Virtual PCR

M. Lexa, J. Horak and B. Brzobohaty

Laboratory of Plant Molecular Physiology, Masaryk University Brno, Faculty of Science, Kotlarska 2, 611 37, Brno, Czech Republic and Institute of Biophysics AS CR, Kralovopolska 135, 612 65, Brno, Czech Republic

ABSTRACT
Summary: We present an algorithm that uses public sequence data to predict PCR products. The algorithm is implemented as a CGI script. Output is compared to real-world PCR.

Availability: Perl code and instructions for installation are freely available over the internet at http://www.sci.muni.cz/LMFR/vpcr.html

Contact: m-lexa@sci.muni.cz

INTRODUCTION
PCR has become an indispensable tool in molecular biology. Utilizing oligonucleotide primers and DNA polymerase, it amplifies DNA up to many thousand bases in length (Bej et al., 1991). Success of a particular application mostly depends on the choice of proper primers. Optimally, they will not form dimers or hairpins. Most importantly, they will anneal to DNA sequences of interest, but not to other sequences. A quick similarity search will reveal possible complementary sequences for a primer. However, it will not tell whether one can expect a PCR product because of another primer annealing to the template in close proximity. The abundance of genomic sequences in public databases makes it now possible to simulate amplification results. We describe a virtual PCR software tool that uses public DNA sequence databases to predict PCR products for arbitrary primer pairs.

ALGORITHM AND IMPLEMENTATION
The virtual PCR program (VPCR) exists as a CGI script written in Perl, that after installation can be accessed using a WWW browser. It processes user-given primers and obtains BLAST search results (Altschul et al., 1990) with these primers to identify sequences in public databases that are complementary to any two of them. Finally, it prints out potential PCR products.

In more detail, the process of using the script starts by filling a form that accepts user input. The only strictly required inputs are the primer sequences to be used. By default, the algorithm searches the entire GenBank database for matches and after approximately 60s displays identified PCR products. Most often, the user will find it useful to restrict the search to a particular species. Changing the E-value limit for the BLAST search will change the level of homology required for a successful match, thereby increasing or decreasing the specificity of the VPCR. This is somewhat similar to changing annealing temperature in a real-world reaction. Appropriate search mode and sequence database can be chosen in the input form, with detailed explanation of those given on NCBI BLAST and our webpages.

EXPERIMENTAL VERIFICATION
To demonstrate the potential of the VPCR algorithm, we show its use and how it compares to data from real-world PCR. We have chosen to show two envisioned uses: (i) evaluation of primers to be used in amplification from genomic DNA; and (ii) identification of PCR products with primers amplifying a gene family.

PCR reactions used to evaluate the algorithm were carried out by standard methods using Arabidopsis DNA and these primers:

ARR5a = GTTGATTCTCTCTATCTCTCTCACG
ARR5b = CACACCACCATTTTACATATCTC
ARR7a = GTTGGTGAGGTCATGAGGATGGAGATTC
ARR7b = GTTTTGCTAAGGTCTTGGCCTCTATACAT
GEN1 = CATGTTCTTGCYGTYGATGAY AGT
GEN2 = CCAGTCATKCCAGGCATWSAG
GEN3 = ATAAARAATCTCATCACGCWCCCTC

Single gene-specific primer evaluation
Two pairs of primers previously designed to amplify a region of ARR5 and ARR7 genes were used to test the performance of the algorithm. The resulting VPCR products are displayed and visually compared with a gel of real PCR products in Figure 1. Simulated bands which correspond to experimental bands are marked by arrows in Figure 1A. The algorithm has correctly identified the ARR7 product obtained in a parallel PCR reaction. It found only one additional product of low
Fig. 1. The VPCR algorithm and real-world PCR. (A) Graphical presentation of VPCR output for the following primer sets: ARR7a, b (lane 2), ARR5a, b (lane 3), GEN1,2 (lane 4) and Gen1,3 (lane 5). The width of the bands was chosen to represent the E-value of the BLAST search. Wider bands represent higher homology between primers and template. Arrows show bands present on gels of actual PCR products. (B) Graphical presentation of simulation results as in (A). Only high-specificity bands are shown. (C) Agarose gels of corresponding PCR reactions. Primers as in (A) and (B). Arabidopsis genomic DNA was used as template.

specificity. This product could not be seen on the gel (Figure 1C). With ARR5 primer pair, the algorithm identified the product seen in real PCR. However, a number of additional PCR products, most of which were much longer and had lower specificity values, were found (Figure 1A). Figure 1B shows only simulated products of high specificity. Comparison with the bands given in Figure 1C shows that eliminating the low-specificity bands improves the match between simulation and experiment.

Gene family-specific primer evaluation

A set of degenerate primers previously designed to amplify a conserved region of all ARR genes was used to further test the performance of the algorithm. These were used as pairs GEN1,2 and GEN1,3. The resulting VPCR products are contrasted to real PCR products in Figure 1. Simulated bands which correspond to real PCR products seen on a gel are marked by arrows in Figure 1A. While the algorithm has correctly identified only a small fraction of ARR products, it identified additional PCR products, most of which were longer, low-specificity sequences (Figure 1A).

DISCUSSION

We have constructed and tested a VPCR algorithm based on the BLAST local alignment search of GenBank non-redundant sequences. The corresponding perl CGI script is available on the Internet. To our knowledge, this is a novel resource useful in PCR primer design to evaluate primer specificity. Additionally, people obtaining multiple PCR products may consult VPCR to identify them.

Currently the script utilizes BLAST searches on the NCBI server. However, it is the policy of NCBI to protect their server and prevent frequent queries originating from a single installation. Therefore, we are unable to run the VPCR script for general public from our web pages. Instead, readers are encouraged to download and install their private copy on their own.

In principle, test results corresponded with expectations, but specificity of the prediction must be improved before wider use. Using VPCR in its present form, one should limit output to high-specificity products by using E-value < 10. We see the following areas of improvement to increase the predictive ability of VPCR:

(i) using databases that cover the whole genome of a given organism;
(ii) identification of PCR products across boundaries of GenBank entries;
(iii) replacement of BLAST routines with calculations of primer binding; and
(iv) correcting for dimer and hairpin formation.

With these improvements the algorithm could become a standard component of primer design software.

ACKNOWLEDGEMENTS

This work was supported by the Grant Agency of the Czech Republic (204/99/D024; Matej Lexa; software development) and the Czech Ministry of Education (J07/98:143100008; Bretislav Brzobohaty; real-world PCR). We wish to thank Dr Hana Konecna for primer synthesis and Prof John M. Cheeseman for reading the manuscript.

REFERENCES