Chapter 13

qPrimerDB: A Powerful and User-Friendly Database for qPCR Primer Design

Wei Chang, Yue Niu, Mengna Yu, Tian Li, Jiana Li, and Kun Lu

Abstract

Real-time quantitative polymerase chain reaction (qPCR) is a powerful tool for analyzing and quantifying gene expression, and its primer design is the first and most important step. In order to improve the efficiency and effectiveness of primer design, we designed a database qPrimerDB, based on thermodynamics gene-specific for multispecies qPCR primers design.

In this chapter, we explained the working principle of the database and detailed the step-by-step practical steps with examples. The valuable and time-saving qPrimerDB database is publicly accessible at http://biodb.swu.edu.cn/qprimerdb and will be routinely updated.

Key words qPrimerDB, Quantitative real-time polymerase chain reaction, Free online database, Primer design, High-level efficiency

1 Introduction

Real-time quantitative polymerase chain reaction (qPCR) has become one of the most powerful tools for molecular genetic studies and has been widely used to quantify the gene expression levels, even those with low expression levels, in clinical and biological fields since 1993 [1, 2]. The fluorescence signal is monitored based on the whole classical PCR reaction by adding fluorophore. Then, the initial concentration of desired gene is quantified through the *Ct* (cycle threshold) value, which has a linear relationship with the logarithm of the initial copy number of the template. Nowadays, there are two most population methods to detect the fluorescence signal in qPCR. One is the unspecific detection independent of the target sequence using fluorescent dyes such as SYBR Green, and another one is sequence-specific fluorescent oligonucleotide probes such as TaqMan probes or molecular beacons [3].

Either any method, however, the suitability of the designed primers and probes is one of the most essential factors for the successful qPCR, since the specificity of the qPCR is closely related

Chhandak Basu (ed.), *PCR Primer Design*, Methods in Molecular Biology, vol. 2392, https://doi.org/10.1007/978-1-0716-1799-1_13,

to the annealing of primers to their complementary targets [4], and there are many genes with similar sequences, especially in polyploid species. Moreover, previous studies have found that the transcription levels existed greatly variation among different gene regions in several eukaryotes [5–7].

In order to ensure the accuracy and reliability results of qPCR, the primers need to be designed strictly and efficiently; thus, various computational methods, programs, and databases have been developed. Several local or Web-based programs based on Primer3 [8], an important program for primer design, have been developed for batch primer design, including BatchPrimer3 [9], QuantPrime [10], PCRTiler [11], PRIMEGENS [12], and PrimerMapper [13]. The primers designed by these tools lack sufficient specificity, because the primers are only evaluated by the Nucleotide Basic Local Alignment Search Tool (BLASTN) or sequence similarity searching. To address the abovementioned challenge, some programs based on the thermodynamics specificity checked were published, such as MFPrimer-2.0 [14] and the MapReduce-based method, MRPrimer [15]. As a result, many qPCR primer databases were created, such as PrimerBank [16], MRPrimerW [17], and GETPrime 2.0 [18]. However, they contain only a few important species.

In our study, we provided the most comprehensive, uniform qPCR primer design database to date, qPrimerDB, to researchers. We developed it through an automatic gene-specific qPCR primer design and thermodynamics-based validation workflow. Furthermore, qPrimerDB database provides precomputed primer pair spanning 147 important organisms including as many sequenced genomes as possible, such as human, mouse, yeast, rice, and zebrafish. Our database contains 3,331,426 of the best primer pairs for each gene, based on primer pair coverage, as well as 47,760,359 alternative gene-specific primer pairs, which can be conveniently batch downloaded and rapid retrieval of qPCR primers for selected genes, via a user-friendly interface. The primers designed by this database for 66 randomly selected genes shown specific and accurate results through qPCR assays and gel electrophoresis. Next, we will introduce the use and function of this website in detail.

2 Materials

2.1 Computer and Websites

Please prepare a computer with Internet access during this procedure. The major websites used in this procedure is qPrimerDB (http://biodb.swu.edu.cn/qprimerdb), as well as other websites you may need to use as a supplement include NCBI (http://www.ncbi.nlm.nih.gov/).

2.2 DNA Sequences and Gene ID

DNA Sequences and Gene ID representing the target gene (s) of interest can be based on your own collection or be downloaded from NCBI, as well as a site dedicated to the target species. Sequences should be in a FASTA format.

3 Methods

The qPrimerDB website is a database for qPCR primers. The main page of the website comprises eight functional sections (Fig. 1). On the "Home" page, users can view the site's basic functions and the latest news, for instance, the qPrimerDB website has been updated to V1.2 since 2018, updated and added 304 new organisms, and a total 516 organisms have been included (*see* **Note 1**). The next "Browse" section in the navigation bar, we classify all organisms into four parts "Favorites," "Animals," "Plants" and "Others," help users find target organisms more concisely. In the "Tools" section, we embedded the BLAST program. The "Download" section can help users easily download the target primer sequences in batches (*see* **Note 2**). By selecting "Documents," users can find the manual, pipelines for primer design and database implementation, statistics for each organism, and related resources (*see* **Note 3**). In the "help" section, the main purpose is to help users solve

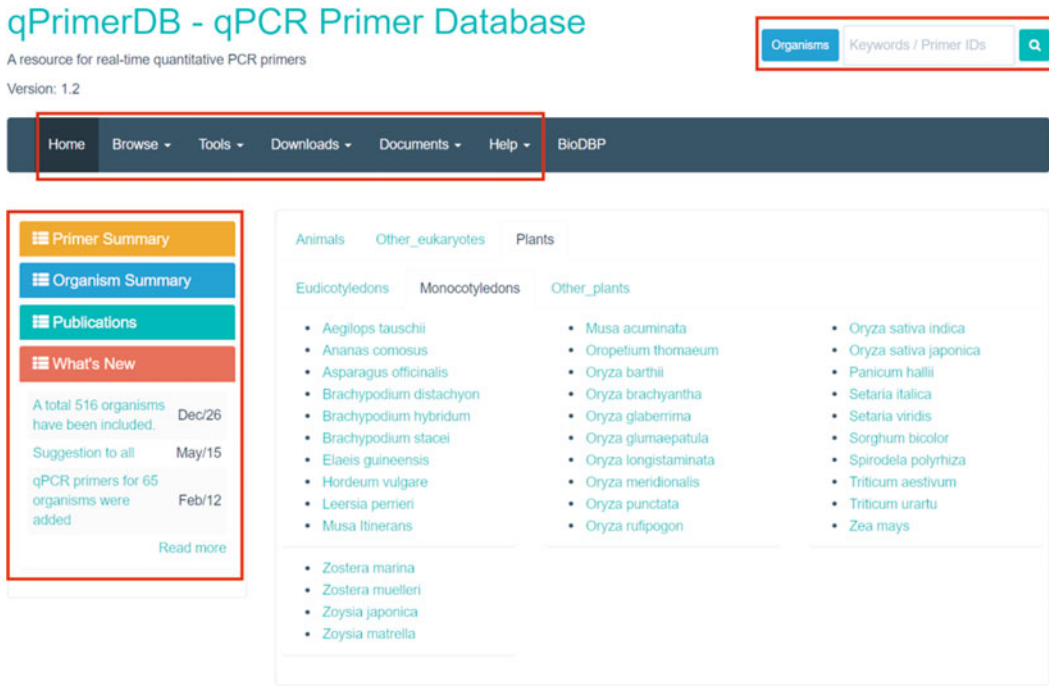

Fig. 1 Browse interface on the "Home" page. A search box is provided in the upper right corner of every page to enable convenient searching of keywords of interest. Functional sections and information summary are picked out in the red box which provides different applications

problems that are difficult to solve in using the website and let us improve the website to make it more user-friendly (*see* **Note 4**). A "search" tool is available on the upper right corner of website page, users input gene names, primer IDs, or keywords of interest to get the target primer sequence. Common biological database links exist and can be analyzed by clicking "BioDBP" (*see* **Note 5**).

In this protocol, we use the *Arabidopsis thaliana* gene *AT1G72390* as a demonstration to explain in detail how to design the qPCR primers step by step and, when appropriate, show optimal parameters.

3.1 Accessing Website

Use your computer or laptop to access the qPrimerDB website via the Internet: http://biodb.swu.edu.cn/qprimerdb.

3.2 Primer Search

There are two ways to help users find the target qPCR primers: input gene ID to search interest primers (Fig. 2) or entering nucleotide sequence manually (or uploading the nucleotide sequence in FASTA format) to search them by BLAST (Fig. 3). The choice of options depends upon the user's resources and format.

3.2.1 Search by Gene ID

A search box is provided with the upper right corner of every page to enable convenient searching for keywords of interest. Option 1—users click "organisms" to find option "*Arabidopsis thaliana*" and enter AT1G72390 into the input box next to "organisms" (*see* **Notes 6** and **7**), and then, click on the green button next to input box and wait for the results (Fig. 2a).

When the page jumps, users will get the best qPCR primer(s) of AT1G72390 provided by the website, Fprimer: GGCTGAA GATTTTCTCTTAGCG and Rprimer: ACTGTTGCA TATCGTTTGCAG (Fig. 2b, *see* **Note 8**).

Option 2 to get this result is users can browse the organism of interest in "home" page, by sequentially clicking "Plants," "Eudicotyledons," and "*Arabidopsis thaliana*" (Fig. 2c), enter AT1G72390 into the input box the upper right corner of website page (Fig. 2d), and finally wait for the page to jump to the same result (Fig. 2e).

After users clicking the blue button below in primerID in Fig. 2b or primerID A.thaliana.005900v1 in Fig. 2e, detailed information for the primers is presented (Fig. 3a). If the user has other requirements for the primers, please click blue button below in GeneID Fig. 2b or GeneID AT1G72390 in Fig. 2e; all primers will be listed in table format (Fig. 3b).

3.2.2 Search by Sequence BLAST

When users have the nucleotide sequence but not sure about the gene ID, click the "tool'" button in the navigation bar and select BLAST (Fig. 4a).

Click the "1. Database" menu and select "*Arabidopsis thaliana*." In terms of parameter setting, the *E*-value is a widely

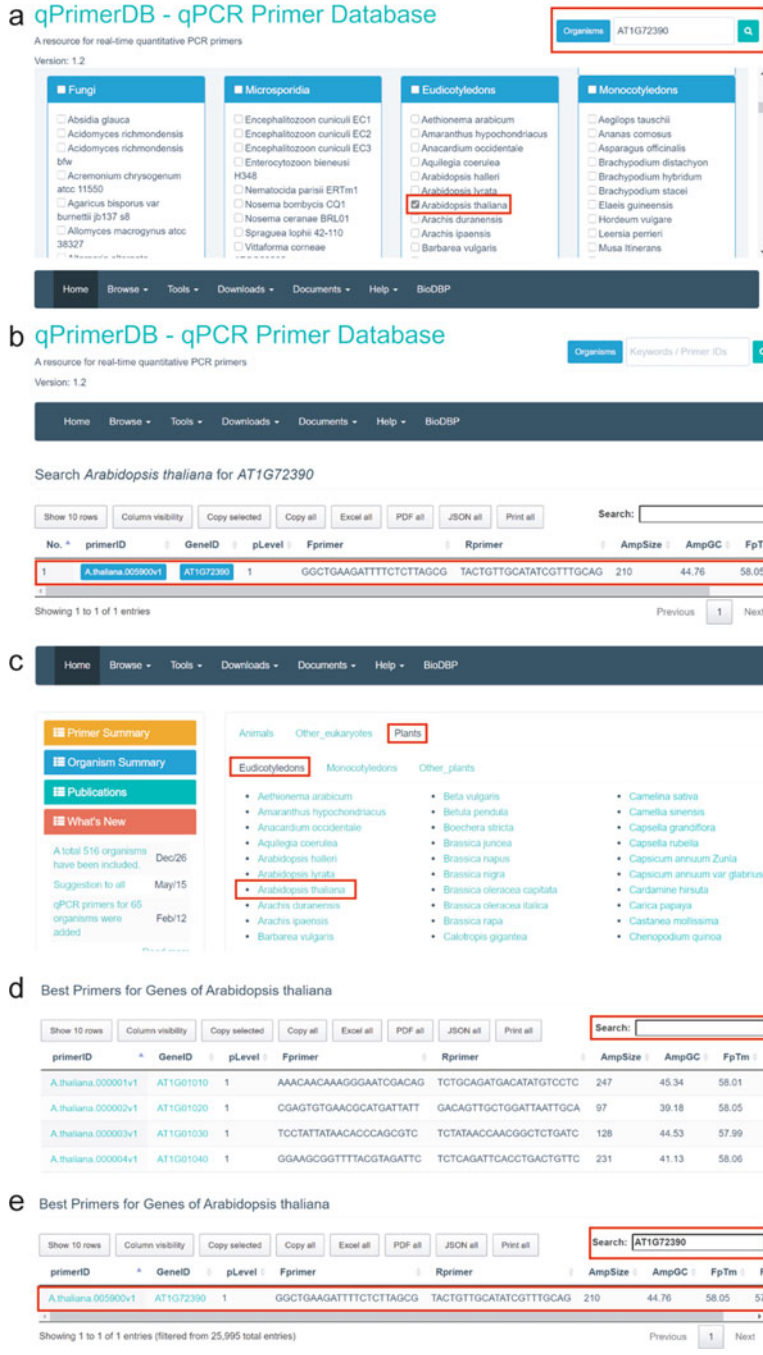

Fig. 2 Screenshots of the navigation bar and browse module in qPrimerDB. (**a**) Use the search box to find the primers. For example, input target gene "AT1G72390" in search box and choose "organisms," "Eudicotyledons," and "*Arabidopsis thaliana.*" (**b**) Best qPCR primer(s) provided by the website after clicking the blue button in (**a**). (**c**) Users can browse the organism of interest, for example, by sequentially clicking "Plants," "Eudicotyledons," and "*Arabidopsis thaliana.*" (**d**) Example of A. thaliana primers in table format. Both the record number per page and the order of each column can be adjusted, as needed. Users can input target gene ID in search box for a more precise lookup. (**e**) After precise lookup, qPrimerDB will provide the user with the best primer(s)

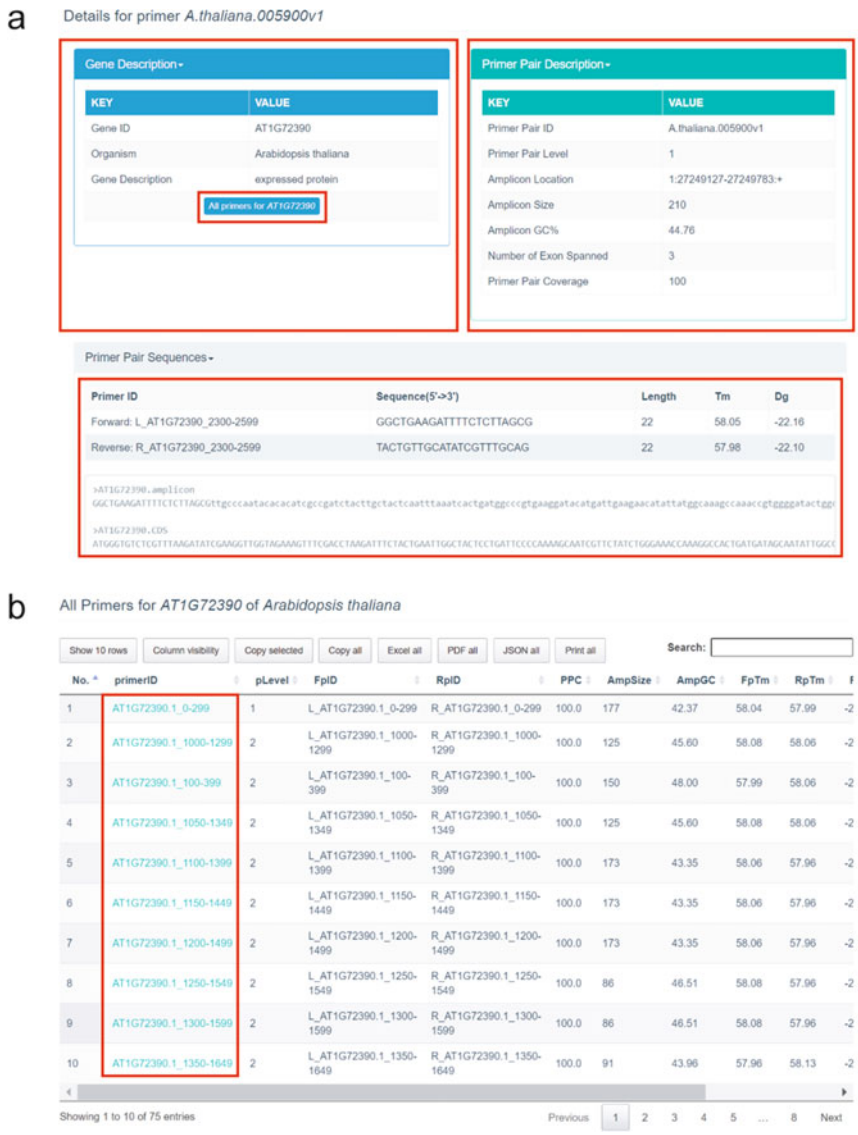

Fig. 3 qPrimerDB primer details page. (**a**) Detailed information for primer ID A.thaliana.005900v1 is presented in three sections: Gene Description (gene ID, organism, gene description, and a blue button "All primers for AT1G72390"); Primer Pair Description (primer pair ID and level, amplicon location, amplicon size, amplicon GC content, number of exons spanned, PPC); and Primer Pair Sequences (T_m values of primers, primer sequences, primer length and amplicon and template coding/mRNA sequences). (**b**) Example of all primers for AT1G72390 in table format. After clicking the blue button in (**a**), all primers will be listed in table format

accepted measure for assessing potential biological relationship and *E*-values ≤ 0.01 normally suggest homologous sequences, so we can use the default parameters, if you do not have any other requirements. The last step is entering sequence manually or uploading the nucleotide sequence in FASTA format to search them by BLAST and click "submit" button (Fig. 4b).

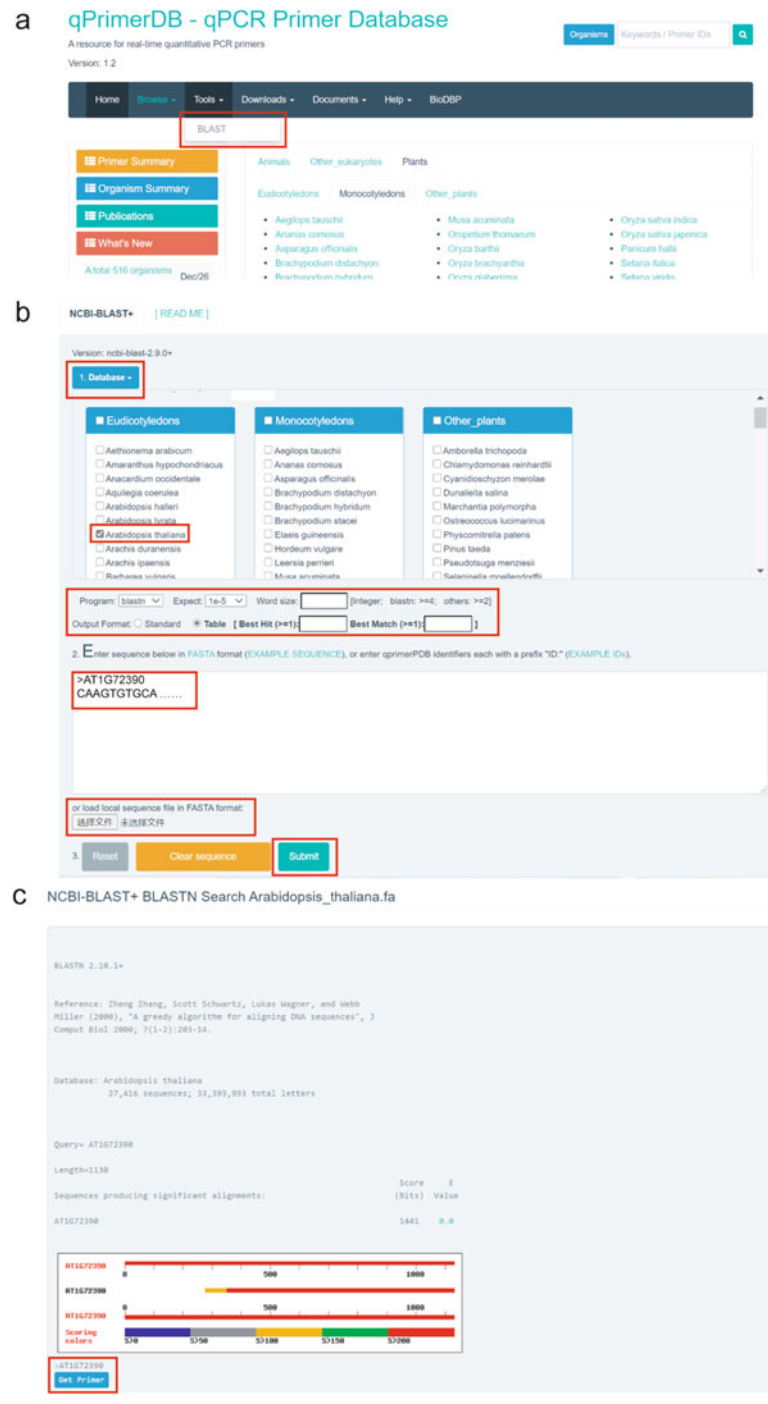

Fig. 4 Screenshots of search primers by sequence BLAST. (**a**) BLAST function in navigation bar "Tool" button. (**b**) BLAST page requires users to select the Database (Animals, Other_eukaryotes, Plants), set the parameters (Program, Expect, Word size and Output Format), and submit the target sequence (or load local sequence file in FASTA format) of the primers to be designed as well, then clicking submit. (**c**) Detailed information for Blast results

After input AT1G72390 sequence, the website will display the results of the BLAST and users can click the blue "get primers" button at the bottom of the page (Fig. 4c); the page then jumps to Fig. 3 (*see* **Note 9**).

3.3 Checking Primers

Detailed information for primerID AT1G72390.1 0-299 is presented in three sections: Gene Description (gene ID, organism, gene description); Primer Pair Description (primer pair ID and level, amplicon location, amplicon size, amplicon GC content, number of exons spanned, primer pair coverage); and Primer Pair Sequences (primer sequences, primer length, T_m values of primers, and amplicon sequence) (Fig. 3a, *see* **Note 10**). Users can click the blue bottom "All primers for AT1G72390" to get more primers of interest, if have other requirements about qPCR primers design (Fig. 3b, *see* **Note 11**).

4 Notes

1. qPrimerDB database will be updated regularly based on user feedback, while nonspecific primers will be removed.
2. In qPrimerDB, all qPCR primers and the best primers designed for each organism are, respectively, compressed into two zip files. Due to unstable Internet, users also can download two primer files from the "Downloads" page and then pick gene-specific primers in personal computer at any time.
3. The primer pairs for each gene are divided into three levels based on PPC and the binding stability of the binding site, and it can be scanned by "Pipeline" from "Documents."
4. On the Help page, users can scan frequently asked questions (FAQs) and answers. It also shows a Contact Us tab in the Help menu, where users can submit some comments or suggestions about the database establishment.
5. The BioDB Platform is a collaborative project and developed by scientists from professional fields of Molecular Biology, Genomics, Microbiology, Bioinformatics, and Computer Science.
6. Please send a feedback form to us if the qPrimerDB did not cover the organisms of your interest, and we will add the gene-specific qPCR primer pairs for you.
7. Some gene qRT-PCR primers may not be included in our database for the time being, please design your own primers according to **Note 9**. You are also welcome to email us to improve the quality of our database.
8. Website also provides multiple version support for some important organisms, such as cotton, and please make your choice.

9. For different annotation files of the same species with different names, users can input FASTA file to convert gene ID according to BLAST function.
10. qRT-PCR primer design for all the template fragments with the following parameters suggestion: amplicon size 80–300 bp; amplicon GC content 40–60%, with an optimal GC content of 50%; primer length 18–28 nt, with an optimal length of 22 nt; and melting temperature (T_m) 58–64 °C, with an optimal T_m of 60 °C and maximum T_m difference, per primer pair, of less than 3 °C.
11. Users are invited to submit their qPCR detection results (from melting curve analysis and/or gel electrophoresis) to the database designer, especially the specificity of the experimentally examined primer pairs. Such information will be used for further improving qPrimerDB.

References

1. Higuchi R, Fockler C, Dollinger G et al (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Nat Biotechnol 11:1026–1030
2. Lu K, Li T, He J, Chang W et al (2018) qPrimerDB: a thermodynamics-based gene-specific qPCR primer database for 147 organisms. Nucleic Acids Res 46(D1):D1229–D1236
3. Rodríguez-Lázaro D, Hernández M (2013) Real time PCR in food science: introduction. Curr Issues Mol Biol 15:25–38
4. Rosadas C, Cabral-Castro MJ, Vicente AC et al (2013) Validation of a quantitative real-time PCR assay for HTLV-1 proviral load in peripheral blood mononuclear cells. J Virol Methods 193:536–541
5. Arhondakis S, Clay O, Bernardi G (2008) GC level and expression of human coding sequences. Biochem Biophys Res Commun 367:542–545
6. S'emon M, Mouchiroud D, Duret L (2005) Relationship between gene expression and GC-content in mammals: statistical significance and biological relevance. Hum Mol Genet 14:421–427
7. Rao YS, Chai XW, Wang ZF, Nie QH, Zhang XQ (2013) Impact of GC content on gene expression pattern in chicken. Genet Sel Evol 45:9
8. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115
9. You FM, Huo N, Gu Y, Luo M, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9:253
10. Arvidsson S, Kwasniewski M, Riano-Pachon DM, Mueller-Roeber B (2008) QuantPrime—a flexible tool for reliable high-throughput primer design for quantitative PCR. BMC Bioinformatics 9:465
11. Gervais AL, Marques M, Gaudreau L (2010) PCRTiler: automated design of tiled and specific PCR primer pairs. Nucleic Acids Res 38: W308–W312
12. Kushwaha G, Srivastava GP, Xu D (2015) PRIMEGENSw3: a web-based tool for high-throughput primer and probe design. Methods Mol Biol 1275:181–199
13. O'Halloran DM (2016) PrimerMapper: high throughput primer design and graphical assembly for PCR and SNP detection. Sci Rep 6:20631
14. Qu W, Zhou Y, Zhang Y, Lu Y, Wang X, Zhao D, Yang Y, Zhang C (2012) MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity. Nucleic Acids Res 40:205–208
15. Kim H, Kang NN, Chon KW, Kim S, Lee NH, Koo JH, Kim MS (2015) MRPrimer: a MapReduce-based method for the thorough design of valid and ranked primers for PCR. Nucleic Acids Res 43:1–10

16. Wang X, Spandidos A, Wang H, Seed B (2012) PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. Nucleic Acids Res 40:D1144–D1149
17. Kim H, Kang N, An K, Koo J, Kim MS (2016) MRPrimerW: a tool for rapid design of valid high-quality primers for multiple target qPCR experiments. Nucleic Acids Res 44: W259–W266
18. David FPA, Rougemont J, Deplancke B (2017) GETPrime 2.0: gene- and transcript-specific qPCR primers for 13 species including polymorphisms. Nucleic Acids Res 45:D56–D60