

Original article

Selection of reference genes in peanut seed by real-time quantitative polymerase chain reaction

Shengjuan Jiang,* Yujun Sun & Songhua Wang

College of Life Science, Anhui Science and Technology University, Anhui Fengyang 233100, China

(Received 10 February 2011; Accepted in revised form 28 June 2011)

Summary Real-time quantitative polymerase chain reaction (qPCR) is a powerful tool for quantifying gene expression, combining both high sensitivity and specificity with efficient signal detection. Choosing the ‘correct’ normalisation control is essential and critical for obtaining reliable expression data. In this article, ten reference genes from *Arachis hypogaea* were analysed as internal controls for gene expression assay by using qPCR in two tissues of the seed, including embryo and cotyledon. The results showed that phospholipase D and actin were the most stable genes, whereas tubulin B4 and polyubiquitin 10 should be avoided when doing gene expression analysis because of their high variabilities in those tissue types. Although phospholipase D was scored as the most stable gene, it should be carefully validated for each particular experiment because it was easily regulated by different conditions. Taken together, actin was concluded as the optimal reference gene for studying gene expression in the mature peanut seed.

Keywords Gene expression, normalisation, peanuts, polymerase chain reaction, reference gene, seeds.

Introduction

Real-time quantitative polymerase chain reaction (qPCR) is a highly sensitive and reproducible technique with a wide dynamic range of amplification that allows quantitation of target genes. Two quantification approaches, absolute quantification and relative quantification, can be applied. Typically, in the relative quantification assay, a reference gene is amplified together with the target genes to calibrate their expression levels for different experimental conditions (Huggett *et al.*, 2005). The amount of target gene in each specimen is reported as the ratio of the target gene relative to the reference gene, with change in sample gene expression compared with a reference gene. Theoretically, reference genes should be essential for maintenance of cellular function and vitality and should be stably expressed in different tissues and cells (Yoo *et al.*, 2009). Usually, reference genes include structural genes, metabolism-related genes, protein translation-related genes, cytoskeleton maintenance genes, protein turnover genes and so on. However, a growing body of evidence indicates that transcript levels of these commonly used housekeeping genes may vary considerably in different tissue types or under different experimental conditions, and that a ‘universal control gene’ does not exist (Wong & Medrano, 2005). To avoid the misinterpretation of

the results, the evaluation and selection of stably expressed reference genes are essential for the quantitative expression analyses of the target genes. Bioinformatics tools, such as geNorm (Vandesompele *et al.*, 2002) and Normfinder (Andersen *et al.*, 2004), have emerged to perform the choice of an optimal set of reference genes for each experimental system.

Seeds are essential for flowering plant reproduction because they contain the developing embryo and energy resources that sustain the growth during germination before photosynthesis begins. In legumes, food reserves stored in embryo and cotyledon make seeds important as a food source for both human and animal consumption (Gallardo *et al.*, 2008). Peanut (*Arachis hypogaea* L.) is one of the most important legumes crops in the world. It, containing 45–53% oil and 24–29% protein in its seed, is among the most commonly used dietary crops for their good nutritional value and balanced amino acid content. However, it contains multiple seed allergens, which are responsible for triggering IgE-mediated fatal issued anaphylactic reactions in both children and adults (Shreffler *et al.*, 2006). The study on the peanut shows that mRNA and proteins of the allergens accumulate specifically in the seed (Kang *et al.*, 2007). The expression level of the allergens in the peanut seed can be examined by qPCR technique, which could lead to the applications, such as breeding seeds with more nutrients and reduced allergens (Brandon *et al.*, 2007). Accurate quantitation of the target genes in the

*Correspondent: E-mail: jiangsjahstu@126.com

seeds is heavily dependent on the optimal selection of the validated stable reference genes for data normalisation (Fink *et al.*, 2008). However, qPCR data have been normalised poorly in many molecular analyses, especially in the plant sciences (Gutierrez *et al.*, 2008; Paolacci *et al.*, 2009; Maroufi *et al.*, 2010; Yang *et al.*, 2010; Ye *et al.*, 2010). No statistically validated reference genes have been reported in peanut seed. In the present study, the expression stability of ten candidate reference genes was measured in peanut seeds, and geNorm tool and Normfinder software were used to establish reliable reference genes that were suitable to use in qPCR experiments performed on peanut seeds. Overall, the aim of the study was to identify of appropriate reference genes for allergen expression testing.

Materials and methods

Plant materials

Peanut (*Arachis hypogaea* L.) seeds (LUHUA cultivar, the representative of the cultivated lines in China) were kept by our laboratory at -80°C for 1 week before use. After sterilised with 70% (v/v) ethanol for 40~60 s and followed with treatment of 5% (v/v) sodium hypochlorite (NaOCl) for 20 min on the surface, the seeds were washed with sterile water to remove residual NaOCl. The cotyledon and the other parts of embryo (including embryonic bud, embryonic axes and embryonic root, 'embryo' for short) of the mature seeds from different seed lots (three biological replications) were collected, respectively, and frozen in liquid nitrogen immediately for RNA isolation.

RNA extraction

Total RNA of the embryo and cotyledon was extracted by CTAB/NaCl method, respectively (Azevedo *et al.*, 2003). To eliminate the possible DNA contamination, the RNA samples were treated with DNase I (Promega, Madison, WI, USA) at 37°C for 30 min and purified by ethanol precipitation. Each RNA sample was free of genomic DNA as determined by PCR. Only RNA samples with the $A_{260}/A_{280} > 2.0$ were used for the next experiments. The concentration of RNA was calculated from the absorbance at 260 nm, and 1% agarose gels were run to visualise the integrity of the RNA before and after DNase I treatment. All RNA samples were measured and adjusted to an equal concentration to homogenise RNA input in the subsequent cDNA synthesis reaction.

Primer design

Ten candidate reference genes, which were commonly used for normalisation and considered to be essential for

cellular function, such as cytoskeleton proteins, ribonucleoproteins, enzymes and translation-related proteins, were selected for qPCR analysis (Table 1). The primer pairs of the genes were manually designed with primer Premier 5.0 (PREMIER Biosoft International, Palo Alto, CA, USA) and Oligo 6 (Molecular Biology Insights, Cascade, CO, USA) according to the mRNA sequences in the GenBank database.

Verification of amplified products

Reverse transcriptions of about 800 ng RNA extracted from the embryo and cotyledon, respectively, were performed by using PrimeScriptTM RT reagent Kit (Perfect Real Time) (TaKaRa, Otsu, Shiga, Japan) with oligonucleotide dT primer and random 6 mers according to the instructions of the kit. All the cDNA samples were stored at -20°C and diluted 1:6 with EASY Dilution in the RT reagent kit before being used as template in the standard PCR and qPCR assay. In a series of initial experiments, the primer pairs were used to amplify the target genes via standard PCR and the product sizes were checked on a 1.2% agarose gel to confirm the primers' specificity. All the PCR products were gel-purified by using DNA gel extraction kit, ligated into the T/A cloning pGEM-T easy vector (Promega) at 4°C overnight and transformed into Top10 competent cells. Positive clones were selected and sequenced with ABI PRISM 3730 DNA sequencer. The sequences were searched for certification in GenBank with BLAST analysis.

Real-time quantitative polymerase chain reaction

Real-time quantitative PCR was implemented by using SYBR[®] Premix Ex TaqTM II (Perfect Real Time) (TaKaRa) and Applied Biosystems 7500 Real-Time PCR System (ABI, Carlsbad, CA, USA). The detection was made in the MicroAmpTM 8-strip tube with optical flat caps (ABI) in a final volume of 50 μL reaction mixture containing 25 μL SYBR[®] Premix Ex TaqTM II solution, 2 μL primers solution with the final concentration of 0.4 μM , 1 μL ROXTM Reference Dye II solution and 4 μL optimal RT reactant (cDNA template) and DEPC-treated water. The qPCR amplification was carried out at 95°C for 30 s for initial denaturation, followed by 40 cycles of 95°C for 5 s, 60°C for 34 s. Dissociation curves for each amplicon were then analysed to verify the specificity of each amplification reaction and the dissociation curve was obtained by heating the amplicon from 60 to 95°C . Fluorescence detection was performed at the annealing/extension step (60°C , 34 s). The qPCR experiments were repeated at least three times independently with a new RNA extraction from different peanut seeds each time, and no-template controls and non-RT controls

Table 1 Gene symbol, description and primer sequences of the ten candidate reference genes

Gene symbol	GenBank No	Gene description	Upstream primer 5'-3' Downstream primer 5'-3'	Product size/efficiency
Tub α	EE127259	Alpha tubulin 1	TGCCACCATCAAGACCAAG TCAATGCGACCAAACACCT	180 bp/97%
Tub β	EE125998	Tubulin B4	GCGAGGAGCACGGCATAGA CGCCAGACTGTCCGAAAAC	216 bp/104%
ACT	DQ873525	Actin	TACCAGATGGACAGGTTATCACAAT TGGAAACCACCACTCAAGACAAT	186 bp/102%
PLD	AY274834	Phospholipase D	AGTGGGAGAACAAAGATGATAG TTGGGATAAGATGAACAGCA	162 bp/93%
MDH	DQ889582	Malate dehydrogenase	CAGGTGCTGCTGGAATGA CAACGGAAACAAAGAATC	163 bp/92%
Rpl24	DQ889568	Ribosomal protein L24	CTCACTTCACTGCTCCGT GCACCTCATCGTCTTTAC	117 bp/123%
pUbq	DQ887067	Polyubiquitin 10	AGGAGTCAACCCTTCATC TTTTGGCCTTCACATTATC	129 bp/105%
Ubq	DQ887087	Ubiquitin	CTCGCGACTACAACATCC CACTCCTTCCTAAGCCTCT	197 bp/109%
EF1 α	EE125479	Elongation factor 1-alpha	GACTGCCACACCTCACACATT TTTTACCATAACCAGCATCAC	127 bp/103%
EF2	EE126343	Elongation factor EF-2	ATGGGAAAAGGCGAGGAACAG AACAGCAACACGCACAACAGGT	174 bp/106%

were included for each primer pair to avoid the non-specific amplification.

Data analysis

Data were analysed with 7500 SDS 1.4.0 software (ABI). To generate a baseline-subtracted plot of the logarithmic increase in fluorescence signal (ΔRn) vs. cycle number, baseline data were collected between cycles 3 and 15. All amplification plots were analysed with an Rn threshold of 0.2 to obtain Ct values (the number of amplification cycles needed to reach a fixed threshold in the exponential phase of the PCR reaction). All analyses were based on the Ct values of the PCR products. Based on five dilutions of cDNA template of embryo ranging from 1 to 2^{-4} , the amplification efficiency for each gene was calculated.

Following PCR data collection, the geNorm Visual Basic for Applications applet (Version 3.5) for Microsoft Excel was used to rank the expression variability of reference genes in all samples. Raw Ct values were transformed to relative quantities by using the comparative Ct method where the highest relative quantity for each gene was set to 1 for input into geNorm software (Vandesompele *et al.*, 2002). Each gene expression stability (M) value was calculated with the procedures outlined in the user's manual, and a larger M value indicated that the gene was less stable. Reference genes were ranked by geNorm through the elimination of the worst-scoring candidate control gene (the one with the highest M value) and recalculating of new M values for

the remaining genes. After this procedure was completed, two candidate genes were always ranked best because expression ratios were required for gene stability measurements. To estimate how many reference genes should be used, a pairwise variation of two sequential NF ($V_{n/n+1}$) was estimated as the standard deviation of the logarithmically transformed NF_n/NF_{n+1} ratios, reflecting the effect of including an additional gene. A large variation meant that the added gene had a significant effect on the normalisation and should preferably be included for the calculation of a reliable normalisation factor. Ideally, a pairwise variation of 0.15 was accepted as the cut-off, below which the inclusion of an additional control gene was not required for reliable normalisation.

The NORMFINDER software (Version 0.953 MOMA, Skejby, Århus N, Denmark) was used to identify the optimal normalisation gene among a set of candidate reference genes. It ranked the set of candidate genes according to the stability of their expression in a given sample set. The lowest stability value represented the most stable gene within the examined gene set.

Results

Analyses of the reference genes

Based on review of the relevant literatures, potentially useful reference genes, which were used for normalisation or routinely used as controls for RT-PCR, were selected. All the qPCR assays yielded specific products

with expected sequences. Checked by ethidium bromide-stained agarose gel, only the qPCR products of Rpl24 had the primer dimers. Changing the primers and adjusting the temperatures did not make the results better. Each melting curve of the qPCR reactions except that of Rpl24 gave a single sharp peak, which showed the high specificity of the products.

To compare the different RNA transcript levels, the Ct values were compared directly. The Ct values were inversely correlated with the amount of template nucleic acid present in the reaction. The analysis of absolute Ct values of the ten genes indicated that the Ct values ranged from 17.2 to 27.6, while the majority of these values were between 18 and 26 (Fig. 1). The Ubq and pUbq transcripts were very abundant, with their Ct values reaching the threshold fluorescence only after 17 and 19 amplification cycles, respectively. The less abundant transcripts included EF1 α , EF2, Rpl24, Tub α and MDH. The least abundant transcripts were PLD, ACT and Tub β .

Analyses of qPCR data were usually based on the assumption that all samples had similar PCR efficiencies. The amplification efficiency of the selected primer pair was confirmed by amplifying the cDNA derived from the embryo, and the qPCR efficiency of the genes except Rpl24 was between 92% and 109% (Table 1). Rpl24 gene had a high qPCR amplification efficiency (123%), which exceeded the credibility interval because there existed primer dimers.

Evaluation on the candidate reference genes

The geNorm analysis was used to select the best reference genes among the ten candidate genes and estimate the number of genes required to calculate a robust normalisation factor. The results showed that the studied genes except Tub β (M value 3.083) reached a

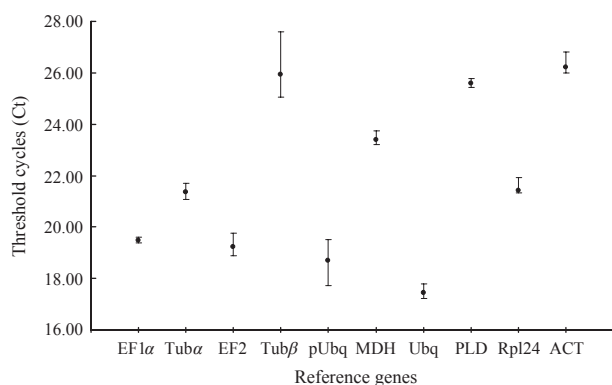


Figure 1 Absolute Ct values of the candidate reference genes in different tissues of peanut seeds. The medians are indicated by dots, and the error bars represent the minimum and maximum Ct values.

high expression stability with the final M values below 1.5 (the default geNorm cut-off value). PLD (M value 0.686) and ACT (M value 0.691) with the lowest M value were the most stable genes, while Tub β and pUbq (M value 1.281) had the lowest expression stability. So PLD and ACT were regarded as the best reference gene pair (Fig. 2). Then, pairwise variation was calculated by geNorm to determine the fewest number of reference genes necessary for accurate normalisation. The results suggested that the use of two most stable expression genes was sufficient for reliable data normalisation because the pairwise variation value of $V_{2/3}$ was below 0.15 (Fig. 3).

Normfinder, which used different algorithms, was applied to identify the optimal normalisation gene among the same candidates. The results indicated that PLD and ACT were identified as the stable genes, whereas Tub β and pUbq ranked poorly.

Discussion

Real-time quantitative PCR has become a widely used tool to examine gene expression levels. Although qPCR appears optimal for performing accurate, reproducible and convenient gene expression profiling experiments, it is becoming increasingly clear that no one gene can be defined as an omnipotent reference for a wide variety of biologic systems. Choosing the 'correct' normalisation reference gene is therefore essential and critical for obtaining reliable expression data (Nolan *et al.*, 2006). The ideal normalisation gene should have low variability from sample to sample, be critical to the activity of the cell and be consistently expressed in the primary tissue

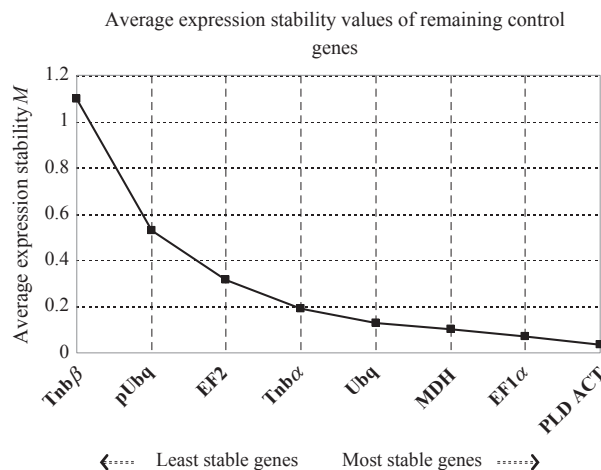


Figure 2 Ranking of the candidate genes as calculated by geNorm in different tissues of peanut seeds. A lower average expression stability M value indicates more stable expression. The least stable genes are on the left, and the most stable genes are on the right.

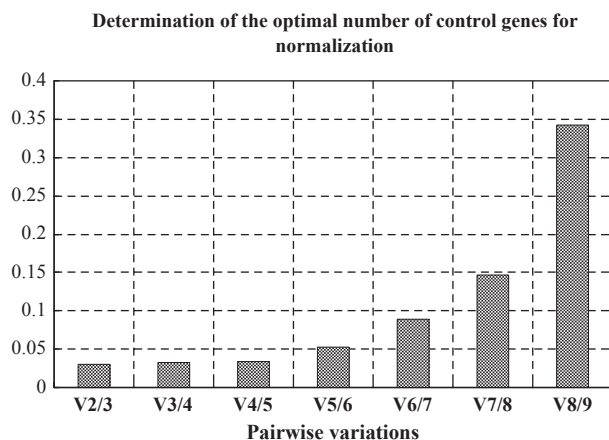


Figure 3 Determination of the optimal number of control genes for normalisation. Pairwise variations ($V_{n/n+1}$) were analysed between the normalisation factors NF_n and NF_{n+1} by geNorm software to determine the number of control genes required for accurate normalisation. V values below 0.15 indicate that no further gene is needed for the calculation of a reliable normalisation factor.

used in the experimental system. To identify the best reference gene in peanut seeds, qPCR assays were performed and the data were analysed by geNorm and Normfinder which were designed to rank candidate reference genes on the basis of their expression levels in the samples.

The ten candidate genes in the seed can be divided into three groups by Ct values analysis. The first group, which includes ubiquitin, polyubiquitin 10, elongation factor and ribosomal protein L24, is correlated with the synthesis and degradation of protein. Ubiquitin is a small protein that exists in all eukaryotic cells. It performs the myriad functions through conjugation to a large range of target proteins and regulating the degradation of specific proteins in a cell (Li & Ye, 2008). Elongation factor and ribosomal protein are involved in the synthesis of proteins. The complex synthesis and degradation mechanisms are regulated to maintain the homeostasis status of the mature peanut seed. The second group consists of members of cytoskeleton, such as tubulin and actin, which are constitutive expression genes. The third group includes malate dehydrogenase and phospholipase D, which participate in the cell metabolic process. In the mature seeds, the high expressions of these genes might be related to the seed vitality.

The results calculated by the two softwares suggest that the phospholipase D and actin are the stable reference genes. PLD, which is first discovered and cloned in plants, is the major family of enzymes that hydrolyses membrane phospholipids in plants. The PLD family in plants is comprised of multiple PLDs with

distinctively different biochemical, molecular and structural properties, which is more complex than that in the animal and fungal systems. Furthermore, PLD is rapidly regulated in plant tissues under various stress conditions, such as freezing injuries, ageing and biotic stresses (Li *et al.*, 2009). Although PLD is scored as the most stable gene in this article, it should be carefully validated for each particular experimental condition because of its expression variability. Actin, which is one of the main components of cellular cytoskeleton, is present in the cell in distinctly different morphological structures. Although actin cytoskeleton is highly dynamic, our results show that it remains steady in different tissues of peanut seed (the relatively stable state during its life history). Some researches on the reference genes in leguminosae plants also confirm the validity of actin as the reference gene. Jian *et al.* (2008) finds that ACT2/7 could be used as the reference gene in different tissues of soybean. Hu *et al.* (2009) presents that a combination of ACT11, UKN1 and UKN2 would be appropriate as the reference genes among different tissues of soybean cultivar Kennong18. Gutierrez *et al.* (2010) shows that ACT1 is regarded as one of the most suitable reference genes in faba bean. Thus, on the evidence, actin is concluded as the optimal reference gene for studying gene expression in the mature peanut seed.

The qPCR method will be further employed to identify the most stable reference genes under different experimental conditions in our future studies of peanut seed. Although actin is commonly used as the reference gene in many qPCR reaction systems, identifying new useful reference genes in the genome becomes possible and necessary, with the growing availability of plant genome sequences and functional genomic platforms.

Acknowledgments

The work was supported by Key Program of Excellent Talents Foundation of the Education Department of Anhui Province in China (2009SQZR112ZD) and Foundation of Anhui Science and Technology University (ZRC2008215). The work described has not been published before and its publication has been approved by all co-authors. The authors have no conflict of interest. We also appreciate comments and suggestions made by editors and anonymous reviewers.

References

- Andersen, C.L., Jensen, J.L. & Ørntoft, T.F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research*, **64**, 5245–5250.
- Azevedo, H., Lino-Neto, T. & Tavares, R.M. (2003). An improved method for high-quality RNA isolation from needles of adult maritime pine trees. *Plant Molecular Biology Reporter*, **21**, 333–338.

- Brandon, H.L., Javier, A.W., Tomokazu, K., Anhthu, Q.B., John, J.H. & Robert, B.G. (2007). Using genomics to study legume seed development. *Plant Physiology*, **144**, 562–574.
- Fink, T., Lund, P., Pilgaard, L., Rasmussen, J.G., Duroux, M. & Zachar, V. (2008). Instability of standard PCR reference genes in adipose-derived stem cells during propagation, differentiation and hypoxic exposure. *BMC Molecular Biology*, **9**, 98.
- Gallardo, K., Thompson, R. & Burstin, J. (2008). Reserve accumulation in legume seeds. *Comptes Rendus Biologies*, **331**, 755–762.
- Gutierrez, L., Mauriat, M., Pelloux, J., Bellini, C. & Van Wuytswinkel, O. (2008). Towards a systematic validation of references in real-time rt-PCR. *Plant Cell*, **20**, 1734–1735.
- Gutierrez, N., Giménez, M.J., Palomino, C. & Avila, C.M. (2010). Assessment of candidate reference genes for expression studies in *Vicia faba* L. by real-time quantitative PCR. *Molecular Breeding*, DOI: 10.1007/s11032-010-9456-7.
- Hu, R., Fan, C., Li, H., Zhang, Q. & Fu, Y.F. (2009). Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Molecular Biology*, **10**, 93.
- Huggett, J., Dheda, K., Bustin, S. & Zumla, A. (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes and Immunity*, **6**, 279–284.
- Jian, B., Liu, B., Bi, Y., Hou, W., Wu, C. & Han, T. (2008). Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Molecular Biology*, **9**, 59.
- Kang, I.H., Srivastava, P., Ozias-Akins, P. & Gallo, M. (2007). Temporal and spatial expression of the major allergens in developing and germinating peanut seed. *Plant Physiology*, **144**, 836–845.
- Li, W. & Ye, Y.H. (2008). Polyubiquitin chains: functions, structures, and mechanisms. *Cellular and Molecular Life Sciences*, **65**, 2397–2406.
- Li, M., Hong, Y. & Wang, X. (2009). Phospholipase D- and phosphatidic acid-mediated signaling in plants. *Biochimica et Biophysica Acta*, **1791**, 927–935.
- Maroufi, A., Van Bockstaele, E. & De Loose, M. (2010). Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR. *BMC Molecular Biology*, **11**, 15.
- Nolan, T., Hands, R.E. & Bustin, S.A. (2006). Quantification of mRNA using real-time RT-PCR. *Nature Protocols*, **1**, 1559–1582.
- Paolacci, A.R., Tanzarella, O.A., Porceddu, E. & Ciaffi, M. (2009). Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Molecular Biology*, **10**, 11.
- Shreffler, W.G., Castro, R.R., Kucuk, Z.Y. et al. (2006). The major glycoprotein allergen from *Arachis hypogaea*, Ara h 1, is a ligand of dendritic cell-specific ICAM-grabbing nonintegrin and acts as a Th2 adjuvant *in vitro*. *Journal of Immunology*, **177**, 3677–3685.
- Vandesompele, J., De Preter, K., Pattyn, F. et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, **3**, 34.1–34.11.
- Wong, M.L. & Medrano, J.F. (2005). Real-time PCR for mRNA quantitation. *BioTechniques*, **39**, 75–85.
- Yang, Y., Hou, S., Cui, G., Chen, S., Wei, J. & Huang, L. (2010). Characterization of reference genes for quantitative real-time PCR analysis in various tissues of *Salvia miltiorrhiza*. *Molecular Biology Reports*, **37**, 507–513.
- Ye, X., Zhang, L., Dong, H. et al. (2010). Validation of reference genes of grass carp *Ctenopharyngodon idellus* for the normalization of quantitative real-time PCR. *Biotechnological Letters*, **32**, 1031–1038.
- Yoo, W.G., Kim, T.I., Li, S. et al. (2009). Reference genes for quantitative analysis on *Clonorchis sinensis* gene expression by real-time PCR. *Parasitology Research*, **104**, 321–328.