Computers in Biology and Medicine 42 (2012) 438–445

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

BootstRatio: A web-based statistical analysis of fold-change in qPCR and RT-qPCR data using resampling methods

Ramon Clèries [a,b,*], Jordi Galvez [a], Meritxell Espino [c], Josepa Ribes [a,b], Virginia Nunes [c,d,e], Miguel López de Heredia [c,d]

[a] Catalan Plan for Oncology–Institut Català d'Oncologia—IDIBELL, Hospital Duran i Reynals, L'Hospitalet de Llobregat, Barcelona 08907, Spain
[b] Department of Clinical Sciences, Universitat de Barcelona, Barcelona, Spain
[c] LGM-IDIBELL, Hospital Duran i Reynals, L'Hospitalet de Llobregat, Barcelona, Spain
[d] Centro de Investigación en Red de Enfermedades Raras (CIBERER), L'Hospitalet de Llobregat, Barcelona, Spain
[e] Secció de Genètica, Departament de Ciències Fisiològiques II, Facultat de Medicina, Universitat de Barcelona, Barcelona, Spain

ARTICLE INFO

Article history:
Received 12 January 2011
Accepted 19 December 2011

Keywords:
Real-time PCR
Bootstrap
Permutation tests
Gene expression ratios
Simulation

ABSTRACT

Real-time quantitative polymerase chain reaction (qPCR) is widely used in biomedical sciences quantifying its results through the relative expression (RE) of a target gene versus a reference one. Obtaining significance levels for RE assuming an underlying probability distribution of the data may be difficult to assess. We have developed the web-based application BootstRatio, which tackles the statistical significance of the RE and the probability that $RE > 1$ through resampling methods without any assumption on the underlying probability distribution for the data analyzed. BootstRatio perform these statistical analyses of gene expression ratios in two settings: (1) when data have been already normalized against a control sample and (2) when the data control samples are provided. Since the estimation of the probability that $RE > 1$ is an important feature for this type of analysis, as it is used to assign statistical significance and it can be also computed under the Bayesian framework, a simulation study has been carried out comparing the performance of BootstRatio versus a Bayesian approach in the estimation of that probability. In addition, two analyses, one for each setting, carried out with data from real experiments are presented showing the performance of BootstRatio. Our simulation study suggests that Bootstratio approach performs better than the Bayesian one excepting in certain situations of very small sample size ($N \leq 12$). The web application BootstRatio is accessible through http://regstattools.net/br and developed for the purpose of these intensive computation statistical analyses.

1. Introduction

Real-time quantitative polymerase chain reaction (qPCR) is widely used in research and diagnostics as a method to reliably quantify nucleic acid amount due to its robustness, easy procedures, reproducibility and the lower sample amount needed in comparison with other methods. When used in combination with retrotranscription (RT-qPCR) it allows determining gene expression. Quantification results are based on either the relative expression (RE) of a target gene versus a reference one or an absolute quantification based on internal or external calibration curves [1]. RE is widely used by researchers as it avoids the complications of generating calibrating material and it is measured as the ratio between the mean target gene expression and that of the reference one. MIQE guidelines for publication of RT-qPCR data suggest that data analysis procedures and statistical methods to assign significance to the data should be indicated when publishing [2]. In the literature, few statistical methods have been developed for the statistical data analysis of RT-qPCR [3–8]. Obtaining significance levels for the RE through statistical modeling entail assuming an underlying probability distribution of the data that may be difficult to assess, specially when data is based on small sample size ($n < 20$) on both target and reference samples. In these situations, resampling methods may be used to assess percentiles, and therefore, statistically significance of a statistical estimator such as the mean, median or standard error of the data [9]. Successful biological applications of these techniques such as computing confidence intervals [10], clustering [11], robust estimation of statistics [12] and non-linear regression [13] have been previously described. These methods can be also useful to efficiently calculate very low *P*-values from a large number of resampled measurements [14].

In this paper we apply bootstrap and permutation tests to assess the statistical significance of the gene expression ratios of

* Corresponding author at: Catalan Plan for Oncology–Institut Català d'Oncologia—IDIBELL, Hospital Duran i Reynals, L'Hospitalet de Llobregat, Barcelona 08907, Spain. Tel.: +34932607417.
E-mail address: r.cleries@iconcologia.net (R. Clèries).

0010-4825/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2011.12.012

RT-qPCR without any assumption on the underlying probability distribution of our data. The web application BootstRatio has been developed to perform statistical analyses of gene expression ratios either when data has been already normalized against a control sample or when control samples are provided. A simulation study has been carried out comparing the performance of the method presented here versus a Bayesian one, this last requiring certain prior distribution for the data to be assumed. In addition, we also show the results of the analysis of real-data, one for each of the possible data sets indicated above, showing the performance of the method.

2. Statistical methods

The Bootstrap method: Analysis of gene expression ratios with no control sample (samples already normalized against a control sample).

Let G be the gene of interest for which we have a sample of m observed expression ratio values $S_G=\{R_{G1},...,R_{Gm}\}$ assuming $R_{Gi}\geq 0$ $\forall i=1,...,m$. We can estimate the mean ratio as $\widehat{\mu}_{R_G}\overline{X}_{R_G}=(1/m)\sum_{i=1}^{m}R_{Gi}$. Our interest is to assess whether $\mu_{R_G}>1$ and therefore, to assess its statistical significance estimating $P(\mu_{R_G}>1)$ through Bootstrap [9]. The bootstrap is a general technique for assessing uncertainty in estimation procedures using computer simulation through resampling from the original data [9]. Bootstrap is a solution [9] when there is doubt that the usual distributional assumptions and asymptotic results are valid and accurate [15,16].

BootstRatio resamples a dataset for specified number of times, say L times. BootstRatio calculates $\overline{X}_{R_G}$ in each resample of the dataset, therefore it computes $\overline{X}^{(j)}_{R_G}$ for the jth resample where-$j=\{1,...,L\}$. At the end of this iterative process, Bootstratio estimates $P(\mu_{R_G}>1)$through computing the number of times that $\overline{X}_{R_G}>1$ out of L times. Specifically, it computes $\widehat{P}(\mu_{R_G}>1)=\sum_{j=1}^{L}I(\overline{X}^{(j)}_{R_G}>1)/L$ that estimates $P(\mu_{R_G}>1)$, where $I(\overline{X}^{(j)}_{R_G}>1)=1$ when $\overline{X}_{R_G}>1$ and $I(\overline{X}^{(j)}_{R_G}>1)=0$ otherwise.

The permutation test: Analysis of gene expression ratios when control samples are provided in the data set.

Let G be the gene of interest for which we have (at least) two conditions, a treatment one T, to be compared with a control one, C, where measures of each condition have been extracted from two groups of patients. A sample of m values has been obtained from the control group, $X_{C1},...,X_{Cm}$, and another sample of size n has been obtained from the treated one, $X_{T1},...,X_{Tn}$. Let $E[X_{Ci}]=\mu_C$ and $E[X_{Tj}]=\mu_T$ be the theoretical mean expressions for control and treated groups, respectively. Our interest is to estimate $R_{GT}=(\mu_T/\mu_C)$ through $\widehat{R}_{GT}=(\overline{X}_T/\overline{X}_C)$ where $R_{GT}>1$ indicates if the treated sample shows higher expression than the control one. On the contrary, $R_{GT}<1$ indicates that treated sample has lower expression than the control one, whereas $R_{GT}=1$ indicates no differences between the expressions of treated and control

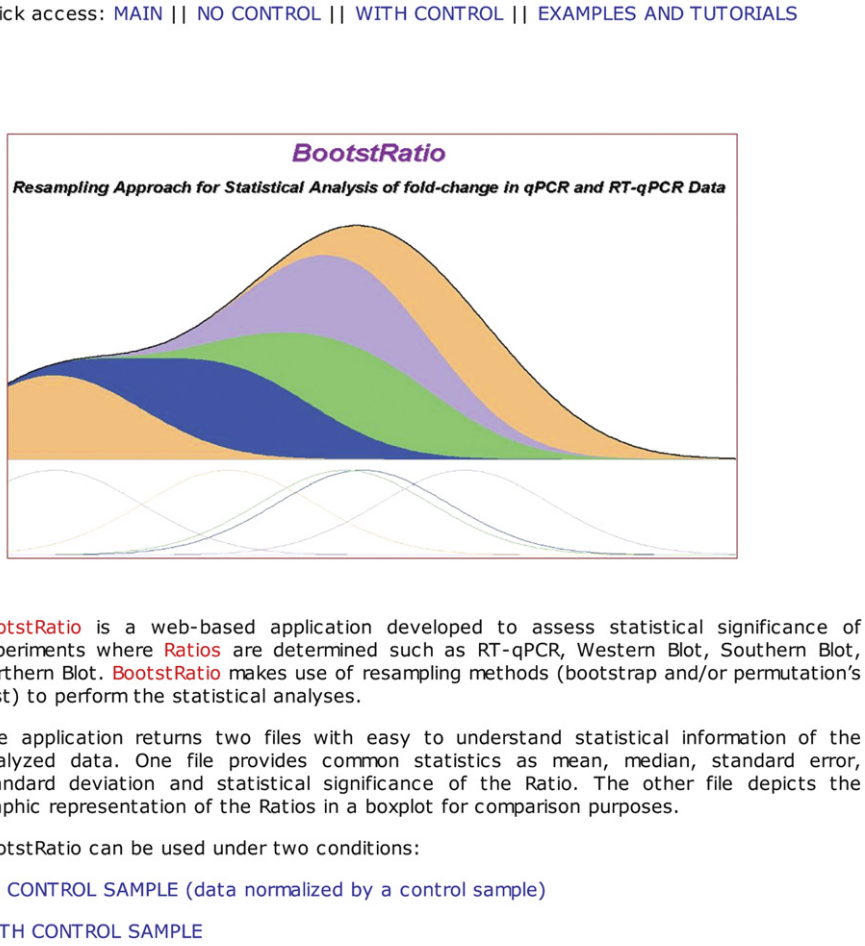


Fig. 1. Snapshot of the main page of BootstRatio web application where the user must select the type of analysis: (I) Unique condition for each gene with no control sample (II) several conditions for each gene using a control sample.

samples. Our interest is to determine whether $R_{GT} \neq 1$ by means of testing whether probabilities $P(R_{GT} > 1) < \alpha$ or $P(R_{GT} < 1) < \alpha$ for a certain significance level α, usually $\alpha \in \{0.1, 0.05, 0.01, 0.001, 0.0005\}$. A permutation's test approach [9,15,16] is used when two conditions or more arise from the data.

The term permutation test refers to rearrangements of the data [9]. Bootstratio generates a number of permutations of the Treatment and Control labels, say L, and it randomly assigns these labels to the observed values. The sampling distribution of the test statistic $\widehat{R}_{GT}$ is computed by forming these L permutations and estimating $P(R_{GT} > 1)$through computing the number of times that $\widehat{R}_{GT} > 1$ out of L. An extended description of these statistical methods can be found in the supplementary material file on journal's website.

2.1. BootstRatio web application

The BootstRatio web application can be found at http://regstattools.net/br. The method described above has been implemented in a website allowing two types of analysis, which requires the user to upload a different ASCII file in each case. In the first type of analysis we have a unique condition for each gene and the user must provide a dataset with expression ratios for each gene. The dataset contains two columns, one column refers to the name of the gene and the other is the gene expression ratio. In the second type of analysis we have several conditions for each gene analysis including a control one; the user must provide a dataset with three columns. The first one refers to the name of the gene, the second refers to the type of condition, which should include a control one as well as one or more treated types, whereas the third column is the expression values.

A snapshot of the web application can be found in Fig. 1. The user must select the type of analysis and also provide an e-mail address (see Fig. 2) where the results of the statistical analysis will be sent. The application also produces two files, web accessible, that comprise the results. The first file includes a table with statistics extracted from the analysis, whereas the second file includes a report in PDF format where the user may find boxplot graphs of the resampling distribution of $\widehat{\mu}_{R_G}$ or $\widehat{R}_{GT}$, depending on the type of analysis and a summary table with main statistics. Tutorials and example datasets can be downloaded from BootstRatio web-site. The web-application has been programmed using R (www.r-project.org), PHP and HTML. R code for the analysis can be obtained from the corresponding author.

2.2. Simulation study: BootstRatio versus Bayesian approach

A simulation study has been carried out in order to assess the performance of the resampling method presented here. We have compared this method with a Bayesian approach under two situations: (A) Performance with three different number of replicates and (B) assuming the presence of random noise. The estimation of the probability that an expression ratio is greater than 1 is a new feature compared to other methodologies [1–8,10–14], and it only can be compared to a Bayesian methodology [17] in a similar line.

The simulation procedure consists in generating a sample of $N=60$, $N=20$ and $N=12$ observations from simulated distributions of R_{GT} and assess the estimation of $P(R_{GT} > 1)$ using BootstRatio as well as the Bayesian method. To compare the performance between methods we have calculated the Relative Error in the $P(R_{GT} > 1)$ (REP) using the Bootstratio method, $REP(\%) = (|P(R_{GT} > 1) - P(R^{Bt} > 1)|/P(R_{GT} > 1))x100$, where $P(R_{GT} > 1)$ is the true probability of R_{GT} greater than one (known value) and $P(R^{Bt} > 1)$ is the relative frequency that R^{Bt}, the Bootstratio estimate of R_{GT}, is greater than one. In the Bayesian method, the REP(%) formula changes $P(R^{Bt} > 1)$ for $P(R^{BY} > 1)$, where $P(R^{BY} > 1)$ is the relative frequency that R^{BY}, the Bayesian estimate of R_{GT}, is greater than one. A detailed description of the simulation method can be found in the supplementary material file on journal's website.

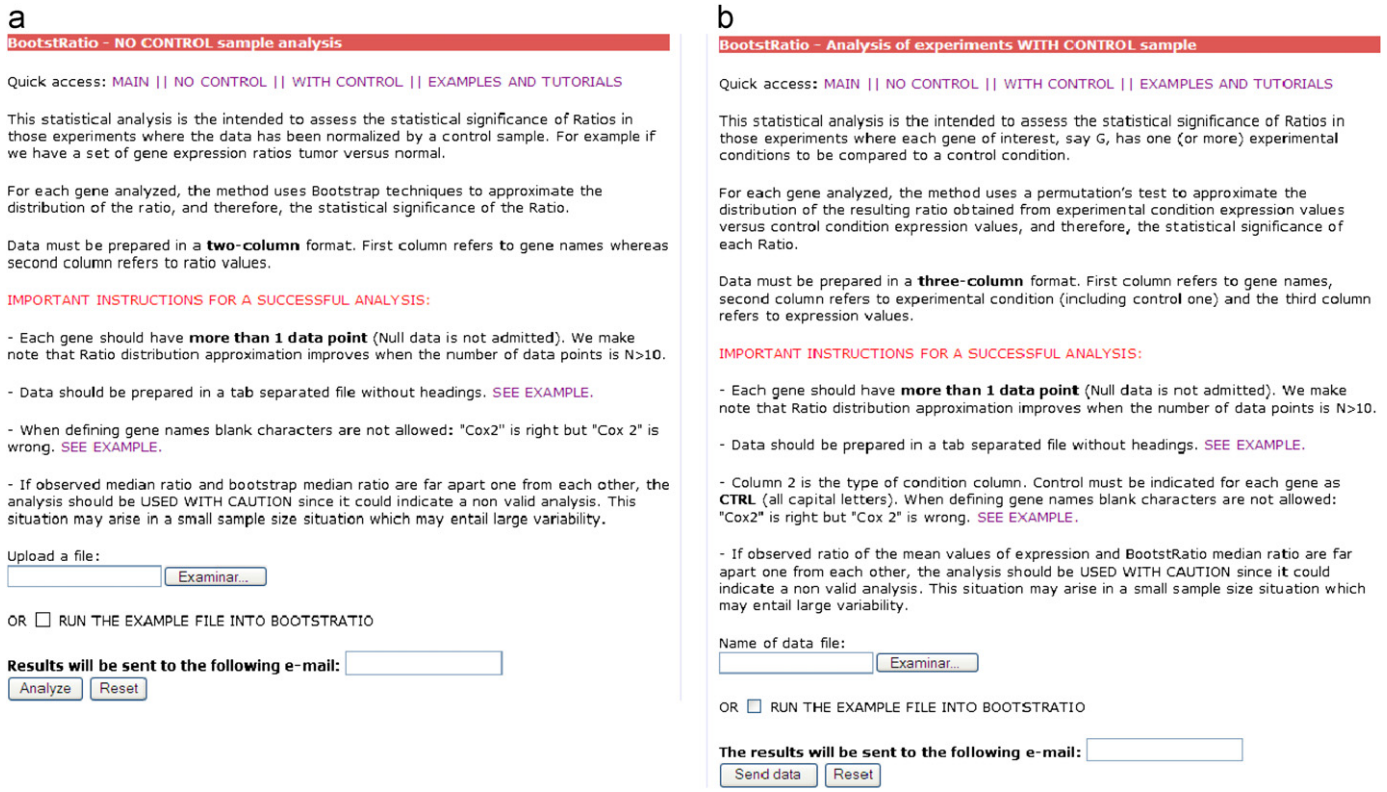


Fig. 2. Snapshots of the BootstRatio web application: (a) No control sample analysis (unique condition for each gene) versus (b) with control sample analysis (several conditions for each gene will be compared to a control sample).

Table 1
Results of the simulation study comparing the performance of the BootstRatio method respect to the Bayesian method. For each sample size analyzed ($N=60$, $N=20$ and $N=12$) we compared the true probability $P(R_{GT}>1)$ with its BootstRatio and Bayesian estimates, $P(R^{Bt}>1)$and $P(R^{BY}>1)$, assessing their relative error in the estimation of the true probability (REP). Cells in italic show the smallest REP when comparing the BootstRatio approach and the Bayesian method in each one of the simulation settings considered.

		Probability distribution of X_T distribution of XT Gamma(α=4500, β=1/5000)				***Probability distribution of XT Gamma(α=4500, β=1/5000) Uniform(1,2)***			
Sample size	***True probability***	***Bootstrap (Bootstratio)method***		***Bayesian (MCMC) method***		***Bootstrap (Bootstratio)method***		***Bayesian (MCMC) method***	
	$P(R_{GT}>1)$	*REP* (%)	$P(R^{Bt}>1)$	REP (%)	$P(R^{BY}>1)$	REP (%)	$P(R^{Bt}>1)$	*REP* (%)	$P(R^{BY}>1)$
$N=$**60**	0.25	9.64	0.27	*3.52*	*0.24*	*3.52*	*0.24*	6.16	0.27
	0.50	3.76	0.48	*3.26*	*0.49*	*4.24*	*0.52*	10.24	0.55
	0.75	*1.16*	*0.74*	2.38	0.73	*2.51*	*0.73*	7.84	0.69
$N=20$	0.25	43.52	0.14	*39.52*	*0.15*	*24.48*	*0.31*	36.48	0.34
	0.50	*37.76*	*0.31*	57.60	0.21	*14.24*	*0.57*	32.24	0.66
	0.75	*7.84*	*0.69*	11.84	0.66	*5.05*	*0.71*	13.49	0.85
$N=12$	0.25	79.52	0.05	*71.52*	*0.07*	32.48	0.33	*16.48*	*0.29*
	0.50	65.76	0.17	*61.76*	*0.19*	48.24	0.74	*46.24*	*0.73*
	0.75	*5.17*	*0.71*	13.19	0.65	*5.05*	*0.71*	9.17	0.68
		Gaussian random noise							
		Probability distribution of X* T (Note: XT= X* T+Random noise) Gamma(α=4500, β=1/5000)				***Probability distribution of X* T (Note: XT= X*T+Random noise) Uniform(1,2)***			
		Bootstrap (Bootstratio)method		***Bayesian (MCMC) method***		***Bootstrap (Bootstratio)method***		***Bayesian (MCMC) method***	
		REP (%)	$P(R^{Bt}>1)$	REP (%)	$P(R^{BY}>1)$	REP (%)	$P(R^{Bt}>1)$	REP (%)	$P(R^{BY}>1)$
$N=$**60**	0.25	20.40	0.30	*16.44*	*0.29*	*3.52*	*0.24*	4.52	0.26
	0.50	*28.24*	*0.64*	34.24	0.67	*4.26*	*0.52*	10.24	0.55
	0.75	2.84	0.77	*0.16*	*0.75*	*2.49*	*0.73*	9.17	0.68
$N=$**20**	0.25	*51.48*	*0.12*	55.52	0.11	*24.56*	*0.31*	36.49	0.34
	0.50	*47.76*	*0.26*	63.74	0.18	*14.24*	*0.57*	32.24	0.66
	0.75	*7.85*	*0.69*	15.81	0.63	*4.79*	*0.71*	13.51	0.85
$N=$**12**	0.25	*71.44*	*0.07*	79.40	0.05	36.56	0.34	*12.48*	*0.28*
	0.50	*45.74*	*0.27*	47.68	0.26	48.28	0.74	*44.24*	*0.72*
	0.75	*1.49*	*0.76*	4.40	0.72	*4.88*	*0.71*	10.51	0.67

MCMC: Markov Chain Monte Carlo Methods; REP: Relative Error in the $P(R_{GT}>1)$;$P(R^{Bt}>1)$ is the relative frequency that R^{Bt}, the Bootstratio estimate of R_{GT}, is greater than one. $P(R^{BY}>1)$ is the relative frequency that R^{BY}, the Bayesian estimate of R_{GT}, is greater than one.

2.3. Real data example: unique condition for each gene with No control sample (samples already normalized against the control)

The dataset consist on the relative expression ratios of 12 S/MT-RNR1, MT-CO2/COX2, and MT-ATP6 mitochondrial genes in prostate cancer [17]. Ratios corresponded to the relative expression of tumor samples versus normal samples for each of the patients (see Example dataset 1 on http://regstattools.net/br). A total of 19 relative expression ratios for each gene were used.

2.4. Real data example: several conditions for each gene with a control sample (when control samples are provided in the data set)

Total RNA was isolated from mice kidneys with the trizol method (Invitrogen). Purity and concentration of RNA were assessed using a spectophotometrer (NanoDrop). 1 μg of RNA was retrotranscribed using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) following manufacturer conditions for 10 min at 25 °C followed by 120 min at 37 °C. cDNAs were stored at −20 °C. qPCR was performed on a custom preplated TaqMan Gene Expression Assays set in a 384-well plate (Applied Biosystems) (See Appendix). The amount of RNA was calculated through the $2^{-\Delta\Delta Ct}$ method [18] using Cyclophillin A (PPIA) as reference gene. Samples were from 3 months-mice in a mix background (C57Bl6-129). The two experimental conditions tested were lithiasic (presence of calculi, P) and non lithiasic (NP) *slc7a8*$^{-/-}$ mice [19]. As control, samples from wild type animals (WT, males and females) were used.

Genes analyzed were *arf1*, *odc1* and *s100a11* (see Example dataset 2 on http://regstattools.net/br) [19]. For each gene analyzed, a total of 25, 16 and 17 expression values were used for the NP, P and control (reference) conditions, respectively. Therefore, the sampling distributions of 2 expression ratios were calculated for each gene (one for NP condition versus control condition, and another one for P condition versus control condition).

3. Results

3.1. Simulation study: BootstRatio versus Bayesian approach

Table 1 shows results of the simulation study comparing the Bootstratio approach with the Bayesian one in the estimation of $P(R_{GT}>1)$ highlighting which one of the methods perform better in each situation. The upper part of the table shows the REP as well as $P(R^{Bt}>1)$ and $P(R^{BY}>1)$ assuming a Gamma probability distribution and a Uniform probability distribution for the numerator (X_T) of the simulated ratio. The Bayesian approach performed slightly better than the Bootstratio one when the simulated distribution of X_T was Gamma ($N=60$ and $P(R_{GT}>1)=0.25$ and $P(R_{GT}>1)=0.5$; $N=20$ and $P(R_{GT}>1)=0.25$; $N=12$ and $P(R_{GT}>1)=0.25$ and $P(R_{GT}>1)=0.5$). Therefore, it performed better in 5 out of 9 situations when the simulated distribution of X_T is Gamma. However, when the simulated distribution of X_T was Uniform, the Bootstratio approach performs better than the Bayesian one in 7 out of 9 situations. Although we found that the Bootstratio approach performed better than the Bayesian method in 11 out of 18 times, we make note that the Bayesian method performed slightly better than BootstRatio when sample size was set to $N=12$ (4 out of 6 times).

When random noise is added to the distribution of X_T (lower part of the Table), the Bootstratio method performed better in 7 out of 9 times when the simulated distribution of X_T is Gamma with random noise. The same results were observed when the simulated distribution of X_T was Uniform with random noise. Therefore, under the condition of random noise added to the data, the Bootstratio approach performs better than the Bayesian one in 14 out of 18 times.

3.2. Real data example application: unique sample for each gene with no control sample (samples already normalized against the control)

Fig. 2(a) shows the web page where the user can perform this type of analysis. The table of statistics returned by BootstRatio web application is shown in Table 2. These statistics are the observed mean ratio (Mean.Obs), median (Median.Obs), standard error (SE.Obs), gene sample size (N.Obs) and the median (Median.Boot), and standard deviation (SD.Boot) of the bootstrap median ratio of the samples, respectively. The two following columns of this table refer to $\widehat{P}(\mu_{R_G}>1)$, column Prob.Ratio > 1, whereas column Prob.Ratio < 1 refers to $1-\widehat{P}(\mu_{R_G}>1)$. The last five columns may allow the user to easily assess the level of significance of the bootstrap ratios. Note that none of the genes was statistically significant to levels below 10%, however, the

Table 2
Statistics of expression ratios extracted from BootsRatio web application for the unique condition for each gene example dataset from Abril et al., 2008.

Gene	Mean.Obs	Median.Obs	SE.Obs	N.Obs	Median.Boot	SD.Boot	pvalue (Ratio < 1)	pvalue (Ratio > 1)	$p<0.1$	$p<0.05$	$p<0.01$	$p<0.001$	$p<0.0005$
12s/MT-RNR1	0.905	0.841	0.085	19	0.897	0.150	0.245	0.755	N	N	N	N	N
MT-ATP6	0.914	0.948	0.099	19	0.905	0.172	0.292	0.708	N	N	N	N	N
MT-CO2/COX2	0.900	0.932	0.111	19	0.877	0.183	0.242	0.758	N	N	N	N	N

Gene: Name of the Gene.
Mean.Obs: observed mean of the gene expression ratio.
Median.Type: observed mean of the gene expression ratio.
SE.Obs: standard error of the observed gene expression ratio.
N.Obs: sample size of the observed expression ratio.
Median.Boot: median value of the bootstrap ratio.
SD.Boot: standard deviation of the bootstrap ratio.
Prob.Ratio > 1: is the relative frequency of Median.Boot > 1. It approximates to a p-value.
Prob.Ratio < 1: is the relative frequency of Median.Boot < 1. It approximates to a p-value.
$p<0.1$: indicates if any of Prob.Ratio > 1 or Prob.Ratio < 1 values are lower than 0.1.
$p<0.05$: indicates if any of Prob.Ratio > 1 or Prob.Ratio < 1 values are lower than 0.05.
$p<0.01$: indicates if any of Prob.Ratio > 1 or Prob.Ratio < 1 values are lower than 0.01.
$p<0.001$: indicates if any of Prob.Ratio > 1 or Prob.Ratio < 1 values are lower than 0.001.
$p<0.0005$: indicates if any of Prob.Ratio > 1 or Prob.Ratio < 1 values are lower than 0.0005.

bootstrap analysis shows that there is a high probability to observe a decreased expression (see column Prob.Ratio <1) of these genes. We did not observe large differences comparing these results with those reported in the previous paper of [17]. Using BootstRatio we have obtained the following probabilities $\widehat{P}(\mu_{R_{12s/\text{MT-RNR1}}}<1)=0.755$, $\widehat{P}(\mu_{R_{\text{MT-CO2/COX2}}}<1)=0.708$ and $\widehat{P}(\mu_{R_{\text{MT-ATP6}}}<1)=0.758$ whereas in the previous analysis using a Bayesian method these probabilities were 0.6740, 0.6844 and 0.6647, respectively. Fig. 3(a) shows the boxplot representation of the three genes analyzed depicting their tendency towards a decreased expression. Ratios above one indicates increased expression whereas those ratios below one shows decreased expression of the analyzed gene.

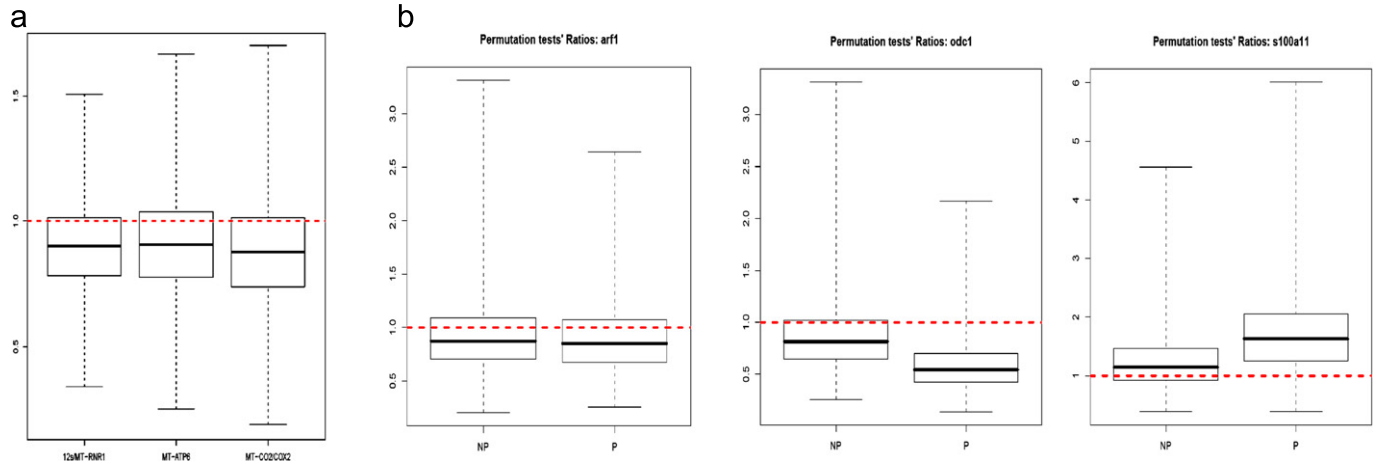
3.3. *Real data example application: several samples for each gene with a control sample (when control samples are provided in the data set)*

Fig. 2(b) shows the web page where the user can perform this type of analysis. Table 3(a) shows the table of descriptive statistics of the observed data whereas Table 3(b) shows the statistics obtained from the permutation's test analysis. If we consider that conditions with type I error of $\alpha=0.05$ – $\widehat{P}(R_G>1)<0.05$ – are statistically significant, we observe that only gene *Odc1* expression ratio with condition P can be considered to be statistically significant with a sampling median ratio of 0.544 and $\widehat{P}(R_{odc1-P}<1)=0.0390$. The gene *S100a11* expression ratio with condition P could be also considered to be statistically significant if we set a type I error of $\alpha=0.10$ – which is not commonly assumed – $\widehat{P}(R_{s100a11-P}<1)=0.089$. Although a*rf1* gene expression ratio cannot be considered to be statistically significant, our approach shows that there might be a probability that a significant lower expression than control conditions can be observed for NP condition, $\widehat{P}(R_{arf1-NP}<1)=0.649$ and for P condition, $\widehat{P}(R_{arf1-P}<1)=0.671$, respectively. Fig. 3(b) depicts the boxplots for each gene and experimental conditions.

4. Conclusions

Real-time PCR is an easy to perform methodology, provides the necessary accuracy and produces reliable as well as rapid quantification results which require a reproducible methodology and adequate mathematical models for data analysis. Other statistical methods have been described for this type of analysis [3–8] which may entail assuming an underlying probability distribution of the data. As the gene expression ratio is a positive function which may depend on the ratio of two random variables with positive support, a probability distribution with positive support such as Gamma might be used for parametric inference [20]. On the other hand, most parametric hypothesis tests are based on asymptotic distributions relying on large sample size, although data sample size is frequently smaller than would be considered optimal ($N<20$). Bayesian simulation methods [21] have been shown to be an alternative to these tests in this situation [17]. In our simulation study we can suggest a similar conclusion when sample sizes are small. When $N=12$ the Bayesian method performed slightly better than the Bootstrap one except when random noise was added. It might suggest that the choice of the prior distribution may affect in certain situations even when this prior distribution is non informative. In this line, the choice of the probability distribution for the expression ratio can still be a critical point on the analysis because inference strongly depends on these parametric assumptions and the erroneous choice of this distribution may lead to not reliable conclusions.

Resampling techniques can be used although we must assume independence of the samples [9]. Advantages of these methods are that they not require the usual normality assumption to be met, and that it can be effectively utilized even with small sample sizes. Since assumption of a probability distribution for biological data can be complex, these methods have been widely used in the bioinformatics field [14]. Another advantage of these techniques is that we can obtain a large sample distribution of the statistic of interest, in our case, the median ratio, and therefore, probabilities – estimated by means of relative frequencies – and other statistics can



The x-axis refers to gene where as the y-axis refers to the bootstrap median of expression ratios

The x-axis refers to experimental condition for each geneanalyzed where as the y-axis refers to the bootstrap median of expression ratios between sample type and sample control

Note: The limits of each box plot represents the rank of the data. The lower and upper edges of the box are the first and third quartiles, therefore 50% of the data occurs in this range. The thick-dark horizontal segment represents the median.

Fig. 3. Boxplots of the Bootstrap Median Expression Ratios: (a) Unique condition for each gene example dataset from Abril et al., 2008; (b) Several conditions for each gene example dataset.

Table 3
Statistics of expression ratios extracted from BootsRatio web application for the several conditions for each gene example dataset[a]: (a) Observed values; (b) Results of the permutation's test sampling.

(a)

Gene	Type	Mean.Type	Median.Type	SE.Type	Mean.Ctrl	Median.CtrlSE.Ctrl	SE.Ctrl	Ratio.Mean.Obs	N.Type	N.Ctrl
arf1	NP	0.981	0.890	0.107	1.033	0.981	0.070	0.950	25	17
arf1	P	1.003	0.934	0.080	1.033	0.981	0.070	0.971	16	17
odc1	NP	0.902	0.609	0.173	1.316	1.137	0.246	0.685	25	17
odc1	P	0.682	0.381	0.157	1.316	1.137	0.246	0.518	16	17
s100a11	NP	1.219	1.127	0.134	1.033	0.926	0.070	1.180	25	17
s100a11	P	1.584	1.174	0.301	1.033	0.926	0.070	1.533	16	17

(b)

Gene	Mean.Ratio	Median.Ratio	SD.Ratio	pvalue(Ratio <1)pvalue(Ratio >1)	pvalue(Ratio >1)	$p<0.1$	$p<0.05$	$p<0.01$	$p<0.001$	$p<0.0005$
arf1	0.927	0.871	0.322	0.351	0.649	N	N	N	N	N
arf1	0.911	0.852	0.330	0.329	0.671	N	N	N	N	N
odc1	0.865	0.815	0.307	0.272	0.729	N	N	N	N	N
odc1	0.579	0.544	0.215	0.039	0.961	Y	Y	N	N	N
s100a11	1.246	1.147	0.461	0.664	0.337	N	N	N	N	N
s100a11	1.716	1.628	0.629	0.912	0.089	Y	N	N	N	N

Gene: Name of the Gene; **Type:** Sample type;
Mean.Type: mean of the type sample; **Median.Type:** median of the type sample;
SE.Type: standard error of the type sample;
Mean.Ctrl: mean of the control sample;**Median.Ctrl:** median of the control sample
SE.Ctrl: standard error of the observed sample;
Ratio.Mean.Obs: type mean divided by control mean;
N.Type: sample size of type sample;**N.Ctrl:** sample size of control sample;
Mean.Ratio: mean value of the ratio samples' mean type divided by samples' mean control;
Median.Ratio: median value of the ratio between samples' mean type divided by samples' mean control;
SD.Ratio: standard deviation of the ratio between samples' mean type divided by samples' mean control;
Prob.Ratio $>$ **1:** is the relative frequency of Median.Ratio.Boot <1. It approximates to a p-value;
Prob.Ratio <1: is the relative frequency of Median.Ratio.Boot <1. It approximates to a p-value;
$p<$**0.1:** indicates if any of Prob.Ratio >1 or Prob.Ratio <1 values are lower than 0.1;
$p<$**0.05:** indicates if any of Prob.Ratio >1 or Prob.Ratio <1 values are lower than 0.05;
$p<$**0.01:** indicates if any of Prob.Ratio >1 or Prob.Ratio <1 values are lower than 0.01;
$p<$**0.001:** indicates if any of Prob.Ratio >1 or Prob.Ratio <1 values are lower than 0.001;
$p<$**0.0005:** indicates if any of Prob.Ratio >1 or Prob.Ratio <1 values are lower than 0.0005.

[a] The table shown above has been subdivided into two tables In order to depict the returned table by BootstRatio application.

be computed in this sample. Although resampling and estimation requires intensive computation, we also present a web application, BootstRatio, to perform this type of analysis with the advantage that the user will get the results on a format that is easily transformed into publication data and easily understandable by non-experts in statistics. Besides, this approach could be extended to other type of analysis in which the final results are provided as ratios, such as western, southern and northern blots.

The permutation test and bootstrap approach are a practical solution for the statistical analysis of gene expression ratio determined by real-time PCR. The web application BootstRatio is easily accessible, freely available and developed for the purpose of the intensive computation statistical analyses.

Authors' contributions

RC developed the statistical model, programmed its R code and analyzed the data. RC and JR designed the web application and drafted the article. MLH carried out real-time PCR experiments, analyzed the data, provided assistance with the design of the web application and finalized the draft. JG programmed the web application and provided assistance to adapt the R code to the web server. ME and VN carried out real-time PCR experiments, provided oversight of the work and co-designed experiments, discussed analyses, interpretation, and presentation.

Acknowledgments

This work was partially supported by CIBER de Enfermedades Raras, an initiative of the ISCIII Institute of Health Research of the Spanish Governement. This work was also supported in part by SAF2009-12606-C02-02 and 09SGR1490 to V. Nunes by MICINN and Generalitat de Catalunya, respectively. Study sponsors had no involvement in the study design, in the collection, analysis and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

Appendix A

List of TaqMan assays used. The link on the Ref TaqMan column directs to Applied Biosystem's web site at the specific gene expression assay. Reference gene is indicated in bold characters.

Gene	Ref TaqMan
S100a11	Mm00845129_g1
Ppia (control)	**Mm02342430_g1**
Arf1	Mm01946109_uH
Odc1	Mm01964631_g1

Appendix B. Supplementary materials

Supplementary materials associated with this article can be found in the online version at doi:10.1016/j.compbiomed.2011.12.012.

Reference

[1] M.W. Pfaffl, M. Hageleit, Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT-PCR, Biotechnol. Lett. 23 (2001) 275–282.
[2] S.A. Bustin, Why the need for qPCR publication guidelines?—The case for MIQE, Methods 50 (2010) 217–226.
[3] M.W. Pfaffl, A new mathematical model for relative quantification in real-time RT-PCR, Nucleic Acids Res. 29 (9) (2001) e45.
[4] P.Y. Muller, H. Janovjak, A.R. Miserez, Z. Dobbie, Processing of gene expression data generated by quantitative real-time RT-PCR, Biotechniques 32 (2002) 1372–1379.
[5] M.W. Pfaffl, G.W. Horgan, L. Dempfle, Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR, Nucleic Acids Res. 30 (2002) e36.
[6] S.N. Peirson, J.N. Butler, R.G. Foster, Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis, Nucleic Acids Res. 31 (2003) e73.
[7] R. Gilsbach, M. Kouta, H. Bonisch, M. Bruss, Comparison of in vitro and in vivo reference genes for internal standardization of real-time PCR data, Biotechniques 40 (2006) 173–177.
[8] J.S. Yuan, A. Reed, F. Chen, C.N. Stewart Jr, Statistical analysis of real-time PCR data, BMC Bioinformatics 7 (2006) 85.
[9] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, New York, 1993.
[10] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evolution 39 (1985) 783–791.
[11] M.K. Kerr, G.A. Churchill, Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments, Proc. Natl. Acad. Sci. USA 98 (2001) 8961–8965.
[12] S. Imoto, T. Higuchi, S. Kim, E. Jeong, S. Miyano, Residual bootstrapping and median filtering for robust estimation of gene networks from microarray data, in: Proceedings of Computational Methods in Systems Biology '04, LNBI, vol. 3082, 2004, pp. 149–160.
[13] P.D.W. Kirk, P.H. Stumpf, Gaussian process regression bootstrapping: exploring the effects of uncertainty in time course data, Bioinformatics 25 (10) (2009) 1300–1306.
[14] M.J. Li, P.C. Sham, J. Wang, FastPval: a fast and memory efficient program to calculate very low P-values from empirical distribution, Bioinformatics 26 (2010) 2897–2899.
[15] E.J.G. Pitman, Significance tests which may be applied to samples from any population, J. R. Stat. Soc. Suppl. 4 (1937) 119–130.
[16] E.J.G. Pitman, Significance tests which may be applied to samples from any population, J. R. Stat. Soc. Suppl. 4 (1937) 225–232.
[17] J. Abril, M.L. de Heredia, L. Gonzalez, R. Cleries, M. Nadal, E. Condom et al., Altered expression of 12 S/MT-RNR1, MT-CO2/COX2, and MT-ATP6 mitochondrial genes in prostate cancer, Prostate 68 (2008) 1086–1096.
[18] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method, Methods 25 (2001) 402–408.
[19] L. Feliubadalo, M.L. Arbones, S. Manas, J. Chillaron, J. Visa, M. Rodes, et al., Slc7a9-deficient mice develop cystinuria non-I and cystine urolithiasis, Hum. Mol. Genet. 12 (2003) 2097–2108.
[20] E.W. Stacy, A generalization of the Gamma distribution, Ann. Math. Stat. 33 (3) (1962) 1187–1192.
[21] A. Gelman, J.B. Carlin, H. Stern, R.B. Rubin, Bayesian Data Analysis, Second ed., Chapman & Hall, London, 2003.