PCRTiler: automated design of tiled and specific PCR primer pairs

Alain L. Gervais*, Maud Marques and Luc Gaudreau

Département de Biologie, Université de Sherbrooke, 2500 boul. de l'Université, Sherbrooke, Qc, J1K 2R1, Canada

ABSTRACT

Efficiency and specificity of PCR amplification is dependent on several parameters, such as amplicon length, as well as hybridization specificity and melting temperature of primer oligonucleotides. Primer design is thus of critical importance for the success of PCR experiments, but can be a time-consuming and repetitive task, for example when large genomic regions are to be scanned for the presence of a protein of interest by chromatin immunoprecipitation experiments. We present here a webserver that allows the automated design of tiled primer pairs for any number of genomic loci. PCRTiler splits the target DNA sequences into smaller regions, and identifies candidate primers for each sub-region by running the well-known program Primer3 followed by the elimination of primers with a high cross-hybridization potential via BLAST. Tiling density and primer characteristics are specified by the user via a simple and user-friendly interface. The webserver can be accessed at http://pcrtiler.alaingervais.org:8080/PCRTiler. Additionally, users may download a standalone Java-based implementation of this software. Experimental validation of PCRTiler has demonstrated that it produces correct results. We have tiled a region of the human genome, in which 96 of 123 primer pairs worked in the first attempt, and 105 of 123 (85%) could be made to work by optimizing the conditions of the PCR assay.

INTRODUCTION

The selection of candidate primer pairs for PCR experiments can be a time-consuming process. When designing a primer pair, oligonucleotides should be chosen with similar melting temperatures (T_m). This is to prevent non-specific hybridization of the primer with the higher T_m, since the highest possible annealing temperature is dependent on the lowest T_m of the primer pair. Additional concerns, such as primer GC content, length, 3′ stability, possible primer dimers and secondary structures must be taken into account. Primer specificity is particularly important in quantitative PCR (qPCR) experiments, which measure a fluorescence value proportional to the total amount of amplified DNA material in a sample. Quantitative analysis is only possible if a single specific amplicon is produced per target genome.

When designing multiple primer pairs, additional constraints must be satisfied. In order to maximize the number of PCR assays that can be conducted at the same time, primer pairs must be designed so that the same experimental conditions are appropriate for all pairs.

Typically, a biologist will design a primer pair by running the Primer3 (1) program, which generates multiple primer pair candidates that satisfy the previously mentioned constraints, and then manually use BLAST (2,3) to ensure that each primer will only hybridize to one locus in the target genome. This process may have to be repeated multiple times in an error-prone routine that involves much copy/pasting. A webserver maintained by the NCBI, called Primer-BLAST, allows the design of a single primer pair at a time. Users wanting to design multiple primer pairs still have to do it manually.

THE PCRTILER WEBSERVER

We have developed the PCRTiler webserver to automate the design of multiple specific primer pairs covering one or multiple genomic loci. Overlapping primer pairs and multiple input sequences are supported by the webserver. PCRTiler handles all aspects of the selection of candidate primer pairs using Primer3, and implements the specificity check using BLAST. An overview of the primer pair design process is given in Figure 1. Other webservers use Primer3 to design primer pairs. For example, MutScreener (4) specializes in the design of PCR primers to be used in sequencing experiments. BatchPrimer3 (5) specializes in designing primers flanking microsatellites or near single

*To whom correspondence should be addressed. Tel: +1 819 821 8000; Fax: +1 819 821 8049; Email: alain.gervais@USherbrooke.ca
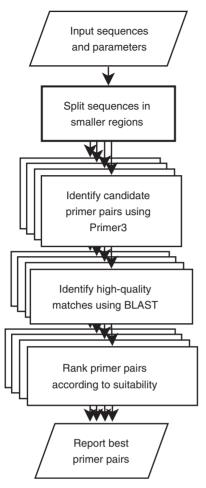
Figure 1. Overview of the PCRTiler oligonucleotide design process. The input sequences are first split into smaller regions according to the parameters defined by the user. A primer pair targeting each of these sub-regions will be designed. PCRTiler maximizes utilization of server resources by designing as many primer pairs in parallel as the server has processors. Candidate primer pairs are identified using Primer3, and BLAST is used to exclude primers with a high cross-hybridization potential (see text). A report is generated when all design tasks are finished.

nucleotide polymorphisms. To our knowledge, PCRTiler is the only webserver allowing the batch design of tiled and specific primer pairs.

Required input and output provided

PCRTiler requires primarily the target DNA sequences, the name of the corresponding organism and the tiling parameter. This tiling parameter can be specified as either a number of total primer pairs to design for each region of interest, or as a tiling distance. The user can alter the default ranges of allowed primer melting temperatures and amplicon lengths. Depending on the tiling parameter, the target sequence is split into subregions, for each of which Primer3 is invoked to suggest candidate primer pairs. By default, a thousand primer pairs are requested from Primer3, but users can override this setting. BLAST is then used to identify potential hybridization sites in the genome of the target organism. A description of the

parameters involved in evaluating primer hybridization sites is given in the next section. During processing, the user is presented with a self-refreshing progress report webpage, until the server provides a list of the best primer pairs (downloadable in CSV or TXT format) and a visual representation of their position on the original sequence. Alternative primer pairs and the raw Primer3 and BLAST results are also provided for inspection. Figure 2 is a screenshot of a typical PCRTiler run used to design primer pairs on a region about 1.5 kb wide of *Mycobacterium tuberculosis*. User-specified parameters were: distance between primer pairs (200 bp), amplicon length (100–150 bp) and primer T_m (60°C and 63°C). This resulted in an output of seven suggested primer pairs.

In order to help users to quickly grasp how to use the webserver and start submitting their own tiling requests, we have included buttons on the main page to automatically fill out the request form with four different sets of test inputs. Pre-computed results of these demonstration inputs are immediately available.

Selection of the best candidate primer pair

All matches for both oligonucleotides of each candidate primer pair are mapped to genome coordinates using BLAST. Not all BLAST hits for an oligonucleotide sequence are considered as potential hybridization sites. To evaluate primer cross-hybridization potential, we compute the total number of mismatches of the primer sequence with each BLAST hit. Similarly, we compute the number of mismatches within 5 bp of the 3'-end of the primer. By default, sites with four or more total mismatches, including two mismatches at the primer 3'-end are not considered potential cross-hybridization sites. These two parameters can be modified by the user.

Using the list of potential hybridization sites of both oligonucleotides of a primer pair within the whole genome of the selected organism, the number of possible amplicons is determined, assuming that amplification between primer sites with a distance of <3000 bp and appropriate strandedness is possible. Amplified fragments resulting from close mispriming of a single primer are also detected. This arbitrary distance threshold is about 15 times greater than the usual amplicon size of a typical qPCR assay (100–200 bp). Amplification of such large fragments is highly unlikely. Users can modify this amplification distance threshold parameter.

Candidate primer pairs are ranked according to the computed specificity metrics. Briefly, the score of a primer pair is inversely proportional to the number of possible amplicons and inversely proportional to the number of hybridization sites for each primer of the pair. The scoring function gives a lot more weight to the number of possible amplicons, since a primer pair with multiple amplicons should never be used. A small bonus proportional to the total number of mismatches to the most similar unintended hybridization site is used to discriminate primer pairs with an equal number of

amplicons and cross-hybridization sites. The exact scoring formula used is:

$$\text{score} = \cfrac{1}{(10\,000 \times (A_c - 1)) + (F_h - 1) + (R_h - 1)} + \cfrac{F_m + R_m}{2 \times (F_l + R_l)} \tag{1}$$

where A_c is the amplicon count for that primer pair, F_h is the number of hybridization sites for the forward primer, F_m is the number of total mismatches with the most similar unintended hybridization site in the genome for the forward primer, F_l is the length of the forward primer, and all R_x variables are the equivalent metrics for the reverse primer.

Genomes supported

PCRTiler includes a mechanism to synchronize its genomes with GenBank (6). The website features one-click addition and removal of genomes to the list of supported genomes, and handles the details of the transaction with the GenBank genome repository and BLAST database creation. At this time, there are 1169 genomes supported by PCRTiler, not counting viral genomes.

DESIGN AND IMPLEMENTATION

To maximize the utilization of the server resources, PCRTiler has been implemented as a multi-threaded application that designs as many primer pairs concurrently as the server has processors. Independent tiling requests are queued until the currently executing tiling job is finished. Users providing an email address will be notified when their request has finished processing. Others will have to use the link provided on the submission confirmation page to view their result.

To promote fair use of the system, the total number of primer pairs that can be designed in a single request is limited to 200, and the maximum duration of a tiling job is set to three hours. Users exceeding those limits can still use PCRTiler, either by installing the standalone PCRTiler application on their personal computer, installing the server version and disabling the limit, or splitting their large request into smaller regions.

PCRTiler will gracefully recover from server restarts. As soon as new tiling requests are submitted to the server, they are compressed and then saved to disk. In the event that the server is restarted, PCRTiler will transparently recover the queued tiling requests, preserving their original order, and resume execution of the run that was aborted.

Figure 2. Screenshot of a typical PCRTiler result page. The first section of the result page includes information about parameters of the request, such as the job name, the targeted organism, the tiling parameter and the desired T_m range. The second section gives the best primer pairs for each sub-region. The user can export this list in comma-separated values or tab-delimited format. The standalone result file can be opened by the standalone PCRTiler application. The third section provides a graphical view of the positions of the suggested oligonucleotides on the provided sequence. Forward and reverse oligonucleotides are highlighted. The fourth section gives an alternative description of the sequence and the primers designed in a markup language suitable for use in a sequence editing software such as Genamics Expression. Please note that this section is hidden (collapsed) in this screenshot. The last section allows the user to delete their result file from the server before the automatic expiry delay, currently set at 14 days.

Software and hardware requirements

PCRTiler requires the Java Runtime Environment (JRE) v1.6.0 and Tomcat 6 running on a computer using the Linux operating system. It should theoretically also run on any combination of platforms and operating systems for which implementations exist for the JRE, Tomcat 6, Primer3 and BLAST binaries, but this has not been tested and therefore is unsupported. During testing, we have validated that it behaves properly when viewed with the latest versions of Firefox, Safari and Internet Explorer.

The performance of PCRTiler is primarily dependent on the available memory. In our experience, for acceptable performance, you need enough memory for the BLAST database (800 MB for *Homo sapiens*, 5 MB for most bacteria), plus a maximum of 1 GB for PCRTiler. Therefore, 2 GB of memory should be enough. This amount of memory is commonly included in recent workstation computers and laptops. PCRTiler is a multi-threaded application, so it will make use of all available CPU cores, accelerating primer design proportionally to the number of cores. The PCRTiler server currently runs Mandriva Linux 2010 on a dedicated Quad-core Intel machine clocked at 2.4 Ghz with 4 GB of RAM. Including the BLAST databases of all 1169 genomes, PCRTiler requires <15 GB of hard disk space.

Standalone version

In addition to the server version, we provide a standalone Java-based application, which includes a graphical user interface and the same one-click genome management feature as the server version. It also handles all aspects of downloading genomes from GenBank and the creation of BLAST databases. Since the standalone and server versions share much of the same code base, they both provide the same functionality. However, the standalone version uses the resources of the client computer. Using the standalone version is the easiest option for most users who want to run PCRTiler locally. Please note that the standalone version does not require Tomcat. To date, it has been shown to work correctly on Linux i386, Windows Vista and Windows XP.

Data retention

PCRTiler results are kept on the server for 14 days. However, users have the option of deleting their result file from the server immediately using the appropriate button on the result page. Users that would like to hold on to a PCRTiler result for a longer time period can download the raw result file from the website, which can be viewed using the standalone version of PCRTiler.
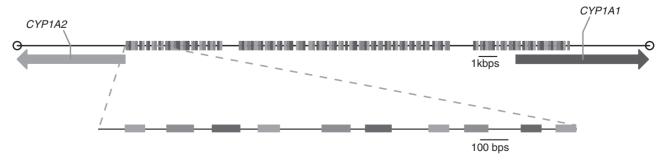
EXPERIMENTAL VALIDATION

We have validated the output from PCRTiler by tiling the 23.4 kbps wide intergenic region of the human genome separating genes *CYP1A1* and *CYP1A2*. Our laboratory is currently investigating the regulation of the expression of these genes, and the initial impetus to design PCRTiler was to simplify the design of primer pairs at this locus. A PCRTiler run was initiated selecting a target amplicon length of 80–120 bp, a primer T_m of 60–63°C and a tiling distance of 200 bp, which led to 123 suggested primer pairs. Figure 3 provides a graphical representation of the promoter region and the primer pairs designed. Note the presence of two regions where PCRTiler was unable to design primer pairs. An initial attempt to manually design oligonucleotides in these two regions also failed to produce a specific amplicon. This indicates that regions for which PCRTiler does not suggest primer pairs are potentially problematic.

Each of the suggested primer pairs was tested by qPCR on human genomic DNA. Under identical standard amplification conditions, 96 primer pairs led to satisfactory amplification products. With minor modification of qPCR conditions a total of 105 of 123 (85%) of primer pairs could be made to work. Successful amplification was defined as a qPCR assay showing a dissociation curve with a single sharp peak at a temperature above 80°C or a single sharp peak at a lower temperature and which produced a single band of the expected size when the product was migrated on an agarose gel. Reactions with a low dissociation temperature were analyzed by gel electrophoresis to exclude the formation of primer dimers. Table 1 summarizes the results obtained. Primer sequences and validation results are provided in Supplementary Table S1.

OPEN-SOURCE SOFTWARE LICENSE

Both the server and standalone versions of the software are published under the GNU General Public License,

Figure 3. The *CYP1A1–CYP1A2* intergenic region that was used to validate a large-scale PCRTiler result. PCRTiler was used to compute primer pairs in this region, designing one primer pair every 200 bp. The intergenic region shown measures 24.3 kb. Each vertical bar (123) represents an amplicon targeted by a primer pair designed by PCRTiler. A closeup view of the first 10 amplicons is shown underneath.

Table 1. Validation of a PCRTiler primer pair design run on the human *CYP1A1–CYP1A2* intergenic region

Result	Before optimization, n (%)	After optimization, n (%)
Successful amplification[a]	96 (78)	105 (85)
Unsuccessful amplification	27 (22)	18 (15)
Total	123 (100)	123 (100)

*See text for the definition of a successful amplification.

version 3. Source code, compiled binaries and installation instructions for both versions are available from the website. This website is free and open to all users and there is no login requirement. PCRTiler has proven to be a useful tool in our lab, and we hope that the scientific community will benefit from it.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Rozen,S. and Skaletsky,H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365–386.
2. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
3. Johnson,M., Zaretskaya,I., Raytselis,Y., Merezhuk,Y., McGinnis,S. and Madden,T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.*, **36(Web Server issue)**, 5–9.
4. Yao,F., Zhang,R., Zhu,Z., Xia,K. and Liu,C. (2006) MutScreener: primer design tool for PCR-direct sequencing. *Nucleic Acids Res.*, **34(Web Server issue)**, 660–664.
5. You,F.M., Huo,N., Gu,Y.Q., Luo,M.C., Ma,Y., Hane,D., Lazo,G.R., Dvorak,J. and Anderson,O.D. (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics*, **9**, 253–253.
6. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2010) GenBank. *Nucleic Acids Res.*, **38(Database issue)**, 46–51.