into a single site in the host genome [36,38,39].

DNA repair proteins and their role in T-DNA integration

Because DSBs provide an important pathway for T-DNA integration and VirD2 most probably does not act as a DNA ligase *per se*, this suggests a role for host factors in T-DNA integration. Using a yeast-based system in which T-DNA integration into the yeast genome can be

directed to either a homologous recombination (HR) [52,53] or a non-homologous (illegitimate) recombination (NHR) pathway [54], the role of various non-homologous end-joining (NHEJ) proteins was examined [55,56]. It was first shown that Ku70, Rad50, Mre11, Xrs2, Lig4 and Sir4 are all required for T-DNA integration by NHR (Table 1; [55]). However, when T-DNA integration occurred by HR, only Rad51 and Rad52, but not Rad50, Mre11, Xrs2, Lig4 or Ku70, were essential. Furthermore, in the absence of Ku70, a dsDNA-binding protein, T-DNA integration occurred only via the HR pathway (Table 1; [55]), whereas in the absence of Rad52, an ssDNA-binding protein, T-DNA integration was possible only by NHR [56]. When both the HR and the NHR pathways were blocked, by mutation of the *KU70* and *RAD52* genes, T-DNA integration was inhibited completely [56]. It was thus suggested that Ku70 and Rad52 are the key enzymes for T-DNA integration and that, in yeast, the pathway for T-DNA integration can be determined by the activity of specific DNA repair genes [56].

In plants, T-DNA integration occurs mainly through NHR, even when the T-DNA shares high homology with the host genome, suggesting that NHEJ proteins are the key players in the process of T-DNA integration. Indeed, two *Arabidopsis* mutants that were hypersensitive to radiation, *rad5* and *uvh1*, were also deficient in T-DNA integration [57]. Although the *uvh1* mutant can still be transformed stably by *Agrobacterium* [58–60], the hypersensitivity of *uvh1* to bleomycin, an antibiotic that causes single-stranded breaks (SSBs) and DSBs through the formation of an intermediate iron complex [61] and the close homology between the *Arabidopsis AtRAD5* and the yeast *RAD51* genes lend further support for the involvement of DSBs and NHEJ proteins in T-DNA integration *in planta*.

Two additional *Arabidopsis* proteins, a DNA ligase AtLIG4 and a dsDNA-binding protein AtKU80 that are involved in DSB repair via NHEJ were analyzed for their role in T-DNA integration but with no conclusive results. *AtLIG4* was dispensable for T-DNA integration in *Arabidopsis* [23] using a root tumor formation assay [62] but was reported to be either required [23] or dispensable [23] for T-DNA integration using an *in planta* transformation procedure that targets the host germ-line cells [63]. Differences between host factors that are required for the root tumor formation assay and *in planta* transformation by the floral dip method might reflect differences in the integration pathways between the target plant tissues (i.e. roots [62] and ovum cells [63]). Indeed, several resistant to *Agrobacterium* transformation (*rat*) mutants of *Arabidopsis* identified by the root tumor formation assay [62] remained susceptible to *in planta* transformation [64].

The involvement of AtKU80 in integration is even less conclusive; it was reported to be both required [22] and dispensable [25] for T-DNA integration, using the same *in planta* transformation assay. These apparently contradictory results might simply arise from the nature of the *in planta* transformation, which is performed under relatively uncontrolled conditions. However, recent results indicate that an *Arabidopsis* mutant in the *AtKU80* gene is completely resistant to *Agrobacterium* infection in the root tumor formation assay (J. Li *et al*., unpublished), supporting the role of AtKU80 in T-DNA integration.

In addition to plant DNA repair enzymes, chromatin structure and DNA packaging proteins can also affect the integration process, as was evident from the *rat* phenotype (i.e. resistant to *Agrobacterium* infection [62,64]) of many *Arabidopsis* mutants disrupted in various histone- and histone deacetylase-encoding genes [62,24]. Consistent with this, a recent genome-wide T-DNA insertional analysis revealed a bias for T-DNA integration in favor of intergenic regions, promoters and 5′ and 3′ untranslated regions [26,27]. Thus, transcription and DNA repair events, which result in temporary unpacking of DNA, might serve as 'hot points' of attraction for T-DNA integration, whereas the NHEJ proteins perform the actual integration process at these sites.

A possible model for DSB-mediated T-DNA integration

We integrated the current knowledge on the functions of DSBs, ds T-DNA intermediates and NHEJ proteins into a unified model for T-DNA integration in plant and yeast cells (Figure 4). In this model, the invading T-strand can undergo minor degradation at its unprotected 3′-end on the way to the nucleus, but its 5′-end remains intact as a result of shielding by the covalently attached VirD2 protein. Once in the nucleus, the T-DNA is converted rapidly into a dsDNA molecule; this process can result in the further loss of several bases from the T-DNA 3′-end as a result of the random-priming nature of the complementation process. This ds T-DNA intermediate will undergo integration into the host DNA via the NHR or HR pathways.

In the NHR pathway, which occurs in plants or in yeast in which the HR pathway has been disabled, VirD2 is replaced by a AtKU70–AtKU80 heterodimer that binds to both ends of the dsDNA molecule, stabilizes them and recruits the DNA-dependent protein kinase (DNA-PK). The role of DNA-PK in DSB repair is still unknown, but it might regulate the repair process through phosphorylation [65]. In parallel, transcription and/or other DNA maintenance reactions result in DNA unpacking, possibly through modification of H2A, and in the formation of a DSB. Here, also, the AtKU70–AtKU80–DNA-PK complex is recruited to the DSB site. Following the processing of the DNA ends, which might result in the addition or removal of a few base pairs, the T-DNA is ligated by the XRCC4–AtLIG4 (or perhaps LIG1 and/or LIG2) complex [65] to the DSB. Alternatively, several dsDNA molecules can become ligated to each other by the XRCC4–AtLIG4 complex, before their integration into the host genome. The XRCC4–AtLIG4 complex then mediates integration of these conjoint ds T-DNA intermediates into the DSB site. Although necessary for T-DNA integration and DSB repair, the exact function of yeast Rad50 and Mre11 or their plant homologs in this process is still unknown [65].

Alternatively, ds T-DNA intermediates might integrate into DSBs in the host genome by HR. Although this integration pathway occurs mainly in yeast, plants might also, albeit infrequently, utilize an HR-like mechanism. In this scenario, the DNA ends are first recognized by Rad52 (or its plant homologs) and not by the Ku70–Ku80

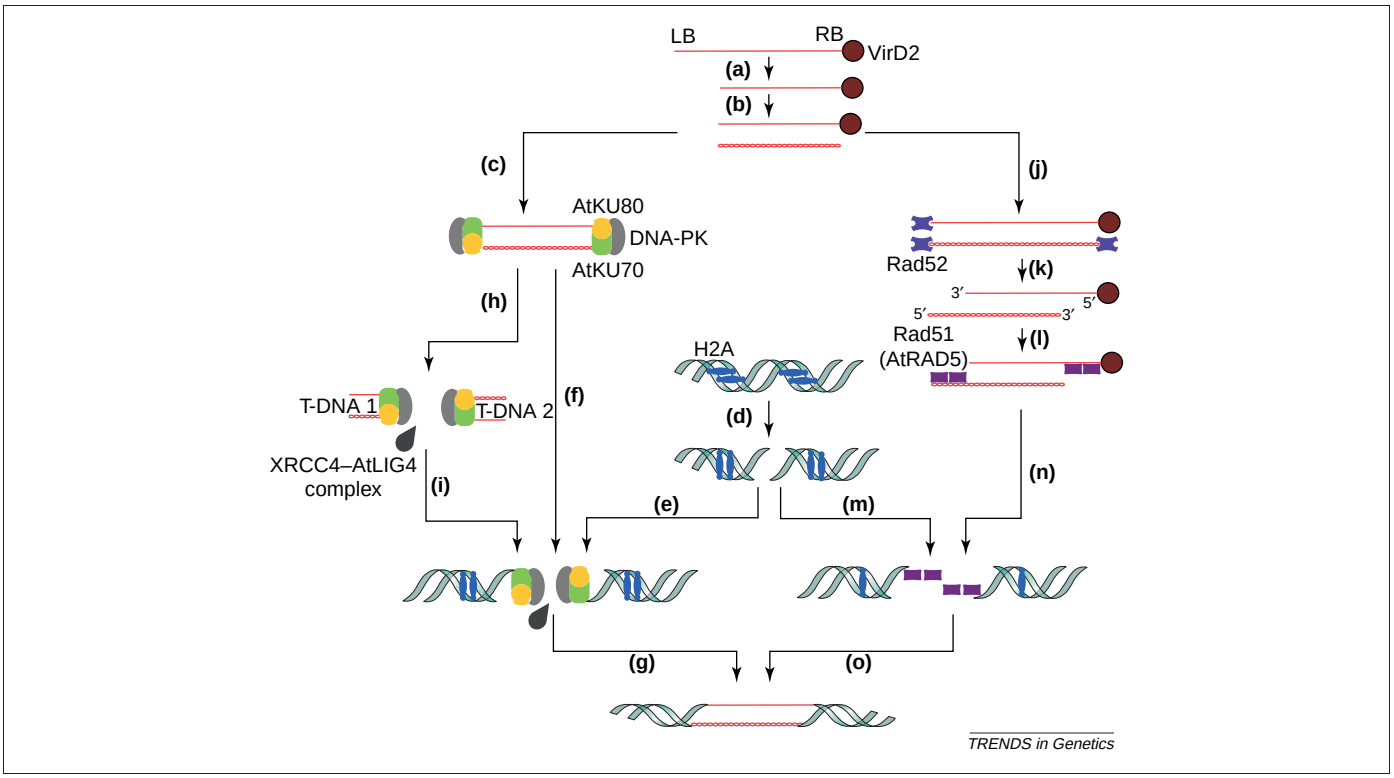

Figure 4. A unified model for T-DNA integration into double-strand (ds) breaks (DSBs) in plant and yeast cells. **(a)** A T-strand can undergo minor degradation at its unprotected 3′-end on the way to the host cell nucleus, which can lead to loss of part of the 3′-end of the T-DNA sequence due to imperfect complementation. **(b)** Within the nucleus, the T-strand is are converted into a ds T-DNA intermediate, which can be processed by two alternative pathways. In the first pathway observed mostly in plants, integration occurs by a non-homologous (illegitimate) recombination (NHR). **(c)** VirD2 is replaced with the AtKU70–AtKU80 heterodimer at both ds T-DNA ends, to which DNA-dependent protein kinase (DNA-PK) is recruited. **(d)** In parallel, DNA unpacking, via H2A modification, and formation of a DSB in the target DNA leads to recruitment of the AtKU70–AtKU80–DNA-PK complex to the DSB site **(e)**, initiating DSB DNA repair, to which ds T-DNA intermediate is attracted **(f)**. **(g)** The XRCC4–AtLIG4 complex probably mediates integration of the ds T-DNA into the DSB site. **(h)** In addition, several ds T-DNA molecules (indicated here as T-DNA 1 and T-DNA 2) can be joined together by the AtKU70–AtKU80–DNA-PK and XRCC4–AtLIG4 machinery and **(i)** only later be attracted to and **(g)** integrate into the DSB site. In an alternative pathway, known to occur in yeast cells but also possible in plants, ds T-DNA intermediates can also integrate by homologous recombination (HR). **(j)** In this scenario, DNA ends are first recognized by Rad52, or its potential plant homolog, with or without the displacement of VirD2 and processed by the Rad52(Rad50)–Mre11 complex, leaving 5′-end overhangs **(k)**. Rad51, or its plant homolog AtRAD5, is then polymerized on the ssDNA overhangs of both **(l)** the T-DNA intermediate and **(m)** the target DNA. The nucleoprotein filaments at the ends of the integrating dsDNA search for microhomologies to DSBs in the host genome and **(n)** anneal to these regions **(o)** mediating their ligation into DSB.

heterodimer. Rad52 probably binds to ss ends [66] rather than ds ends, whereas VirD2 might still be protecting the 5′-end of the T-strand. The Rad52-bound DNA ends are then processed by the Rad52(Rad50)–Mre11 nucleolytic complex, resulting in 5′-end overhangs. Rad51, or its plant homolog AtRAD5, is then polymerized onto the ssDNA overhangs to form nucleoprotein filaments. Concurrently, the ends of the DSBs in the host genome undergo similar modification by the Rad52(Rad50)–Mre11 complexes and associate with Rad51, becoming ‘competent’ for dsDNA integration. The nucleoprotein filaments at the ends of the integrating ds T-DNA intermediate search for microhomologies to DSBs in the host genome and anneal to these regions, the 5′-end overhangs are filled-in, and the dsDNA is ligated into the DSB. This mechanism explains why the 5′-end of the T-DNA, which is protected by VirD2, is more conserved than its 3′-end during integration [18,67,68].

Future prospects

In the past several years, T-DNA integration research has shifted from analyzing integration junction sequences to identifying the molecules that are involved in the integration process. Using yeast and *Arabidopsis* as models, several host genes that are involved in T-DNA integration have been identified, although many others are still awaiting discovery. This research trend is expected to continue, leading not only to the identification of all of the proteins involved in the integration process but also to elucidation of the precise molecular mechanism leading to successful integration. In addition, understanding the molecular pathway(s) of T-DNA integration will lead to the development of new tools and approaches for controlling this process. Such tools might then be used for efficient gene-targeting- and gene-replacement techniques, which are highly valuable for plant research and biotechnology [69,70].

Acknowledgements

We apologize to colleagues whose original research has not been cited due to the lack of space. The work in our laboratories is supported by grants from NIH, NSF Functional Genomic Initiative and NSF 2010 Program, USDA, BARD and BSF to V.C, and by grants from BARD and HFSP to T.T.

References

1 Stachel, S.E. and Zambryski, P.C. (1989) Generic trans-kingdom sex? *Nature* 340, 190–191
2 Gelvin, S. (2003) Improving plant genetic engineering by manipulating the host. *Trends Biotechnol.* 21, 95–98

3 Kelly, B.A. and Kado, C.I. (2002) *Agrobacterium*-mediated T-DNA transfer and integration into the chromosome of *Streptomyces lividans*. *Mol. Plant Pathol.* 3, 125–134
4 Piers, K.L. *et al.* (1996) *Agrobacterium tumefaciens*-mediated transformation of yeast. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1613–1618
5 de Groot, M.J. *et al.* (1998) *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. *Nat. Biotechnol.* 16, 839–842
6 Gouka, R.J. *et al.* (1999) Transformation of *Aspergillus awamori* by *Agrobacterium tumefaciens*-mediated homologous recombination. *Nat. Biotechnol.* 17, 598–601
7 Kunik, T. *et al.* (2001) Genetic transformation of HeLa cells by *Agrobacterium*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 1871–1876
8 Gelvin, S.B. (2000) *Agrobacterium* and plant genes involved in T-DNA transfer and integration. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 51, 223–256
9 Gelvin, S.B. (2003) *Agrobacterium*-mediated plant transformation: the biology behind the 'gene-jockeying' tool. *Microbiol. Mol. Biol. Rev.* 67, 16–37
10 Tzfira, T. and Citovsky, V. (2000) From host recognition to T-DNA integration: the function of bacterial and plant genes in the *Agrobacterium*-plant cell interaction. *Mol. Plant Pathol.* 1, 201–212
11 Zupan, J. *et al.* (2000) The transfer of DNA from *Agrobacterium tumefaciens* into plants: a feast of fundamental insights. *Plant J.* 23, 11–28
12 Tzfira, T. and Citovsky, V. (2002) Partners-in-infection: host proteins involved in the transformation of plant cells by Agrobacterium. *Trends Cell Biol.* 12, 121–129
13 Zambryski, P.C. (1992) Chronicles from the *Agrobacterium*-plant cell DNA transfer story. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 43, 465–490
14 Citovsky, V. *et al.* (1992) *Agrobacterium*–plant cell interaction: induction of *vir* genes and T-DNA transfer. In *Molecular Signals in Plant-Microbe Communications* (Verma, D.P.S., ed.), pp. 169–198, CRC Press, Inc
15 Scheiffele, P. *et al.* (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
16 Christie, P.J. and Vogel, J.P. (2000) Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.* 8, 354–360
17 Tzfira, T. *et al.* (2000) Nucleic acid transport in plant-microbe interactions: the molecules that walk through the walls. *Annu. Rev. Microbiol.* 54, 187–219
18 Tinland, B. *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
19 Rossi, L. *et al.* (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
20 Pansegrau, W. *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
21 Mysore, K.S. *et al.* (1998) Role of the *Agrobacterium tumefaciens* VirD2 protein in T-DNA transfer and integration. *Mol. Plant Microbe Interact.* 11, 668–683
22 Friesner, J. and Britt, A.B. (2003) Ku80- and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. *Plant J.* 34, 427–440
23 van Attikum, H. *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
24 Mysore, K.S. *et al.* (2000) An *Arabidopsis* histone H2A mutant is deficient in *Agrobacterium* T-DNA integration. *Proc. Natl. Acad. Sci. U. S. A.* 97, 948–953
25 Gallego, M.E. *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
26 Alonso, J.M. *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657
27 Rosso, M.G. *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
28 Mayerhofer, R. *et al.* (1991) T-DNA integration: a mode of illegitimate recombination in plants. *EMBO J.* 10, 697–704
29 Gheysen, G. *et al.* (1991) Illegitimate recombination in plants: a model for T-DNA integration. *Genes Dev.* 5, 287–297
30 Rodenburg, K.W. *et al.* (1989) Single-stranded DNA used as an efficient new vehicle for transformation of plant protoplasts. *Plant Mol. Biol.* 13, 711–719
31 Argos, P. *et al.* (1986) The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* 5, 433–440
32 Tinland, B. and Hohn, B. (1995) Recombination between prokaryotic and eukaryotic DNA: integration of *Agrobacterium tumefaciens* T-DNA into the plant genome. *Genet. Eng.* 17, 209–229
33 Shurvinton, C.E. *et al.* (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
34 Narasimhulu, S.B. *et al.* (1996) Early transcription of *Agrobacterium* T-DNA genes in tobacco and maize. *Plant Cell* 8, 873–886
35 Ziemienowicz, A. *et al.* (2000) Plant enzymes but not *Agrobacterium* VirD2 mediate T-DNA ligation *in vitro*. *Mol. Cell. Biol.* 20, 6317–6322
36 Krizkova, L. and Hrouda, M. (1998) Direct repeats of T-DNA integrated in tobacco chromosome: characterization of junction regions. *Plant J.* 16, 673–680
37 De Buck, S. *et al.* (1999) The DNA sequences of T-DNA junctions suggest that complex T-DNA loci are formed by a recombination process resembling T-DNA integration. *Plant J.* 20, 295–304
38 De Neve, M. *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
39 De Block, M. and Debrouwer, D. (1991) Two T-DNAs co-transformed into *Brassica napus* by a double *Agrobacterium tumefaciens* infection are mainly integrated at the same locus. *Theor. Appl. Genet.* 82, 257–263
40 Relic, B. *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
41 Bakó, L. *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
42 Bakkeren, G. *et al.* (1989) Recovery of *Agrobacterium tumefaciens* T-DNA molecules from whole plants early after transfer. *Cell* 57, 847–857
43 Windels, P. *et al.* (2003) T-DNA integration in *Arabidopsis* chromosomes. Presence and origin of filler DNA sequences. *Plant Physiol.* 133, 2061–2068
44 Gorbunova, V.V. and Levy, A.A. (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Res.* 25, 4650–4657
45 Salomon, S. and Puchta, H. (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J.* 17, 6086–6095
46 Kirik, A. *et al.* (2000) Species-specific double-strand break repair and genome evolution in plants. *EMBO J.* 19, 5562–5566
47 Orel, N. and Puchta, H. (2003) Differences in the processing of DNA ends in *Arabidopsis thaliana* and tobacco: possible implications for genome evolution. *Plant Mol. Biol.* 51, 523–531
48 Tzfira, T. *et al.* (2003) Site-specific integration of *Agrobacterium* T-DNA via double-stranded intermediates. *Plant Physiol.* 133, 1011–1023
49 Chilton, M-D.M. and Que, Q. (2003) Targeted integration of T-DNA into the tobacco genome at double-strand breaks: new insights on the mechanism of T-DNA integration. *Plant Physiol.* 133, 956–965
50 Leskov, K.S. *et al.* (2001) When X-ray-inducible proteins meet DNA double strand break repair. *Semin. Radiat. Oncol.* 11, 352–372
51 Kohler, F. *et al.* (1989) Enhancement of transformation rates in higher plants by low-dose irradiation: are DNA repair systems involved in incorporation of exogenous DNA into the plant genome? *Plant Mol. Biol.* 12, 189–199
52 Bundock, P. *et al.* (1995) Trans-kingdom T-DNA transfer from

Agrobacterium tumefaciens to *Saccharomyces cerevisiae*. *EMBO J.* 14, 3206–3214
53 Risseeuw, E.P. *et al.* (1996) Integration of an insertion-type transferred DNA vector from *Agrobacterium tumefaciens* into the *Saccharomyces cerevisiae* genome by gap repair. *Mol. Cell. Biol.* 16, 5924–5932
54 Bundock, P. and Hooykaas, P.J.J. (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
55 van Attikum, H. *et al.* (2001) Non-homologous end-joining proteins are required for *Agrobacterium* T-DNA integration. *EMBO J.* 20, 6550–6558
56 van Attikum, H. and Hooykaas, P.J.J. (2003) Genetic requirements for the targeted integration of *Agrobacterium* T-DNA in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 31, 826–832
57 Sonti, R.V. *et al.* (1995) *Arabidopsis* mutants deficient in T-DNA integration. *Proc. Natl. Acad. Sci. U. S. A.* 92, 11786–11790
58 Nam, J. *et al.* (1998) *Agrobacterium tumefaciens* transformation of the radiation hypersensitive *Arabidopsis thaliana* mutants *uvh1* and *rad5*. *Mol. Plant Microbe Interact.* 11, 1136–1141
59 Preuss, S.B. *et al.* (1999) Radiation-sensitive *Arabidopsis* mutants are proficient for T-DNA transformation. *Mol. Gen. Genet.* 261, 623–626
60 Chateau, S. *et al.* (2000) Competence of *Arabidopsis thaliana* genotypes and mutants for *Agrobacterium tumefaciens*-mediated gene transfer: role of phytohormones. *J. Exp. Bot.* 51, 1961–1968
61 Myung, K. and Kolodner, R.D. (2003) Induction of genome instability by DNA damage in *Saccharomyces cerevisiae*. *DNA Repair (Amst.)* 2, 243–258
62 Zhu, Y. *et al.* (2003) Identification of *Arabidopsis rat* mutants. *Plant Physiol.* 132, 494–505
63 Ye, G.N. *et al.* (1999) *Arabidopsis* ovule is the target for *Agrobacterium in planta* vacuum infiltration transformation. *Plant J.* 19, 249–257
64 Mysore, K.S. *et al.* (2000) *Arabidopsis* ecotypes and mutants that are recalcitrant to *Agrobacterium* root transformation are susceptible to germ-line transformation. *Plant J.* 21, 9–16
65 van Gent, D.C. *et al.* (2001) Chromosomal stability and the DNA double-stranded break connection. *Nat. Rev. Genet.* 2, 196–206
66 Ristic, D. *et al.* (2003) Rad52 and Ku bind to different DNA structures produced early in double-strand break repair. *Nucleic Acids Res.* 31, 5229–5237
67 Brunaud, V. *et al.* (2002) T-DNA integration into the *Arabidopsis* genome depends on sequences of pre-insertion sites. *EMBO Rep.* 3, 1152–1157
68 Tinland, B. (1996) The integration of T-DNA into plant genomes. *Trends Plant Sci.* 1, 178–184
69 Puchta, H. (2003) Towards the ideal GMP: homologous recombination and marker gene excision. *J. Plant Physiol.* 160, 743–754
70 Puchta, H. (2002) Gene replacement by homologous recombination in plants. *Plant Mol. Biol.* 48, 173–182
71 Wang, K. *et al.* (1987) Site-specific nick occurs within the 25 bp transfer promoting border sequence following induction of vir gene expression in *Agrobacterium tumefaciens*. *Science* 235, 587–591
72 Filichkin, S.A. and Gelvin, S.B. (1993) Formation of a putative relaxation intermediate during T-DNA processing directed by the *Agrobacterium tumefaciens* VirD1,D2 endonuclease. *Mol. Microbiol.* 8, 915–926

Elsevier.com – Dynamic New Site Links Scientists to New Research & Thinking

Elsevier.com has had a makeover, inside and out. Designed for scientists' information needs, the new site, launched in January, is powered by the latest technology with customer-focused navigation and an intuitive architecture for an improved user experience and greater productivity.

Elsevier.com's easy-to-use navigational tools and structure connect scientists with vital information – all from one entry point. Users can perform rapid and precise searches with our advanced search functionality, using the FAST technology of Scirus.com, the free science search engine. For example, users can define their searches by any number of criteria to pinpoint information and resources. Search by a specific author or editor, book publication date, subject area – life sciences, health sciences, physical sciences and social sciences – or by product type. Elsevier's portfolio includes more than 1800 Elsevier journals, 2200 new books per year, and a range of innovative electronic products. In addition, tailored content for authors, editors and librarians provides up-to-the-minute news, updates on functionality and new products, e-alerts and services, as well as relevant events.

Elsevier is proud to be a partner with the scientific and medical community. Find out more about who we are in the About section: our mission and values and how we support the STM community worldwide through partnerships with libraries and other publishers, and grant awards from The Elsevier Foundation.

As a world-leading publisher of scientific, technical and health information, Elsevier is dedicated to linking researchers and professionals to the best thinking in their fields. We offer the widest and deepest coverage in a range of media types to enhance cross-pollination of information, breakthroughs in research and discovery, and the sharing and preservation of knowledge. Visit us at Elsevier.com.

Elsevier. Building Insights. Breaking Boundaries.