Chapter 12

Identification of Gene Copy Number in the Transgenic Plants by Quantitative Polymerase Chain Reaction (qPCR)

Poonam Kanwar, Soma Ghosh, Sibaji K. Sanyal, and Girdhar K. Pandey

Abstract

Transgenic events are defined as exogenous DNA insertion in the genome through genetic transformation. It is a powerful means for the improvement of crop plants and to understand the gene function. Multiple DNA insertion events may occur at one or several chromosomal locations. One of the important tasks, after validation of the transformation of transgenic plants, is the identification of single copy in the transgenic. This means the insertion of exogenous DNA fragment only in a single locus in the genome. Southern blot hybridization is a convincing and reliable method, for estimation of copy number in transgenic lines but it is cumbersome and time-consuming process. One of the other well-known methods is quantitative polymerase chain reactions (qPCR), a simple and rapid method to identify copy number from a population of independent transgenic lines. In comparison to the Southern hybridization method, qPCR is simpler to perform, requires less DNA, lesser time and does not require any labeled probes. This method utilizes specific primers to amplify target transgenes and endogenous reference genes. Designing an appropriate and specific primer pair is a very crucial part of the estimation of the gene copy number. In this chapter, we have illustrated a detailed methodology for identification of the gene copy of the transgenic plants.

Key words qPCR, SYBR Green, Primer, Gene copy number, Transgenic plant

1 Introduction

Transgenic plants are extensively used to understand gene function(s) and also for other biotechnological studies. Introduction of exogenous DNA into the plant genome leads to the generation of transgenic plants and number of copies of exogenous DNA inserted in to the genome of the transgenic plant is defined as transgene copy number. Effectiveness of any transgenic depends on the copy number. Multiple transgene copies could lead to higher expression of the gene as well as result in transgene silencing. Therefore, transgene copy number (TCN) determination is usually an essential part of transgene studies and gene function analysis.

Poonam Kanwar and Soma Ghosh contributed equally.

Chhandak Basu (ed.), *PCR Primer Design*, Methods in Molecular Biology, vol. 2392, https://doi.org/10.1007/978-1-0716-1799-1_12,

The conventional method for TCN determination is Southern blot hybridization. Southern blot hybridization-based TCN determination is both costly, time-consuming, and requires microgram quantities of high-quality DNA. A robust alternative to Southern blot analysis is quantitative PCR (qPCR), for a high-throughput determination of TCN [1, 2]. qPCR is a simple modification of PCR to quantify the amount of target DNA by introducing fluorescent or intercalating dyes to detect PCR product as it accumulates in real-time during PCR cycles. It gives absolute or relative quantification of any DNA, including chromosomal DNA, mitochondrial DNA (mtDNA), chloroplast DNA (ctDNA/chDNA), or cDNA generated by reverse transcription of RNA. It has emerged as the method of choice for fast, affordable, and efficient estimation of copy number [3, 4].

There are two methods used for quantification of copy number via qPCR: external standard curve-based method and the ΔCt method involving an internal reference gene [5]. In this chapter, we will discuss the methodology for the estimation of transgene copy number based on internal reference gene methods. In this method, qPCR involves amplification of a test locus with unknown copy number and a reference locus with known copy number. Primer designing is a crucial part for precise and accurate estimation of transgene copy number with qPCR. Here, the details of methodology and statistical model will be discussed for the estimation of transgene copy based on a standard curve generated for the endogenous reference gene. The data quality control issue of transgene copy estimation for a robust and reproducible result is discussed at the end of this chapter.

2 Materials

2.1 Plant Material

CIPK9 overexpressing transgenic plants developed in *Arabidopsis thaliana* Colombia-0 (Col-0). Wild-type (WT) Col-0 plants were used as control.

2.2 Material for Genomic DNA Isolations

15 days old Arabidopsis seedlings grown in half-strength Murashige and Skoog (MS) media.

2.3 Genomic DNA Isolation

DNA isolation buffer (5 mM Tris–HCl at pH 8.0, 100 mM NaCl, 50 mM EDTA, 10% SDS, 10 mM β-mercaptoethanol), Tris-saturated phenol, chloroform: isoamyl alcohol (24:1), isopropanol, 70% ethanol.

2.4 Genomic DNA PCR

10X Taq buffer, 2.5 mM $MgCl_2$, 2.5 mM dNTPs, 0.2 pmol primers, Taq polymerase, double-distilled milli-Q water (ddMQ), genomic DNA and PeQSTAR Thermocycler.

2.5 Primer Design

Reference nucleotide sequence: Genomic DNA sequence of 4-hydroxyphenylpyruvate dioxygenase (*4HPPD*) (AT1G06570); and hygromycin phosphotransferase (*hptII*) from pCAMBIA1300 vector.

Sequence analysis tools: The Arabidopsis Information Resource (TAIR); www.arabidopsis.org.

Primer design tools: Primer designing tool – National Center for Biotechnology Information (NCBI); www.ncbi.nlm.nih.gov.

2.6 Genomic DNA qPCR

Biological samples: Genomic DNA in different serial dilution.

Chemicals: SYBR Green I master mix (Roche Molecular Biochemicals). Milli-Q water, primers for (*4HPPD* and *hptII*).

Instrumentation: Nanodrop (Thermo Fisher SCIENTIFIC NANODROP^{C}), qPCR-compatible 96-well plates (Agilent), Agilent adhesive plate seals, Agilent AriaMx qPCR, a centrifuge with rotor adapted for microtiter plates cycler instrument (AOSHENG mini-plate centrifuge).

3 Methods

3.1 Plant Material

1. CIPK9 overexpressing Arabidopsis plants were generated using a binary vector, which is derived from the pCAMBIA1300 vector under the control of 2X CaMV 35S promoter and NOS terminator. Strandard transformation protocol was used to transform the Col-0 to generate the transgenic plants.
2. Three transgenic CIPK9 overexpressing lines, numbered #2, #5, and #9 were selected for the analysis of the TCN in this study.

3.2 Isolation of Genomic DNA of Arabidopsis Seedlings

1. Seeds of CIPK9 overexpressing transgenic lines and Col-0 were plated in 1/2 MS media containing 1% sucrose, vernalized for 2 days at 4 °C in the dark and subsequently grown at 22 °C in 16 h in light/8 h dark cycle. Genomic DNA was extracted after 15 days from both transgenic and Col-0 seedlings.
2. Seedling (0.01–0.1 g) was homogenized in a 1.5 ml micro-centrifuge tube. 500 µl of DNA isolation buffer was added to the homogenized tissue. This was then incubated at room temperature for 15 min, followed by spinning at 10,625 × *g* for 2 min. at 4 °C and then the supernatant was transferred to a new micro-centrifuge tube. Finally 200 µl Tris-saturated phenol (Basic) was added to the supernatant, mixed well by vortexing for 1 min., and incubated at room temperature for 2 min.
3. After this, an equal volume of chloroform: isoamyl alcohol (24:1) was added and mixed very gently (to avoid shearing of the DNA) by inverting the tube until the phases were

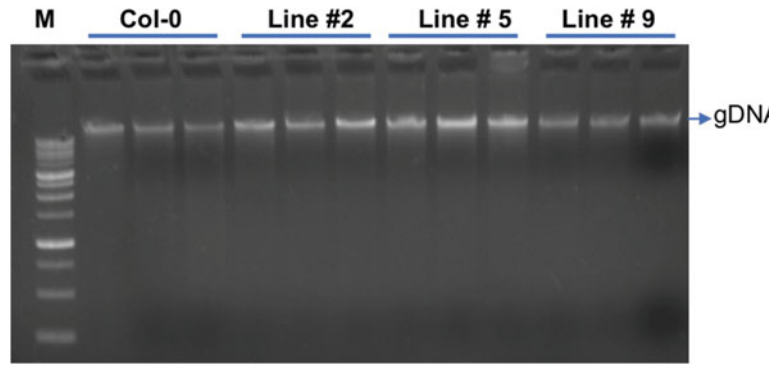

Fig. 1 2μg of gDNA was loaded in 0.8% agarose gel to assess the quality of isolated gDNA of Col-0 and transgenic lines

completely mixed; and then centrifuged at 10,625 × *g* for 3 min. at 4 °C. The upper aqueous phase was transferred to a new tube.

4. To this, 2 volumes of isopropanol was added and mixed gently until the DNA precipitated. The tube was incubated in ice for 15–20 min to aid the DNA precipitation. The tube was then centrifuged at 12,470 × *g* for 10 min. at 4 °C. DNA precipitate was washed two times with 70% ethanol.
5. The DNA pellet was dried at 37 °C for 10–20 min. Then it was resuspended in Milli-Q water with RNase A (10 mg/ml) and then the tube was kept at 37 °C for 1 h for digestion of RNA.
6. Concentration of DNA was determined by measuring the absorbance at 260 nm using a Nanodrop. These gDNA samples can be stored at 4 °C for the short term or −20 or −80 °C for the long-term period. Figure 1 shows the 2 μg of gDNA (overexpression lines of CIPK9 along with Col-0) loaded on 0.8% agarose gel for determining the quality of gDNA.

3.3 Genomic DNA PCR

1. For quantification of copy number from a transgene, we have selected the *hptII* gene in the pCAMBIA-based binary vector. *hptII* gene sequence was taken from reference nucleotide sequence (NCBI) and primers was designed for normal PCR based on the sequence. Forward 5′-GCCTGAACTCACCG CGACG-3′ and reverse 5′-CTCATCGAGAGCCTGCGCG-3′ primers of the *hptII* were used for genomic DNA PCR to detect the transgene in the putative transgenic plants and Col-0 was used as a negative control. The amplified product was analyzed on 0.8% agarose gel. Amplification of the *ACTIN* gene was used as control for PCR (Fig. 2). The *ACTIN* gene PCR amplification was performed using the primer pair 5′-ATGGCTGAGGCTGATGATATT-3′ and 5′-TTAGAAAC ATTTTCTGTGAAC-3′ (*see* **Note 1**).

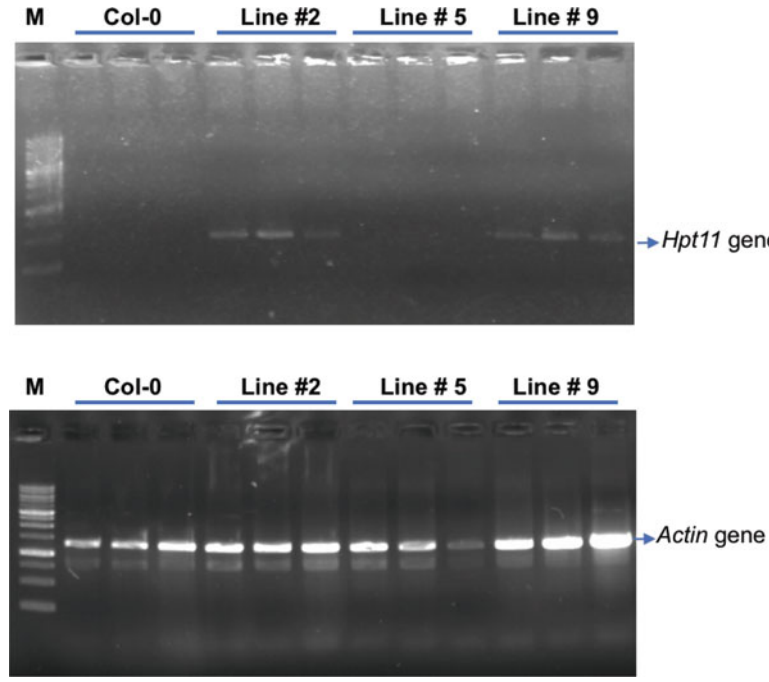

Fig. 2 Genomic PCR of the transgenic lines of CIPK9 along and Col-0 performed with primers for *Hpt11*. *Actin* PCR was performed for all the samples as a loading control

2. The following reaction mixture was used for the PCR reaction: 10× Taq buffer, 1.5 μl; 2.5 mM $MgCl_2$, 2 μl; 2.5 mM dNTP, 1.5 μl; 0.2 pmol forward primer, 0.7 μl; 0.2 pmol reverse primer, 0.7 μl; Taq polymerase: 0.1 μl (0.5–1.0 U/50 μl of reaction mix); ddH_2O to a final volume of 15 μl. Target DNA: 100 ng (genomic DNA).
3. PCR reactions were carried out using PeQSTAR Thermocycler machine according to the following protocol: Initial denaturation at 94 °C, 4:00 min. 1 cycle. Denaturation: 94 °C, 0:30 s. Annealing: 58–65 °C, 0:30 s 27–40 cycles and Extension: 72 °C, 1:00 min. Final extension: 72 °C, 5:00 min. 1 cycle. Based on our results, CIPK9 transgenic line # 2 and # 9 were selected for the detection of the copy number since line number #5 did not yield amplicon with *hptII* primers.

3.4 Primer Design for TCN Analysis

1. For quantification of copy number of a transgene, we again selected the *hptII* in the pCAMBIA1300 vector. *4HPPD* was used as reference gene (internal control) since it is a single-copy gene in the genome of Arabidopsis [6–8]. Full-length genomic DNA sequence of the *4HPPD* was searched, analyzed, and downloaded from online available public databases TAIR (*see* **Notes 2** and **3**).

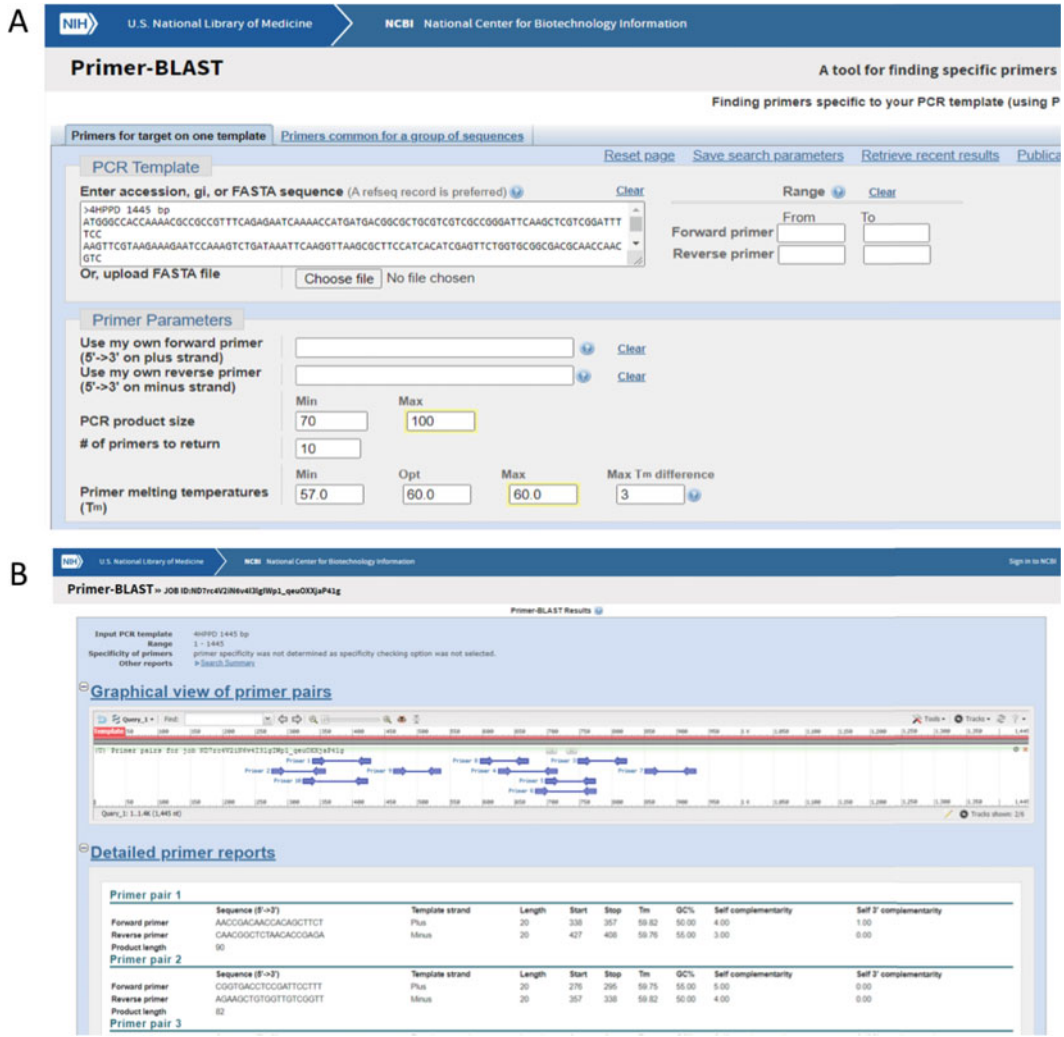

Fig. 3 Screenshot of the NCBI primer designing tool. (**a**) Screenshot showing different options for manual modification of various important primer parameters to obtain an optimum combination. Various primer parameters such as Tm, GC content, primer length, and amplicon size can be modified according to the requirement of user. (**b**) Screenshot showing display of the primer search outcome in the NCBI primer designing tool

2. Primer designing tools of the NCBI were used, and the following important factors were considered for designing qPCR primers: GC content of 50–60%, amplicon of 75–100 bp is ideal since short PCR products are typically amplified with higher efficiency than longer ones. Also, the primer length should be 18–24 bp for qPCR analysis (Fig. 3).
3. It is important to assess the specificity of the primer. Therefore, the BLAST tool available with NCBI and TAIR was used to align the primers with a selected reference sequence and

another similar sequence. This step is done to ensure that the respective primer(s) bind only to the desired region on the reference sequence and not to any other nonspecific sequence in the genome.

4. The *4HPPD* gene primer pair used in this study is 5′-CGGC TCTTGTCGTTCCTTCT -3′ and 5′- TGGAGAAAGCT GACTCTGCG-3′. The *hptII* primer pair used in this study is 5′-CCTGACCTATTGCATCTCCCG −3′ and 5′-CCTCCGC GACCGGTTGTA-3′.

3.5 qPCR for TCN Analysis

1. DNA from each sample were diluted in the different concentrations. The dilutions were 100 ng/μl, 10 ng/μl, 1 ng/μl, 100 pg/μl, 10 pg/μl, 1 pg/μl, 0.1 pg/μl, and 0.01 pg/μl.
2. Total reaction volume for each sample is 10 μl and contains 1 μl of genomic DNA, 10 pmol of each primer, 1 μl of SYBR Green I master mix.
3. A master mix containing all of the components was prepared and distributed in 96 plates. After this, the plate was sealed. The sealed plates were then briefly centrifuged at 500 × *g*, at room temperature to ensure that all liquid was at the bottom of wells (*see* **Note 4**).
4. qPCR was performed using the protocol: 95 °C for 10 min. followed by 35 cycles with 95 °C for 10 s, 55 °C for 30 s, and fluorescent detection step at 55 °C (automated for the qPCR machine).
5. Genomic qPCR was performed with each primer pair of the *hptII* and 4*HPPD* with all the dilutions of the CIPK9 transgenic lines and Col-0 (*see* **Notes 4** and **5**).

3.6 Quantification of Copy Number

1. qPCR was done using either 4*HPPD* primers or *hptII* primers for three biological repeats and three technical repeats for two independent CIPK9 overexpressing Arabidopsis plants (Line #2 and Line #9).
2. There was no change in the Ct values in lower dilutions (ranging from 10 to 0.01 pg/μl); therefore, we used Ct values of four serial dilutions (ranging from 100 ng/μl to 100 pg/μl) to obtain the standard curves for each gene (for each line). The correlation coefficients of the standard curves in the range between 0.96 and 0.99. The standard curves are shown in Fig. 4 (*see* **Notes 6** and **7**).
3. qPCR efficiency was determined using the slope and intercept values from the standard curve. In our case, we obtained an efficiency of more than 90% for each primer set. For calculation, we used the qPCR efficiency calculator from Thermo Scientific

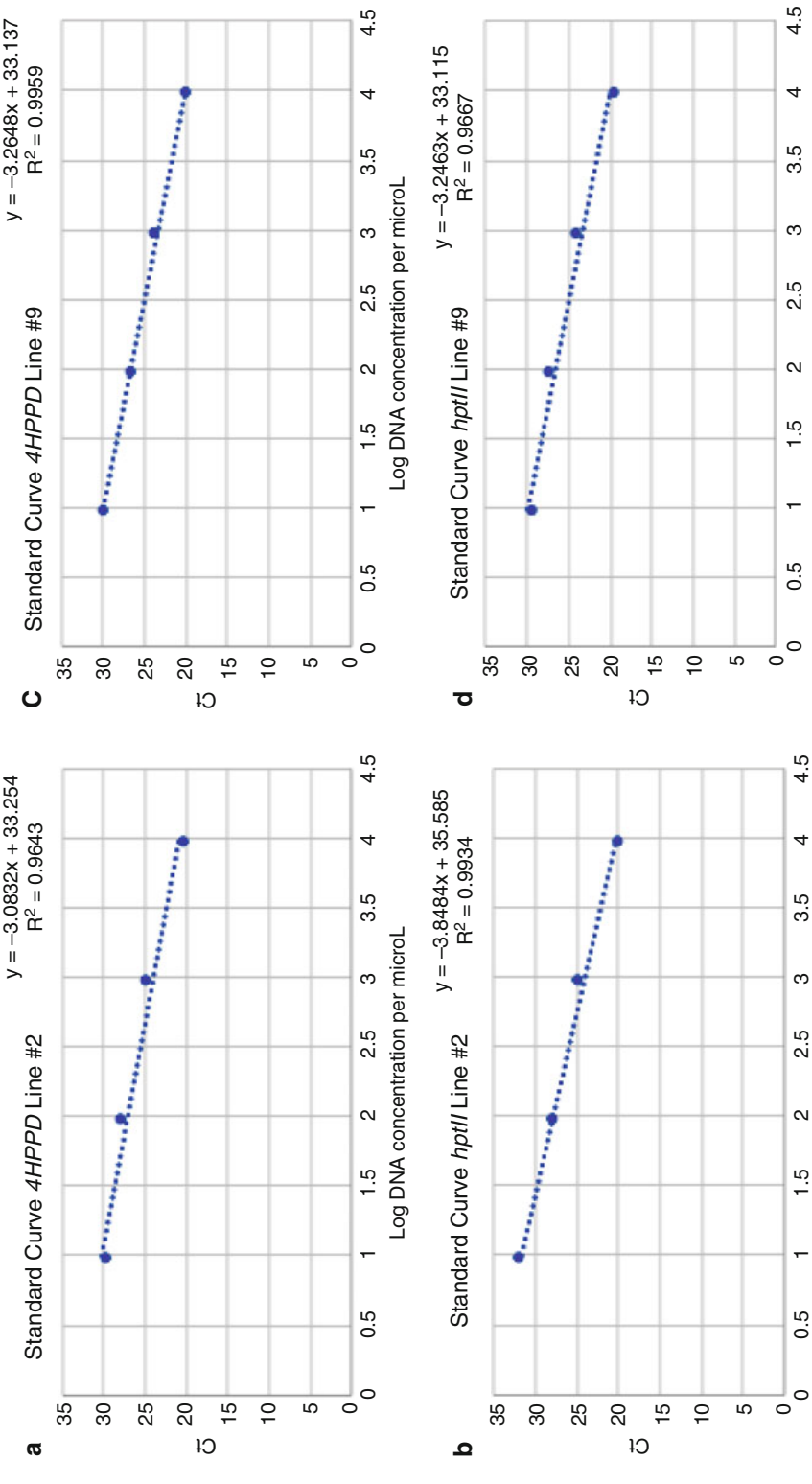

Fig. 4 Standard curves generated through the Ct values of Line #2 and Line #9. (**a**) and (**b**) Standard curves of Ct value for *4HPPD* and *hptII* for CIPK9 overexpressing transgenic Line #2. Different serial dilutions of genomic DNA (100 ng, 10 ng, 1 ng, and 100 pg). (**c**, **d**) Standard curves of Ct value for *4HPPD* and *hptII* for CIPK9 overexpressing transgenic Line #9. Different serial dilutions of genomic DNA (100 ng, 10 ng, 1 ng, and 100 pg)

(www.thermofisher.com/in/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-resources-library/thermo-scientific-web-tools/qPCR-efficiency-calculator.html).

4. For calculation of copy number, we followed the protocol from Weng et al. [9]. The slope and intercept values derived from the standard curve were used. The average Ct values for the four dilutions were used. All these values were plotted in an equation

$$\frac{X}{R} = 10^{\left(\frac{Cx-Ix}{Sx} - \frac{Cr-Ir}{Sr}\right)}$$

Cx and Cr are average Ct values of *hptII* and *4HPPD* for a particular line, Ix and Ir are intercepts for *hptII* and *4HPPD*, respectively, and Sx and Sr are slopes for *hptII* and *4HPPD*, respectively. The value of X/R is doubled to obtain the copy number [9]. Following this formula, we found that Line #2 in our experiment has 1 copy of CIPK9 and Line #9 has 2 copies of CIPK9.

4 Notes

1. DNA quality is an important factor for TCN analysis. So, it should be checked using a normal PCR reaction and any housekeeping gene primers to ascertain the yield.
2. Designing primers is a very crucial part of the TCN analysis. It may be challenging to design qPCR primers that contain a DNA sequence that is not unique within the genome. For the reference genes, it should have only single copy of the gene in the genome. We used the *4HPPD* gene in our analysis.
3. Since CIPK9 is also an endogenous gene, using this we would not have been able to identify the variation in copy number. So, we used *hptII* (present in the binary vector) primers for the quantification of the copy number in the CIPK9 overexpressing transgenic lines.
4. qPCR requires biological and technical triplicates for each run. It is also important to tighten the seal of the PCR plate because evaporation in a well may result in the loss of signal in the respective well in the test PCR plate. Also, it is necessary to determine visually if evaporation occurred in some of the wells to rule out the possibility of failed amplification. The other possible reasons for failed PCR might be nonoptimal PCR conditions due to primers or reaction components.
5. Contamination while executing the qPCR reactions might result in positive signal in the negative control. In this case, remaking all critical components for the qPCR reaction will

help in resolving the problem. The specificity of the assay can be assessed by melting curve analysis (needs to be performed in each run). Nonspecific amplification can be detected by the presence of different melting temperature curves either above or below the specific one.

6. The slope of the standard curve should be in the range of 3.0–3.4. Also, the regression coefficient should be in the range of 0.95–0.99. The PCR efficiency should be between 90% and 110%. These factors should be carefully calculated before copy number calculation. If they are not in range, careful optimization is to be performed.
7. We have used four dilutions. Using higher-order dilutions (diluting the DNA further) will flaten the curve and as a result, will affect the properties of the standard curve.

5 Conclusion

1. We have described here a quantitative real-time PCR assay in Arabidopsis for estimation of exogenous *hptII* copy numbers by comparison with the endogenous reference gene coding for *4HPPD*.
2. Using this method is a definite advantage over the standard Southern blot analysis. This can be done by comparing the Ct signals of the target gene and an endogenous control gene.
3. This method is easier to handle, highly reliable, and useful in the identification of copy numbers in transgenic lines.

Acknowledgement

Research work in G.K.P.'s lab is partially supported by grants from the University of Delhi (IoE/FRP grant), Department of Biotechnology (DBT), Department of Science and Technology (DST), and Council of Scientific and Industrial Research (CSIR), India. S.G. acknowledges CSIR for Junior Research Fellowship.

References

1. Bubner B, Baldwin IT (2004) Use of real-time PCR for determining copy number and zygosity in transgenic plants. Plant Cell Rep 23:263–271
2. Song P, Cai CQ, Skokut M et al (2002) Quantitative real-time PCR as a screening tool for estimating transgene copy number in WHISKERS™-derived transgenic maize. Plant Cell Rep 20:948–954. https://doi.org/10.1007/s00299-001-0432-x
3. Mason G, Provero P, Vaira AM, Accotto GP (2002) Estimating the number of integrations in transformed plants by quantitative real-time PCR. BMC Biotechnol:2. https://doi.org/10.1186/1472-6750-2-20
4. Schmidt MA, Parrott WA (2001) Quantitative detection of transgenes in soybean [Glycine max (L.) Merrill] and peanut (Arachis hypogaea L.) by real-time polymerase chain reaction. Plant

Cell Rep 20:422–428. https://doi.org/10.1007/s002990100326

5. Yuan JS, Burris J, Stewart NR et al (2007) Statistical tools for transgene copy number estimation based on real-time PCR. BMC Bioinformatics. https://doi.org/10.1186/1471-2105-8-S7-S6
6. Garcia I, Rodgers M, Pepin R et al (1999) Characterization and subcellular compartmentation of recombinant 4-hydroxyphenylpyruvate dioxygenase from arabidopsis in transgenic tobacco. Plant Physiol 119:1507–1516. https://doi.org/10.1104/pp.119.4.1507
7. Honda M, Muramoto Y, Kuzuguchi T et al (2002) Determination of gene copy number and genotype of transgenic Arabidopsis thaliana by competitive PCR. J Exp Bot 53:1515–1520. https://doi.org/10.1093/jxb/53.373.1515
8. Kihara T, Zhao CR, Kobayashi Y et al (2006) Simple identification of transgenic Arabidopsis plants carrying a single copy of the integrated gene. Biosci Biotechnol Biochem 70:1780–1783. https://doi.org/10.1271/bbb.50687
9. Weng H, Pan A, Yang L et al (2004) Estimating number of transgene copies in transgenic rapeseed by real-time PCR assay with HMG I/Y as an endogenous reference gene. Plant Mol Biol Report 22:289–300. https://doi.org/10.1007/BF02773139