ELSEVIER

Soil Biology & Biochemistry 39 (2007) 1956–1967

Soil Biology & Biochemistry

www.elsevier.com/locate/soilbio

An empirical approach to target DNA quantification in environmental samples using real-time polymerase chain reactions

Robert H. Gulden[a,**], David Levy-Booth[b], Rachel Campbell[a], Jeff R. Powell[c], Miranda M. Hart[c], Kari E. Dunfield[b], Jack T. Trevors[d,*], K. Peter Pauls[a], John N. Klironomos[c], Clarence J. Swanton[a]

[a]*Department of Plant Agriculture, University of Guelph, Guelph, Ont., Canada N1G 2W1*
[b]*Land Resource Science, University of Guelph, Ont., Canada N1G 2W1*
[c]*Department of Integrative Biology, University of Guelph, Ont., Canada N1G 2W1*
[d]*Department of Environmental Biology, University of Guelph, Ont., Canada N1G 2W1*

Received 26 October 2006; received in revised form 19 February 2007; accepted 20 February 2007
Available online 26 March 2007

Abstract

Recent advances in molecular techniques have allowed for the routine examination of nucleic acids in environmental samples. Although current methodologies are very sensitive, accurate target DNA quantification from environmental samples remains challenging. To facilitate high-throughput DNA quantification from environmental samples, we developed a novel DNA quantification method based on a non-linear curve-fitting approach to extract additional information from quantitative PCR amplification curves and used the fitted parameters to develop multiple regression standard equations for target DNA quantification. A 3-parameter sigmoidal function performed superior to a 4-parameter Weibull function for generating the multiple regression standard equations. In a verification experiment, target DNA was quantified in a series of 'unknown' samples in three soils using this approach and the results were compared to target DNA values determined using corrected and uncorrected Ct-based (threshold cycle) methods. For each method, the deviations from the expected target DNA content were determined. Results clearly showed that over all DNA concentrations, target DNA content determined by the non-linear curve-fitting method was more accurate and more precise than values predicted by all other methods. Analysis of variance conducted on the predicted DNA contents also revealed fewer statistical artifacts with the non-linear curve fitting method compared to the conventional Ct-based methods. The novel approach described here is accurate, inexpensive, and very amenable for automation and high-throughput applications.

Keywords: DNA quantification; Critical threshold; Genetically modified; Multiple regression; Non-linear regression; Polymerase chain reaction; Quantitative PCR; Real-time PCR; SAS; Soil DNA

1. Introduction

Recent advances in DNA extraction and quantification techniques have made it possible to quantify low amounts of target DNA in environmental samples with high levels of sensitivity and accuracy. These tools can be used to successfully address many scientific questions in areas such as molecular ecology, functional genomics, and nucleic acid cycling in the environment (e.g., England et al., 1997; Gulden et al., 2005; Lerat et al., 2005). Real-time quantitative polymerase chain reaction (qPCR) provides high sensitivity, accuracy, repeatability, and a linear response over a large dynamic range (Freeman et al., 1999; Rutledge and Côté, 2003) making it among the best of the quantitative techniques available at this time. However, accurate quantification of target DNA content in environmental samples presents challenges to the analyst that are not present in DNA samples from other sources,

**Also to be correspond to. Tel.: +1 519 824 4120x52512; fax: +1 519 763 8933.
*Corresponding author. Tel.: +1 519 824 4120x53367; fax: +1 519 837 0442.
E-mail addresses: rgulden@uoguelph.ca (R.H. Gulden), jtrevors@uoguelph.ca (J.T. Trevors).

0038-0717/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.soilbio.2007.02.014

including: high spatial variability within and among samples, the presence of PCR inhibiting substances such as clay, humic and fulvic acids, and large ranges in non-target DNA quantities (reviewed by Wilson, 1997).

Real-time qPCR quantifies amplification of target DNA through incorporation of fluorophores into the amplicon during annealing and extension. During real-time PCR, fluorescence takes the shape of a logistic curve with a lower asymptote equating to background fluorescence and an upper asymptote or plateau equating to the exhaustion of *Taq* polymerase or consumables in the PCR. Typically, only a single data point near the lower asymptote is used to quantify the amount of target DNA initially present in the qPCR. The fractional cycle at which the fluorescence curve crosses an arbitrary threshold fluorescence value is used to determine the initial target DNA content of an unknown sample using a standard curve constructed from a dilution series of known amounts of target DNA. The key assumptions of the critical threshold (Ct) or crossing point method are identical amplification efficiencies at different target DNA concentrations and a linear relationship between the Ct over a range of target DNA concentrations. In our studies on environmental samples, we observed that the first assumption is often violated and the same has been reported elsewhere (Liu and Saint, 2002a). For accurate target DNA determination from environmental samples, an estimate of the efficiency of DNA recovery also is required. The addition of a known quantity of an exogenous reference DNA species prior to DNA extraction and subsequent correction for recovery of the reference has been shown to reduce the variability associated with real-time qPCR results from environmental samples (Möller and Jansson, 1997; Park and Crowley, 2005). This however, requires multiplexing or multiple qPCRs for each sample. Moreover, assumptions such as equal extraction and equal amplification efficiencies between the target and the reference DNA must be met for this approach to be accurate. Competitive real-time qPCR has been developed to minimize differences in amplification efficiencies between the target and reference DNA (e.g., Widada et al., 2001). However, much of this work has been conducted with a focus on reverse transcription real-time qPCR, the limitations of which were recently reviewed by Huggett et al. (2005).

To overcome limitations of the Ct method for target DNA determination, a number of alternative mathematical approaches have been suggested (Pfaffl, 2001; Bar et al., 2003; Ramakers et al., 2003; Rutledge and Côté, 2003; Peirson et al., 2003; Tichopad et al., 2003; Rutledge, 2004; Swillens et al., 2004; Larionov et al., 2005). Individual fluorescence curves generated during a real-time qPCR contain substantially more information than is utilized by the Ct method. Liu and Saint (2002b), Bar et al. (2003), and Ramakers et al. (2003) have employed curve fitting methods to estimate the qPCR efficiency of individual reactions during the early stages of the exponential phase of fluorescence above the background and estimate the initial target DNA content without a standard curve. However, these methods also have been developed primarily for reverse transcription real-time qPCR and although these methods can account for differences in PCR efficiency, they do not account for the efficiency of DNA recovery from environmental samples.

In this study, we developed an empirical approach to estimate plant (*Zea mays* L.) target DNA content in soil samples that is economical and suitable for high-throughput applications, and we compared this approach to Ct-based methods for target DNA quantification from environmental samples. This novel approach was based on three main steps: (i) fitting a non-linear regression model to entire fluorescence curves generated by qPCR in standard and unknown samples, (ii) generating a standard equation containing only parameters significant for explaining the target DNA content from standard samples, and (iii) using the standard equation and the appropriate curve fit parameters from unknown samples from the same matrix to estimate target DNA content in these samples.

2. Materials and methods

2.1. Soil and DNA sources

The experiments were carried out on three different soils which were obtained from Elora, Ont., Canada (43°41′N, 80°26′W; sand 26.1%, silt 60.1%, clay 13.8%, pH 7.3,O.M. 5.0%), Woodstock, Ont., Canada (43°8′N, 80°45′W; sand 40.4%, silt 45.0%, clay 14.6%, pH 7.1, O.M. 4.0%), and Lethbridge, Alta., Canada (49°36′N, 1120°36′W; sand 37.1%, silt 38.2%, clay 24.8%, pH 8.0, O.M. 4.0%), Canada. Prior to use, the soils were air dried and confirmed to be free of target DNA, which facilitates generation of the standard equation. However, to simulate the variability from field samples, the soils were neither homogenized nor sieved before starting the experiments.

Genomic plant DNA was extracted from transgenic corn leaf tissue using the CTAB method described by Doyle and Doyle (1987) without modification. Before extraction, the tissue was flash frozen in liquid nitrogen, lyophilized, and ground to a fine powder. After DNA extraction, the quantity and quality of genomic DNA was determined using a spectrophotometer (260/280 nm) and gel electrophoresis (1% agarose). Two genes in the corn genome were used as target sequences, the indigenous 10-kD *Zein* gene (Einspanier et al., 2004; Gulden et al., 2005) which codes for a storage protein and the recombinant *CP4 EPSPS* gene which codes for the 5-enolpyruvylshikimate-3-phospate synthase present in genotypes resistant to the herbicide glyphosate (Barry et al., 1997; Lerat et al., 2005). The copy number of the target genes was determined using the genomic weights provided by Arumuganathan and Earle (1991), with the assumption of 2 copies of the *CP4 EPSPS* gene (heterozygous) per genome and 2 copies of the 10-kD *Zein* gene (homozygous) per genome.

To correct for recovery efficiency of the plant target genes from soil samples, a 620 bp fragment of the green fluorescent protein (*GFP*) gene (*Aequorea victoria* L.) (Prasher et al., 1992) was used as a reference. The partial *GFP* gene had been cloned into the *TOPO TA* vector and transformed into *Escherichia coli* JM109 (Lerat et al., 2005). Transformed bacteria were grown in liquid LB medium supplemented with 50 μM ampicillin at 37 C for 16 h. The *TOPO T4* plasmids were extracted using the Mini Prep kit (Qiagen, Mississauga, Ont.). DNA quality was examined as above and plasmid copy number was determined using standard conversions for a 4520 bp double stranded plasmid (1 ng = 205,286,724 copies).

2.2. Standard curves in soil and verification experiment

To generate the standard curves the Ct-based methods, 10-fold dilution series of corn genomic and plasmid DNA extract were prepared in sterile, ultrapure water. For the generation of standard curves for soil-based Ct methods and the multiple regression standard equation of the novel approach, 10^1–10^8 gene copies of target recombinant *CP4 EPSPS* and indigenous 10 kD-*Zein* corn genes were added to 0.25 gm samples from each soil in 96-well bead plates of the Ultra-Clean soil DNA kit (MoBio Laboratories, Solano Beach, CA). For the preparation of a standard curve for the exogenous control gene, 10^3–10^{10} plasmid copies containing the *GFP* gene fragment were added to a separate set of 0.25 gm soil samples. Two independent samples of each DNA concentration were prepared in each soil and subjected to DNA recovery. This experiment was repeated two times and the data were combined.

In a separate verification experiment, samples of each soil were supplemented with one of three quantities (10^2, 10^4, and 10^6 gene copies) of corn target genes (*CP4 EPSPS* and *Zein*) and one of three quantities (10^2, 10^4, and 10^6 copies) of *GFP* plasmid DNA. The experiment was a complete factorial design containing three replicates of all treatment combinations. The verification experiment was repeated two times independently and the data were combined. For each soil, real-time PCR analysis for each run of the verification experiment and the standards derived from soil was conducted on the same PCR plate.

To investigate variation associated with PCR replication over time (plate replicates) when DNA contents are estimated using the same standards from a different, independent PCR run, the soil DNA extracts from the standard curve experiment in the Elora soil were subjected to four independent PCR analyses for both corn genes over a period of about 3 months. For this and the verification experiments, each biological replicate was repeated twice during each PCR (i.e., two technical replicates).

2.3. DNA recovery and real-time PCR

DNA recovery and real-time quantitative PCR (qPCR) assays for these DNA targets were described in detail by Lerat et al. (2005) and Gulden et al. (2005). In brief, total DNA was recovered using the 96-well format Ultra-Clean soil DNA extraction kit (MoBio Laboratories, Solana Beach, CA). The protocol was modified by adding the floculant, $NH_4Al(SO_4)_2$ (Braid et al., 2003), the nucleic acid–protein interaction inhibitor, aurintricarboxylic acid (Blagodatskaya et al., 2003), and a single glass bead at the beginning of the procedure (Lerat et al., 2005). Total recovered DNA from 0.25 gm air-dried soil was eluted into a 100 μl volume of the supplied storage buffer.

Recombinant *CP4 EPSPS* and indigenous 10-kD *Zein* in corn and the *GFP* fragment on the plasmid were targeted for real-time PCR quantification. Primers, conditions, and amplicon detection for all real-time qPCR assays were described previously by Lerat et al. (2005) and Gulden et al. (2005). Detection of the *EPSPS* and 10-kD *Zein* used molecular beacons with FAM and HEX fluorophores for detection, respectively, while the *GFP* assay was based on SYBR Green I (Lerat et al., 2005). Target DNA in 1 μl subsamples was amplified in 20 μl total volume using an iCycler equipped with iCycler IQ Optical System Software v3.1 (BioRad, Hercules, CA). To clearly identify positive fluorescence curves in samples containing low target DNA concentrations, amplification was conducted for 54 cycles. In addition to conducting real-time qPCR on all soil samples, real-time qPCR was also conducted on the 10-fold DNA dilution series in water used to prepare the soil samples. Appropriate positive and negative controls were included for each qPCR run.

2.4. Curve fitting and data analysis

Critical threshold values adjusted to the same threshold fluorescence and baseline subtracted PCR raw data for each real-time qPCR were exported to MS Excel and prepared for further statistical analysis with SAS (The SAS Institute, Cary, NC). The basic SAS code required for this analysis has been included in Appendix A. Raw baseline subtracted fluorescence data of each qPCR from the ten-fold dilution series in the soils was fit to the following two functions which describe the logistic fluorescence curves generated by real-time qPCR using the NLIN procedure in SAS (Appendix A).

$$3-\text{parameter sigmoidal}, \quad y=\frac{a}{1+e^{-((x-x_0)/b)}}, \tag{1}$$

where a is the maximum fluorescence, b is the slope of the linear portion of the sigmoidal curve, x is the PCR cycle (1–54), and x_0 is the fractional cycle at which fluorescence is one-half of maximum (a).

$$4-\text{parameter Weibull}, \quad y=a\left[1-e^{-\left((x-x_0+b\ \ln\ 2^{1/c})/b\right)^c}\right], \tag{2}$$

where a is the maximum fluorescence, b and c are co-dependent shape parameters that describe the sigmoidal

curve, x is the PCR cycle (1–54), and x_0 is the fractional cycle at which fluorescence is one-half of maximum (a).

The sigmoidal model has been used before for qPCR fluorescence curve analysis (Liu and Saint, 2002b; Rutledge, 2004) and the Weibull function was chosen for its ability to describe these types of curves (Luko, 1999). After non-linear regression analysis, the predicted values of a representative subset of divergent fluorescence curves were graphed with the raw data and Durbin–Watson statistics and the coefficient of determination (R^2) (Kalseth, 1987) were compared. To determine which curve fitting method was most appropriate for subsequent target DNA determination, the average variation expressed as a percentage of the grand mean that was associated with each parameter was determined for all positive fluorescence curves generated from the standard curve experiment in soil. The curve-fitting method that provided the least variation among all parameter means was used for the generation of a standard equation using multiple regression.

2.5. DNA quantification

The content of target corn 10-kD *Zein* and recombinant *EPSPS* in 'unknown' soil samples from the verification experiment were calculated using five different methods. target DNA in soil samples was quantified using (i) a novel multiple regression standard equation (MRSE) approach and results from this approach were compared to results generated by (ii) the Ct method standard curves generated from the dilution series in water (Ct_{water}), (iii) the Ct method corrected for recovery efficiency with the exogenous *GFP* reference gene using the water-based standard curves for both genes ($Ct_{water\ corr}$), (iv) the Ct method using the standard curve generated in soil (Ct_{soil}), and (v) the Ct method corrected for DNA recovery efficiency with the exogenous *GFP* reference gene using standard curves generated in soil for both genes ($Ct_{soil\ corr}$). The third method is the standard Ct–based method often used, where extraction and amplification efficiency of the target DNA is normalized using the recovery and amplification efficiency values generated for a single concentration of a reference gene. Method (iv) was thought to be an improvement over method (iii) and methods (ii) and (v) were included for comparison only.

To generate the standard equation for the MRSE method and the Ct-based standard curves, all fitted parameter values and Ct values from negative PCRs as determined by visual inspection of the fluorescence data, were removed. For the MRSE approach, standard equations were developed for each target gene in each soil using the curve parameters fit for the 3-parameter sigmoidal function. To develop these equations, linear and quadratic components of all parameters were regressed against the $\log_{10}$ values of the dilution series prepared in soil using multiple regression analysis in SAS (see Appendix A). To eliminate over-parameteriztion of MRSEs, non-significant parameter components were dropped and multiple regression was repeated until standard equations containing only the significant combination of components of the parameters were generated for each gene in each soil. The Ct values used for the Ct-based methods were based on a common fluorescence threshold for which Ct values were determined by the iCylcer software before export for further analysis. Ct-based standard curves were developed by regressing Ct values against the $\log_{10}$ DNA concentrations from the dilution series prepared in water and soil using linear regression analysis. MRSEs and Ct standard curves were generated using individual subsamples rather than subsample means to avoid artificial inflating of the coefficient of determination (R^2) through the removal of variability.

In the verification experiment, all samples were subjected to non-linear regression and fit to the sigmoid function to generate the parameter estimates required for DNA quantification using MRSEs. Prior to calculation of DNA quantities using the MRSEs, PCRs with parameter estimates indicative of negative PCRs (e.g., a values that were negative or small positive values (<50), b values that were negative, or x_0 values outside the range of the number of possible PCR cycles) were removed. For the remaining PCRs, target DNA content was determined by using the appropriate parameter estimates and the MRSEs generated for the specific soil and target DNA combination. The same curve-fitting procedure and MRSEs were used to estimate the target DNA content of the PCR results from the multiple PCR runs of the standard curve experiment DNA extracts that were conducted over a period of 3 months for the Elora soil. For the Ct-based methods, target DNA contents of unknown samples from the verification experiment were determined using Ct values and the linear standard curves developed for each target DNA species. PCRs indicative of negative results such as Ct values >45 were not included in this analysis. These were the same PCRs excluded from the MRSE analysis. Ct-based linear standard equations generated from the dilution series prepared in water were used for estimating target DNA content using Ct_{water}. For the $Ct_{water\ corr}$ method, the DNA quantities of each target gene were adjusted by dividing the Ct_{water} copy number by the fraction of recovery (based on copy number) of the exogenous *GFP* reference gene before converting the corrected values back to $\log_{10}$. For the Ct_{soil} method, target DNA was quantified using the linear Ct-based standard curves generated from the dilution series of each gene in each soil. For the $Ct_{soil\ corr}$ approach, the Ct_{soil} values were corrected using the *GFP* recovery determined from the soil-based Ct linear standard curve as described for the $Ct_{water\ corr}$ method. All subsequent data analyses were conducted on $\log_{10}$ values of the means of the technical replicates (i.e., a single value for each biological replicate).

After estimating the target DNA content from the verification experiment using the five different methods, the deviations between the estimated and the known target DNA contents of each sample were determined for each

target gene in each soil. The deviations from the expected values measured the accuracy of the method while the error associated with each means was a measure of the precision. For this comparison and the ANOVA described below, the treatments containing 10^2 *GFP* copies were excluded from the $Ct_{water\ corr}$ and $Ct_{soil\ corr}$ methods as such low exogenous reference gene concentrations are impractical. For each method of DNA quantification, an ANOVA was conducted to determine the partitioning of the Type III sums squares (SS) to each main factor (i.e., corn target DNA content, *GFP* content which altered the ratio of total DNA to target DNA, soil type, target gene, and extract or biological replicate) of the verification experiment. The ANOVA was also conducted on the deviations from the expected values to determine the influence of total DNA (Ct_{water} and Ct_{soil}) and reference gene concentrations for correction ($Ct_{water\ corr}$ and $Ct_{soil\ corr}$) on the accuracy of the predicted target DNA content was determined for all methods. For the four independent repetitions of the standard curve PCR analyses, the contribution to total variation (Type III SS) and the significance of each main effect (target DNA content, plate or replication in time, extract or biological replication, and pipette or technical replication) were determined by ANOVA.

3. Results

3.1. Curve fitting and parameter estimate use for MRSE

Although the Weibull function performed superior to the 3-parameter sigmoidal function at describing the baseline-subtracted fluorescence curves obtained from qPCR (data not shown), the parameter estimates for the sigmoidal function were used to develop the multiple regression standard equations (MRSEs) for determining target DNA content in unknown samples. The Weibull function provided highly variable estimates of the fit parameters for similar curves and this precluded its use for the development of the MRSEs. For example, the mean standard deviations for the a, b, c, and x_0 parameters of the Weibull function for the 10 kD-*Zein* gene in Elora soil standard curves were 57.4%, 93.2%, 55.1%, and 31.5%, respectively, when expressed as a percentage of the mean. In comparison, the means of the standard deviations associated with the a, b, and x_0 parameters for the sigmoidal function for the same gene in the same soil samples were 24.4%, 2.6%, and 22.1%, respectively. The selected fluorescence curves shown in Fig. 1 (HEX-based 10-kD *Zein* PCRs) were typical of real-time qPCR results from soil samples, where (i) the upper asymptote is rarely reached, irrespective of fluorescence chemistry, (ii) the slopes and the maximum fluorescence values decrease as the initial target DNA content decreases, and (iii) slopes of fluorescence curves with similar target DNA content are not always parallel. Negative curves (i.e., curve 5) were easily identified by their parameter estimates (Fig. 1, Table 1). Results for the FAM-based *CP4 EPSPS* and the SYBR Green I-based *GFP* PCRs were similar, albeit with a different ordinate scale (data not shown).

The empirical equations developed for DNA quantification from real-time qPCR analysis are shown in Table 2. Greater R^2 statistics indicate that the MRSEs were superior to the standard curves generated in soil using the Ct method. In some cases (e.g., *CP4 EPSPS* Woodstock and *GFP* Lethbridge), the MRSE substantially improved the proportion of the data explained by the regression equation compared to the Ct-based linear

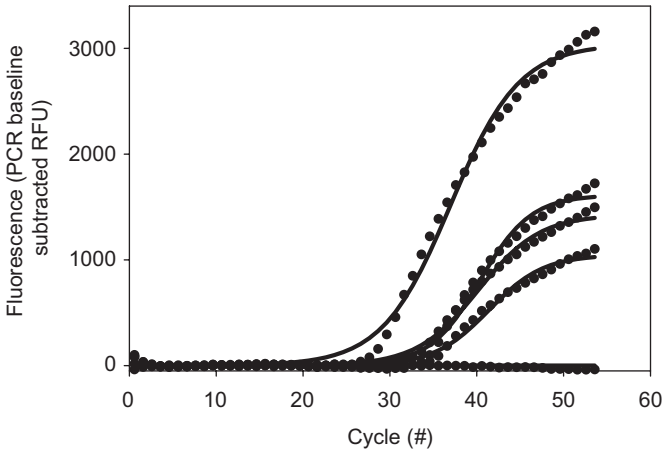

Fig. 1. Raw background subtracted fluorescence data (points) and predicted values (lines) fit to a 3-parameter sigmoidal function for five selected fluorescence curves for the corn 10-kD *Zein* gene.

Table 1
Curve fit parameters for the 3-parameter sigmoid function for the five curves shown in Fig. 1

Curve No.	Target DNA content (gene copies)	Parameter			Durbin–Watson statistic	R^2
		a	b	x_0		
1	10^7	3040.3	3.95	37	0.1437	0.995
2	10^5	1611.7	2.91	40.4	0.1861	0.993
3	10^6	1427.2	3.52	40.2	0.2060	0.994
4	10^4	1047.5	3.14	41.4	0.2293	0.992
5	0	6.04	−1.05	32.2	0.3234	0.031

The target DNA content, the Durbin–Watson and R^2 fit statistics are indicated.

Table 2
Multiple regression standard equations (MRSE) in three soils and threshold cycle (Ct) standard curves in three soils and water for three genes in three different soils

Target gene/soil type	MRSE		Ct_{soil}	
CP4 EPSPS gene				
Elora	$y^a = 26.28-0.576x_0+0.76b$	(0.876)	$y = 19.53-0.418Ct$	(0.832)
Woodstook	$y = 20.41-0.476x_0+1.44\ b$	(0.891)	$y = 16.86-0.349Ct$	(0.668)
Lethbridge	$y = 30.44-0.668x_0+0.53\ b$	(0.903)	$y = 20.89-0.456Ct$	(0.873)
Zein gene				
Elora	$y = 57.43+0.010a-2.49x_0+0.028x_0^2$	(0.915)	$y = 15.33-0.288Ct$	(0.775)
Woodstook	$y = 18.64+0.003a-0.38x_0+0.075b^2$	(0.938)	$y = 16.02-0.306Ct$	(0.870)
Lethbridge	$y = 55.66+0.002a-2.22x_0+0.020x_0^2+0.471b$	(0.966)	$y = 15.32-0.275Ct$	(0.864)
GFP gene				
Elora	$y = 14.14-0.422x_0$	(0.976)	$y = 11.66-0.402Ct$	(0.943)
Woodstook	$y = 2.984-0.009x_0^2+5.044b-0.91316b^2$	(0.970)	$y = 11.82-0.395Ct$	(0.907)
Lethbridge	$y = 7.889-0.009x_0^2+0.867b$	(0.978)	$y = 10.70-0.377Ct$	(0.548)
			Ct_{water}	
CP4 EPSPS			$y = 10.82–0.274Ct$	(0.999)
Zein			$y = 12.20–0.324Ct$	(0.994)
GFP			$y = 7.72–0.267Ct$	(0.980)

R^2 is indicated in parentheses.
[a]Where y is the $\log_{10}$ DNA concentration in gene copies per 0.25 gm soil.

equations. In general, the Ct-based linear standard equations performed poorly when derived from the dilution series added to soil (Ct_{soil}) (Fig. 2). High variation and a tendency for non-linearity as indicated by the bell-shaped residual plot most prominent for the Woodstock standard curve, contributed to the poor R^2 values for Ct_{soil}. Similar results were observed for the *EPSPS* and *GFP* genes (data not shown). PCRs for the lowest target DNA concentration were negative for the corn genes in all soils. As expected, the R^2 values for Ct-based standard curves developed from the water dilution series (Ct_{water}) were high for all three-target genes (>0.98).

The curve fit parameters that contributed significantly to the MRSEs differed between target genes and among soils. For the *CP4 EPSPS* gene (FAM beacon), only linear components of the x_0 and b values were significant in predicting DNA content in all three soils. The significant parameter components for the MRSEs for the HEX-based 10 kD-*Zein* and SYBR Green I-based *GFP* target genes were less consistent among soils. These results indicated varying interactions between fluorescence chemistries and origin of environmental samples.

3.2. Comparison of methods

In the verification experiment, the accuracy and precision of DNA content in soil samples using four different Ct-based methods and the novel MSRE approach was examined by determining the deviations from the expected values for each corn target DNA concentration (Fig. 3). To avoid bias, this comparison was carried out on the PCR data from the same subsamples for all methods. Visual censoring resulted in the exclusion of the same technical replicates for the MRSE and the Ct-based methods which typically were negative curves at the low target DNA concentrations (data not shown). In each soil and for each corn gene (*CP4 EPSPS* and 10 kD-*Zein*), the results were similar and therefore precision and accuracy were summarized for each target DNA concentration only. Low deviations from the expected values for all target DNA concentrations clearly indicate that the overall performance of the MRSE method was superior in accuracy and precision for determining the target DNA content compared to all other methods. Using the MRSE method, both accuracy and precision were excellent for the larger target DNA contents (10^4 and 10^6 gene copies). At low target DNA contents, accuracy declined to within one order of magnitude. For all Ct-based methods, accuracy was more variable across the examined target DNA concentrations. Ct_{water} and $Ct_{water\ corr}$ tended to underpredict target DNA content at high concentrations by wide margins, while the soil-based methods (Ct_{soil} and $Ct_{soil\ corr}$) overpredicted the target DNA content at low concentrations (10^2 copies) by about 4 orders of magnitude. Correcting the target DNA estimates with the *GFP* reporter gene for $Ct_{water\ corr}$ improved the accuracy at higher target DNA concentrations with a concomitant loss in accuracy at the low target DNA concentrations. Correction in $Ct_{water\ corr}$ did not improve precision of the estimates. Differences in the accuracy and precision between Ct_{soil} and $Ct_{soil\ corr}$ were minimal.

Target DNA contents from environmental samples derived from real-time qPCR analysis typically would be subject to further statistical analysis. The subsequent statistical analysis of the estimated target DNA contents and their deviations showed that the MRSE method

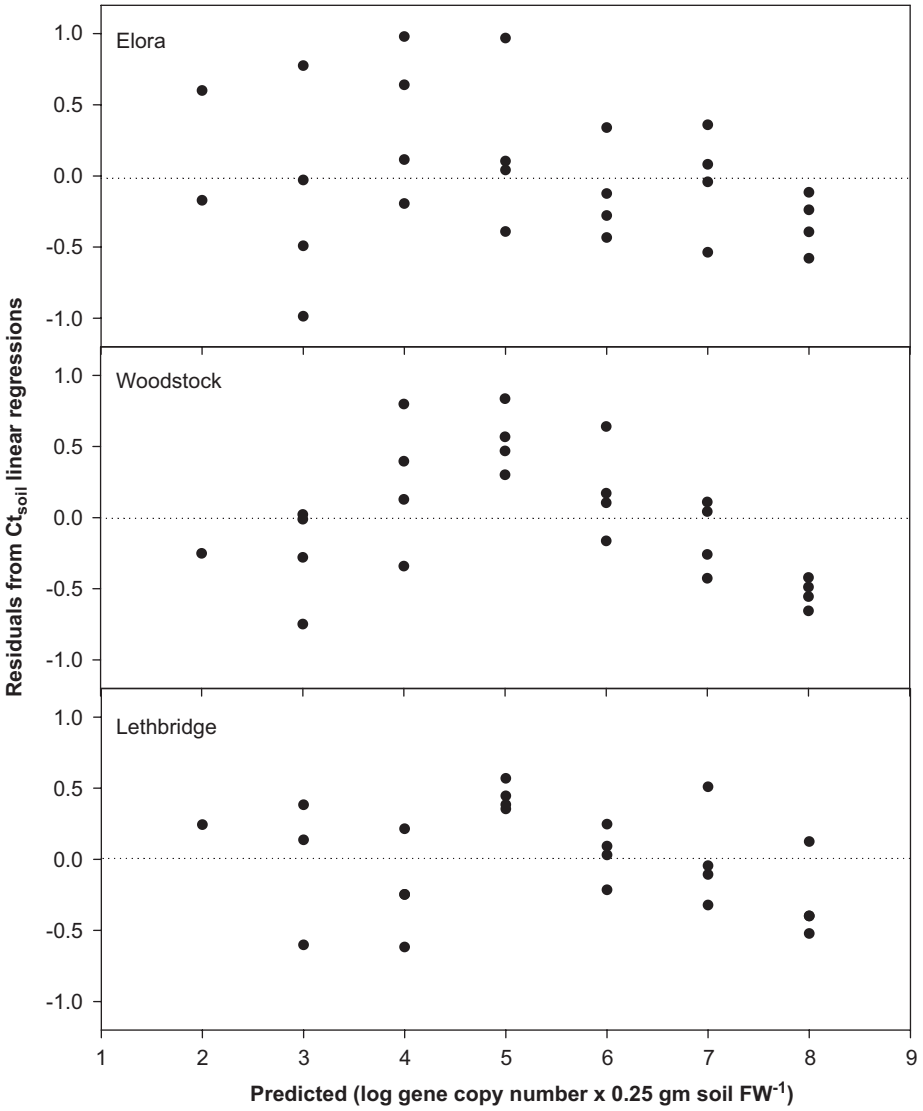

Fig. 2. Predicted vs. residuals plots from the linear regressions analysis generating the standard curves for the Ct_{soil} method for the 10 kD-*Zein* gene in each soil.

resulted in the lowest coefficient of variation in the ANOVA and also the greatest partition of the sums squares to the target DNA main effect (Table 3). For the Ct-based methods, the effects of normalization using the exogenous *GFP* reference gene were similar to those observed in the accuracy and precision analysis, where correction improved the results significantly using $Ct_{water\ corr}$, but did not improve the results of $Ct_{soil\ corr}$. This analysis also showed that the level of the *GFP* reference gene significantly affected the estimation of the corn target DNA using all Ct-based methods (Table 3), but was not significant when using the MRSE approach. In $Ct_{water\ corr}$, mean accuracy was greater when corrected using the highest *GFP* concentration, while in $Ct_{soil\ corr}$, correction with the medium *GFP* concentration resulted in greater accuracy over all soils, genes and target DNA concentrations (data not shown). Comparison of independent PCR replications of the Elora soil standard curve DNA extracts over time (plate replicates) also showed reduced variability using the MRSE approach compared to the Ct_{soil} method (Table 4). Using both methods, the plate replicates contributed more to total variation than the biological and technical replicates; however, relative to the variation contributed by the DNA content this source of variation was low. Results for the *EPSPS* gene were similar, although the magnitude of improvement was only 2-fold for the MRSE method compared to the Ct method (data not shown).

4. Discussion

Our results clearly showed that the multiple regression standard equation approach (MRSE) provided superior accuracy and precision in estimating target DNA content in soil compared to the corrected and uncorrected Ct-based methods. The increased information from entire fluores-

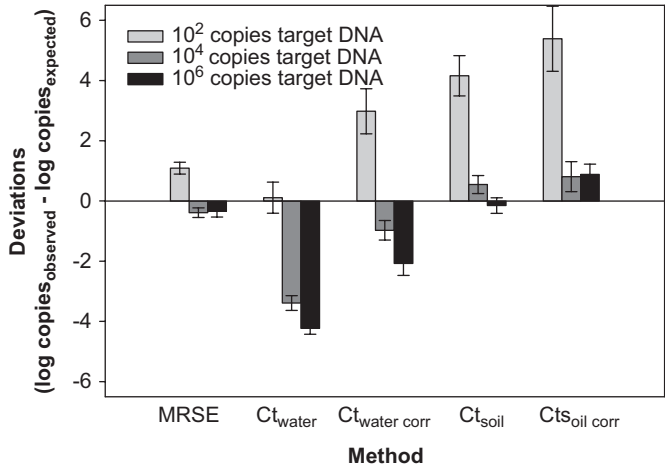

Fig. 3. Mean deviations (accuracy) and standard errors of mean deviations (precision) from the expected corn target DNA concentrations determined by five different methods (for the three levels of target DNA in MRSE, Ct_{water}, and Ct_{soil} $n = 30$–49; for $Ct_{water\ corr}$ and $Ct_{soil\ corr}$ $n = 25$–32). Means for different soil types and target genes were combined.

Table 3
Partitioning of the Type III sums squared (SS) of each main effect for each method of DNA estimation in the verification experiment

Factor	Method[a]				
	MRSE	Ct_{water}	$Ct_{water\ corr}$	Ct_{soil}	$Ct_{soil\ corr}$
Target DNA content	49.35	13.06	16.24	9.73	15.49
GFP content	1.03	6.85	18.16	7.23	8.43
Soil	8.05	11.73	2.61	10.21	7.73
Gene	0.33	0.30	0.01	4.49	5.02
Biological Rep. (Extract)	3.40	0.13	1.25	0.01	0.06
c.v. (%)	22.6	119.3	56.4	34.7	34.4
P_{GFP}	0.2510	0.0084	0.0001	0.006	0.0057

The main effects examined included target DNA content, the ratio of total to target DNA (GFP effect), the variation associated with the three soils, the two corn target genes, and biological replicates. The experimental c.v. and significance of the GFP effect are indicated.

[a]Sums square (SS)%.

cence curves that was used to generate and estimate target DNA content in soil samples clearly contributed to these observations. The improved performance of this method for estimating DNA content in soil samples also resulted in fewer statistical artifacts in the data. Several known sources contribute to the variability of estimating DNA content in environmental samples (Wilson, 1997). These include variation in the efficiency of DNA recovery, the ratio of total to target DNA content, the PCR efficiency due to coextraction of inhibitors, and run-to-run changes in PCR efficiency due to slight differences in the master mix or the performance of the qPCR thermocycler. Estimation of and correction for these sources of error are possible, but may require additional PCRs for each sample which is time consuming, increases costs, and may be impractical for high-throughput applications. The novel approach described here integrated these sources of variation during

Table 4
Partitioning of the Type III sums squared (SS) and *P*-values of each main factor contributing to soil DNA content determined using the MRSE and Ct_{soil} methods for the 10 kD-*Zein* standard curves in the Elora soil run on four independent PCR plates

Factor	Method			
	MRSE		Ct_{soil}	
	SS (%)	*P*-value	SS (%)	*P*-value
Target DNA content	93.11	0.0001	90.86	0.0001
Plate Replicate	0.41	0.0948	3.05	0.0001
Biological Rep. (Extract)	0.01	0.6561	0.04	0.4676
Technical Rep. (Pipette)	0.01	0.7161	0.07	0.3416
c.v. (%)	6.4		6.8	

The main effects examined included target DNA content, replication in time (plate), biological replication (extract effect) and technical replication (pipette effect). The experimental c.v. is indicated.

the construction of the standard equation which eliminated the need for multiple PCRs on individual samples or other means of correction. In these experiments, we determined the relative contribution of several sources of error to estimating DNA contents in soil including the variation associated with different total to target DNA ratios (*GFP* vs. corn target DNA), biological replication (extraction efficiency), technical replication (pipetting error), and replication of PCRs over time (run-to-run variation) and found that the MRSE method was more effective at reducing the magnitude of these sources of variation compared to Ct-based methods. The variation associated with replication over time was greater than that associated with biological and technical replicates, despite the use of a commercially available master mix intended to reduce this source of error. However, the error associated with sequential PCR runs may have been confounded by repeated freezing and thawing of the samples and variation in the performance of the qPCR thermocycler over time. Nevertheless, estimating target DNA content using the MRSE approach also reduced the significance of this source of variation compared to the Ct_{soil} method.

The poor accuracy of the Ct_{soil} method at low target DNA contents may have been caused by the inadequate linear fit of the Ct-based standard curves in soil compared to those generated in water. The reduced fit of the Ct_{soil} standard curve was caused by a combination of inherently greater variation among biological replicates extracted from soil compared to those prepared in water, non-linear recovery of target DNA (e.g., 10 kD-*Zein* gene in Woodstock soil), or increased PCR inhibition at lower target DNA concentrations. However, Lerat et al. (2005) found no evidence of coextraction of PCR inhibitors using this extraction protocol in the Elora soil. Linearity of the DNA recovery efficiency over a range of concentrations is a fundamental assumption for corrected and uncorrected Ct-based methods and violation of this assumption appears to have contributed to our results.

The corrected and uncorrected Ct methods using water-based standard curves were poor at estimating target DNA content in soil. The consistent underestimation of target DNA content by the Ct_{water} method was primarily the result of the inability of this method to account for DNA recovery efficiency from soil. Although accuracy was improved with $Ct_{water\ corr}$, non-linear and variable recovery of target DNA from environmental samples or differences in the amplification efficiencies in water compared to soil samples likely contributed to the erroneous target DNA content estimates. In the verification experiment, the water-based Ct standard curve predicted very low recovery of the exogenous *GFP* reference gene ($< 1\%$ in 80% of the samples), which likely also contributed to the lack of a clear improvement in accuracy and precision through correction for recovery. The poor recovery of the *GFP* reference gene suggests that this DNA species behaved very differently from the genomic plant DNA in soil. Park and Crowley (2005) observed a marked improvement in relative accuracy when correcting target DNA recovery with a *GFP* containing vector as an exogenous reference gene, however, similar to our results, normalization did not affect the precision of their estimates.

To overcome differences in PCR efficiencies between the exogenous reference gene and the target DNA due to sequence differences, techniques such as competitive real-time qPCR have been developed and used on DNA recovered from the environment (e.g., Möller and Jansson, 1997) and these have been coupled with novel mathematical approaches to further reduce variation in the data (Vu et al., 2000). However, the elimination of differences in qPCR efficiency among target and reference DNA alone may not eliminate inaccuracies in target DNA estimates that may be arising from non-linear DNA recovery from environmental matrixes as suggested by our data.

Alternative mathematical approaches to estimating target DNA content have been developed. For example, the method suggested by Ramakers et al. (2003) uses a selected portion (about 6 cycles) of the early exponential phase of the fluorescence amplification curve which was linearized and used to estimate PCR efficiency and initial DNA content of the sample. Similar to the MRSE approach, this method also relies on a single PCR for each sample, but uses only a small portion of the amplification curve. The MRSE approach indicated that for a number of standard equations, the slope (b) of the amplification curves did not significantly contribute to explaining the DNA content of the standard samples. It is not clear, how the method by Ramakers et al. (2003) would perform in such instances.

The standard equations were generated in soil samples free of target DNA. It is unclear what level of contamination of target DNA would be tolerated by the MRSE approach before a significant loss of accuracy or precision in the target DNA estimates. The acceptable level of contamination would likely depend on the range of target DNA to be quantified. It is advisable to minimize target DNA contamination in samples of the environmental matrix used to generate the MRSEs and the level of contamination should be determined. In this experiment, cell lysis was not considered as DNA extracts were used to generate the standards, however, a standard equation can easily be constructed for microbial target DNA quantification from intact cells.

For generating the MRSE, we used the same sigmoid function that was applied to real-time qPCR fluorescence data by Liu and Saint (2002b) and Rutledge (2004). The method by Rutledge (2004) calculated an initial fluorescence value (F_0) from the curve fit parameters and minimized this value for each curve through iterative curve fitting with the successive removal of the final data point of the fluorescence data between each iteration. Then, F_0 was multiplied by a correction factor to determine the initial target DNA content in the PCR. In contrast, we used the curve fit parameters directly to generate a standard equation without the need for iterative curve fitting and calibrating a correction factor. It is not known whether the same accuracy could have been achieved using the method suggested by Rutledge (2004) and developing an empirically derived correction factor which our results suggest to be non-linear to account for DNA recovery and other sources of error. Rutledge (2004) reported inherent difficulties in this function to fit the upper inflection point and plateau phase of the PCR fluorescence data. We observed the same; however, the plateau phase was rarely reached in our experiments, particularly in samples with low-target DNA content. The absence of a distinct plateau on fluorescence curves from environmental samples also precluded the use of amplitude normalization as a means for removing variation (Larionov et al., 2005). In this study, we found no need to right censor the data, as described by Rutledge (2004), because the slopes appeared to be fit well, the error of maximum fluorescence appeared consistent, and the plateau parameter played only a minor role in the MRSEs. Although the Weibull function overcame these limitations to a large extent, the lack of independence of some parameters in this model (Juckett and Rosenberg, 1993) restricted its usefulness for the development of a MSRE. Meaningful parameters of the sigmoid function allowed for quick visual detection of negative curves from further analysis.

For high-throughput applications, statistical outlier detection is essential. There are several possibilities to achieve and automate this process. For example, the standard errors associated with individual parameters determined for the MRSE can be used to exclude curves of unknown samples with parameter values that do not fall within a predetermined range (e.g. ± 3 standard deviations from the mean). Such a procedure can easily be automated using SAS or spreadsheet software. We did not evaluate this method in these experiments. Using the standard equations to prepare a training set for outlier detection is similar to the method described by Bar et al. (2003) who used a training set of qPCRs to calibrate a Gaussian

distribution based on the amplification efficiency of individual qPCRs. With Ct-based methods, aberrant curves cannot be identified easily without visualization due to the Ct value's lack of information describing individual amplification curves.

The extraction efficiency of this modified commercial kit was high which also has been reported elsewhere (Park and Crowley, 2005; Mumy and Findlay, 2004). In the low target DNA treatments (10^2 copies) in this study, a maximum of only a single copy of target DNA for all genes would have been present in the 1 μl DNA extract volume subjected to qPCR analysis under perfect recovery conditions and therefore all PCRs for target DNA concentrations of 10^1 copies per 0.25 gm soil in standards were negative. Many of the qPCRs (about 30–40%) in the low target DNA treatment (10^2 copies per 0.25 gm soil) were confirmed positive for target DNA in all soils indicating high DNA recovery using this method.

Differences among significant parameters for the MRSEs may have been influenced by the target genes, qPCR kinetics, or soil characteristics. Using molecular beacons to detect target sequences compared to SYBR Green I reduced the variability in the significant parameters for MRSEs among the three soils. Whether this was due to the target DNA or differences in the qPCR kinetics is not known. Nevertheless, such subtle interactions between target genes, soil type, and qPCR kinetics cannot be detected with Ct-based methods.

In summary, empirical evidence presented here showed that the novel approach for estimating target DNA content in environmental samples improved the accuracy of estimation and reduced unwanted variability and statistical artifacts in the data compared to conventional Ct-based methods. The approach advocated here consists of non-linear curve fitting of a sigmoidal function to whole fluorescence curves generated from real-time PCR for standard and unknown samples. A standard equation (MRSE) was then generated using only parameters significant for describing the DNA content of standard samples and this equation and the appropriate curve fit parameters from unknown samples were used to estimate target DNA content in these samples. This method is suitable for high-throughput applications and is easily automated. Using this approach, specific standard equations are necessary for accurate estimation of target DNA content for different matrixes (e.g., soil types) and therefore this method is most suitable for experiments with high sampling frequencies from the same environmental matrix.

Acknowledgments

Funds for this project were provided by the NSERC Strategic Grant Program (STPGP 258065-02) (Canada).

Appendix A

SAS code for non-linear curve fitting and multiple regression are shown in Table A1.

Table A1
SAS code for non-linear curve fitting and multiple regression

SAS code	Non-linear curve fitting
data persist;	– Data input step.
input cycle AA BB CC DD …; cards;	cycle = PCR cycle, AA = the fluorescence value for PCR curve *A* at cycle *x*
0.59 38.3 26.11 113.92 109.4 … 1.59 −6.91 −27.31 34.84 15.17 2.59 −13 −36.63 −3.46 2.49 … 51.58 1449.4 1553.75 1925.78 2101.3 52.58 1505.73 1561.42 1965.51 2150.29 53.58 1567.84 1586.32 2031.22 2132.9	– Data block.
proc nlin data = persist method = marquardt;	– Non-linear regression using Marquardt's method.
title 'AA'; parameters *a* = 800 xo = 30 *b* = 2;	– Parameters and starting values; *a* = max fluorescence; xo = threshold cycle at which max fluorescence is reached; *b* = slope of the exponential phase
model AA = *a*/(1 + exp(−1*(cycle–xo)/*b*));	– 3-parameter sigmoidal function using the parameter starting points and the fluorescence values (AA) at each amplification cycle (cycle)
ods select ParameterEstimates; run; quit;	– Omits all output other than parameter estimates.
…	– Repeat this procedure for each PCR curve to be fit.

Table A1 (*continued*)

SAS code	Non-linear curve fitting
SAS code	Multiple regression standard equation (MRSE)
data persist;	– Data input step.
input soil$ DNA *a* xo *b*;	– Variables in data block.
*ct*2 = *ct***ct*; *a*2 = *a***a*; xo2 = xo*xo; *b*2 = *b***b*; cards;	– Calculating quadratic parameter estimates.
Elora 8.0 1359.8 35.15 3.41 Elora 8.0 1653.9 35.04 3.46 Elora 8.0 2735.4 36.77 3.82 …	– Data block.
proc sort; by soil;	– Data must be sorted for 'by' step in subsequent procedure which separates the data analysis accordingly.

proc reg data = persist;
model DNA = *a a*2 xo xo2 *b b*2;
by soil;
run; quit;
	– Multiple linear regression. Use 'trial and error' to reduce to only significant parameter estimates (*a a*2 xo xo2 *b b*2). Paramenter estimates are regressed against target DNA content of standards to generate the standard equation.

proc means mean stddev data = persist; var *a* xo *c*; by soil; run; quit;	– Generation of parameter means and standard deviations that could be used for outlier detection.

References

Arumuganathan, K., Earle, E.D., 1991. Nuclear DNA content of some important plant species. Plant Molecular Biology Report 9, 208–219.

Bar, T., Stahlberg, A., Muszta, A., Kubista, M., 2003. Kinetic outlier detection (KOD) in real-time PCR. Nucleic Acids Research 31, e105.

Barry, G.F., Kishore, G.M., Padgette, S.R., Stallings, W.C., 1997. Glyphosate-tolerant 5-enolpyruvylshikimate-3-phospate synthases. US Patent 5,627,060.

Blagodatskaya, E.V., Blagodatskii, S.A., Anderson, T.H., 2003. Quantitative isolation of microbial DNA from different types of soils of natural and agricultural ecosystems. Microbiology 72, 744–749.

Braid, M.D., Daniels, L.M., Kitts, C.L., 2003. Removal of PCR inhibitors from soil DNA by chemical flocculation. Journal of Microbiological Methods 52, 389–393.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19, 11–15.

Einspanier, R., Lutz, B., Rief, S., Berezina, O., Zverlov, V., Schwarz, W., Mayer, J., 2004. Tracing residual recombinant feed molecules during digestions and rumen bacterial diversity in cattle fed transgene maize. European Food Research & Technology 218, 269–273.

England, L.S., Lee, H., Trevors, J.T., 1997. Persistence of *Pseudomonas aureofaciens* strains and DNA in soil. Soil Biology & Biochemistry 29, 1521–1527.

Freeman, W.M., Walker, S.J., Vrana, K.E., 1999. Quantitative rt-PCR: Pitfalls and potential. Biotechniques 26, 112–125.

Gulden, R.H., Lerat, S., Hart, M.M., Powell, J.R., Trevors, J.T., Pauls, K.P., Klironomos, J.N., Swanton, C.J., 2005. Quantitation of transgenic plant DNA in leachate water: real-time polymerase chain reaction analysis. Journal of Agricultural & Food Chemistry 53, 5858–5865.

Huggett, J., Dheda, K., Bustin, S., Zumla, A., 2005. Real-time RT-PCR normalization: strategies and considerations. Genes and Immunity 6, 279–284.

Juckett, D.A., Rosenberg, B., 1993. Comparison of the Gompertz and Weibull functions as descriptors for human mortality distributions and their intersections. Mechanisms of Ageing & Development 69, 1–31.

Kalseth, T.O., 1987. Cautionary note about R^2. The American Statistician 39, 279–285.

Larionov, A., Krause, A., Miller, W., 2005. A standard curve based method for relative real time PCR data processing. BMC Bioinformatics 6 (62).

Lerat, S., England, L.S., Vincent, M.L., Pauls, K.P., Swanton, C.J., Klironomos, J.N., Trevors, J.T., 2005. Real-time polymerase chain reaction quantification of the transgenes for roundup ready corn and roundup ready soybean in soil samples. Journal of Agricultural & Food Chemistry 53, 1337–1342.

Liu, W.H., Saint, D.A., 2002a. A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. Annals of Biochemistry 305, 291.

Liu, W.H., Saint, D.A., 2002b. Validation of a quantitative method for real time PCR kinetics. Biochemical & Biophysical Research Communications 294, 347–353.

Luko, S., 1999. A review of Weibull distributions and selected engineering applications. SAE Transactions 108, 398–412.

Möller, A., Jansson, J.K., 1997. Quantification of genetically tagged cyanobacteria in Baltic Sea sediment by competitive. PCR Biotechniques 22, 512–518.

Mumy, K.L., Findlay, R.H., 2004. Convenient determination of DNA extraction efficiency using an external DNA recovery standard and quantitative-competitive PCR. Journal of Microbiological Methods 57, 259–268.

Park, J.W., Crowley, D.E., 2005. Normalization of soil DNA extraction for accurate quantification real-time PCR and of target genes by DGGE. Biotechniques 38, 579–586.

Peirson, S.N., Butler, J.N., Foster, R.G., 2003. Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. Nucleic Acids Research 31, e73.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research 29, e45.

Prasher, D.C., Eckenrode, V.K., Ward, W.W., Pendergast, F.G., Cormier, M.J., 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. Gene 111, 229–233.

Ramakers, C., Ruijter, J.M., Deprez, R.H.L., Moorman, A.F.M., 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neuroscience Letters 339, 62–66.

Rutledge, R.G., 2004. Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. Nucleic Acids Research 32, e178.

Rutledge, R.G., Côté, C., 2003. Mathematics of quantitative kinetic PCR and the application of standard curves. Nucleic Acids Research 31, e93.

Swillens, S., Goffard, J.C., Marechal, Y., d'Exaerde, A.D., El Housni, H., 2004. Instant evaluation of the absolute initial number of cDNA copies from a single real-time PCR curve. Nucleic Acids Research 32, e56.

Tichopad, A., Dilger, M., Schwarz, G., Pfaffl, M.W., 2003. Standardized determination of real-time PCR efficiency from a single reaction set-up. Nucleic Acids Research 31, e122.

Vu, H.L., Troubetzkoy, S., Nguyen, H.H., Russell, M.W., Mestecky, J., 2000. A method for quantification of absolute amounts of nucleic acids by (RT)-PCR and a new mathematical model for data analysis. FASEB Journal 14 (Supplement S), A1234.

Widada, J., Nojiri, H., Kasuga, K., Yoshida, T., Habe, H., Omori, T., 2001. Quantification of the carbazole 1,9a-dioxygenase gene by real-time competitive PCR combined with co-extraction of standards. FEMS Microbiology Letters 202, 51–57.

Wilson, I.G., 1997. Inhibition and facilitation of nucleic acid amplification. Applied & Environmental Microbiology 63, 3741–3751.