Determination of gene copy number and genotype of transgenic *Arabidopsis thaliana* by competitive PCR

Munechika Honda¹,⁴, Yasunori Muramoto¹, Tsuyoshi Kuzuguchi¹, Sadanori Sawano¹, Masayuki Machida² and Hiroyuki Koyama³

¹Laboratory of Plant Biotechnology and Microbiology, Gifu Prefectural Institute for Bio-industrial Technology, 3481-2 Kamihachiya, Hachiya-cho, Minokamo, Gifu, 505-0004, Japan
²Institute of Molecular and Cell Biology, National Institute of Advanced Industrial Science and Technology, 1-1 Higashi, Tsukuba, Ibaraki, 305-8566, Japan
³Laboratory of Plant Cell Technology, Faculty of Agriculture, Gifu University, 1–1 Yanagido, Gifu, 501-1193, Japan

Received 3 September 2001; Accepted 22 January 2002

Abstract

A simple and rapid method is described for determining the integrated T-DNA copy number and the genotype in transgenic *Arabidopsis thaliana* by two-step competitive PCR. First, the amount of genomic DNA in the extracts, obtained from an individual *A. thaliana* transformant, was accurately determined by the 1st competitive PCR using a known single copy gene, 4HPPD (4-hydroxyphenylpyruvate dioxygenase), as a target. Second, the number of T-DNA copies per genome was estimated by quantifying the *NPTII* gene, which was involved in the T-DNA, by the 2nd competitive PCR using exactly the same amount of genomic DNA for each sample. The estimated copy number and genotype obtained by this procedure were identical to those determined by Southern blot analysis and segregation analysis.

Key words: *Arabidopsis thaliana*, gene copy number, genotype, quantitative PCR.

Introduction

Isolation of homozygous transgenic plants having a single copy of the transformed gene is important for transgenic studies and for molecular breeding. However, technical problems exist in the conventional methods used to isolate such transgenic plants. For instance, Southern blot analysis, which is generally used for determining the copy number of the introduced fragment, not only requires a relatively large amount of DNA but also is labour-intensive and time-consuming. Furthermore, Southern blot analysis may not be accurate enough to determine copy numbers greater than two, although multiple copies integration are often found with *Agrobacterium*-mediated transformation (Chan et al., 1992; Grevelding et al., 1993; Aldemita and Hodges, 1996). The genotype of transgenic plants is usually estimated by a segregation test using the seed progenies obtained. This takes a long time and requires much labour for cultivation. Overall, there is a need for an alternative procedure that reduces labour and cost for transgenic studies to be undertaken.

The polymerase chain reaction (PCR) is one of the most sensitive methods for detecting the integrated gene in the transgenic plant genome, and thus it can reduce the amount of DNA required for analysis (McGarvey and Kaper, 1991). This, in turn, reduces the number of samples needed for extracting DNA. Quantitative PCR (Q-PCR) has been successfully used for determining the copy number of genomes of quite a number of species, including humans (Gilliland et al., 1990; Deng et al., 1993). Using competitive PCR, one of the Q-PCR techniques, a rapid and reliable method has recently been developed for determining the copy number of the integrated DNA in a filamentous fungus, *Aspergillus oryzae* (Honda et al., 2000).

In the present study, the competitive PCR technique was applied to determining the copy number of the integrated T-DNA and the genotype in transgenic *A. thaliana*. The copy number and the genotype determined by this method were identical to those estimated by Southern blot analysis and a segregation test.

¹ To whom correspondence should be addressed. Fax: +81 574 25 3804. E-mail: honda@bio.rd.pref.gifu.jp
Approximately 100 ng of genomic DNA, which can be isolated from a leaf segment, is sufficient for analysing both indexes.

Materials and methods

Plant materials and growth conditions

Transgenic *A. thaliana* having the *NPTII* (neomycin phosphotransferase II) gene (Koyama et al., 2000) was used in this study (Fig. 1A). The primary transgenic plants were obtained by *Agrobacterium*-mediated transformation with pAtCS-121-Hm, made by replacing intron-*GUS* in pLG-121-Hm (Ohta et al., 1990) by mitochondrial citrate synthase from carrot (Takita et al., 1999). Homozygous lines at the third (T3) generation were used for estimating the copy number, and the second generation (T2) consisting of null, heterozygous and homozygous transgenic plants, were used for estimating genotype by competitive PCR as described below.

For estimating the copy number and the genotype using competitive PCR, plants were grown on vermiculite at 23°C under artificial lighting (150 μmol m⁻² s⁻¹) for a 12 h day for 4 weeks. A mature leaf (approximately 50 mg fresh weight) was harvested for isolating the genomic DNA using Isoplant kit (Nippon gene, Japan). For Southern blot analysis, approximately 20 plants were grown in 100 ml flask containing 1/2 liquid MS medium with 1% sucrose for 2 weeks, and harvested for the genomic DNA extraction (approximately 1 g fresh weight). Segregation tests were carried out by growing the seed progeny of T2 transgenic plants on a 1/2 solid MS medium with 1% sucrose containing hygromycin (50 μg ml⁻¹) and kanamycin (100 μg ml⁻¹).

Vector constructions for competitors

PCR competitors, which are longer (*NPTII* MIMIC) or shorter (4HPPD cDNA) than the target DNA in the genome, were constructed as follows. The *NPTII* MIMIC was made by inserting a 75 bp PsI–PstI fragment into the PsI site of the 838 bp the *NPTII* gene fragment subcloned on pUC18 (Fig. 1B). The 75 bp PsI–PstI fragment, which contains the MCS of the pGEM-T Easy plasmid vector (Promega, USA), was prepared by PCR using PCR primers carrying the nucleotide sequence specific to pGEM-T Easy with the PsI recognition sequence at the 5' end [MSCF PsI (5'-AGCC TGCAGATTTAGGTGACACTATAGAAT-3') and MCSR PsI (5'-CGACTGCTAATACGACTCATATAGGCC-3')] (Fig. 1B). The *NPTII* gene fragment was obtained by PCR using primers specific to the *NPTII* gene [NPTII-XbaI (5'-GCAATTC-AGGAGAGGGCTATTGCGCTAT-3') and NPTII-HindIII (5'-GCAATAAGCTTAACCTCTGTTAAAGCGATA-3')], then subcloned onto pUC18.

The *A. thaliana* 4-hydroxyphenylpyruvate dioxygenase (4HPPD) gene was chosen as a gene of a single copy in a genome based on the information from GenBank (AF047834; Garcia et al., 1999). The 4HPPD cDNA fragment was prepared by PCR from the *A. thaliana* cDNA pool [4HPPD Forward (5'-GCTCTAGCTCTTCCTGGGGTGTGCGGACTGCTAATACGACTCATATAGGCC-3') and 4HPPD Reverse (5'-TGGACTTGCTGTTAGTGCCTTCTGGGGTGTGCGGACTGCTAATACGACTCATATAGGCC-3')], subcloned onto pGEM-T Easy, and then used as a 4HPPD competitor. Both vectors used as competitors were highly purified by the CsCl centrifugation and quantified by absorbance at 260 nm. Then each series of the dilution competitor solutions was quantified by the DNA-binding bisbenzimide H33258 fluorochrome method to increase the reliability of the quantification.

Fig. 1. Schematic drawings of the plasmid vector used for transformation of *A. thaliana* with the citrate synthase gene isolated from *Daucus carota* (DcCS) (A) and the vector used for quantitative PCR with *NPTII* MIMIC as a competitor (B). Arrowheads indicate the positions of primers used for the competitive PCR. LB, left border; RB, right border; pNOS, nopaline synthase gene promoter; tNOS, nopaline synthase gene terminator; pCaMV35S, cauliflower mosaic virus 35S promoter; *NPTII*, neomycin phosphotransferase II; HPTII, hygromycin phosphotransferase II; XbaI, HindIII, BamHI, restriction enzyme; MIMIC, artificial PsI–PstI fragment (75 bp).

Southern blot analysis

Genomic DNA was isolated from each *A. thaliana* transgenic line using the Isoplant kit, and then quantified spectrophotometrically (*A*₂₆₀–*A*₂₈₀). Approximately 10 μg of each DNA was digested with appropriate restriction enzymes (Takara Shuzo, Japan) that cut (XbaI, HindIII and BamHI(TS)) within the T-DNA, or did not cut (SacI) within the T-DNA (Fig. 1A) overnight. The fragments were separated by electrophoresis on 0.8% agarose gels (Sigma, USA) and transferred onto a positively charged nylon membrane (Roche, USA) by capillary transfer using the salt transfer protocol by the manufacture. Hybridization was carried out with the *NPTII* fragment that had been DIG-labelled with the DIG DNA labelling kit (Roche). The immunological detection was performed using the CSPD (ready to use, Roche), then the membrane was exposed to a CCD camera system (Lumi-ImagerTM F1, Boehringer Mannheim, Germany) for 10 min.

Competitive PCR

All PCR were carried out using the Expand High Fidelity PCR system (Roche) with 27 cycles at 94°C for 15 s, 60°C for 30 s and 72°C for 3 min. Each reaction mixture consisted of...
PCR buffer, 200 μM dNTPs, appropriate primers and templates in a total volume of 10 μl. Ten μl of each PCR reaction was electrophoresed on a 1.5% agarose gel. The image of the ethidium bromide fluorescence of each amplified band was captured by the CCD camera system under UV irradiation and the amount of each band was calculated from both intensity and area of the band using Lumi Analyst™ software (Boehringer Mannheim). The 1st competitive PCR was performed to determine the exact amounts of genomic DNA in the extracts by quantifying 4HPPD contents. The target gene, which was involved in approximately 20 ng of the genomic DNA (4HPPD quantified), was amplified with the 4HPPD specific primers in the presence of 0–0.65 attomoles of the double strand 4HPPD competitor (4HPPD cDNA plasmid DNA). The 2nd competitive PCR was performed for estimating a copy number of the integrated T-DNA by quantifying the NPTII gene using 20 ng of genomic DNA, which was accurately determined by the 1st competitive PCR. A quarter attomoles of the double strand NPTII competitor (pUC18 NPTII MIMIC plasmid DNA) was added to each reaction, and then the target DNA was amplified using the NPTII gene specific primers (NPTII F47 (5'-TGGAGAGGCTATTCGGCTAT-3') and NPTII R788 (5'-AACTCGTCAAGAAGGCGATA-3')).

Results and discussion

Competitive PCR is known as one of the Q-PCR techniques, although the reliability of quantification by this method can be limited by a number of factors affecting the amplification efficiency (Harada and Yamada, 1992; Freeman et al., 1999). These include the amount of templates, the primer sequences and positions, the large difference of the length or the nucleotide sequence between the competitor DNA and the target DNA. Competitor DNAs were constructed to minimize such factors affecting the amplification, by introducing a small artificial DNA fragment into the target gene for the NPTII competitor (Fig. 1) or by isolating a cDNA fragment lacking a small intron for the 4HPPD competitor. Differences in length between the target and the competitor for NPTII and 4HPPD were 75 bp and 108 bp, respectively, which corresponded to 10% and 8.5% to that of amplicon sizes, respectively. On the other hand, recent advances of the real time PCR suggest that the exponential phase of PCR is more reliable than the plateau phase of PCR for the quantification of DNA (Ingham et al., 2001), whereas such a plateau phase is usually used for the determination of DNA in a competitive PCR. In these experimental conditions, the PCR attain to the plateau phase about 35 cycles. Thus, 27 cycles were employed to increase the reliability of the quantification of DNA by this method.

Another important factor that affects the reliability of estimating the copy number of the integrated DNA in a genome is the accuracy in quantification of the sample genomic DNA (Honda et al., 2000). A conventional spectrophotometrical method (A260/A280) was used for the quantification of the sample genomic DNA. This method is reliable for quantifying the concentration of a highly purified DNA, but may not allow reliable quantification for a genomic DNA. Thus, the amount of the sample genomic DNA content was first determined accurately using competitive PCR (the 1st competitive PCR) by quantifying a known single copy gene of A. thaliana, 4HPPD as an internal standard. Figure 2 shows the ratio of the fluorescence intensity of a band corresponding to 4HPPD genomic DNA (upper) over that corresponding to 4HPPD competitor (lower) plotted versus various amounts of 4HPPD competitor added to the reaction containing approximately 20 ng of the sample genomic DNA, which was quantified spectrophotometrically. Linear regression equations with correlation coefficients greater than 0.95 were obtained in all cases. On the other hand, the genome size of A. thaliana is
125 mega base pairs per haploid (The A. thaliana Genome Initiative, 2000) and a theoretical weight of 1 pmol of 1 kbp DNA is 0.66 μg (Sambrook et al., 1989). Thus, it could be inferred that each 20 ng of A. thaliana genomic DNA contains 0.24 attomole of the single copy target gene (e.g. 4HPPD). Using this theoretical value (20 ng genomic DNA = 0.24 attomole of the target) and the amount of competitor DNA added to the reactions, which gave the logarithmic ratio of fluorescence equal to zero in Fig. 2, the amounts of sample genomic DNA could be determined accurately.

Using Southern blot analysis, the transgenic plant carrying single copy (TS) or double copies (TD) of integrated T-DNA were selected from the T3 transgenic population having integrated T-DNA homozygously at a single locus which had been prepared in previous work (Koyama et al., 2000). A series of restriction enzymes was used to increase the reliability of Southern blot analysis (Fig. 3). Under this experimental condition, the TS and TD lines showed one and two signals, respectively, in all cases. These results indicated that the TD line carries two copies of T-DNA in the genome. On the other hand, it was possible that a single signal of Southern blot analysis would be caused by the insertion of a single copy of T-DNA or the insertion of two T-DNAs in a tandem inverted orientation (Ingham et al., 2001). However, the hybridizing band of the TS line with the SacI (no cutting site within T-DNA) treatment was shorter than that with the BamHI (one cutting site within T-DNA). Thus, it could be inferred that TS carries a single copy of T-DNA in the genome. Using these transgenic plants, the 2nd competitive PCR for quantifying NPTII gene copy number in the genome was performed.

The 2nd competitive PCR was carried out using 20 ng of each genomic DNA as a template, which had been accurately determined by the 1st competitive PCR, and with 0.24 attomole of the NPTII MIMIC as a competitor in each reaction. Under this condition, the expected ratios of band intensity of the target amplicon to the competitor amplicon are one for TS and two for TD. The observed values derived from the 2nd competitive PCR were very close to their theoretical values (Table 1). Thus, it is concluded that the copy number of the integrated T-DNA can be accurately determined by this two-step competitive PCR method. This two-step competitive PCR can be used for determining the genotype of transgenic plants from a small amount of genomic DNA, which can be extracted from a small leaf at an early growth stage. A series of seed progenies derived from heterozygous transgenic plants, containing a single copy of the integrated T-DNA, which were obtained from the same primary transgenic plants of TS, was grown. Genomic DNA was isolated from a leaf segment (approximately 50 mg fresh weight) obtained from a 4-week-old plant, and the genotype determined by the same PCR protocols that were used for estimating the copy number of TS and TD (Table 2). In this experiment, the theoretical copy number (the amount of integrated DNA) of the homozygous plants and the heterozygous plants was 1 and 0.5, respectively. Three transgenic plants were analysed, among which two were found to be heterozygotic and the other was homozygotic. One each of the homozygote and the heterozygote were grown, and the genotype was determined by a segregation test using seed progenies. As shown in Table 2, all the genotypes that were estimated by the competitive PCR in this experiment were identical to those obtained from the segregation test.

As described previously, the copy number of transgenic plants using small amounts of genomic DNA (approximately 100 ng) can be determined precisely by this
method. Many useful and informative procedures for the screening of transgenic plants, such as inverse PCR (Does et al., 1991), randomly primed PCR (Trueba and Johnson, 1996) and a modified method based on the amplification of a region flanking the transgene in a restriction fragment (Sertini et al., 1999) have been reported, but these procedures may not be utilized for determining the genotype. By contrast, the method outlined here is useful for estimating the genotype of transgenic plants (Table 2) and this property would appear to be an advantage when applying this method to molecular breeding.

Recently, it was reported that the transgene copy number of transgenic plants could be accurately determined by a real time PCR method (Ingham et al., 2001). In this report, the average of calculated coefficient of variation (ACV), which reflected the accuracy of determination, was less than 10% for single or double copies of the integrated T-DNAs. The ACV values of the present study’s experimental conditions are less than 10% (Tables 1, 2), suggesting that the accuracy of this method is comparable to the real time PCR method. In general, real time PCR methods need to use a real time PCR apparatus, which is much more expensive than a regular PCR apparatus. In this experiment, similar accuracy of the determination of integrated T-DNA copy number was obtained using a regular PCR apparatus. This can be one of the benefits, when using this method to determine the integrated copy numbers by Q-PCR techniques.

It has been reported that multiple T-DNA copies (more than two) are often introduced by Agrobacterium-mediated transformation. To determine such multiple copies accurately by this method, it is necessary to modify the experimental conditions. As described above, the accurate determination of integrated copy numbers by this method would need to use an exponential phase of PCR. In this experiment, the conditions were optimized, such as the ratio of genomic DNA to competitor DNA in the reactions and the number of PCR reactions (27 cycles), to distinguish the copy numbers between 0, 0.5, 1, and 2 per haploid. Under this condition, the PCR reactions for the target (T-DNAs) and the competitor amplicon could be at the exponential phase. However, if the sample genomic DNA contained multiple copies, it may increase the amounts of the target DNA in the reactions and, in turn, attain the saturation of amplification of the target DNA. Thus, to determine multiple copy numbers accurately by this method, it is necessary to dilute the amounts of sample genomic DNA in the reactions. Since the ACV values of these experimental conditions were less than 8% (the logarithmic ratio of fluorescence equal to approximately zero), up to 10 copies of T-DNAs may therefore be determined accurately using such a modification.

In conclusion, this method, which reduces the labour and the time involved compared to the conventional methods, can be used to select homozygous plants carrying a single copy of the integrated T-DNA. In this experiment, the two-step PCR was used to increase the reliability. As described above, the reliability of this procedure is strongly affected by the accuracy of the quantification of the genomic DNA, and thus the 1st competitive PCR was performed to determine the amounts of the genomic DNA. However, this step may be replaced with other methods such as the DNA-binding bisbenzimide H33258 fluorochrome method without loss of reliability (data not shown). Such a modification would also reduce the labour of the authors’ method. The method can be readily applicable for different plant species by simply designing primers for their target genes. Transgenic rice plants, carrying the HPTII (hygromycin phosphotransferase II) gene, have been analysed by using HPTII competitors, and it was found that both the

Table 1. *Estimation of the gene copy number (NPTII) in transgenic A. thaliana by competitive PCR*

<table>
<thead>
<tr>
<th>Factors/lines</th>
<th>TD (double copies)</th>
<th>TS (single copy)</th>
<th>Null</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of fluorescence intensity*</td>
<td>2.02±0.12</td>
<td>1.03±0.08</td>
<td>0</td>
</tr>
<tr>
<td>Copy number in a genome (Q-PCR)</td>
<td>2</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Copy number in a genome (Southern blot analysis)</td>
<td>2</td>
<td>1</td>
<td>–</td>
</tr>
</tbody>
</table>

*Each value represents the average ± SD of three independent experiments. Ratio of fluorescence intensity of a band corresponding to genomic DNA over that corresponding to competitor DNA.

Table 2. *Estimation of the genotype (NPTII) in transgenic A. thaliana by competitive PCR*

<table>
<thead>
<tr>
<th>Factors/lines</th>
<th>Homozygote (TS)</th>
<th>Heterozygote (TS)</th>
<th>Null</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of fluorescence intensity*</td>
<td>1.10±0.08</td>
<td>0.45±0.04</td>
<td>0</td>
</tr>
<tr>
<td>Segregation test (survival/total seeds)</td>
<td>73/73</td>
<td>51/67</td>
<td>0/76</td>
</tr>
</tbody>
</table>

*Each value represents the average ± SD of three independent experiments. Ratio of fluorescence intensity of a band corresponding to genomic DNA over that corresponding to competitor DNA.
copy number and the genotype judged by this method were perfectly identical to those obtained by the conventional procedures (data not shown). Thus, it is expected that the method presented here provides an alternative procedure for estimating the integrated gene copy number and the genotype in transgenic plants.

Acknowledgements

We thank Dr K Ando for providing cDNA libraries of *Arabidopsis thaliana*.

References


