

---

# Identification of T-DNA Integration Sites: TAIL-PCR and Sequence Analysis

# 19

Jaehyuk Choi, Junhyun Jeon, and Yong-Hwan Lee

---

## 19.1 Introduction

*Agrobacterium tumefaciens* is a bacterial plant pathogen causes crown gall disease. *A. tumefaciens* can transfer a DNA segment from its tumor-inducing plasmid (Drummond et al. 1977). The transferred DNA (called T-DNA) is enclosed in-between the so-called left and right borders (LB and RB). Its capability of gene transfer has been widely used as a genetic engineering tool in plants (Tinland 1996). *Agrobacterium*-mediated transformation (AMT) has also been introduced into other organisms such as bacteria, fungi, algae, mammals (Piers et al. 1996; de Groot et al. 1998; Cheney et al. 2001; Kunik et al. 2001; Kelly and Kado 2002). Especially, AMT was quickly applied to more than 100 fungal species due to highly efficient transformation (Michielse et al. 2005). One of the fungal species for which AMT was intensively employed was *Magnaporthe oryzae*, which

is a causal agent of rice blast disease (Rho et al. 2001). Thus far, more than 200,000 AMT transformants have been generated by the *Magnaporthe* community (Xu et al. 2006). A mutant library consisting of 21,070 transformants was generated by our colleagues and the high-throughput screening system was employed to yield more than 180,000 data points for both genotypes and phenotypes of the transformants (Jeon et al. 2007). Analyses using Southern blot, thermal asymmetric inter-laced (TAIL)-PCR, and sequencing revealed that more than 70 % of the tested transformants had single copy of T-DNA integration and 1,110 T-DNA insertion sites were identified in the genome of *M. oryzae* (Choi et al. 2007).

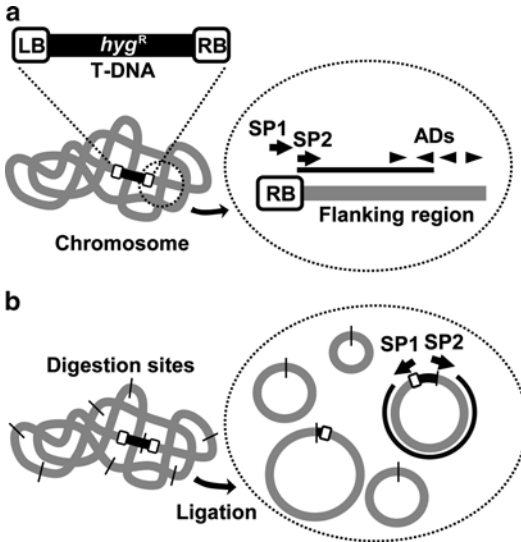
Here, we provide a detailed protocol for analyzing T-DNA insertion patterns. The isolation of T-DNA and its flanking sequences is a key step in identification of T-DNA insertion sites. TAIL- and inverse-PCRs were mainly used for the purpose (Liu and Whittier 1995; Ochman et al. 1988). TAIL-PCR is a hemi-specific PCR method where specific (SP) and random primers were used in combination (Fig. 19.1a). This PCR method directly isolates the target region without any pre- or posttreatment (Liu and Whittier 1995). Thus, TAIL-PCR is suitable for amplifying flanking regions from a large number of samples in parallel. One issue with TAIL-PCR is that random arbitrary degenerate primers can generate nonspecific products. Multi-rounds of PCRs are required to increase specific products. Unlike TAIL-PCR, inverse-PCR protocols involve enzyme digestion

---

J. Choi (✉)  
Division of Life Sciences, College of Life  
Sciences and Bioengineering, Incheon National  
University, Incheon, Republic of Korea  
e-mail: [jaehyukc@incheon.ac.kr](mailto:jaehyukc@incheon.ac.kr)

J. Jeon, Ph.D.  
Department of Agricultural Biotechnology, Seoul  
National University, Seoul, Republic of Korea

Y.-H. Lee, Ph.D.  
Center for Fungal Genetic Resources, Seoul National  
University, Seoul, Republic of Korea  
e-mail: [yonglee@snu.ac.kr](mailto:yonglee@snu.ac.kr)

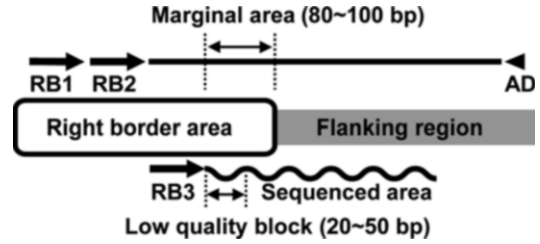


**Fig. 19.1** Principles of PCRs to isolate T-DNA and flanking sequences. Schematic diagram of TAIL-PCR (a) and inverse-PCR (b). *LB* left border, *RB* right border, *SP* specific primer, *AD* arbitrary degenerate primer, and *hyg<sup>R</sup>* hygromycin resistance gene cassette

of transformant's DNA and ligation steps, which require more time and labor (Ochman et al. 1988). However, it specifically finds targets because only specific primers are used for the rescue of flanking regions (Fig. 19.1b). Taken together, TAIL-PCR is suitable for large-scale isolation of flanking regions and inverse PCR is good for improving specificity.

## 19.2 Materials

1. Extracted genomic DNA.
2. Specific primers (SPs): Both right and left borders (RB and LB, respectively) can be used when SPs are designed. At least three sets of SPs are required for one side border. The first specific primer (e.g., RB1) is located inside the border area and amplification goes toward end of the border. We called the area from the third specific primer (e.g., RB3) to the end of border as "marginal area" (Fig. 19.2) and at least 70 bp of marginal area are recommended. If the marginal area is too short, no border sequence is detected when isolated flanking sequences are BLAST searched with NCBI sequence database. Moreover, longer



**Fig. 19.2** Location of SPs and marginal area in RB. Arrows indicate SPs and the arrowhead the AD primer. The line from RB2 to AD is the isolated PCR product by TAIL-PCR. *RB* right border, *AD* arbitrary degenerate primer

marginal areas are needed for the LB because truncation (i.e., loss of border sequence) happens frequently during T-DNA integration and LB is more susceptible for truncation than RB (Tinland 1996).

3. Arbitrary degenerate primers (ADs): ADs are mixed primers consisting of different specific primers. For example, the degenerate primer sequence "NGTCGASWGANA" consist of 64 specific primers ( $4(N) \times 2(S) \times 2(W) \times 4(N) = 64$ ). We use "64" as degeneracy of this primer. ADs with higher degeneracy lead to more products but also the number of unwanted products will be increased. Different composition of nucleotides indicates different melting temperatures.
4. 96-well ELISA plate: This plate is used for storage of diluted genomic DNA.
5. 96-well reaction plate: Also called 96-well PCR plates.

## 19.3 Methods

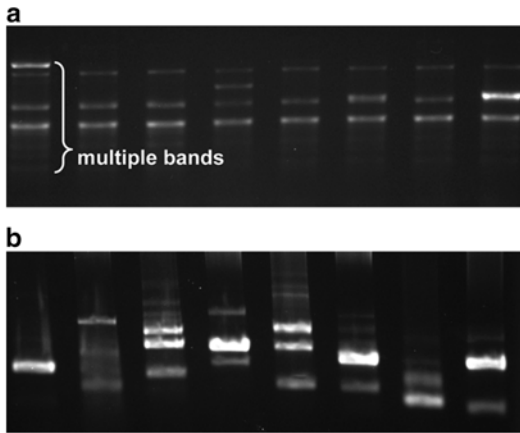
### 19.3.1 Large-Scale TAIL-PCR

1. Dilute genomic DNA in the 96-well ELISA plate to a final concentration as 10–30 ng/ $\mu$ L in the PCR. Add 3  $\mu$ L of the diluted genomic DNA to the 96-well reaction plate (i.e., PCR tubes).
2. Prepare the PCR mixture (Table 19.1). Except template genomic DNA, all the other reagents are mixed together in a 2 mL microcentrifuge

**Table 19.1** Composition of the TAIL-PCR mixture

Components (units)	Final concentration	Original concentration	×1 (μL)	×100 (μL)
Template DNA (ng/μL)	10–30	70–200	3	
dNTP mixture (mM)	0.20 each	2.5 each	1.6	160
10× Reaction buffer	1×	10×	2	200
MgCl <sub>2</sub> (mM)	0.12	1.5	1.6	160
SP (μM)	0.2	10	0.4	40
AD (μM)	3	10	6	600
Sterilized distilled water	–	–	5.2	520
Taq polymerase (U)	0.01	1	0.2	20
Total			20	2,000

- tube and distributed in a 96-well reaction plate. The final concentration of SP is adjusted to 0.2 μM in primary reactions, respectively. AD is used at 3–4 μM according to its degeneracy. We tested different commercial Taq DNA polymerases and found no difference in the efficiency of rescuing T-DNA flanks. Concentration of dNTP or MgCl<sub>2</sub> can differ according to polymerases.
3. First round PCR. We used the following PCR cycling parameters: (1) 94 °C for 3 min; (2) 5 cycles of 94 °C for 10 s, 62 °C for 1 min, and 72 °C for 1 min; (3) 2 cycles of 94 °C for 30 s, 25 °C for 3 min (ramping to 72 °C for 3 min), and 72 °C for 2.5 min; (4) 15 cycles of 94 °C for 30 s, 65 °C for 1 min, 72 °C for 2.5 min, 94 °C for 30 s, 65 °C for 1 min, 72 °C for 2.5 min, 94 °C for 30 s, 44 °C for 1 min, and 72 °C for 2.5 min; and (5) 72 °C for 7 min. The “ramping” option increases temperature slowly to help AD primers to anneal as much as possible. If a PCR machine doesn’t have this option, one can alternatively add three steps of different temperatures (e.g., 40 °C for 1 min, 50 °C for 1 min, and 60 °C for 1 min). The annealing temperature for the SP (i.e., 65 °C here) should be the actual temperature of your SPs. They are different according to the composition and length of SPs.
  4. Load 3 μL of the primary PCR products on a 1.5 % agarose gel. If the multiple bands with same sizes were shown in all lanes (Fig. 19.3a), the PCR worked well. At this stage, the targeted specific products are less than nonspecific products, which cause multiple bands.
  5. Dilute the primary PCR products threefold. For example, if 17 μL of the primary PCR product is left, add 34 μL of water to the PCR product. Prepare a new 96-well ELISA plate and add 147 μL of water to each well. Transfer 3 μL of diluted primary PCR products to the wells (i.e., 50-fold dilution). Mix well by pipetting.
  6. The 150-fold diluted PCR products are used as template DNA for the secondary PCR. Because 3 μL of template DNA is used in a 20 μL reaction volume, the primary PCR product is diluted 1,000-fold finally (51/17 × 150/3 × 20/3). All the conditions are the same as for the primary PCR except the increased concentration of second SPs from 0.2 to 0.4 μM.
  7. Second round PCR. We used the following PCR cycling parameters: (1) 94 °C for 3 min; (2) 5 cycles of 94 °C for 10 s, 62 °C for 1 min, and 72 °C for 1 min; (3) 15 cycles of 94 °C for 30 s, 66 °C for 1 min, 72 °C for 2.5 min, 94 °C for 30 s, 65 °C for 1 min, 72 °C for 2.5 min, 94 °C for 30 s, 44 °C for 1 min, and 72 °C for 2.5 min; and (5) 72 °C for 7 min.
  8. Load 3 μL of the secondary PCR products on agarose gel. Single or multiple bands can be shown but size of the bands should be different in all lanes (Fig. 19.3b). If similar pattern of bands is observed in all lanes like the primary PCR products, we may not obtain positive results from these PCR products. Then it may be better to try another PCR



**Fig. 19.3** TAIL-PCR products loaded on the agarose gels. First (a) and second (b) round of PCR products

from the beginning using different AD primers or lowering the annealing temperature.

9. Perform purification of PCR products for better sequencing results. Purification is optional but improves sequencing quality significantly. Lots of primers (SPs and ADs) remain after the second round PCR and can interrupt sequencing reaction. To save time and effort, enzyme purification using ExoSAP-IT® (USB, Cleveland, OH, USA) is suggested. Add 1  $\mu$ L of the ExoSAP-IT® enzyme to each well of 96-well reaction plate and incubate at 37 °C for 20 min. Inactivate the enzyme at 80 °C for 15 min before sequencing.
10. Purified PCR products are sequenced starting from the third SPs (e.g., RB3), which also improves selectivity of the correct targets.

### 19.3.2 Identification of T-DNA Integration

1. Following BLAST search with downloaded sequences, three major types can be found: (1) border + flanking area, (2) flanking only, and (3) border only or with vector backbone. The portions of three types were 68 %, 28 %, and 4 %, respectively, when we screened

~2,000 sequences (Choi et al. 2007). “Type 1” sequences can be used for determination of insertion positions. “Type 2” sequences often are obtained when the “marginal area” is too short or border truncation is severe. “Type 3” sequences arise from irregular integration of T-DNA such as tandemly repeated T-DNA and read-through of vector backbone. This irregular integration could happen frequently (10–20 %) (Meng et al. 2007).

2. Determine T-DNA insertion site. Ideally, “type 1” sequences contain border and flanking area without any unmatched region (named “gap”) between them. This type is called “complete junction” (Choi et al. 2007). The boundary between border and flanking area is named as “T-DNA insertion site.” However, some “type 1” sequences have a gap between border and flanking area (called “incomplete junction”). Low sequencing quality, irregular integration of T-DNA, and multiple T-DNA integration might cause this gap. In this case, we replace the gapped area with a virtual flanking region and determine the new boundary as the T-DNA insertion site. However, such prediction with a virtual flanking region sometimes introduces an error in determining of the insertion site, depending on the length of the virtual region. Even in a mutant having complete junctions at both borders, the mutant may have deletion, addition, or translocation of genomic DNA. Such genomic change allows one inserted T-DNA in a mutant had more than two values of the insertion positions. For example, when determined positions at both borders have 10-bp difference in a mutant, the mutant is regarded as having genomic deletion during T-DNA integration. However, the computer program counts it as two different mutants with 10-bp difference in T-DNA insertion position. Such genomic deletion was frequently (78 %) observed in our analysis for T-DNA insertion sites (Choi et al. 2007). Thus, a new concept for T-DNA insertion site should be employed in counting the number of T-DNA mutants not to overestimate it. We defined “T-DNA tagged

location” as position information with a buffering range (Choi et al. 2007). The buffering range is set as 35 bp because most errors (>98 %) were limited to that length. That is, when one T-DNA insertion site is determined, the next one should be apart at least 35 bp from the first one.

### 19.3.3 Patterns of T-DNA Integration in Fungi

1. *More truncation in LB.* Integrated T-DNAs are often observed with partial loss of its border sequences (Mayerhofer et al. 1991). The phenomenon is called “truncation” (Tinland 1996). It is a well-known feature which happens during T-DNA integration in plants (Forsbach et al. 2003). As shown in plants, LBs are truncated more often than RBs in *M. oryzae*: more than 60 % of LBs were truncated while less than 10 % of RBs were truncated (Choi et al. 2007; Meng et al. 2007; Li et al. 2007). Conservation of RB is explained by protection of the bacterial virulence protein, VirD2. When T-DNA is transported to the host, virulence proteins, VirD2 and VirE2, guide the T-DNA to the nucleus (Tinland 1996). Because VirD2 binds covalently to the 5' end of T-DNA, right borders are protected against nucleolytic degradation. Therefore, due to more truncation of LB, a longer marginal area is suggested for LB sequences in primer design of the TAIL-PCR.
2. *Actual microhomology exists only in LB.* Microhomology means shared nucleotides between border and flanking sequences. In general, this is more frequently observed in LB than RB border (Tinland 1996; Mayerhofer et al. 1991; Forsbach et al. 2003). However, because two random sequences can make microhomology at a certain ratio ( $1/4^n$ , where  $n$  means the number of shared nucleotides), the portion ( $T/4^n$ , where  $T$  means total number of the analyzed sequence samples) should be subtracted from the observed frequency to

calculate actual microhomology. In our analysis, microhomology was found in 85 % of LB and 31 % of RB samples (Choi et al. 2007). However, actual microhomology existed only in LB and ranged from 2 to 6 bp (Choi et al. 2007). Short homology in LB might be an anchoring region for T-DNA integration via illegitimate recombination model (Tinland 1996).

3. *Deletion, duplication, addition, and rearrangement can occur in the genome of a host.* The host genome sequence is also changed during T-DNA integration. Deletion of target sites is prevalent in fungi (~80 %) and plants (~90 %) (Choi et al. 2007; Forsbach et al. 2003). The deleted lengths are mostly short (less than 30 bp) but it can be up to ~2 kb. Duplication (addition of the same sequence with the other flanking region), addition (filler DNA), and rearrangement of the genomic sequences is observed at a low frequency. In general, the frequency is lower in fungi than plants.
4. *T-DNA insertion is frequently observed in the physically bended area of the genome.* The insertion positions of T-DNA integration might be favorable at highly bendable areas (Choi et al. 2007; Zhang et al. 2007). Peaked bendability was observed within 100 bp distance from the insertion positions (both sides), suggesting that the regions are structurally bendable or flexible for T-DNA integration (Choi et al. 2007). Bended structures of DNAs are usually found in promoter areas where transcription factors recognize the position for transcription initiation (Perez-Martin et al. 1994). This might explain why T-DNA insertions were often found in the promoter regions (Choi et al. 2007; Forsbach et al. 2003).

**Acknowledgements** This work was supported by the National Research Foundation of Korea grant funded by the Korea government (2008-0061897, 2013-003196 (to Y.H.L.), and 2013R1A1A1A1010928 (to J.C)), and the Next-Generation BioGreen 21 Program of Rural Development Administration in Korea (PJ00821201 (to Y.H.L)).

## References

- Cheney D, Metz B, Stiller J (2001) *Agrobacterium*-mediated genetic transformation in the macroscopic marine red alga *Porphyra yezoensis*. *J Phycol* 37:11
- Choi J, Park J, Jeon J, Chi MH, Goh J, Yoo SY, Park J, Jung K, Kim H, Park SY, Rho HS, Kim S, Kim BR, Han SS, Kang S, Lee YH (2007) Genome-wide analysis of T-DNA integration into the chromosomes of *Magnaporthe oryzae*. *Mol Microbiol* 66:371–382
- de Groot MJA, Bundock P, Hooykaas PJJ, Beijersbergen AGM (1998) *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. *Nat Biotechnol* 16:839–842
- Drummond MH, Gordon MP, Nester EW, Chilton MD (1977) Foreign DNA of bacterial plasmid origin is transcribed in crown gall tumors. *Nature* 269:535–536
- Forsbach A, Schubert D, Lechtenberg B, Gils M, Schmidt R (2003) A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome. *Plant Mol Biol* 52:161–176
- Jeon J, Park S-Y, Chi M-H, Choi J, Park J, Rho H-S, Kim S, Goh J, Yoo S, Choi J, Park J-Y, Yi M, Yang S, Kwon M-J, Han S-S, Kim BR, Khang CH, Park B, Lim S-E, Jung K, Kong S, Karunakaran M, Oh H-S, Kim H, Kim S, Park J, Kang S, Choi W-B, Kang S, Lee Y-H (2007) Genome-wide functional analysis of pathogenicity genes in the rice blast fungus. *Nat Genet* 39:561–565
- Kelly BA, Kado CI (2002) *Agrobacterium*-mediated T-DNA transfer and integration into the chromosome of *Streptomyces lividans*. *Mol Plant Pathol* 3:125–134
- Kunik T, Tzfira T, Kapulnik Y, Gafni Y, Dingwall C, Citovsky V (2001) Genetic transformation of HeLa cells by *Agrobacterium*. *Proc Natl Acad Sci U S A* 98:1871–1876
- Li GH, Zhou ZZ, Liu GF, Zheng FC, He CZ (2007) Characterization of T-DNA insertion patterns in the genome of rice blast fungus *Magnaporthe oryzae*. *Curr Genet* 51:233–243
- Liu YG, Whittier RF (1995) Thermal asymmetric inter-laced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* 25:674–681
- Mayerhofer R, Koncz-Kalman Z, Nawrath C, Bakkeren G, Cramer A, Angelis K, Redei GP, Schell J, Hohn B, Koncz C (1991) T-DNA integration: a mode of illegitimate recombination in plants. *EMBO J* 10:697–704
- Meng Y, Patel G, Heist M, Betts MF, Tucker SL, Galadima N, Donofrio NM, Brown D, Mitchell TK, Li L, Xu JR, Orbach M, Thon M, Dean RA, Farman ML (2007) A systematic analysis of T-DNA insertion events in *Magnaporthe oryzae*. *Fungal Genet Biol* 44:1050–1064
- Michielse CB, Hooykaas PJJ, van den Hondel CA, Ram AFJ (2005) *Agrobacterium*-mediated transformation as a tool for functional genomics in fungi. *Curr Genet* 48:1–17
- Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* 120:621–623
- Perez-Martin J, Rojo F, de Lorenzo V (1994) Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. *Microbiol Mol Biol Rev* 58:268–290
- Piers KL, Heath JD, Liang X, Stephens KM, Nester EW (1996) *Agrobacterium tumefaciens*-mediated transformation of yeast. *Proc Natl Acad Sci U S A* 93:1613–1618
- Rho HS, Kang S, Lee YH (2001) *Agrobacterium tumefaciens*-mediated transformation of the plant pathogenic fungus, *Magnaporthe grisea*. *Mol Cells* 12:407–411
- Tinland B (1996) The integration of T-DNA into plant genomes. *Trends Plant Sci* 1:178–184
- Xu J-R, Peng Y-L, Dickman MB, Sharon A (2006) The dawn of fungal pathogen genomics. *Annu Rev Phytopathol* 44:337–366
- Zhang J, Guo D, Chang YX, You CJ, Li XW, Dai XX, Weng QJ, Zhang JW, Chen GX, Li XH, Liu HF, Han B, Zhang QF, Wu CY (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