BMC
Bioinformatics

METHODOLOGY ARTICLE

Open Access

Shape based kinetic outlier detection in real-time PCR

Davide Sisti*[†1], Michele Guescini*[†2], Marco BL Rocchi[1], Pasquale Tibollo[2], Mario D'Atri[1] and Vilberto Stocchi[2]

Abstract

Background: Real-time PCR has recently become the technique of choice for absolute and relative nucleic acid quantification. The gold standard quantification method in real-time PCR assumes that the compared samples have similar PCR efficiency. However, many factors present in biological samples affect PCR kinetic, confounding quantification analysis. In this work we propose a new strategy to detect outlier samples, called SOD.

Results: Richards function was fitted on fluorescence readings to parameterize the amplification curves. There was not a significant correlation between calculated amplification parameters (plateau, slope and y-coordinate of the inflection point) and the Log of input DNA demonstrating that this approach can be used to achieve a "fingerprint" for each amplification curve. To identify the outlier runs, the calculated parameters of each unknown sample were compared to those of the standard samples. When a significant underestimation of starting DNA molecules was found, due to the presence of biological inhibitors such as tannic acid, IgG or quercitin, SOD efficiently marked these amplification profiles as outliers. SOD was subsequently compared with KOD, the current approach based on PCR efficiency estimation. The data obtained showed that SOD was more sensitive than KOD, whereas SOD and KOD were equally specific.

Conclusion: Our results demonstrated, for the first time, that outlier detection can be based on amplification shape instead of PCR efficiency. SOD represents an improvement in real-time PCR analysis because it decreases the variance of data thus increasing the reliability of quantification.

Background

In the last few years, real-time quantitative polymerase chain reaction (real-time PCR) has become the technique of choice for absolute or relative quantification of gene expression due to its rapidity, accuracy and sensitivity [1-3]. Furthermore, recent advances in the sequencing of the human genome, mRNA and miRNA expression profiling of numerous cancer types, disease-associated polymorphism identification and the expanding availability of genomic sequence information for human pathogens have led to marked growth in molecular diagnostics [4-6].

The gold standard quantification method (Ct method) in real-time PCR assumes that the compared samples have similar PCR efficiencies. However, quantification by real-time PCR is very sensitive to slight differences in PCR efficiencies among samples. Indeed, a small difference of 5% in PCR efficiency will result in a three-fold difference in the amount of DNA after 25 cycles of exponential amplification. Many factors present in samples as well as co-extracted contaminants can inhibit PCR, confounding template amplification and analysis [7-10]. This is a major problem when working with biological samples. Severe inhibition will lead to false-negative results, whereas a slight to moderate inhibition can result in an underestimation of the affected sample's DNA concentration [11]. Furthermore, amplification efficiency can fluctuate as a function of non-optimal assay design, enzyme instability, or the presence of inhibitors [12]. Although a variety of methods have been developed to quantify template DNA [11,13-17], very few allow simultaneous evaluation of template quantity and quality

* Correspondence: davide.sisti@uniurb.it, michele.guescini@uniurb.it[1]
Dipartimento DiSUAN, Sezione di Biomatematica, Università degli Studi di Urbino "Carlo Bo", Campus Scientifico Sogesta; Località Crocicchia - 61029 Urbino, Italy
[2] Dipartimento di Scienze Biomolecolari, Sezione di Ricerca sull'Attività Motoria e della Salute, Università degli Studi di Urbino "Carlo Bo", Via I Maggetti, 26/2 - 61029 Urbino, Italy
† Contributed equally
Full list of author information is available at the end of the article

without the addition of an internal positive control that is co-amplified with the target of interest. Hence Bar and co-workers proposed a method (called KOD) based on amplification efficiency calculation for the early detection of non-optimal assay conditions [18,19]. This approach is extremely straightforward and effective, but it is based on a PCR amplification efficiency calculation for which there is still not a method fully accepted by the scientific community. A large number of studies have attempted to calculate amplification efficiency assuming that PCR is inherently exponential in nature. Based on the assumption of the log-linearity region, constant amplification efficiency is calculated from the slope of linear regression in that window [20-23]. An alternative approach is based on the observation that PCR trajectory can be effectively modelled by the sigmoid function [14,24] allowing PCR efficiency to be estimated using non-linear regression fitting [15,25,26]. Recently, a simplified approach called "linear regression of efficiency" has allowed us to estimate amplification efficiency by applying linear regression analysis to the fluorescence readings within the central region of amplification profile [27]. Notably, it has been demonstrated that estimates of PCR efficiency vary widely according to the approach that has been adopted [28].

Very recently, Tichopad et al. [29] introduced a new quality control test for quantitative PCR; in this procedure the first derivative maximum and the second derivative maximum were estimated using a logistic fitting on the PCR trajectory. This approach allowed them to monitor the first half of the curve using two parameters.

Our study aims to develop a quality test tool, which is not based on amplification efficiency estimation, in order to detect samples that do not show an amplification kinetic similar to those of standard samples. In this work, a non-linear fitting of Richards equation was used to parameterize PCR amplification profiles from a large sample set. The subsequent calculation of the variance of the estimated parameters and the development of a statistical measure based on the Mahalanobis distance allowed us to develop the SOD method (Shape based kinetic Outlier Detection). The SOD analysis of inhibited amplifications and the comparison of this method with KOD were investigated in detail.

Methods
Quantitative Real-Time PCR
The DNA standard consisted of a pGEM-T (Promega) plasmid containing a 104 bp fragment of the mitochondrial gene NADH dehydrogenase 1 (MT-ND1) as insert. This DNA fragment was produced by the ND1/ND2 primer pair (forward ND1: 5'-ACGCCATAAAACTCT-TCACCAAAG-3' and reverse ND2: 5'-TAGTAGAA-

GAGCGATGGTGAGAGCTA-3'). This plasmid was purified using the Plasmid Midi Kit (Qiagen) according to the manufacturer's instructions. The final concentration of the standard plasmid was estimated spectophotometrically by averaging three replicate A_{260} absorbance determinations.

Real-time PCR amplifications were conducted using LightCycler® 480 SYBR Green I Master (Roche) according to the manufacturer's instructions, with 500 nM primers and a variable amount of DNA standard in a 20 μl final reaction volume. Thermocycling was conducted using a LightCycler® 480 (Roche) initiated by a 10 min incubation at 95°C, followed by 40 cycles (95°C for 5 s; 60°C for 5 s; 72°C for 20 s) with a single fluorescent reading taken at the end of each cycle. Each reaction combination, namely starting DNA and inhibitor agent, was performed in triplicate and repeated in two separate amplification runs. All the runs were completed with a melt curve analysis to confirm the specificity of amplification and lack of primer dimers. *Ct* (fit point method) was determined by the LightCycler® 480 software version 1.2 and exported into an MS Excel data sheet (Microsoft) for analysis after background subtraction (available as Additional file 1). For *Ct* (fit point method) evaluation, a fluorescence threshold manually set to 0.4 was used for all runs.

Estimation of PCR efficiency
The raw PCR data were used to calculate amplification efficiency. The PCR efficiency for each individual sample was derived from the slope of the regression line in the window of linearity [20]. Baseline correction and window of linearity identification were carried out using the latest version of LinRegPCR (v11.0) [23]. PCR efficiencies were estimated from four sample sets: standard amplification curves, standard amplification curves with the addition of tannic acid read-outs, standard amplification curves with the addition of IgG read-outs and standard amplification curves with the addition of quercitin read-outs. The window of linearity calculated from all the data sets encompassed the fluorescence threshold of 0.4 chosen for the quantitative analysis.

Mathematical model of KOD
The mathematical model of KOD, based on efficiency, was proposed by Bar et al. [18]. Briefly, this was done comparing PCR efficiency of a sample (x_{eff}) with the efficiencies of standard curve samples. A test sample is classified as an outlier if $|z| > 1.96$ with $z = \frac{x_{eff} - \mu_{eff}}{\sigma_{eff}}$, where μ_{eff} is the efficiency mean and σ_{eff} is the standard deviation

of the efficiency of standard curve samples. Alternatively, it is to be considered that the statistic $\left(\frac{x-\mu}{\sigma}\right)^2$ is distributed as a χ^2 with one degree of freedom; if $\chi^2 > 3.84$, we can reject the null hypothesis at $\alpha = 0.05$.

Mathematical model of SOD

Shape based kinetic outlier detection (SOD) was based on the shapes of the amplification curves. In order to fit fluorescence raw data, nonlinear regression fitting of 5-parameter Richards function, an extension of the logistic growth curve, was used [11,25].

$$F_x = \frac{F_{\max}}{\left[1 + e^{\left(-\frac{1}{b}(x-c)\right)}\right]^d} + F_b \tag{1}$$

where x is the cycle number, F_x is the reaction fluorescence at cycle x, F_{max} is the maximal reaction fluorescence, F_b is the background reaction fluorescence and b, c and d represents the estimated coefficients. Nonlinear regressions for 5-parameter Richards functions were performed determining unweighted least squares estimates of parameters using the Levenberg-Marquardt method.

The shape parameters used were the plateau value of amplification curve (F_{max}), tangent straight line slope in inflection point (m) and y-coordinate of inflection point (Y_f) (Additional file 2).

The y-coordinate of inflection point (Y_f) was calculated as follows:

$$Y_f = F_{\max}\left(\frac{d}{d+1}\right)^d \tag{2}$$

and the tangent straight line slope (m) was estimated as:

$$m = \frac{F_{\max}}{b}\left(\frac{d}{d+1}\right)^{d+1} \tag{3}$$

Normal distribution of F_{max}, Y_f and m parameters, obtained from standard samples, was checked using the Kolmogorov-Smirnov test for normality; the significance of the correlation between these parameters and input DNA concentrations, expressed as *Log(DNA)*, was tested with a *t* test as follows:

$$t = r\sqrt{\frac{n-2}{1-r^2}} \tag{4}$$

where r is the Pearson coefficient and n the sample size ($n = 72$). The multivariate normality of the adopted reference set was evaluated according to Rencher AC [30] (Additional file 3). In addition, the asymmetry (*Asym*) of the amplification curves was estimated as follows:

$$Asym = 1 - \frac{2Y_f}{F_{\max}} \tag{5}$$

replacing Y_f and F_{max}, Eq. 5 can be simplified as:

$Asym = 1 - 2\left(\frac{d}{d+1}\right)^d$. In agreement with this equation the curve is symmetric (that is Asym = 0) when d = 1, or 2*Yf = Fmax. On the contrary, when d>1 we have 2*Yf<Fmax (the curve is asymmetric) hence Asym>0.

Statistical model of SOD

After developing a method to estimate three different shape-parameters (F_{max}, Y_f, m), the next step was to set a criterion to identify test samples that deviated from expected values. This was done using sample vector $y = \begin{pmatrix} F_{\max} \\ Y_f \\ m \end{pmatrix}$ which can be calculated for each experimental amplification; if **y** belongs to a multivariate normal distribution, with mean vector $\mu = \begin{pmatrix} \mu_{F\max} \\ \mu_{Yf} \\ \mu_m \end{pmatrix}$ and Σ the corresponding variance-covariance matrix, the *(y-μ)'Σ*⁻¹*(y-μ)* value (Mahalanobis distance) has asymptotic χ^2 distribution, with 3 degrees of freedom. The Mahalanobis distance is based on correlations between variables through which different patterns can be identified and analyzed. It is a useful way of determining the similarity of an unknown multivariate sample set to a known one. It takes into account the correlations of the data set and is not dependent on the scale of measurements. Mean vector and variance-covariance matrix were calculated from

shape parameters of standard curve samples. Then if $\chi^2 >$ 7.81, we can reject the null hypothesis (with $\alpha = 0.05$) and establish that the shape of the amplification curve is different from the shape of the standard curve samples, considering all three parameters [30]. All elaborations and graphics were obtained using Excel (Microsoft), Statistica 6.0 (Statsoft) and Statistical Package for Social Sciences (SPSS 13.0).

Results
Standard curve SOD analysis
The SOD model relies on the assumption that in order to achieve a reliable quantification, the amplification curves of unknown samples should not be significantly different from those of the standard curve. We introduced the idea that the amplification kinetic can be monitored by the shape of the amplification curve. The shape of amplification curves was parameterized using the nonlinear regression fitting of the Richards function on the fluorescence readings [11]. This mathematical procedure allowed us to obtain the five parameters characteristic of the Richards equation. These values were subsequently used to calculate the slope of the tangent at the inflection point (m), the y-coordinate of the inflection point (y_f) and the maximum fluorescence value (F_{max}) of the reading. Finally, these three parameters allowed us to create a "fingerprint" for each amplification curve.

Based on this assumption, the parameters m, y_f and F_{max} of the amplifications used to build a standard curve should not be significantly different from one another and should not be correlated with input DNA. To verify this assumption, a standard curve was generated over a wide range of input DNA (3.14×10^7-3.14×10^2; Fig. 1; Additional files 1). Table 1 shows the mean, SD, and Kolmogorov-Smirnov test from a total of 72 runs. These results demonstrated that m, y_f and F_{max} were normally distributed, even though they showed a different dispersion. Subsequently, the relationship between m, y_f and F_{max} and the Log of the starting DNA template was studied. As shown in Fig. 2, there was not a significant corre-

lation between the Log of input DNA and these parameters (F_{max}: $R^2 = 0.017$ $p = 0.28$; y_f: $R^2 = 0.033$ $p = 0.12$; m: $R^2 = 0.030$ $p = 0.14$). In fact, determination coefficients (R^2) quantified only a very low proportion of parameter variances less than 3,3%.

In order to objectively define an amplification profile as an outlier, we introduced the variable Log(Nob/Nexp), which estimates errors from quantification analysis using the Ct method. This variable relies on the residues estimated as the difference between calculated molecules, using the Ct method (Log of Number of Observed Molecules, referred to as LogNob), and input DNA molecules (Log of Expected Molecules, referred to as LogNexp; in fact LogNob-LogNexp = Log(Nob/Nexp)). The ratio Log(Nob/Nexp) showed a normal distribution satisfying the assumption of homoscedasticity (Additional file 4). It is thus possible to determine a 95% confidence interval (CI) for the variable Log(Nob/Nexp). These residues showed a normal distribution regardless of the starting DNA template, with the average equal to zero and the standard deviation constant ($\sigma = 0.041$). In our database, out of a total of 72 runs used to construct the standard curve, 6 runs showed the ratio Log(Nob/Nexp) out of the CI (Additional file 5). Subsequently, PCR efficiency (Eff) was also estimated for each amplification curve; the LinRegPCR software [20,23] was used to fit the data points in the optimal range of the PCR exponential phase to obtain an automated evaluation of Eff (Table 1).

To determine how well outlier samples can be identified by KOD and SOD, we applied these statistical analyses to the runs of the standard curve; in particular we found that KOD identified 2 runs over the χ^2 threshold value of 3.84 while SOD revealed 3 runs out of the CI (Additional file 5). These outliers are probably false-positives due to the definition and intrinsic properties of the 95% CI.

Inhibitor effects on real-time amplification
Tannic acid oxidizes to form quinones which covalently bind to *Taq* DNA polymerase inhibiting its activity [31]. Real-time amplification plots from 3.5×10^4 DNA molecules in the presence of increasing concentrations (0-0.1 mg per mL) of tannic acids were obtained. All the quantification values were obtained using the *Ct* method. The

Table 1: One-Sample Kolmogorov-Smirnov test of calibration curve.

	Eff	yf	m	Fmax
N = 72				
Mean (S.D.)	1.88 (0.02)	23.89 (2.86)	8.61 (1.20)	46.41 (6.07)
K-S value (*p*)	0.99 (0.28)	1.25 (0.09)	0.75 (0.63)	1.13 (0.15)

Means (standard deviation) and Kolmogorov-Smirnov (K-S) test value (probability) of the following parameters: LineReg efficiency (*Eff*), ordinate value of inflection point (*y_f*), slope of tangent straight line in inflection point (*m*) and plateau value (*Fmax*).
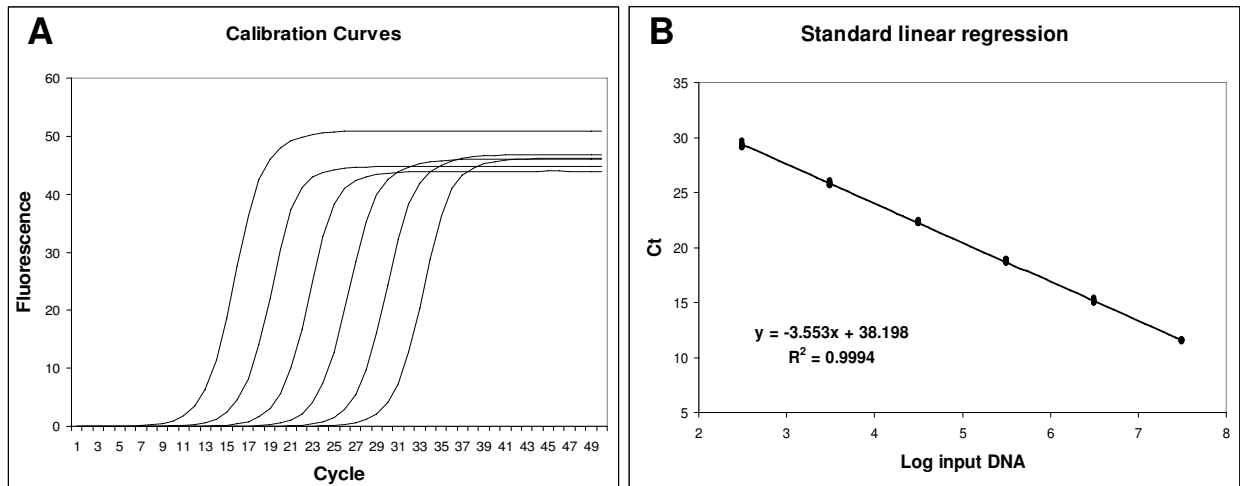
Figure 1 Linear regression analysis of standard samples. The amplification profiles were produced by averaging the fluorescence readings of twelve replicate reactions (A). Linear regression obtained plotting Log input DNA versus *Ct* (B).
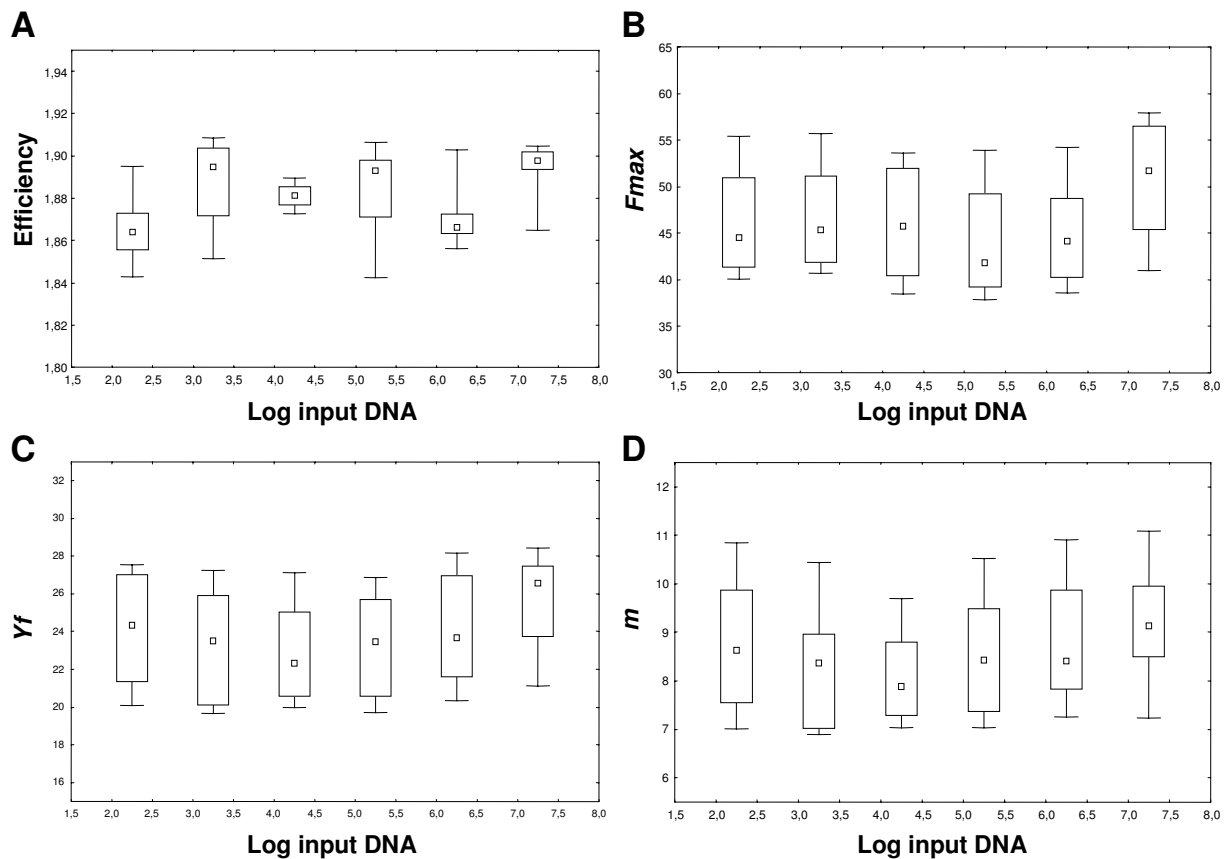

Figure 2 Efficiency and shape parameter values of standard curve samples. The plots of efficiency (A), *Fmax* (B), *Yf* (C) and *m* (D) were shown; we reported in abscisse the Log transformation of input DNA and in ordinate the parameter value. The square represents the median, the length of the box shows the interquartile range and the whiskers indicate the min-max values of the estimated parameters.
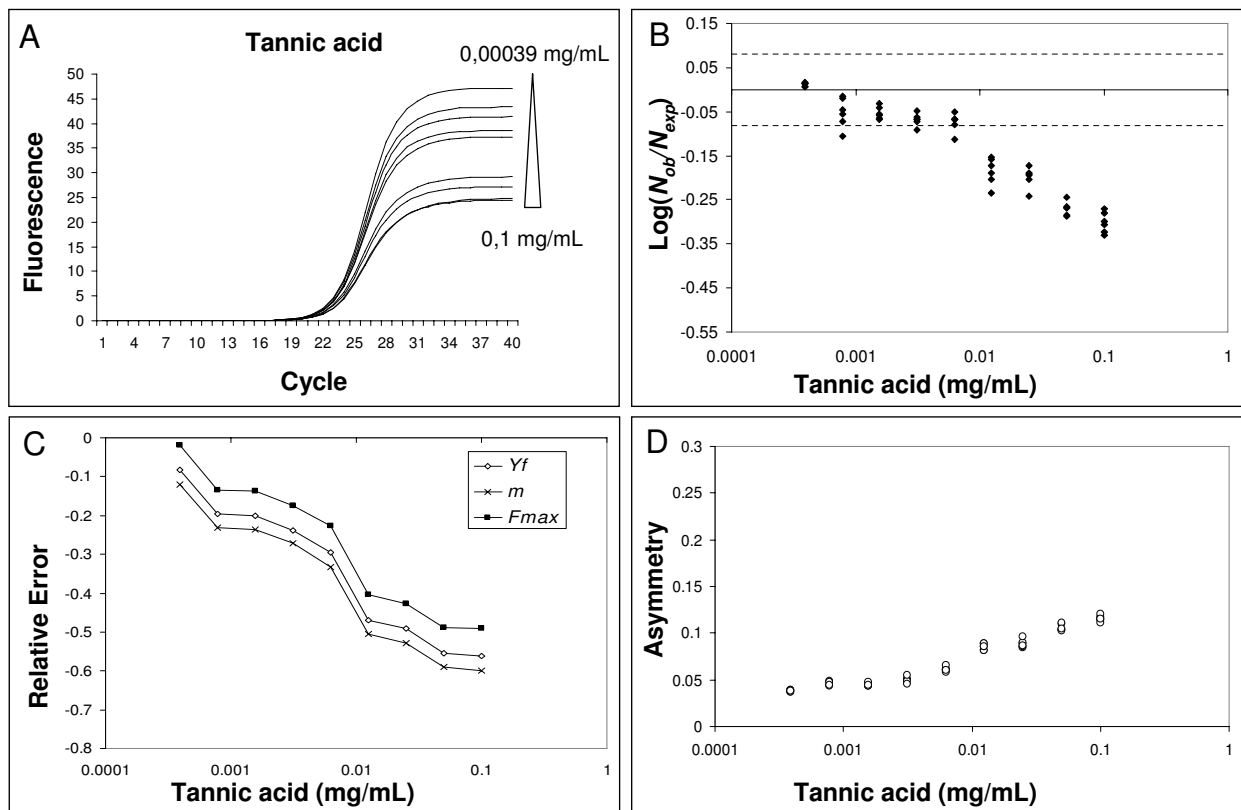
Figure 3 Effect of tannic acid inhibition on amplification curve shape. Left upper panel: amplification profiles obtained from samples with equal starting number of template molecules and increasing inhibitor concentrations. For each inhibitor concentration only an amplification curve was plotted (instead of all 6 replicates). Values over and under triangle indicator (at the upper right of the Fig.) show the lowest and the highest inhibitor concentration used (A). Right upper panel: effect of PCR inhibition on the ratio $Log(N_{ob}/N_{exp})$ in the presence of equal starting number of template molecules and increasing inhibitor concentration. The ratio $Log(N_{ob}/N_{exp})$ represents the residues obtained from linear regression of calibration curves where $LogN_{ob}$ is the number of calculated molecules using Ct method and N_{exp} is the number of expected molecules. Each symbol represents a single run. The abscisse axis is the mean and the dotted lines are the 95% confidence interval of the $Log(N_{ob}/N_{exp})$ ratio calculated from standard curve runs (B). Left lower panel: variation of F_{max}, Y_f and m versus increasing inhibitor concentration. The variation is expressed as Relative Error =
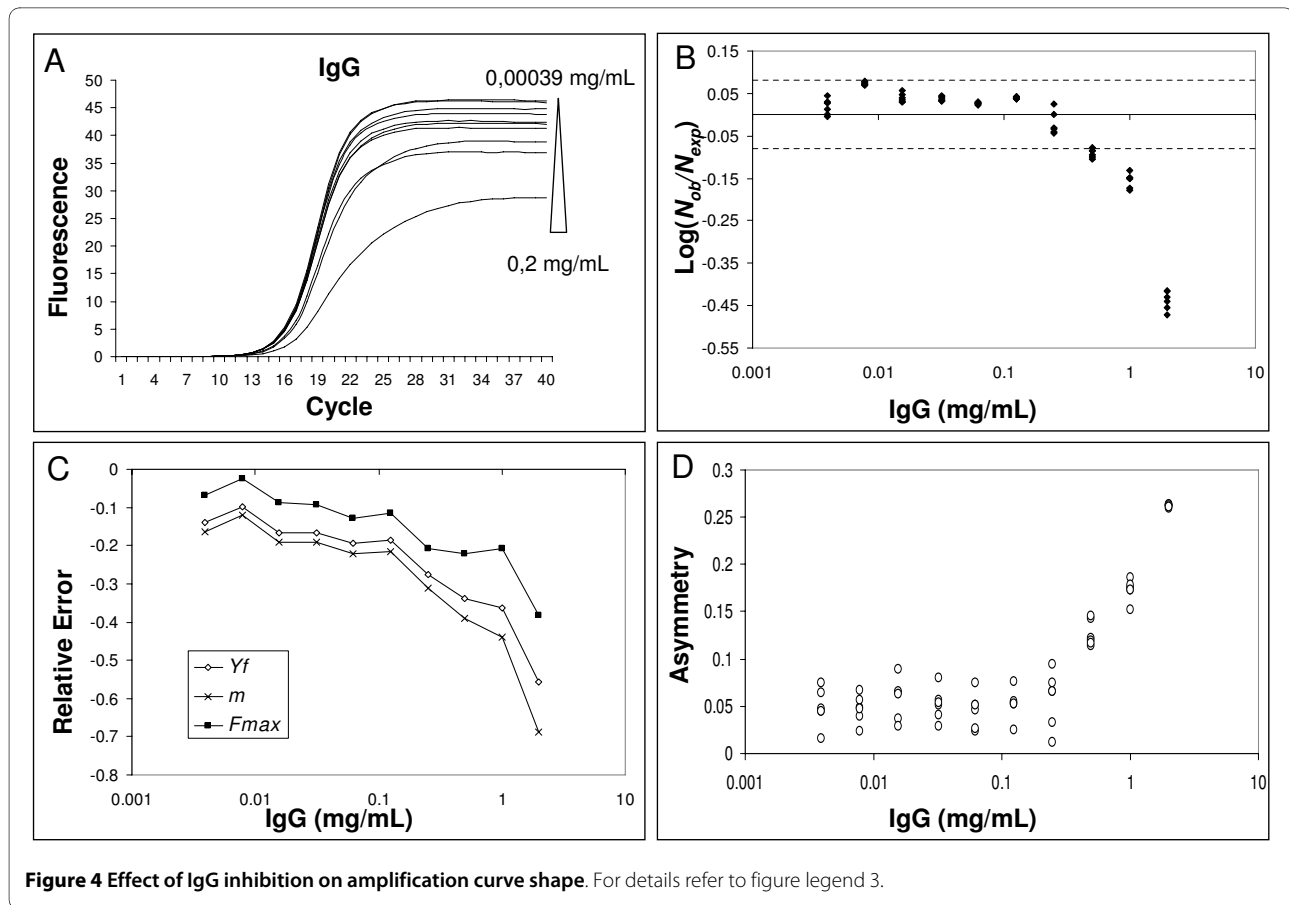
$1 - \dfrac{\overline{x}Parameter_{obs}}{\mu Parameter_{exp}}$; where $\overline{x}_{parameter_{obs}}$ is the mean of parameter calculated for each inhibitor concentration; $\mu_{parameter_{exp}}$ represents the

mean of parameter value from standard curve samples (C). Right lower panel: asymmetry values versus increasing inhibitor concentration. Asymmetry

was computed as the following ratio: $\mathbf{Asym} = 1 - \dfrac{2Y_f}{F_{max}}$ (D).

resulting amplification curves and the corresponding quantifications demonstrate the effects of inhibition on real-time analysis (Fig. 3A and 3B). As the tannic acid concentration increased, the Ct values went up steadily leading to an underestimation of the starting molecules. This quantification error was highlighted when $Log(N_{ob}/N_{exp})$ dropped out the corresponding CI (Fig. 3B). Suppressed amplification was demonstrated by the calculations of efficiency using LinRegPCR procedure (Additional file 5). The observed errors were the result of the progressive reduction of the plateau, linear phase

length and slope of the inhibited curves; together these effects led to increasing Ct values (Fig. 3A) [19,32].

These data led us to investigate the modifications of the parameters m, y_f and F_{max} in response to increasing inhibitor concentrations. Fig. 3C shows the increase in relative error of m, y_f and F_{max} in the presence of increasing tannic acid concentrations. Notably, these results also showed that curve asymmetry (Eq. 5) increased with higher inhibitor concentrations. This in turn demonstrates that not only the slope (m) and plateau (F_{max}) of the curve decreased but also the shape changed moving towards a more and more Richards' type kinetic (Fig. 3D).

Figure 4 Effect of IgG inhibition on amplification curve shape. For details refer to figure legend 3.

Subsequently, we evaluated the effects of IgG and quercitin, molecules known to inhibit PCR, on amplification kinetics [11,32,33]. Both these molecules result in a significant underestimation of starting DNA molecules at high inhibitor concentrations (Fig. 4B and 5B). As shown in Fig. 4 and 5, we always found a change in parameters m, y_f and F_{max} when the quantification error occurred.

Furthermore, the asymmetry analysis showed an interesting singularity in the quercitin effects compared to those of tannic acid and IgG. In fact, quercitin led to kinetic alterations without a significant effect on the curve symmetry (Fig. 5D).
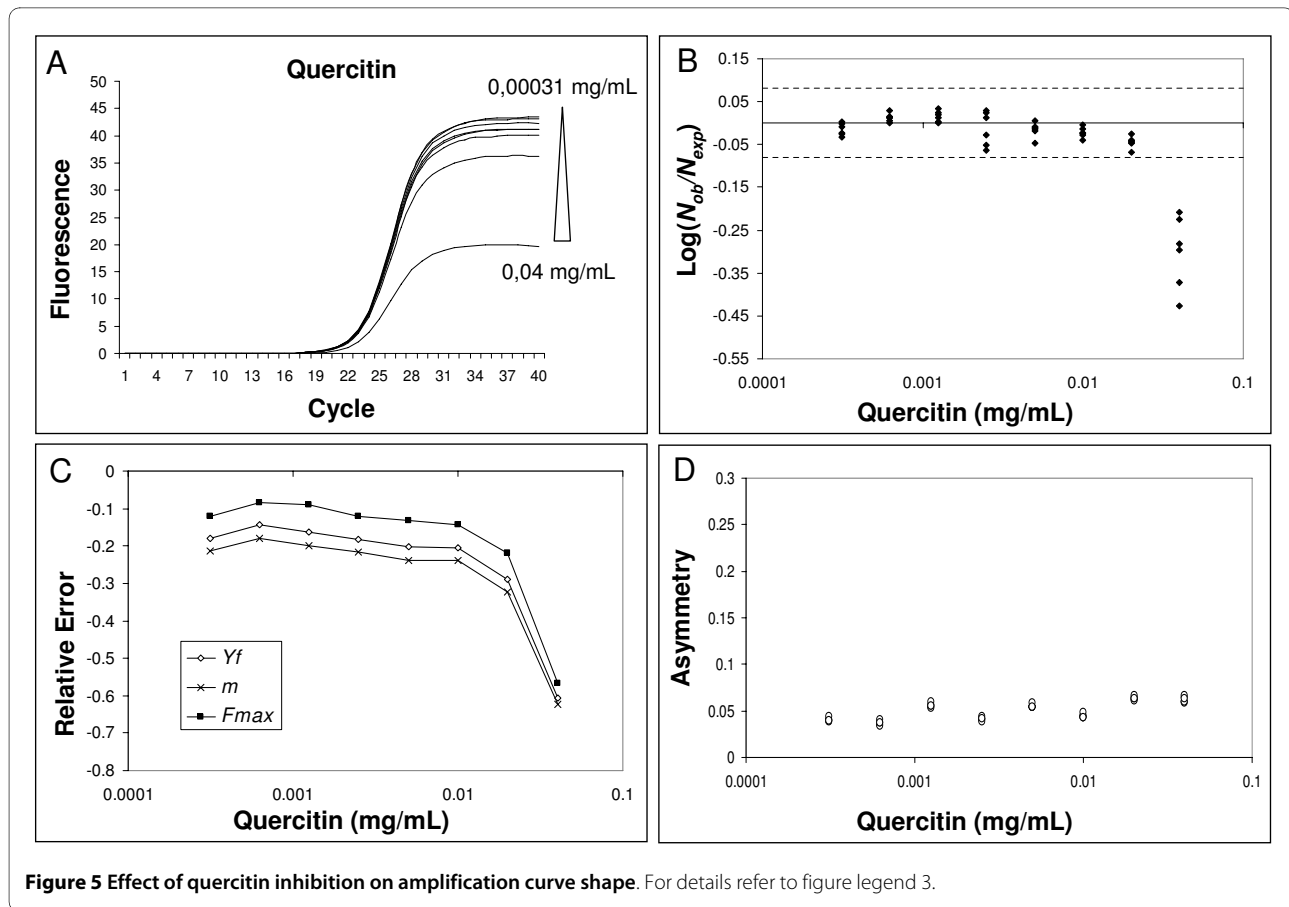
SOD versus KOD analysis
SOD and KOD analyses were used to identify samples with aberrant PCR kinetics, due to inhibitor presence, which might lead to erroneous quantifications. F_{max}, m and y_f values calculated from each amplification curve, obtained in the presence of increasing tannic acid, IgG or quercitin concentrations, were used to estimate the χ^2_{SOD} value. Hence if the χ^2_{SOD} value from an amplification curve was higher than the threshold value 7.81, the quantification was defined as an outlier. PCR efficiencies were also estimated and χ^2_{KOD} values determined from the

same amplifications. Quantification curves with a χ^2_{KOD} values over 3.84 were rejected.

Hence the SOD and KOD performances were evaluated according to their ability to identify an amplification as an outlier when the $Log(N_{ob}/N_{exp})$ ratio is not within 95% CI. The results obtained by SOD and KOD analyses in the presence of increasing tannic acid concentrations are shown in Fig. 6A and 6B. When tannic acid concentrations ranging from 0.1-0.0125 mg/mL were added, all the obtained curves had significant quantification errors (Fig. 6A and 6B; full symbols indicate samples that showed the ratio $Log(N_{ob}/N_{exp})$ below the lower limit of 95% CI). These curves were associated with χ^2_{SOD} values higher than the threshold value of 7.81 (Fig. 6B; the horizontal line shows χ^2_{SOD} threshold value). In this concentration range, KOD analysis appeared to be less powerful than SOD. In fact, KOD found as outliers ($\chi_{KOD}^2 > 3.84$) only 8 of the 24 curves showing a $Log(N_{ob}/N_{exp})$ ratio out of 95% CI (Fig. 6A). There were no outliers under 0.00625 mg/mL tannic acid concentration, with the exception of some amplifications that were randomly out of the CI.

SOD and KOD analyses were also applied to real-time quantifications in the presence of IgG or quercitin as inhibitors. When amplification reactions were conducted

Figure 5 Effect of quercitin inhibition on amplification curve shape. For details refer to figure legend 3.

in the presence of 2-0.5 mg/mL IgG, the suppression of amplification was efficiently revealed by both SOD and KOD, though SOD was more sensitive than KOD. In fact, SOD highlighted 17 outliers versus 15 revealed by KOD out of a total of 17 outliers (in the presence of IgG 17 runs led to a $\text{Log}(N_{ob}/N_{exp})$ out of 95% CI) (Fig. 6C and 6D). Analogous results were also obtained for quercitin. In the presence of 0.04 mg/mL of quercitin, SOD found 6 outliers compared to the 3 revealed by KOD out of a total of 6 outliers (Fig. 6E and 6F; for details of SOD and KOD analysis see Additional file 5).

Finally, we defined as true positives (*TP*) those amplifications showing $\chi^2 >$ threshold value and those that led to a $\text{Log}(N_{ob}/N_{exp})$ ratio out of the 95% CI. Conversely, false positives (*FP*) were defined as samples that showed the $\chi^2 >$ threshold value and a $\text{Log}(N_{ob}/N_{exp})$ ratio within the 95% CI. Consequently, true negatives (*TP*) were those amplifications showing $\chi^2 <$ threshold value that led to a $\text{Log}(N_{ob}/N_{exp})$ ratio within the 95% CI and false negatives (*FN*) those showing $\chi^2 <$ threshold value and $\text{Log}(N_{ob}/N_{exp})$ ratio out of the 95% CI.

Based on these definitions, the *'sensitivity'* of SOD and KOD is represented by the ratio $Sens = \frac{TP}{TP+FN}$ while the

'specificity' is the ratio: $Spec = \frac{TN}{TN+FP}$. Table 2 shows that SOD was more sensitive than KOD in all the tested settings, while SOD and KOD were equally specific in the presence of IgG and quercitin. SOD was also more specific than KOD in the presence of tannic acid.

Discussion

A topic of great interest is the development of hand-free tools for the detection of aberrant amplification profiles in real-time PCR analysis. Real-time PCR has rapidly become the most widely used technique in nucleic acid quantification. Although real-time PCR analysis has gained considerable attention in many fields of molecular biology, it is still troubled by significant technical problems [34]. Hence the present study has focused on the investigation of a new outlier detection approach which is not based on the PCR efficiency estimate but rather on the shape of the amplification profile.

The amplification nature of PCR makes it vulnerable to small differences in efficiencies of compared samples [20]. In fact, the current "gold standard" in real-time PCR analysis, the threshold cycle method (called *Ct* method),
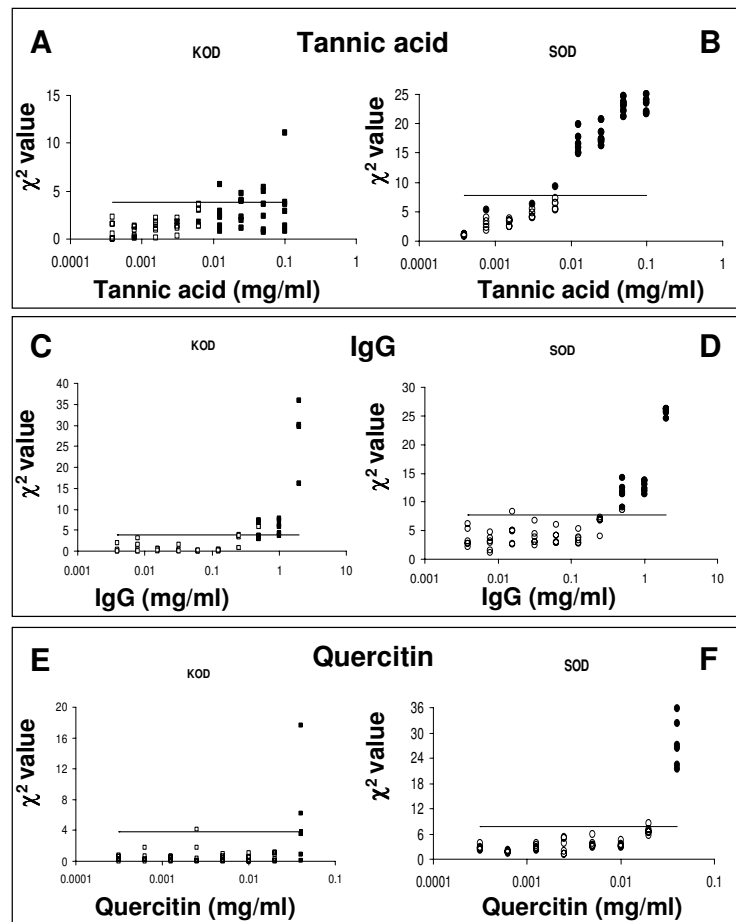
Figure 6 Values of KOD and SOD related of each amplification curve versus *Log* of inhibitor concentration. Symbols (squares and dots) represent the χ^2 values related to each amplification curve obtained in the presence of different inhibitor concentrations. The horizontal continuous lines are the critical values for detecting outliers (left panels: the KOD χ^2 critical value is 3.84; right panels: the SOD χ^2 critical value is 7.81; with $a = 0.05$). Different inhibitors were used: Tannic acid (A-B), IgG (C-D) and Quercitin (E-F). True outliers (represented by black symbols; squares for KOD and dots for SOD) are amplification curves with $Log(N_{ob}/N_{exp})$ ratio out of 95% confidence interval, while white symbols represent acceptable runs with $Log(N_{ob}/N_{exp})$ ratio included in 95% confidence interval. The 95% confidence interval has been obtained from the amplification curves of the standard samples. The black symbols, over the horizontal continuous line, are runs correctly detected as outliers. Conversely, black symbols under this line are undetected outliers.

requires similar PCR efficiencies among compared samples.

However, dissimilarity in PCR efficiency results from different starting material sources, for example, different types of tissues [9]. Such differences might also be found when inhibitors of *Taq* DNA polymerase are present in cDNA samples [35] or in the presence of low quality SYBR green and/or dNTPs [36,37]. Furthermore, the frequency of PCR inhibition [38] and different inhibitory effects even among replicates [39] highlight the need of kinetic quality assessment for each sample. Hence Bar et al. [18] proposed a statistical method, called KOD, to detect samples with dissimilar efficiencies.

KOD searches for outliers based on the main assumption that to obtain a reliable quantification, PCR runs

have to show efficiencies which are not significantly different from each other. This condition is verified comparing the slopes of the straight-line regression calculated in the window-of-linearity after the log-transformation of each read-out fluorescence. In other words, if we return to raw data, the profile of the exponential curves in the window-of-linearity, mustn't be significantly different among compared runs. In the development of the SOD method we extended this concept to the whole curve, and all the runs included in the analysis have to show comparable amplification profiles.

The *Ct* method is based on the analysis of a serially diluted target. An example of this approach is presented in Fig. 1A careful examination of the obtained amplification profiles illustrates the central principle of the SOD

Table 2: Sensitivity and specificity of KOD and SOD analysis.

KOD	Tannic Acid	IgG	Quercitin
Sensitivity	0.30***	0.76***	0.50***
Specificity	0.94	0.96	0.98
SOD	**Tannic Acid**	**IgG**	**Quercitin**
Sensitivity	0.93***	1.00***	1.00***
Specificity	1.00	0.94	0.98

The comparison between PCR amplification profiles in the presence of different inhibitors using KOD and SOD analyses showed that SOD is significantly more sensitive than KOD in all the tested conditions (*** p < 0.001; Z test for difference of proportion), whereas KOD and SOD specificity is similar in the same conditions (p > 0.05; Z test for difference of proportion).

method: all amplification curves are similar in shape and only the profile position is related to target quantity. The first amplification profiles, corresponding to the most concentrated samples, are found on the left, whereas samples with an increasing dilution factor regularly shift towards the right. This observation led us to the insight that an exclusion criterion could be based on the difference in shape rather than efficiency. This is in agreement with the work by Rutledge and Stewart [40] in which these authors described the amplification curve as a function of efficiency. Hence if efficiency determines the shape of a curve, by monitoring the shape of an amplification profile, information concerning the efficiency of amplification can be obtained.

Firstly, a "fingerprint" for each amplification curve using m, y_f and F_{max} resulting from the fitting of the Richards equation on raw data was obtained. Subsequently, these parameters were used to obtain the variance-covariance matrix in order to calculate the Mahalanobis distance [30]. This statistical measure is based on correlations among variables through which different patterns can be identified and analysed. In particular, the SOD analysis made use of the Mahalanobis distance to determine the similarity of an unknown sample compared to the standard set. This approach was very useful because it allowed us to evaluate not only the variance of single parameters (m, y_f and F_{max}), but also to quantify the reciprocal co-variations among m, y_f and F_{max}.

F_{max} was considered in the development of SOD because this parameter demonstrates successful amplification and usually, in suboptimal amplification conditions, the read-outs do not reach characteristic F_{max} values [9]. Examining our database, it was noted that F_{max} showed high variance, thus it slightly affects χ^2_{SOD} alone, but F_{max} had a significant impact on the variance-covariance matrix. The parameter m describes the slope of the curve in the inflection point [11]. In our model, the higher the value of m, the higher the amplification rate is.

However, this estimator does not directly indicate the amplification efficiency understood as the proportion between current and previous product amounts [38]. Finally, the asymmetry of amplification profiles was monitored by the relationship between F_{max} and y_f. It has been demonstrated that absolutely symmetrical PCR curves seldom occur, justifying the introduction of a five-parameter fit [25]. Furthermore, in our previous work [11], it was demonstrated that the amplification reaction may deviate from a symmetric sigmoid curve to an asymmetric sigmoid (well described by Richards equation) in the presence of suboptimal efficiency. In fact, the goodness of fit of the logistic model progressively decreased with lower efficiency suggesting a change of PCR curve amplification shape [32].

The correlation analysis between m, y_f and F_{max} obtained from the standard curve and input DNA demonstrated that these shape parameters are concentration-independent. This supports our experimental hypothesis that all the amplification curves of the standard curve are similar in shape and only the profile position determines target quantity. In the presence of PCR inhibition, it was found that increasing concentrations of tannic acid and IgG resulted in decreasing F_{max} and m values, while asymmetry increased with higher inhibitor concentrations (when asymmetry increases, y_f decreases more than the corresponding F_{max}; Fig. 3 and 4). It may be that tannic acid inhibition is simply due to fluorescence quenching since we found a dramatic decrease in F_{max} and a slide curve slope decrease. However, we also showed that fluorescence asymmetry increased demonstrating that tannic acid produced an amplification kinetic distortion. The addition of quercitin to PCR amplifications produced very interesting data. In fact, we found decreased F_{max} and m values in the presence of high inhibitor concentrations, however this flavonid did not induce an asymmetric modification of the curves (Fig. 5D). The reported data clearly demonstrate that the SOD method can iden-

tify non-optimal PCR kinetics resulting from different inhibition models. Furthermore, the results obtained in the presence of quercitin highlight the importance of using a multivariate approach.

When comparing SOD to KOD performance, it was found that SOD was more sensitive than KOD in all the tested settings. SOD and KOD were equally specific in the presence of IgG and quercitin, whereas SOD was more specific than KOD in the presence of tannic acid.

Furthermore, the SOD method presents several advantages over KOD; SOD is completely hand-free. Indeed, it is not necessary for the user to identify a window of analysis as in the KOD method, and more importantly, SOD does not rely on a constant efficiency value avoiding all the problems connected with its determination [28,40,41]. As previously reported, variable PCR efficiency determination can lead to different results contributing to erroneous and spread quantifications [19]. Moreover, log-transformation of fluorescence data that could be responsible for bias in the analysis are avoided.

The SOD method has been developed for the chemistry Sybr Green, and the application of this procedure to other chemistries such as TaqMan, needs to be evaluated extensively.

Very recently, Tichopad et al. [29] proposed a new KOD procedure based on Malahanobis statistic [30]. In this study the first derivative maximum and the second derivative maximum were estimated using a logistic fitting on the central portion of the PCR trajectory. Using these two parameters these authors proposed monitoring only the first half of the curve. On the contrary, the SOD method is based on the possibility of describing the whole PCR trajectory using Richards equation. SOD represents a continuation and an extension of the application of Richards equation to real-time PCR readings [11]. We think that the SOD method introduces original concepts that are not found in the recently developed method described by Tichopad et al. [29]. SOD takes advantage of the possibility of describing the shape of the whole PCR trajectory through the combination of the parameters m, y_f and F_{max} while the method by Tichopad et al. [29] focuses on two key points of the trajectory: the maximum of the first and second derivative. Furthermore, in the SOD method we used quite a different metric approach. Although other multivariate methods are available for similar tasks (support vector machines, K-means cluster), we used asymptotic distribution of the Mahalanobis distance because it is a logical extension of the KOD method, which is based on univariate normal distribution.

Conclusion

We demonstrated for the first time that a comparison of the shape variation of an amplification profile with the shape of standard profiles can be used to exclude aberrant samples from *Ct* analysis. This allows us to avoid the spread of results and therefore increases the potential of quantification analysis.

Hence we propose SOD as a hand-free quality control method in real-time PCR analysis with applications in any field of molecular diagnostics.

Additional material

Additional file 1 Fluorescence data and fitting elaboration of standard sample amplifications (standard curve) and amplifications obtained in the presence of: tannic acid, IgG and quercitin.

Additional file 2 Analytical solutions for the y value of the inflection point *(Y$_F$)* and the slope of tangent straight-line *(m)* crossing the inflection point.

Additional file 3 A) Chi-square distribution of the squared distances about the population mean vector (D2 = *(y-µ)'Σ^{-1}(y-µ)*) with 3 degrees of freedom. B) Scatter plots of all pairs of variables F_{max}, Y_f and m.

Additional file 4 P-P plot of the variable *Log(N$_{ob}$/N$_{exp}$)*.

Additional file 5 KOD and SOD elaborations of standard sample amplifications (standard curve) and amplifications obtained in the presence of: tannic acid, IgG and quercitin.

Abbreviations
Ct: threshold cycle; IgG: immunoglobulin G; SOD: shape based kinetic outlier detection; KOD: kinetic outlier detection; Asym: Asymmetry.

Authors' contributions
MG and DS carried out the design of the study, participated in data analysis, developed the *SOD* method and drafted the manuscript. MBLR participated in data collection and analysis and critically revised the manuscript. PT carried out the real-time PCR. DM participated in data collection. VS participated in the design of the study and critically revised the manuscript. All authors read and approved the final manuscript.

Author Details
^{1}Dipartimento DiSUAN, Sezione di Biomatematica, Università degli Studi di Urbino "Carlo Bo", Campus Scientifico Sogesta; Località Crocicchia - 61029 Urbino, Italy and ^{2}Dipartimento di Scienze Biomolecolari, Sezione di Ricerca sull'Attività Motoria e della Salute, Università degli Studi di Urbino "Carlo Bo", Via I Maggetti, 26/2 - 61029 Urbino, Italy

References
1. Gingeras TR, Higuchi R, Kricka LJ, Lo YM, Wittwer CT: **Fifty years of molecular (DNA/RNA) diagnostics.** *Clin Chem* 2005, 51(3):661-671.
2. Nolan T, Hands RE, Bustin SA: **Quantification of mRNA using real-time RT-PCR.** *Nature Protocols* 2006, 1(3):1559-1582.
3. VanGuilder HD, Vrana KE, Freeman WM: **Twenty-five years of quantitative PCR for gene expression analysis.** *Bio Techniques* 2008, 44(5):619-626.
4. Gunson RN, Bennett S, Maclean A, Carman WF: **Using multiplex real time PCR in order to streamline a routine diagnostic service.** *J Clin Virol* 2008, 43(4):372-375.
5. Watzinger F, Ebner K, Lion T: **Detection and monitoring of virus infections by real-time PCR.** *Molecular aspects of medicine* 2006, 27(2-3):254-298.
6. Kaltenboeck B, Wang C: **Advances in real-time PCR: application to clinical laboratory diagnostics.** *Advances in clinical chemistry* 2005, 40:219-259.
7. Akane A, Matsubara K, Nakamura H, Takahashi S, Kimura K: **Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification.** *Journal of forensic sciences* 1994, 39(2):362-372.

8. Wilson IG: **Inhibition and facilitation of nucleic acid amplification.** *Applied and environmental microbiology* 1997, **63**(10):3741-3751.
9. Tichopad A, Didier A, Pfaffl MW: **Inhibition of real-time RT-PCR quantification due to tissue-specific contaminants.** *Mol Cell Probes* 2004, **18**(1):45-50.
10. Rossen L, Norskov P, Holmstrom K, Rasmussen OF: **Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solutions.** *International journal of food microbiology* 1992, **17**(1):37-45.
11. Guescini M, Sisti D, Rocchi MB, Stocchi L, Stocchi V: **A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition.** *BMC bioinformatics* 2008, **9**:326.
12. Kainz P: **The PCR plateau phase - towards an understanding of its limitations.** *Biochimica et biophysica acta* 2000, **1494**(1-2):23-27.
13. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods (San Diego, Calif)* 2001, **25**(4):402-408.
14. Liu W, Saint DA: **Validation of a quantitative method for real time PCR kinetics.** *Biochem Biophys Res Commun* 2002, **294**(2):347-353.
15. Rutledge RG: **Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications.** *Nucleic acids research* 2004, **32**(22):e178.
16. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic acids research* 2001, **29**(9):e45.
17. Goll R, Olsen T, Cui G, Florholmen J: **Evaluation of absolute quantitation by nonlinear regression in probe-based real-time PCR.** *BMC bioinformatics* 2006, **7**:107.
18. Bar T, Stahlberg A, Muszta A, Kubista M: **Kinetic Outlier Detection (KOD) in real-time PCR.** *Nucleic acids research* 2003, **31**(17):e105.
19. Kontanis EJ, Reed FA: **Evaluation of real-time PCR amplification efficiencies to detect PCR inhibitors.** *Journal of forensic sciences* 2006, **51**(4):795-804.
20. Ramakers C, Ruijter JM, Deprez RH, Moorman AF: **Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data.** *Neurosci Lett* 2003, **339**(1):62-66.
21. Wilhelm J, Pingoud A, Hahn M: **Validation of an algorithm for automatic quantification of nucleic acid copy numbers by real-time polymerase chain reaction.** *Anal Biochem* 2003, **317**(2):218-225.
22. Wilhelm J, Pingoud A, Hahn M: **SoFAR: software for fully automatic evaluation of real-time PCR data.** *Bio Techniques* 2003, **34**(2):324-332.
23. Ruijter JM, Ramakers C, Hoogaars WM, Karlen Y, Bakker O, Hoff MJ van den, Moorman AF: **Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data.** *Nucleic acids research* 2009, **37**(6):e45.
24. Liu W, Saint DA: **A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics.** *Anal Biochem* 2002, **302**(1):52-59.
25. Spiess AN, Feig C, Ritz C: **Highly accurate sigmoidal fitting of real-time PCR data by introducing a parameter for asymmetry.** *BMC bioinformatics* 2008, **9**:221.
26. Qiu H, Durand K, Rabinovitch-Chable H, Rigaud M, Gazaille V, Clavere P, Sturtz FG: **Gene expression of HIF-1alpha and XRCC4 measured in human samples by real-time RT-PCR using the sigmoidal curve-fitting method.** *Bio Techniques* 2007, **42**(3):355-362.
27. Rutledge RG, Stewart D: **A kinetic-based sigmoidal model for the polymerase chain reaction and its application to high-capacity absolute quantitative real-time PCR.** *BMC biotechnology* 2008, **8**:47.
28. Cikos S, Bukovska A, Koppel J: **Relative quantification of mRNA: comparison of methods currently used for real-time PCR data analysis.** *BMC molecular biology* 2007, **8**:113.
29. Tichopad A, Bar T, Pecen L, Kitchen RR, Kubista M, Pfaffl MW: **Quality control for quantitative PCR based on amplification compatibility test.** *Methods* 2010, **50**(4):308-312.
30. Rencher AC: **Methods of Multivariate Analysis.** 2nd edition. Wiley, Printed in US; 2002.
31. Young CC, Burghoff RL, Keim LG, Minak-Bernero V, Lute JR, Hinton SM: **Polyvinylpyrrolidone-Agarose Gel Electrophoresis Purification of Polymerase Chain Reaction-Amplifiable DNA from Soils.** *Applied and environmental microbiology* 1993, **59**(6):1972-1974.
32. Tichopad A, Polster J, Pecen L, Pfaffl MW: **Model of inhibition of Thermus aquaticus polymerase and Moloney murine leukemia virus reverse transcriptase by tea polyphenols (+)-catechin and (-)-epigallocatechin-3-gallate.** *J Ethnopharmacol* 2005, **99**(2):221-227.
33. Nolan T, Hands RE, Ogunkolade W, Bustin SA: **SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations.** *Anal Biochem* 2006, **351**(2):308-310.
34. Murphy J, Bustin SA: **Reliability of real-time reverse-transcription PCR in clinical diagnostics: gold standard or substandard?** *Expert review of molecular diagnostics* 2009, **9**(2):187-197.
35. Chandler DP, Wagnon CA, Bolton H Jr: **Reverse transcriptase (RT) inhibition of PCR at low concentrations of template and its implications for quantitative RT-PCR.** *Applied and environmental microbiology* 1998, **64**(2):669-677.
36. Kubista M, Stahlberg A, Bar T: **Light-up probe based real-time Q-PCR.** *Genomics and Proteomics Technologies Proceedings of SPIE* 2001:53-58.
37. Karsai A, Muller S, Platz S, Hauser MT: **Evaluation of a homemade SYBR green I reaction mixture for real-time PCR quantification of gene expression.** *Bio Techniques* 2002, **32**(4):790-792. 794-796
38. Tichopad A, Dzidic A, Pfaffl MW: **Improving quantitative real-time RT-PCR reproducibility by boosting primer-linked amplification efficiency.** *Biotechnology Letters* 2002, **24**:2053-2056.
39. Rosenstraus M, Wang Z, Chang SY, DeBonville D, Spadoro JP: **An internal control for routine diagnostic PCR: design, properties, and effect on clinical performance.** *Journal of clinical microbiology* 1998, **36**(1):191-197.
40. Rutledge RG, Stewart D: **Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time PCR.** *BMC molecular biology* 2008, **9**:96.
41. Skern R, Frost P, Nilsen F: **Relative transcript quantification by quantitative PCR: roughly right or precisely wrong?** *BMC molecular biology* 2005, **6**(1):10.