

# MultiPriDe: automated batch development of quantitative real-time PCR primers

A. C. Ziesel<sup>1,2,\*</sup>, M. A. Chrenek<sup>1,2</sup> and P. W. Wong<sup>1</sup>

<sup>1</sup>Department of Ophthalmology, Emory University, B5500 Clinic B, 1365 Clifton Road NE, Atlanta, GA 30322, USA and <sup>2</sup>Department of Biological Sciences, University of Alberta, CW405 Biological Sciences Centre, Edmonton, Alberta T6G 2E9, Canada

Received January 1, 2008; Revised March 21, 2008; Accepted March 24, 2008

## ABSTRACT

**Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) is a commonly employed gene expression quantification technique. This requires the development of appropriately targeted oligonucleotide primers, which necessitates the identification of ideal amplicons, development of optimized oligonucleotide sequences under most favorable pre-determined reaction conditions, and management of the resultant target-oligonucleotide pair information for each gene to be studied. The Primer3 utility exists for development of oligonucleotide primers and fills that role effectively. However, the manual process of identifying target sites and individually generating primers is inefficient and prone to user-introduced error, especially when a large number of genes are to be examined. We have developed MultiPriDe (Multiple Primer Design), a Perl utility that accepts batch lists of Gene database identifiers, collects available intron and exon position data critical to qRT-PCR primer development, and supplies these sites as identified targets for the Primer3 utility. This automated 'gene to primer' procedure is coupled with a set of optimized hybridization conditions used by the Primer3 utility to maximize successful primer design. MultiPriDe and assembled repeat libraries are available upon request. Please direct requests to aziesel@emory.edu.**

## INTRODUCTION

Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) is a technique by which expression levels of individual genes can be quantified by measuring the abundance of their mRNA products relative to a standard

(such as 18S rRNA). It represents a robust means of gene expression quantification on its own, and is an ideal companion technique for confirmation or further inspection of results obtained by methods such as SAGE (1) and microarray analysis strategies (2). The basis of qRT-PCR is traditional RT-PCR, cycling temperatures to allow repeated steps of double-stranded nucleic acid melting, primer annealing, polymerase-driven synthesis of new double-stranded nucleic acid. In qRT-PCR, oligonucleotide primers are designed to span intronic sequences, the portions of the genome not represented in mature mRNA products. When these primers bracket an intronic splice site, a small product will be produced when the primers anneal to mRNA during the reaction. An easily recognizable larger product (or in the case of an extraordinarily long intron, no product at all) would be produced should any contaminant genomic DNA persist in the sample used for the reaction (Figure 1). One such detection stratagem, referred to here as SYBR Green, employs a chemical agent that binds to synthesized double-stranded nucleic acid products generated during the course of the qRT-PCR (3). This associated molecule then fluoresces at 520 nm when excited with a wavelength of 497 nm; this emitted wavelength is detected by the qRT-PCR apparatus and recorded over the course of the reaction. Combining information regarding increased fluorescent emission, temperature cycling, initial amounts of RNA used in the qRT-PCR reaction and fluorescence emission of an included standard allows a calculation to be made determining the levels of the targeted RNA species, quantifying its expression in the original sample.

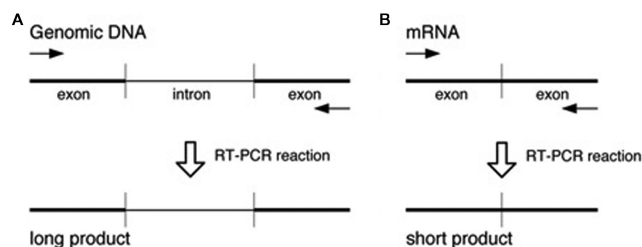
Determining appropriate intron-spanning target sites and developing oligonucleotide primers against them involves several steps, starting with identifying target genes, determining their genomic sequence, their mRNA coding sequence (CDS) and the location of intervening intronic sequences. Oligonucleotide primer development can be done by manually identifying ideal primer sites, but the majority of primer design is currently conducted

\*To whom correspondence should be addressed. Tel: +1 404 778 5531; Fax: +1 404 778 4411; Email: aziesel@emory.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© 2008 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

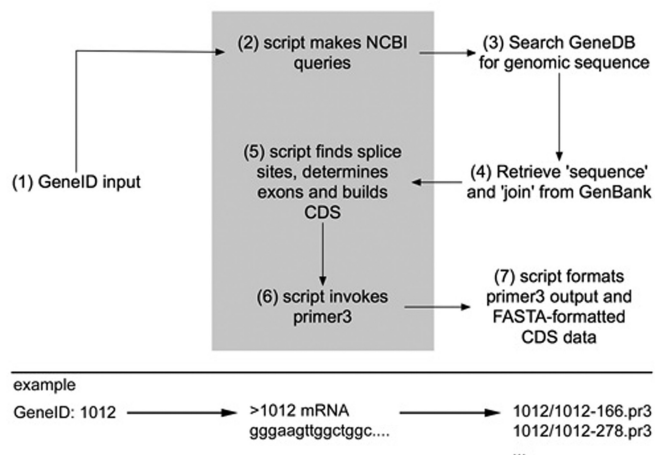


**Figure 1.** RT-PCR primers are chosen in adjacent exons, with the predicted product spanning an intronic splice site. In this way, should any genomic DNA persist in the RNA preparation, any amplification from genomic DNA will produce a long product (A), containing both the targeted sequence and intervening intron. Amplification from mRNA (B) will produce only a short product as the intronic sequence has been spliced out of the mRNA.

with the aid of computational tools, such as MuPlex (4), PriFi (5), the widely adopted Primer3 utility (6) or its alternate interface, Primer3Plus (7). Primer3 is freely available as a standalone package and is available online through a simple web interface. However, for those users disinclined to run Primer3 locally through a command line interface, designing primers against identified targets for each chosen gene individually through a web interface is potentially time-consuming and requires careful management of retrieved information. In addition to the management of these aforementioned data, there is also the issue of developing targets and primers, and selecting reaction conditions that will be most optimally successful. The difficulty of coordinating these data increases with the size of the qRT-PCR project undertaken. We elected to address these issues of target site identification, primer development and optimization, and data organization to ameliorate the process for molecular biology laboratories. To this end, we have developed a set of optimized qRT-PCR primer design conditions and a script that incorporates these conditions into an automated design procedure.

## METHODS

MultiPriDe is designed to streamline and standardize the process starting with target gene identification through to primer design. After user identification of targets, MultiPriDe handles the entire procedure independent of additional user input, and produces organized, easily accessible primer design information. Moreover, as all primer sets are designed with the same optimized reaction conditions, all qRT-PCR reactions should be performable simultaneously or with the same settings from run to run. The steps that MultiPriDe follows are illustrated in Figure 2: starting with initial input, the script identifies and retrieves the required information from NCBI resources (8). It then determines intronic splice sites, which are ideal qRT-PCR target sequences. These ideal targets are then directed to a local install of the Primer3 package that uses our empirically optimized hybridization conditions to develop sets of primers amplifying each identified target. The identified CDS, intronic splice sites and primer design results are available to the user upon completion. We have chosen to employ a locally installed



**Figure 2.** The process MultiPriDe follows, starting with inputted GeneIDs and concluding with identified CDS and primers for each intronic splice site. The processes described in the central gray box are those that are performed by the user's computer.

version of Primer3 for use with MultiPriDe chiefly to allow users to exploit the power of their own laboratory computers rather than further taxing existing publicly available Primer3 servers.

We developed a specific set of parameters for oligonucleotides designed by MultiPriDe, specifically deviating at key points from those default parameters specified by the Primer3 package. First, we chose an annealing temperature range of 55.0°C (minimum), 57.5° (optimum) and 60.0°C (maximum) instead of the default 60.0, 63.0 and 65.0°C range specified in the Primer3 defaults. We chose this range because it typically led to acceptable qRT-PCR reactions when working with mouse and rat RNA samples, and because oligonucleotides designed with our temperature range typically exhibited a narrower range of ideal primer concentrations in reaction (unpublished results). We also adjusted the maximum target oligonucleotide size to 25 nt from 27 nt to compensate for the reduction in annealing temperature. We altered the default settings for maximum self-complementarity and maximum 3' self-complementarity from values of 8.00 and 3.00, respectively, to 5.00 and 2.00, respectively, to reduce primer dimer formation. Finally, we use a specialized repeat library to develop the qRT-PCR primer sets described here. This repeat library is a merger of two of the base libraries offered with the Primer3 package (the 'rodent and simple' and 'human' libraries). We use this merged library for ease of primer development in our own cross-species experiments and it is not intended to confer any special benefit with the exception of simplicity for multi-species batch primer design. Use of our merged library is not necessary with MultiPriDe.

MultiPriDe was written entirely in Perl, and uses the LWP module for certain functions. Local installations of standard Perl 5.8.0, the LWP module [available through the Comprehensive Perl Archive Network (CPAN)] and the Primer3 package (available at [http://www-genome.wi.mit.edu/genome\\_software/other/primer3.html](http://www-genome.wi.mit.edu/genome_software/other/primer3.html)) are required for functionality. The user supplies an

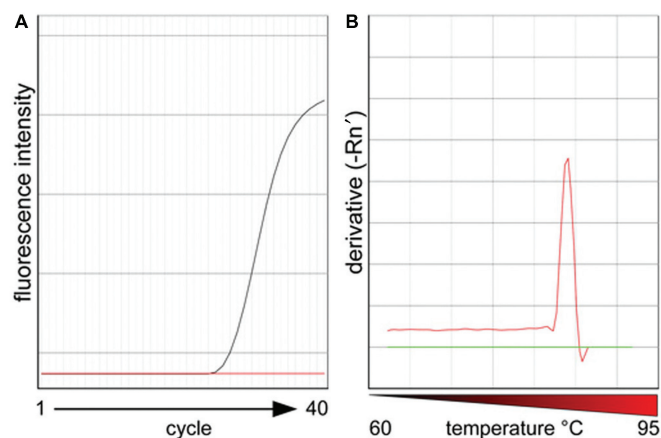
individual or a batch list of GeneID numbers via command line interface. MultiPriDe first queries NCBI, accessing the Entrez Gene records for these numbers. The script then collects the genomic sequence entry and first join data entry for that record. This information is used to identify exons and intronic splice sites in the genomic sequence, and prints the compiled CDS to the screen, along with CDS nucleotide positions for intronic splice sites. These splice sites will be specified as targets for Primer3 oligonucleotide design in subsequent steps. In the case of genes that are composed of a single exon, no target is specified; the entire monoexonic CDS is a potential target site for oligonucleotide design. Next, MultiPriDe invokes a locally installed version of Primer3, and using our hybridization conditions coupled with the intronic splice site positions, develops a maximum of 100 primer pairs spanning each splice site in the CDS. MultiPriDe creates directories named for each GeneID queried, and writes into these the Primer3 output files named according to the convention *GeneID\_splicesite.pr3*. These .pr3 files are readable in any text editor and are equivalent to the output a user may receive when performing an equivalent analysis using a Primer3 server. Those GeneIDs that lack any of the requisite information (an entry, genomic sequence or join data) return an error message describing the nature of the error and an output file containing FASTA-formatted cDNA sequence if available.

An example of our own use of MultiPriDe follows. Cell cultures were prepared of human WERI-Rb1 (gift of Dr J. Boatright, Emory University, Atlanta, Georgia) or rat NRK-52E cells [American Type Culture Collection (ATCC), Manassas, VA, USA] for subsequent RNA extraction. Dishes (60 mm) of cells were grown to monolayer, media was removed and 1 ml of Trizol (Invitrogen, Carlsbad, CA, USA) was added to each dish. The dishes were shaken at 80 r.p.m. for 2 min, then 900  $\mu$ l of lysed cells were removed from the dish. For mouse RNA, a mixture of RNA extracted from multiple tissues was used, including brain, heart, mammary gland, muscle, liver, lung, ovaries, spleen, submandibular gland and uterus. RNA was prepared from both cultured cells and mouse tissue using Trizol according to the manufacturer's recommended protocol with the inclusion of an additional chloroform wash following organic phase separation, quantified using fluorescence methods and then diluted to 100 ng/ $\mu$ l. RNA samples weighing 5  $\mu$ g each were treated with Turbo DNA-free DNase (Ambion, Austin, TX, USA) to remove any contaminating DNA. DNase-treated RNA was then purified and reconstituted using Qiagen RNeasy MinElute columns (Qiagen, Valencia, CA, USA) and RNA was eluted in 100  $\mu$ l of nuclease-free water (2 ng/ $\mu$ l).

Oligonucleotides designed using MultiPriDe were ordered from (MWG Biotech, High Point, NC, USA) as high-purity salt free in 0.01  $\mu$ mol size. Each oligonucleotide was dissolved in nuclease-free water to 100  $\mu$ M. Two forward and two reverse primers from the list of primers identified from our MultiPriDe analysis were ordered for a single splice site for each gene, generating four unique pairs of primers for each gene. For a given multiexonic gene, primer pairs were chosen from MultiPriDe's

generated results with priority placed on (i) splice site proximity to 5' end of transcript, (ii) splice site at least 100 bp from either end of transcript and (iii) at least two pairs of primers were generated for the specific chosen splice site. Individual oligonucleotides were aliquoted into pairs at 1  $\mu$ M concentration (per primer) with the pairs set up as follows: pair 1—F1/R1; pair 2—F1/R2; pair 3—F2/R1; pair 4—F2/R2. This concentration was considered a 5  $\times$  primer pair stock for subsequent testing of primers at an initial 200 nM concentration of each primer in qRT-PCR. For reactions that showed a single dissociation peak in the reactions when template RNA was added but showed primer dimers in the absence of template RNA, primers were retested at 100 nM concentration. For reactions that showed no amplification of fluorescent signal in the presence of template the primers were retested at 400 nM.

qRT-PCRs were performed using a QuantiTect SYBR green one-step RT-PCR kit (Qiagen). Reactions (25  $\mu$ l) were set up with 5 ng of template RNA or an equivalent volume of nuclease-free water as a control and 100, 200 or 400 nM concentration of each of the primers in a pair to be tested. qRT-PCRs were performed in an ABI-7500 real-time thermocycler (Applied Biosystems, Foster City, CA, USA). The reverse-transcriptase reaction began with 15 min at 95°C to activate the polymerase, 50°C for 30 min and then followed with 40 cycles of 95°C for 15 s (strand dissociation), 55 to 60°C for 30 s (primer annealing,



**Figure 3.** A sample RT-PCR using MultiPriDe-developed primers. (A) Demonstrates an amplification curve for these primers, which is a plot of fluorescence intensity on the ordinate axis and the number of cycles on the abscissa axis. The increase in fluorescence is the result of free SYBR Green binding to double-stranded DNA (initially at a very low concentration), which increases roughly 2-fold per amplification cycle until primers and other reagents begin to run out. The fluorescence reaches a maximum at high cycle numbers. A smooth exponential increase in the initial stages of the PCR suggests the amplification of only one product, consistent with the primers being specific. (B) Shows a representative dissociation curve for this reaction, plotting change in fluorescence against change in temperature. As double-stranded DNA product is melted with increasing temperature, it dissociates into two single strands and releases complexed SYBR green. The release of SYBR green results in diminished fluorescence. The central sharp peak seen indicates the homogenous melting behavior of the double-stranded DNA produced by the qRT-PCR, strongly suggesting the production of a single DNA species.



**A**

```
perl — 120x24
a049003:~/remote/multipride alison$ perl multipride.pl
(January 2008) MultiPriDe: a script to design RT primers for individual exons belonging to a specified set of genes.
Please ensure your Primer3 install and repeat library is in the same directory as this script before using.

Enter GeneID numbers separated by commas: 26270
```

**B**

```
perl — 120x24
a049003:~/remote/multipride alison$ perl multipride.pl
(January 2008) MultiPriDe: a script to design RT primers for individual exons belonging to a specified set of genes.
Please ensure your Primer3 install and repeat library is in the same directory as this script before using.

Enter GeneID numbers separated by commas: 26270
>26270 mRNA; assembled from GenBank mRNA NC_000001.9 join data
gccccggttcgggagcttcggccctgcgtaggagcggggtgcaggtgtgggtgctgagccgcccgcctggagggggagacagcttcaggacacgcagggccgcagcgagggccggggccc
tggggatcccgagccatggatgctcccaactccaagcagccctggacagcattaacgagctgcccagagaacatcctgctggagctgttcacgcacgtgcccggcccgagctgctgctg
aactgccgctggtctgcagcctctggcgggacctcatgcacctcatgacctctggaacgcgaagtgcctgcagaggggcttcacccaaggactgggacagcccgctggccgactgg
aaaattcttacttctacggagcctgcataaggaacctcctgcgcgaacccgctgctgaagaggatattgttgcattggcaaatgatttcaatgggtggggaccgctgggaaggtggagagc
ctcctggagcccgagacagattttctgaacccaaagtcagaagattttgtcacatcctacgaatgtgcctcaagtcaccagctggtagaccttgtagccgagggctactgggag
gagctactagacacattccggccggacatcgtggttaaggactggttggctgcagagccgactgtggtgcacactaccaactcaagtgacgtggcctggctgactacttctgtgtg
gcctccttcgagccccacctgtgacctccaacagtggaacatgccacatggacagaggtctctacaccttctcagactacccccggggtgtccgctacatctcttcagcatggg
ggcaggggacacccagtagtgggcaggctggtatgggcccagagtcaccaacagcagcattgtcgctagccccaagatgaccaggaaccaggccctcctccgaggtcagcctgggcagaag
catggacagggagggctgcccacatgcctaccgagctgttgcagattttctgacagctgtccatcctgtgtctgggtcagccagaggttctccaggcaggagctgagcatggggt
gggcagtgagggtccctgtaccagcagctcctgcccgggttaacacctaccagctgtggttaacttactgtcacatagctctgacgtttgtgttaataatgttttcaggccgggactg
tggctcacgcctgtaatccagcactttgggagaccgagcgaggtggatcacgaggtcaggagatagagacatcctggccaacacgggtgaacacctgtctctactaaaaatacaaaaaa
ttagccgggctggtggcgggctgctgtagtcccagctactcgggagggctgtagcagaagaatggcgtgaacccgggaaggcagagcttgtagtgagccgagatcacgccactgcactcca
gcctgggtgacagagcagactcggctcataaataataataataataataataataataataataatgttttcagtaaaa

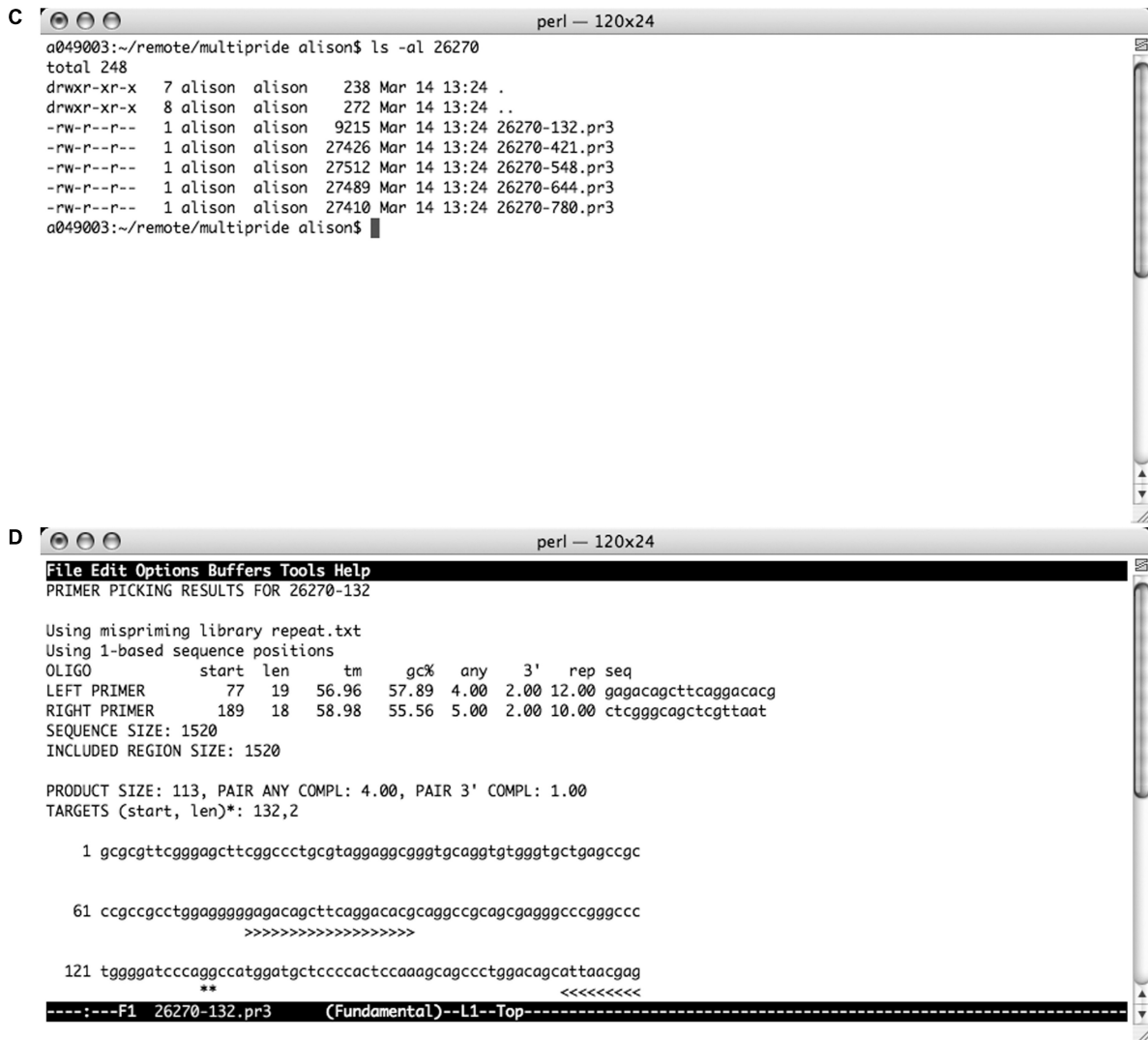
Join positions, first base of each exon: 132 421 548 644 780 1520
```

**Figure 4.** The MultiPriDe user experience. (A) The program is invoked using Perl, and asks for GeneID numbers as input. (B) MultiPriDe prints a FASTA-formatted CDS to the screen, along with the locations of splice sites. (C) The directory produced for a given GeneID, containing the primer design files for each splice site. (D) The text file Primer3 output for an individual splice site.

temperature dependent on the primer pair), and 72°C for 40 s (DNA synthesis). Fluorescence data was collected during the extension step of each cycle. Annealing temperature was 55°C for primers designed to have a melting temperature of 57.5°C, and 60°C for primers designed to have a melting temperature of 63°C. Amplification curves were generated using the supplied Applied Biosystems SDS v1.2 software. Following qRT-PCR, dissociation curves were generated for each sample. The qRT-PCR products were dissociated at 95°C for 5 min, then reassociated at 60°C. Incremental steps of +0.1°C were done with fluorescence data collected at each step and the derivative of the change in fluorescence was plotted by the SDS software.

Amplification plots with and without template RNA were examined for each primer pair. Those that showed high efficiency amplification of signal with template added and no amplification of signal in the absence of template were considered good amplification plots (Figure 3).

A primer pair was considered to have failed if the amplification plot showed amplification of signal from the reaction that contained no template or if the efficiency of the amplification was weak in the presence of template. Dissociation curves with and without template RNA were examined for each primer pair. Those that showed a single peak for the reaction containing template and no peaks for the reaction containing no template were considered to be good dissociation curves; we consider our threshold for success to be quite stringent, and may be more stringent than is required for some purposes (Figure 3). Those that showed multiple peaks for the reaction containing template or any peak including low melting temperature peaks (primer dimers) for the no template reaction were considered to be bad dissociation curves. Each primer pair was graded as passed if both the amplification plot and the dissociation curves were good, or failed if either the amplification plot or the dissociation curve was bad, according to our aforescribed criteria.



**Figure 4. Continued.**

## RESULTS AND DISCUSSION

Figure 4 describes the process involved in using MultiPriDe. Figure 4A shows invocation of the script and its request for GeneID numbers. MultiPriDe then displays the gene's CDS with a FASTA-formatted header, along with the nucleotide positions of the splice sites relative to the CDS. As described earlier, MultiPriDe then makes a directory for each GeneID and places the Primer3 output files in that folder, as shown in Figure 4C. The output files can be read using any text editor; in Figure 4D the Emacs terminal utility is used, showing the familiar Primer3 output format.

A comparison between traditional, manual primer design and primer design conducted using MultiPriDe was undertaken. To manually design a single pair of primers for a single splice site of a single gene from our gene set (human cadherin 13, geneID 1012) we needed 12 min, 45 s to retrieve the genomic sequence, identify and confirm a single splice site, run Primer3 for that single location and save all relevant files. The same gene was

submitted to our script; 1 min 26 s was required to identify and develop 100 pairs of primers for each of this gene's 13 splice sites. This is  $\sim 7$  s per splice site, or roughly 110 times faster than the manual identification and primer development. MultiPriDe designed primers for all splice sites of the same 270 genes in slightly  $<2$  h, running on a 2.0 GHz dual processor Power Mac G5 (Apple Computer, Cupertino, CA, USA), factory condition with the exception of 6 GB of RAM, using a 100 Mbit/s internet connection for retrieving information from NCBI. Of the 270 genes tested, 19 lacked sequence associated with their GeneID, and therefore primers could not be designed. Experimentally, our primer selection and optimized qRT-PCR conditions led to an immediate success rate of 75% among the 251 genes that we tested experimentally; that is, for each targeted gene, 75% of the first batch of selected primers generated acceptable qRT-PCR results on the first attempt. Table 1 describes the breakdown of first round success for each of the three species considered. Rachlin and colleagues (4) describe a 39% chance for

**Table 1.** Summary of results for qRT-PCR amplifications using the first two selected pairs of MultiPriDe-developed primers

Species	Successful qRT-PCR reactions at (200 nM)	Total successful qRT-PCR reactions	Total genes analyzed
Human (%)	39 (53)	55 (74)	74
Mouse (%)	61 (75)	71 (88)	81
Rat (%)	43 (45)	63 (66)	96
All species	143 (57)	189 (75)	251

'Successful qRT-PCR reactions at (200 nM)' includes only those reactions that were successful at 200 nM primer concentration. 'Total successful qRT-PCR reactions' includes those successful at 200 nM, as well as those that were not successful at 200 nM, but were subsequently successful when tested at 100 nM or 400 nM. Percentages are calculated based on total genes analyzed for a given species.

primers designed using their MuPlex utility (which relies on default parameters for oligonucleotide and salt concentrations identical to Primer3's default parameters) occurring in more than 10 genomic locations. Should multiple locations for a given primer occur across multiple exons, that primer would give confounding results when used in qRT-PCR. A number of failures are also due to double peaks that could result from amplification across an alternative splice site; this has not been quantified. We feel these phenomena may account for much of our first-round failures.

qRT-PCR primer design for experiments of even modest scope can quickly become complicated and time-consuming. Manual identification of ideal target sites is tedious and prone to user-introduced error; individual queries directed against a public Primer3 server are equally tedious. We do not attempt to quantify error rate for manual management of sequence and primer design data in this study. Earlier findings described an error rate of 30–33% for multiple-step tasks involving text editors, even amongst highly experienced users (9). MultiPriDe greatly reduces the possibility of user-introduced error as all steps short of the initial GeneID entry are managed entirely by the script. Additionally, the individual files a user may generate when undertaking these tasks manually quickly accumulate and may further

impact organization. MultiPriDe was designed with the interests of time and organization in mind. We are confident that MultiPriDe will be of use to molecular biologists undertaking qRT-PCR experiments of any magnitude.

## ACKNOWLEDGEMENTS

The authors would like to thank Dr J. Boatright and Dr J. Nickerson for their constructive commentary on this project. The authors gratefully acknowledge the support of the NSERC, RPB, NIH (NEI) (P30-EY006360) and the Knights Templar of Georgia. Funding to pay the Open Access publication charges for this article was provided by the Knights Templar of Georgia.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Kuo,B., Chen,Y., Bohacec,S., Johansson,O., Wasserman,W. and Simpson,E. (2006) SAGE2Splice: unmapped SAGE tags reveal novel splice junctions. *PLoS Comput. Biol.*, **2**, e34.
2. Li,R. and Li,C. (2006) Butyrate induces profound changes in gene expression related to multiple signal pathways in bovine kidney epithelial cells. *BMC Genomics*, **14**, 234.
3. Qiagen Inc. (2005) QuantiTect SYBR Green PCR Handbook. Qiagen Inc., Valencia, CA, November 2005 release.
4. Rachlin,J., Ding,C., Cantor,C. and Kasif,S. (2005) MuPlex: multi-objective multiplex PCR assay design. *Nucleic Acids Res.*, **33**, W544–W547.
5. Fredslund,J., Schauser,L., Madsen,L., Sandal,N. and Stougaard,J. (2005) PriFi: using a multiple alignment of related sequences to find primers for amplification of homologs. *Nucleic Acids Res.*, **33**, W516–W520.
6. Rozen,S. and Skaletsky,H. (2000) Primer3 on the WWW for general users and for biologist programmers. In Krawetz,S. and Misener,S. (eds), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, Humana Press, Totowa, NJ, pp. 365–386.
7. Untergasser,A., Nijveen,H., Rao,X., Bisseling,T., Geurts,R. and Leunissen,J.A.M. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.*, **35**, W71–W74.
8. Maglott,D., Ostell,J., Pruitt,K. and Tatusova,T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **33**, D54–D58.
9. Card,S., Moran,T. and Newell,A. (1983) *The Psychology of Human-Computer Interaction*. Erlbaum Associates, Hillsdale, NJ.