

Open Access

SQRestriction: Bio-Informatics Software for Restriction Fragment Length Polymorphism of Batches of Sequences

Julio Perez-Marquez*

Department of Biomedicine and Biotechnology, Alcala de Henares University Campus, Madrid 28871, Spain

Abstract

Objective: The existing informatics applications are specialized in particular aspects of the enzymatic restriction analysis of nucleotides sequences. In order to design an application that fulfills different needs of the investigation in restriction enzyme analysis, I have developed SQRestriction (SQR), a suite that compiles several utilities.

Methods: I applied object-oriented programming to set algorithms in C++ that compare the differences or similarities in the restriction sites of orthologous genes or in nucleotide sequences with single nucleotide polymorphisms, insertions or deletions.

Results: SQR is mainly a tool for the analysis of restriction fragment length polymorphisms but it has also other applications such as cloning of DNA; I demonstrate those functions in examples of different types of nucleotide sequences. I conducted a comparative analysis between SQR and existing software in the web; the main difference is that SQR is a suite of different applications that analyzes either one, two or multiple sequences of nucleotides.

Conclusion: I conclude that SQR is a competitive application since it offers differential features that are useful for academics and in laboratories of molecular biology. The SQRestriction software is open access at

http://www2.uah.es/biologia_celular/JPM/SQR/SQR.html

Keywords: Software; Enzyme restriction; RFLP; Polymorphism

Introduction

A feature of each individual genome is that the length of the fragments of DNA cut with restriction enzymes (a class of endonucleases that recognizes specific sequences of nucleotides) is unique. The technique of Restriction Fragment Length Polymorphism (RFLP) exploits this singularity by comparing the enzyme restrictions of different DNAs by gel electrophoresis [1].

Based on the principles of the RFLP, some variations of the technique have been developed: the Terminal Restriction Fragment Length Polymorphism (TRFLP) compares the differences in the position of restriction sites of ribosomal genes [2]. Another example is the Amplified Fragment Length Polymorphism (AFLP), which combines restriction with the PCR technique and uses primers containing the sequences of nucleotides of specific restriction sites. Finally, when the differences in the restriction fragment lengths are tested in the amplicons produced by the PCR the technique is named CAPS: Cleaved Amplified Polymorphic Sequences [3].

The RFLP and derived techniques have been used for the detection of single nucleotide polymorphisms (SNP) or to recognize the presence of insertions or deletions and the sizing of the variable number of tandem repeats in the DNA. They are tools for genetic mapping, for the genetic fingerprinting as a probe for forensics or paternity, for the predictive heritability of genetic diseases and studies of species biodiversity [4].

The present work shows the broad functionality of a suite of linked applications named SQRestriction (SQR) that compiles different aspects of the analysis of the enzymatic restriction of DNA sequences. The existing software in the web [5,6] can be used to find the restriction sites of DNA sequences; however, the applications are generally focused on very specific aspects of the enzymatic restriction analysis. The main advantage of SQR is that starting with a batch of nucleotide sequences, the restriction analysis can be performed in either a group, in two aligned sequences or in a single sequence according to different requirements. Thus, SQR may be a suitable application for the different applications of the RFLP and the derived techniques described above.

The comparative between SQR and some applications available in the web resulted in the following summary of brief differences: SQR allows the analysis of any nucleotide sequence without previous editing and extends the length of nucleotides that can be compared. For same cases, SQR has a larger number of enzymes for the analysis; includes a list of 242 enzymes that can be eligible individually and the possibility of editing enzymes. Finally, SQR offers several graphic displays of results. Instead of creating independent applications, I wanted to create a multitask software with the potential to be used in any laboratory of molecular biology; an additional purpose in its development was to produce a set of easy-to-use applications with educational utility.

Methods

Software design

I applied object-oriented programming using the C++ Builder 2009 software from Embarcadero technologies to produce the SQRestriction software that runs in the Windows environment. This application has been previously tested for the design of bioinformatics tools [7]. The SQRestriction software is open access at http://www2.uah.es/biologia_

*Corresponding author: Julio Perez-Marquez, Ph.D, Department of Biomedicine and Biotechnology, Alcala de Henares University Campus, Madrid 28871, Spain, Tel: (+34 91) 885 47 97; Fax: (91) 885 47 99; E-mail: julio.pmarquez@uah.es

Received July 24, 2014; Accepted August 20, 2014; Published August 22, 2014

Citation: 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. doi:10.4172/jcsb.1000155

Copyright: © 2014 Perez-Marquez J. 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.

celular/JPM/SQR/SQR.html

Applications in SQRestriction

SQR is a suite of different applications that has been designed to report the results of different kinds of analysis of restriction enzymes in nucleotide sequences. SQR can be run in two languages: English and Spanish and also, to help the user to familiarize with the application, each interface includes a menu of instructions as well as examples of representative nucleotide sequences.

The program starts with the multi-sequence application that consecutively analyzes the restriction sites of pairs of sequences in a batch. From here, two applications can be launched: the align-RFLP application, a tool that analyzes the restriction sites in two aligned sequences, and the application that analyzes the restriction sites of single sequences. Furthermore, SQR displays additional applications such as one that reports the enzyme sites and the length of the restriction fragments in the sequences; another application may be used for nucleotide sequence comparison of two sequences. SQR also counts with different graphical displays, including a diagram showing the appearance of agarose gel electrophoresis after restriction of sequences (Figure 1).

Input of nucleotide sequences in SQR

Any nucleotide sequence can be pasted in the input boxes of all applications. Also, the applications can open any *.txt file that contains nucleotide data or *.SQP files that are created using the save button in the multi-sequence application. The length of the input sequence is limited to only 4500 bases in the multi-sequence application, while the applications of single sequence and the align-RFLP analysis have no limitation; I have tested sequences with lengths of 10000 bases and I found no limitation.

Selection of restriction enzymes in the applications

All applications contain a list of 242 different restriction enzymes. The analysis can be performed by selecting any number of restriction enzymes in the lists. The single sequence tool also allows the edition of restriction enzymes that are not in the list.

Results

The multi-sequence application

The initial interface of the software is the multi-sequence application: a tool that compares various nucleotide sequences by pairs; each sequence included in the sequences A input box is compared to each sequence in the input B box.

The application links to the tools of restriction in one and two sequences (Figure 2). This figure shows the results the of restriction analysis of three orthologous sequences that are included in the application as examples. The results of the analysis are shown in three lists: one list displays the enzymes that cut each pair of sequences in identical positions; a second list shows the enzymes that either cut in only one of the sequences of the pair or that cut in both sequences but at different lengths. Finally, a third list shows the enzymes that do not cut any sequence. These three lists can be exported to excel (Microsoft) and thus SQPrimer features connectivity to other functions and applications of the Windows environment. One utility of this tool would be to compare the differences or similarities in the restriction sites of orthologous genes; thus, the example sequences in this application are fragments of the CLRP gene of rat, human and CHO-cells [8].

The application for alignment and restriction analysis of two sequences

The align-RFLP application aligns two nucleotide sequences and finds all the possible restriction fragment length polymorphisms found between them (Figure 3). An applicability of this tool is to compare the restriction fragment length polymorphisms between genes. Figure 3 shows the results of the analysis of the two orthologous sequences included in the software as example; they are nucleotide segments of the rat and human *clrp* genes. The application displays the aligned sequences, the mismatched nucleotides and the enzymes that cut only one of the sequences (gray box).

From this interface, two additional applications can be launched: one is a graphical representation of the differential restriction of the two sequences analyzed (Figure 4). The second application is a tool that compares the nucleotide composition, molecular weights and lengths of the two sequences (Figure 5).

The single sequence application

The result obtained with the single sequence application of SQR is displayed as a nucleotide sequence along with the position of the restriction sites and the enzymes that cut (Figure 6). This application may be used, for instance, for cloning one DNA sequence; the example sequence in the application is the rat CLRP open reading frame at were enzymes that belong to the polylinkers of vectors as for instance pcDNA, pBlueScript, or pGEMT can be found.

In the interface there is a link to the application that shows the enzyme report: the enzymes that either cut or do not cut the sequence, the position of the enzyme targets in the sequence and the size of the restriction fragments. The interface links to the graphic representations of the restriction sites in either circular or linear sequences (Figure 7); the figure shows examples of the graphical figures of the restriction of the same sequence with one restriction enzyme. Finally, the application links to the gel tool (Figure 8); the figure displays one example of how the distribution of the restriction fragments would be in the electrophoresis on agarose gel of the reaction of several restriction enzymes on the same DNA.







Validation of SQR

To validate SQR I tested the application in different cases of genetic research. Mutations in the Ras proteins are frequent in different types of cancer; I set out to determine whether SQR would be able to find RFLPs in mutations of these genes. For instance, various mutations at codon 12 of the *k-ras* oncogenes have been classically detected by a mutagenic PCR-RFLP method that generates a BstNI site right upstream of

codon 12 in the wild type, but not in *k-ras* mutants [9]. To begin with, the coding sequence of the Human *k-ras* was obtained from the nih database (Accession: CCDS8702.1) and the nucleotide sequence was copied and pasted by right-clicking the input sequence box A of the multi-sequence application of SQR. The wild *K-ras* human sequence was then directly passed to the single sequence application of SQR; all enzymes were selected, and finally the buttons of linear cut and display protein were consecutively clicked; the application confirmed that the

₩ RFLP graph RFLP graph	Print L Close
Rat CLRP	Aati or Stul 565
	Figure 4: The RFLP graphical display.









sequence had one G amino acid at codon 12 (GGT). Consecutively, using a word processor, the codon 12 was changed to GAT (G12D), GTT (G12V), TGT (G12C) or AGT (G12S) and the sequences were pasted in the input box B of the multi-sequence application of SQR. As shown in Figure 9, the software showed the enzymes that cut identically in all pairs of sequences and also the enzymes that did not cut any of the sequences. In addition, SQR found one RFLP in the comparison of the wild type with the G12S mutation that could be detected with three different enzymes: BfaI, MacI and RmaI (Figure 9); this RFLP was confirmed and graphically displayed with the align-RFLP application of SQR. Thus, SQR found that one of the known mutations of *k-ras* produces one RFLP that can be directly detected by enzyme restriction without the need of a PCR based mutagenesis of the wild type.

Another example was used to analyze a well described mutation linked to RFLPs. The Sickle-cell anemia is a hereditary disease caused by a point mutation in the sixth codon (G7V) of the Human beta-globin gene. I proceeded in similar fashion to what is described above and obtained the coding region of that gene in the nih data base (Accession: J00173) and changed the nucleotide 17 from A to T. In this case, I used the align-RFLP application of SQR to enter the wild and mutated sequences. The software obtained the previously described RFLP: the mutated sequence lacks of one MstII restriction site [10]; in addition, SQR also found alternative enzymes that produce the same RFLP, such as DdeI, AocI or SauI.

Discussion

I have developed software that has several utilities for the genetic analysis using restriction enzymes. SQR is useful to clone by finding the restriction sites in DNA that are contained in vectors, to compare the restriction differences of similar sequences that differentiate in the base composition, insertions or deletions of nucleotides and for the analysis of the differences of restriction among batches of nucleotide sequences that belong to different species or to different individuals within a population.

I conducted a comparative analysis between SQR and software that analyzes restriction sites of single nucleotide sequences such as the TACG Restriction Mapping (http://biotools.umassmed.edu/tacg4/, the WebCutter application (http://rna.lundberg.gu.se/cutter2/), or NEBcutter (http://tools.neb.com/NEBcutter2/; [5]). SQR was also compared to the Insilico simulator for *E. coli* sequences (http://insilico. ehu.es/restriction/two_seq/; [6]) for two sequences and the watcut application (http://watcut.uwaterloo.ca/watcut/watcut/template. php?act=snp_new) for multiple sequences. The main difference is that SQR is a suite of different applications that was designed to analyze either only one, two or multiple sequences.

Compared to the Insilico application that analyzes two sequences, I found that this application present a limitation: the maximum length of the sequences is 700 nucleotides, while I have compared sequences

SQRestriction	Compare enzymes in pairs of sequences
Open 🗺	Open 🗺
File Kras_Wid>	File Kras_GAT> Kras_GTT> Kras_TGT>
Sequences A	Sequences B
Image: Wild>ATGACTGAATATAAACTTGTGGTAGTTGGAGCTGGTGGCGTAGGCAAGAG]	☑ Kras_GAT>ATGACTGAATATAAACTTGTGGTAGTTGGAGCTGATGGCGTAGGCGAAGAG ☑ Kras_GTT>ATGACTGAATATAAACTTGTGGTAGTTGGAGCCGTGGCGCATAGCCAAGAG ☑ Kras_GTT>ATGACTGAATATAAACTTGTGGTAGTTGGAGCTGTGGCCGTAGGCAAGAGT ☑ Kras_GTS>ATGACTGAATATAAACTTGTGGTAGTTGGAGCTGTGGCGCAGGGAAGAGT
< H	4
Identical restriction sites and positions	Different restriction sites and positions
Kras_Wild>ALuI-30/-66/ *vs*Kras_GAT>ALuI-30/-66/ Kras_Wild>ALuI-30/-66/ *vs*Kras_GT>ALuI-30/-66/ Kras_Wild>ALuI-30/-66/ *vs*Kras_TGT>ALuI-30/-66/ Kras_Wild>ALuI-30/-66/ *vs*Kras_AGT>ALuI-30/-66/	Kras_Wild>BfaI or MaeI -No Restriction "vs" Kras_AGT> BfaI or MaeI-32/ Kras_Wild> MaeI or BfaI -No Restriction "vs" Kras_AGT> MaeI or BfaI-32/ Kras_Wild> RmaI -No Restriction "vs" Kras_AGT> RmaI-32/
۰. <u>۳</u> . ۲	۲
Enzimes that do not cut pairs of sequences	Kins Wid5 Acl_CCGC-No Restriction "ws" Kras_GAT> Acl_CCGC-No Restriction "Array" Kras_GAT> Acl_CCGG-No Restriction "ws" Kras_GAT> Acl_CCGG-No Restriction "ws" Kras_GAT> Brial or Meat-No Restriction No Restriction "ws" Kras_GAT> Brial or Meat-No Restriction "ws" Kras_GAT> Brial or Meat-No Restriction "ws" Kras_GAT> Brial or Meat-No Restridor No Restridor No Restriction No Restriction No Restridor No Res

that extend more than 10.000 bases with SQR. The main difference between the Watcut application for multiple sequences and SQR is that the first one requires the edition of the sequences before analysis while ours does not; thus, the input of sequences in SQR is easy and even allows the simple copy-paste of nucleotides without previous formatting. Differentially to Watcut, the SQR application allows to select two sequences from the bunch and to perform an alignment and restriction analysis which is ideal for the detection of SNPs in the DNA of two individuals: for instance, between control and mutated animals or between healthy individuals and an individual with a genetic disease.

Differences were also found with specific aspects of the compared applications. There is software that analyzes a fix number and nonselectable number of enzymes [5,6]. As compared to those applications that show the enzymes that can be analyzed, I found that the 242 enzymes included in SQR is an intermediate number between the 411 enzymes analyzed by WebCutter and the 206 enzymes available in TACG Restriction mapping. Compared to all them, SQR has the advantage that it allows the user to select the enzymes individually for analysis. Regarding the graphical display of the results, most applications except NEBcutter, simply display the nucleotides of sequence and the position of the enzymes. In contrast, our application also offers a graphical display that can be used for the presentation of the results. Also, the appearance of the agarose gel that would be obtained in the laboratory after restriction of DNA is provided in the single sequence application; this is ideal for teaching restriction analysis and it is a feature that is exclusive to SQR. Finally, compared to software that analyzes single sequences, the TACG restriction mapping and the Nebcutter applications lack one report of the fragment sizes after the restriction analysis; however, the WebCutter and SQR do show that information.

I have shown that SQR has several utilities for the enzymatic

J Comput Sci Syst Biol ISSN: 0974-7230 JCSB, an open access journal restriction analysis of sequences of nucleotides, including the RFLP analysis of DNA that contains polymorphisms. I demonstrate the applicability of SQR in examples of different types of nucleotide sequences, including oncogenes and genes involved in hereditary diseases. The main limitation of SQR is that it is an executable (*.exe) application and it is not a multi-platform; therefore it only runs in the Windows OS environment, which limits its widespread utilization. Having that limitation in mind, future work includes the translation of the C++ code in SQR to Java and implementing its interfaces in Html5. Future research will be focused on developing new bioinformatics software to cover different aspects of genetic engineering. Because its applicability, I conclude that SQR is a competitive application since it offers particular features that are useful for teaching and in laboratories of molecular biology.

Acknowledgment

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

- Wyman AR, White R (1980) A Highly Polymorphic Locus in Human DNA. PNAS 77: 6754-6758.
- Osborn AM, Moore RB, Timmis KN (2000) An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environmental Microbiology 2: 39-50.
- Konieczny A, Ausubel FM (1993) A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4: 403-10.
- Jeffreys AJ, Wilson V, Thein SL (1985) Individual- specific fingerprints of human DNA. Nature 316: 76-9.
- 5. Vincze T, Posfai J, Roberts RJ (2003) NEBcutter: a program to cleave DNA with

restriction enzymes. Nucleic Acids Res 31: 3688-3691.

- San Millán RM, Martínez-Ballesteros L, Rementeria A, Garaizar J, Bikandi J (2013) Online exercise for the design and simulation of PCR and PCR-RFLP experiments. BMC Res Notes 6: 513.
- Garcia-Segura LM, Perez-Marquez J (2014) A new mathematical function to evaluate neuronal morphology using the Sholl analysis. J Neurosci Meth 226: 103–109.
- Perez-Marquez 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: 166-73.
- Hruban RH, van Mansfeld ADM, Offerhaus GJA, van Weering DHJ, Allison DC, Goodman SN, Kensler TW, Bose KK, Cameron JL, Bost JL (1993) K-ras Oncogene Activation in Adenocarcinoma of the Human Pancreas. A Study of 82 Carcinomas Using a Combination of Mutant-Enriched Polymerase Chain Reaction Analysis and Allele-Specific Oligonucleotide Hybridization. Am J Pathology 143: 545-554.
- Wilson JT, Milner PF, Summer ME, Nallaseth FS, Fadel HE, Reindollar RH, McDonough PG, Wilson LB.(1982) Use of restriction endonucleases for mapping the allele for beta-S-globin. PNAS 79: 3628-3631.