

SQPrimer: The Utility of Designing Homologous Primers for the Genetic Analysis Based on the PCR

Julio Perez-Marquez*

Department of Biomedicine and Biotechnology, University of Alcalá de Henares, Spain

Abstract

Objective: Most bioinformatics applications that design primers for the technique of the PCR analyze one single sequence of DNA as template; only a few applications process various nucleotide sequences. A common feature of existing software is that produces primers that are unique for each template, which is partially useful in the genomic analysis. The objective was to create an application able to find the primers that are identical in multiple nucleotide sequences and the primers that are unique to each sequence.

Methods: I applied object-oriented programming using the C++ Builder 2009 to implement algorithms that find particular short strings (primers) in nucleotide sequences. The software is a set of applications with a simple design that serves as a didactic tool to find and analyze primers. To test the application, I used molecular biology to clone genes in the laboratory.

Results: I have developed a bioinformatics application for the PCR technology named SQPrimer that focus in finding identical primers in several sequences of nucleotides. I have shown the applicability and accuracy of the application in various examples of the genetic analysis. The software was able to 1: design primers at conserved sequences of nucleotides among different species; 2: find and design mutation specific primers in sequences with single nucleotide polymorphisms and 3: design primers to detect length polymorphisms: insertions, deletions or expansion of triplet nucleotide repeats.

Conclusion: SQPrimer is bioinformatics software that adds the capability of designing primers that are identical in a batch of sequences, a utility that can be used in some strategies of the PCR analysis. The software is accessible at http://www2.uah.es/biologia_celular/JPM/SQPP/SQPrimer.html.

Keywords: Bioinformatics; Primer; PCR; Software; Polymorphisms; Gene; DNA

Introduction

Primers are a short chain of nucleotides chemically synthesized in orientation 5' to 3' that are used for amplification of DNA by the polymerase chain reaction (PCR). The PCR technique includes three steps in one cycle: the DNA denaturation, the alignment of two primers to the extremes of the sequence of DNA to be amplified, and the synthesis of DNA using polymerases; the repetition of this cycle (c) produces 2^c molecules of DNA named amplicons. Many genetic applications use the PCR; among them, those that detect variations in the nucleotide sequence of genes. For instance, PCR is used in the analysis of mutations involved in genetic diseases, in forensics and in the studies of differences amongst species. The design of specific and selective primers is a critical factor for a successful DNA amplification by PCR [1,2].

There are several software programs in the web that design specific primers for the PCR [3-8]. Some applications are devoted to specialized tasks such as primer-BLAST [9] that designs primers that do not match to any other DNA, apart from the one of interest. By contrast, there are software applications that obtain primers from a group of nucleotide sequences; among others the Primaclade and BatchPrimer3 software which are based on Primer3 [10,11]. That particular software has an array of different applications; for instance, it designs degenerated primers, finds primers that recognize microsatellites of nucleotide repeats (or SSR, simple sequence repeat) or detects primers that include single nucleotide polymorphism (SNP), a kind of nucleotide variation in the DNA. The existing software has in common that they serve to design primers that are different and specific for one or various DNA templates.

Some areas of the genetic analysis that use the PCR require the design of homologue primers in divergent nucleotide sequences. There are two possible strategies to distinguish by PCR different DNAs with a degree of homology: one is to use primers that are unique to each sequence, the other is to use a combination of primers that are unique to each sequence and homologous to the sequences in the analysis. In the case of sequences that vary in one single nucleotide the second strategy is the only alternative: one of the pair of primers that will be used in the PCR is homologous to all sequences in the analysis. Here I show the design of one application named SQPrimer that is particularly useful in produce the primers that are identical in multiple DNAs together with primers that are unique to each sequence.

In the year 2002 we cloned a cDNA from rat tissue that codifies a putative protein (CLRP) with complex leucine repeats [12]. With concrete examples that use this and other sequences of nucleotides [13], I show that SQPrimer designs primers at conserved regions among different species and also serves to find and to design mutation specific primers in sequences with SNPs by using different strategies of the PCR analysis. SQPrimer is also a tool to design primers to detect

***Corresponding author:** Julio Perez-Marquez, Department of Biomedicine and Biotechnology, University of Alcalá de Henares, Madrid-28871, Spain, Tel: +34918854898; Fax: +34918854799; E-mail: 12092002@mmm.muroran-it.ac.jp

Received September 11, 2014; **Accepted** September 29, 2014; **Published** September 30, 2014

Citation: Perez-Marquez J (2014) SQPrimer: The Utility of Designing Homologous Primers for the Genetic Analysis Based on the PCR. J Comput Sci Syst Biol 7: 229-234. doi:[10.4172/jcsb.1000162](https://doi.org/10.4172/jcsb.1000162)

Copyright: © 2014 Mao J, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

length polymorphisms: insertions, deletions or expansion of triplet nucleotide repeats.

Methods

Algorithms and conditions for the design of primers in SQPrimer

The most important function of SQPrimer contains one algorithm that finds short strings of nucleotides (primers) in the inputted sequence templates. The process to design primers for the PCR requires setting several variables; thus, this function depends on the values of the primer conditions that are indicated by the user in one panel of the application. The variables in the function are: the primer length and the number of guanines+cytosines/primer length (%GC), and the number of repeats of single nucleotide (i.e. AAA...) or dinucleotide (i.e. ATATAT...) that are allowed in its sequence. The algorithm also evaluates the melting temperature (T_m) of the alignment of the primer to the DNA template; regarding to this variable, three different options can be selected by the user: the basic, the simple and the salt 50mM, with equations that have previously described [2,15]:

Basic: $nATx2^\circ + nGCx4^\circ$.

Simple: $64.9^\circ + 41^\circ \times (nGC-16.4)/\text{primer length}$.

Salt: $81.5^\circ + 16.6^\circ \times (\log_{10} [0.05]) + 0.41^\circ \times (\%GC) - 675/\text{primer length}$.

(nAT: number of adenine + thymine; nGC: number of guanine + cytosine; %GC: nGC/primer length; $^\circ$: Celsius degrees).

Basically, the algorithm starts with the declarations of two arrays of the size of the length of the primer; one array is for the sense and the other for the antisense. Another two arrays serve to memorize the position of the primers in the template. Then, the algorithm progressively takes the primers from the sequence template up to its full length. For each primer, the algorithm determines the presence of single or double nucleotide repeats; if the selected primer fits the conditions indicated by the user, then the algorithm determines whether the primer also fits with the T_m . Primers that fulfill all conditions are stored, together with its position, in the arrays. Finally, the algorithm evaluates whether each pair of sense and antisense primers in the arrays are separated in the template the distance that has been specified by the user; if they do, both primers are displayed in the application.

Software design

I applied object-oriented programming using the C++ Builder 2009 application from Embarcadero technologies to produce the SQPrimer software that runs in the Windows environment. This programming tool has been previously tested for the design of bioinformatics applications [14]. The SQPrimer software is open access at http://www2.uah.es/biologia_celular/JPM/SQPP/SQPrimer.html.

Applications within SQPrimer and input of nucleotide sequences: SQPrimer contains two main interconnected windows: the multi-sequence application and the tool that design pairs of sense and antisense primers for single DNA templates. Additionally, there is one primer analysis tool that analyzes the nucleotide composition and the self-complementarity of the primers and there are also two graphical displays (Figure 1). The multi-sequence tool requires the input of various nucleotides sequences in the text box; the nucleotide sequences can be also pasted by the user and, alternatively, the application can open *.txt files or *.SQP files that can previously be saved with the single sequence application. Once the template sequences are included in the

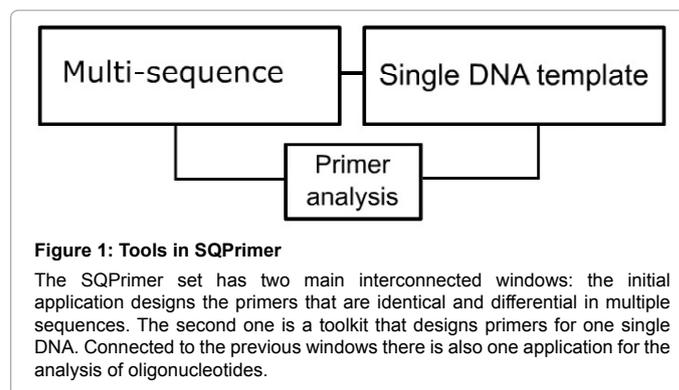


Figure 1: Tools in SQPrimer

The SQPrimer set has two main interconnected windows: the initial application designs the primers that are identical and differential in multiple sequences. The second one is a toolkit that designs primers for one single DNA. Connected to the previous windows there is also one application for the analysis of oligonucleotides.

interfaces, the production of the primers by the software requires two steps: first, to indicate the values primer conditions and then to click the buttons of function. The functions in the two main applications will produce lists of primers that meet the conditions established by the user; those lists can be exported to excel (Microsoft) and thus, SQPrimer features connectivity to other functions and applications of the Windows environment. The design of primers is accompanied with the display of a graphical representation of the position of the primers in the nucleotide templates. Additional features of SQPrimer that help the user to familiarize with the application are instruction menus in each window as well as examples of the nucleotide sequences that are explained in the results. Compared to other software, SQPrimer offers simplicity, as it is required for a didactic tool; for instance, if the software does not produce results using a particular set of primer conditions, a new search can be done in few steps by changing the variables of primer length and/or GC content followed by a click on the button of the function.

The multi-sequence application: Two different functions can be run in this application clicking in the respective buttons: one that designs the primers that are different or identical in various sequences and another that produces primers that detect length polymorphisms. For the first function, the software produces two groups of results: primers that can be found in all the templates (identical) and primers that are unique to each inputted sequence; all of them meet the conditions specified by the user. If the application finds primers that are unique, SQPrimer will automatically display the primers that have one distinctive nucleotide at the extreme 3'. As shown in the results, the multi-sequence application serves to clone orthologous DNAs and is also useful for the detection of SNPs. The function that designs primers to analyze length polymorphisms of the DNA produces pairs of primers that flank nucleotide sequences that differ in either deletions or insertions, including the expansion of trinucleotide repeats.

Cloning of the cDNA of CLRP from different species: To design primers to clone the CLRP cDNA of CHO cells I followed one strategy based in the design with SQPrimer of primers that are homologues in the rat and human nucleotide sequences. We had previously cloned the CLRP gene of *Rattus norvegicus* [GenBank: AF406814.1] and blasted the sequence to obtain information of a similar nucleotide sequence from the Homo sapiens chromosome 5, BAC clone from the database [GenBank: AC005214.1]. One segment of the human cDNA containing CLRP was isolated from human prostate Marathon-Ready cDNA (Clontech) using a PCR strategy based on the sequence homology between these two species. I tested a batch of homologous primers for the rat CLRP cDNA that were designed with SQPrimer to amplify the human cDNA from human prostate and obtained a positive PCR

reaction. The human cDNA of CLRP isolated from human prostate was 2642 nt long (nt: nucleotide).

One positive PCR reaction on human DNA was obtained with the homologous sense 5'-AGGGCATCAGCAGTATTG-3' and antisense 5'-GAGGAAGAGGTTCTGAAG-3' primers from the published rat cDNA (nt 1-18 and 1174-1191, respectively) for 30 cycles of denaturation at 94°C, annealing at 55°C for 30 sec, each and extension at 72°C for 2 min. After sequencing, the extremes of the human cDNA were amplified by rapid amplification of cDNA extremes (RACE) using the nested and polyT adaptor primers included in the Marathon cDNA amplification kit (Clontech laboratories) using the following rat primers: the 5' was amplified with the human antisense primer 5'-CCATCCTCTACTCATAAC-3' (nt 720-438) and the 3' extreme with the sense primer 5'-CATTCTGTACTGCCTCATC-3' (nt 1297-1315) at 94°C for 30 sec, 55°C for 30 sec and 68°C for 3 min. Finally The PCR products obtained in the reactions were gel purified, subcloned in the pGEM-T vector (Promega) and sequenced. The human CLRP cDNA obtained was 2642 nt long.

Results

Cloning of orthologous cDNAs using homologous primers designed by SQPrimer

The applicability and accuracy of SQPrimer was tested in the laboratory in concrete examples of genetic analysis. The steps to clone the CLRP cDNA of CHO cells was as follows: 1-To obtain the human cDNA (described in the methods), 2-To use of SQPrimer to design the primers that are identical in the rat and human sequences, 3-To use different combinations of those homologous primers in PCRs using the cDNA of CHO cells as template. 4-To isolate and purify the PCR products in the agarose gels and clone that cDNA.

Having obtained the CLRP cDNA from rat and human I used these two sequences as template and run one of the functions included in the multi-sequence application in order to find both the primers that are identical in the two templates as well as those primers that are different (unique to each template). The application was used with the following fixed values of the primer conditions: GC=50% ± 2 and basic Tm of 54 ± 2. As expected, the SQPrimer application designed more primers if the restrictive conditions of the variables of the primer are more permissive: no homologous primers were found in the templates at any primer length selected if nucleotide repeats was 0. By contrast, if 2 repeats of one single nucleotide and 2 repeats of dinucleotide were allowed the software designed 4 primers of 18 nt, 2 of 20 nt but none over 21 nt long. Thus, the number of homologous primers designed by the application also increased with the decrease of the primer length.

I purified the total cDNA of CHO cells and made different PCR amplifications using different combinations of these four homologous sense and antisense primers (Figure 2). It should be warned that this strategy may not always produce results since only three, out of the four primers that are identical in rat and human, produced positive PCR reactions. One 315 nt long PCR product was obtained with two sense and antisense primers and was subcloned and sequenced. After confirming nucleotide homology to the rat CLRP cDNA I proceeded to amplify the 5' and 3' extremes to obtain the full cDNA sequence by RACE. The final cDNA product was 2306 nt long and displays 94.7 % homology with the rat orthologous CLRP (data not shown). I conclude that the multi-sequence function of SQPrimer is useful to design primers at conserved regions of nucleotides among different species or DNAs sequences that display a degree of homology.

Detection of alleles by PCR using the unique/identical function of SQPrimer

In one analysis of human DNA samples I found one abundant allele of CLRP (allele A) and one individual that had one allelic sequence with two single nucleotide differences in the open reading frame (allele B). In allele A, the codon at position 950 of the cloned DNA was GCC which encodes Ala and one second codon starting at position 962 was GCA, which also encodes Ala. By contrast, the allele B had ACC at the first codon and GTA in the second codon, which is translated to Thr and Val, respectively (Figure 2). The two allelic sequences of CLRP are included in the examples of the SQPrimer application.

All primers designed by SQPrimer that are unique to one sequence should hybridize differentially the SNPs in the alleles A and B. Because it is generally accepted that among all possible primers the best ones to recognize the SNPs are those that have the differential nucleotide at the extreme 3', I selected two 18 nt long reverse (or antisense) primers, each unique to one of the alleles, that have that feature and that were provided by the function included in SQPrimer. The two reverse primers that were selected are located at the same position in the two DNA templates (Figure 2). To recognize the alleles in agarose gels by their lengths, a combination of 18 nt long primers was used in two PCR reactions per individual. One reaction was carried out with one of the reverse primers and one forward primer that is identical in both sequences; the second PCR reaction was performed with the second reverse primer and a different forward primer at a different position than the one used in the previous reaction, which is also present in both sequences (Figure 2). As shown in the gel of the figure, one individual is homozygous for the allele A and the other is heterozygote and has the alleles A and B of the CLRP gene. In conclusion, the multi-sequence application of SQPrimer was useful in the design of specific primers that detect nucleotide sequences with SNPs.

The functionality of SQPrimer was tested by changing the variables of the design of primers with these two sequences as templates. For instance, the software was run with a fixed value of the primer length of 20 nt. The application found no differential primers if the repetition of one single nucleotide was not allowed; conversely, if that repetition is allowed in the primer sequence, the number of unique primers designed increases with increasing variation in the proportion of GCs and if the presence of two nucleotide repeats is admitted (Table 1). I conclude that SQPrimer produces results that are consistent with the effect of restricting the conditions of the primers.

Using the SQPrimer application to design identical primers that detect length polymorphisms in various sequences

Insertions and deletions of nucleotides are forms of genetic mutations in the DNA; they range from one to a large amount of nucleotides. There is one function in SQPrimer that detects homologous primers in several sequences that may serve to distinguish length polymorphisms. Two different examples were included in the application to test this functionality; starting from the sequence of the CLRP cDNA, I artificially created two sequences using a word processor: one sequence with a single nucleotide insertion and another with one nucleotide deletion. A second example is an insertion that consists on the extension of triplet nucleotide repeats; a kind of mutation that is shared by a group of genes that cause genetic diseases. In this example, the sequences included in the application were: the real coding sequence of the Huntingtin mRNA and one sequence with a repeat of nucleotide triplets that expands 63 nt, which was also constructed artificially. As shown in the Figure 3, the application finds

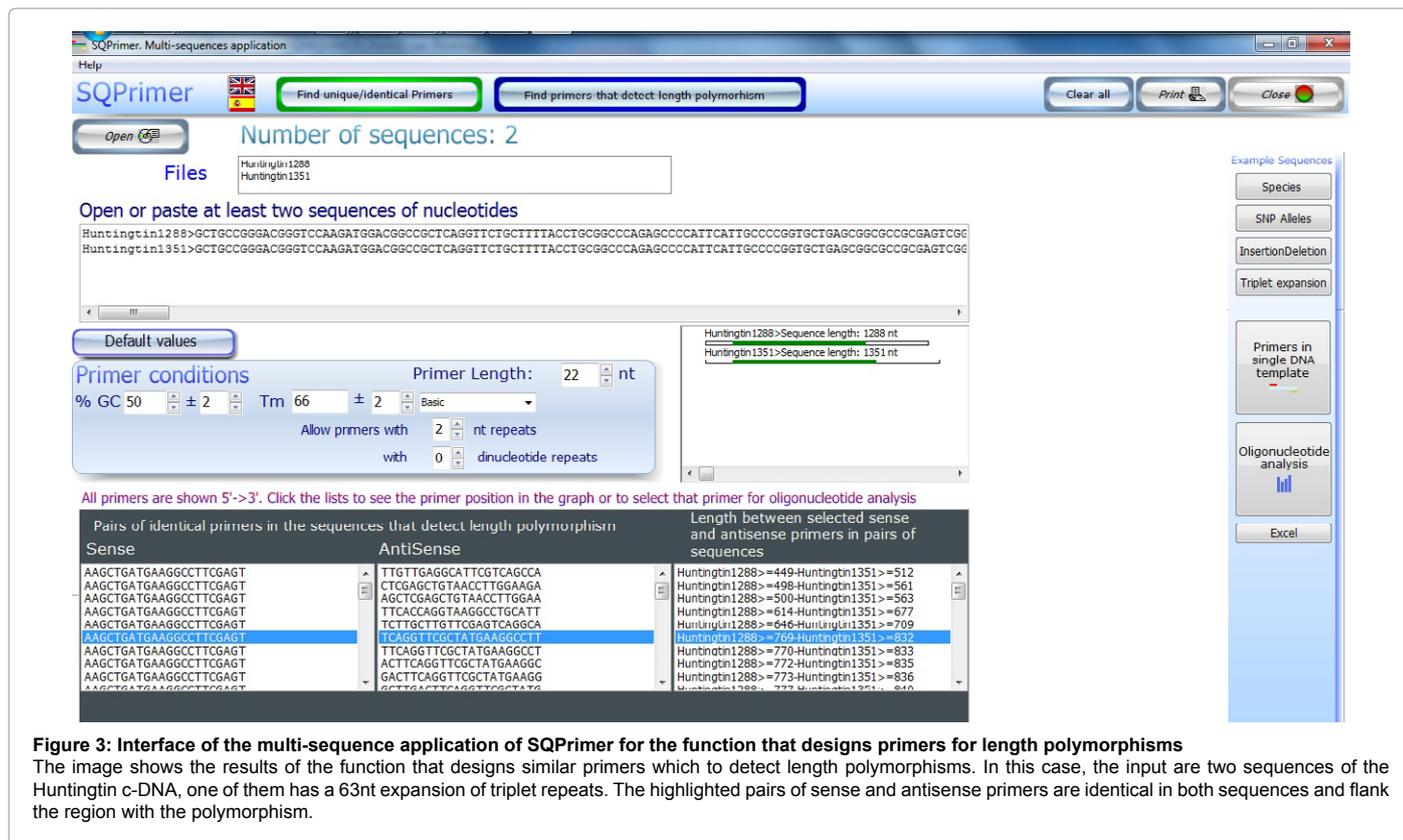


Figure 3: Interface of the multi-sequence application of SQPrimer for the function that designs primers for length polymorphisms
 The image shows the results of the function that designs similar primers which to detect length polymorphisms. In this case, the input are two sequences of the Huntingtin c-DNA, one of them has a 63nt expansion of triplet repeats. The highlighted pairs of sense and antisense primers are identical in both sequences and flank the region with the polymorphism.

Number of dinucleotide repeats=0

	%GC: 62	%GC: 60	%GC: 58
%GC ± 0	0	7	0
%GC ± 5	17	18	8

Number of dinucleotide repeats=2

	%GC: 62	%GC: 60	%GC: 58
± 0	0	24	0
± 5	47	58	35

Table 1: Number of unique primers designed by SQPrimer in the analysis of two alleles of CLRP.

pairs of sense and antisense primers that are identical in the sequences of the two Huntingtin alleles and also displays the different lengths of expected PCR products for each sequence using those homologous primers. In conclusion, this application of SQPrimer designs primers that flank the sites at where the length polymorphism occurs and serve to detect insertions, deletions or expansions of trinucleotide repeats.

Discussion

The PCR method has a large applicability in very different fields such as the detection of pathogens, drug discovery, genetic engineering, genetic diagnoses, mutagenesis, molecular anthropology, genetic phylogeny, etc. In all these fields of research the products of the PCR are used to clone and to sequence DNAs. For any application, the technique of the PCR largely depends on the design of primers that are used in the reaction.

SQPrimer might be useful in fields such as anthropology or phylogeny. I believe that the finding of identical primers in a large number of sequences from different species is one strategy to find

the same sequence in new species. With this approach, SQPrimer showed to be useful to clone the CLRP cDNA of CHO cells; the results indirectly show that *CLRP* is a conserved gene in vertebrates because the homology of sequence of their primers.

SQPrimer proved useful in the design of identical primers for the analysis of different types of polymorphisms. The PCR can determine directly the presence of SNPs between individuals by using a combination of primers that are unique to each sequence and primers that identical in the templates; this application of the PCR does not require additional steps of DNA purification, cloning and sequencing. There are two methods to try to avoid mispriming in the research of sequences with high homology by PCR; one is at the step of the primer design using applications as SQPrimer: it would be advisable to design primers with the highest possible values of the Tm and to select primers with the distinctive nucleotide at the extreme 3'. The second method can be carried out at the laboratory: test those primers in different PCR reactions with increasing melting temperatures.

Several genetic human diseases are caused by deletions and insertions in genes; there is also a group of diseases that are determined by the extension of the number of triplet repeats in particular genes [16]. SQPrimer is focused in the design of identical primers in different DNA sequences that display these kinds of polymorphisms; therefore, the application may be useful in the studies of human genetic mutations that use molecular techniques based on the PCR.

The functionality of SQPrimer can be compared with some existing primer design tools. Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) [3], Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>) [6] or Primer-Blast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) are tools for finding

unique primers in single PCR templates; in contrast, the multi-sequence application of SQPrimer designs the differential and also the homologous primers in several templates. Primaclade (<http://primaclade.org/index.html>) admits a group of nucleotide sequences; the software is developed to design minimally degenerated primers to work reliably and specifically on a number of species [10]; in contrast, SQPrimer does not produce degenerated primers but primers that can directly be used in PCR analysis. Some of the indicated applications require the input of a previously formatted sequence, others include a considerable amount of variables that need to be set in the design of primers; in the case of SQPrimer, any sequence can be pasted in the application and the simplicity in the number of the most important variables of the primer design may facilitate an educational use of the application. Finally, the function of SQPrimer that localizes the primers that are identical in the inputted sequences is a tool that may help researchers that do not know where to expect the length polymorphism in the sequences or to uncover unexpected insertions or deletions of nucleotides in batches of sequences. In conclusion, a common feature of existing software is that it is developed to design primers that are unique for each template; SQPrimer adds the capability of designing primers that are identical in a batch of sequences and this an utility that can be used in some strategies of the PCR analysis.

Present research in my laboratory is focused on developing bioinformatics software that covers different aspects of genetic engineering and has educational utility. SQPrimer is related to bioinformatics software named SQRestriction that serves for various types of restriction analysis of nucleotide sequences [17]. The main limitation of SQPrimer is that it is an executable (*.exe) application and it is not a multi-platform; therefore it only runs in the Windows OS environment. Having that limitation in mind, future work includes the translation of the C++ code in SQR to Java and implementing its interfaces in Html5.

Conclusion

SQPrimer is a bioinformatics tool that designs the primers that are identical in different DNAs together with primers that are unique to each sequence, as it is required for several types of genetic analysis that use the PCR technology. Because its usability, the application may also be an educational tool for teaching the requirements of the design of primers for the PCR. The application was tested in the cloning of orthologous genes and in finding of SNP by PCR. SQPrimer is particularly interesting to design homologous primers that detect length polymorphisms, including the expansion of triplet repeats

Acknowledgement

Thanks to Daniel Pérez Grande for the scientific review of the manuscript. This work was supported by Ministerio de Economía y Competitividad, Spain (grant number: BFU2011-30217-C03-01).

References

1. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, et al. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350-1354.
2. Murphy WJ, O'Brien SJ (2007) Designing and optimizing comparative anchor primers for comparative gene mapping and phylogenetic inference. *Nat Protoc* 2: 3022-3030.

3. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132: 365-386.
4. van Baren MJ, Heutink P (2004) The PCR suite. *Bioinformatics* 20: 591-593.
5. Arányi T, Váradi A, Simon I, Tusnády GE (2006) The BiSearch web server. *BMC Bioinformatics* 7: 431.
6. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, et al. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35: W71-74.
7. Qu W, Zhou Y, Zhang Y, Lu Y, Wang X, et al. (2012) MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity. *Nucleic Acids Res* 40: W205-208.
8. San Millán RM, Martínez-Ballesteros I, Rementeria A, Garaizar J, Bikandi J1 (2013) Online exercise for the design and simulation of PCR and PCR-RFLP experiments. *BMC Res Notes* 6: 513.
9. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, et al. (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13: 134.
10. Gadberry MD, Malcomber ST, Doust AN, Kellogg EA (2005) Primaclade-a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics* 21: 1263-1264.
11. You FM, Huo N, Gu YQ, Luo MC, Ma Y, et al. (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9: 253.
12. Pérez-Márquez J, Reguillo B, Paniagua R (2002) Cloning of the cDNA and mRNA expression of CLRP, a complex leucine repeat protein of the Golgi apparatus expressed by specific neurons of the rat brain. *J Neurobiol* 52: 166-173.
13. (1983) A novel gene containing a trinucleotide repeat that is expanded and unstable on HD chromosomes. The Huntington Disease Collaborative Research Group. *Cell* 72: 971-983.
14. Garcia-Segura LM, Perez-Marquez J (2014) A new mathematical function to evaluate neuronal morphology using the Sholl analysis. *J Neurosci Methods* 226: 103-109.
15. Dieffenbach CW, Lowe TM, Dveksler GS (1993) General concepts for PCR primer design. *PCR Methods Appl* 3: S30-37.
16. Goldberg YP, Andrew SE, Clarke LA, Hayden MR (1993) A PCR method for accurate assessment of trinucleotide repeat expansion in Huntington disease. *Hum Mol Genet* 2: 635-636.
17. Perez-Marquez J (2014) SQRestriction: Bio-Informatics Software for Restriction Fragment Length Polymorphism of Batches of Sequences. *J Comput Sci Syst Biol* 7: 186-192.