

SEQcutter: Web Server to cleave DNA with Restriction Enzymes

Salman Ullah Khan¹ and Khalida Naveed¹

¹Baqai Institute of Information Technology, Karachi, Pakistan

Accepted 25 Jan 2015, Available online 01 Feb 2015, Vol.3 (Jan/Feb 2015 issue)

Abstract

Sequence cutter, is a program a web server that will accept an input DNA sequence and produce a comprehensive report of the restriction enzymes that will cleave the sequence. It produces a variety of outputs including restriction enzyme maps, theoretical digests and links into the restriction enzyme database, Importantly, its table of recognition sites is updated daily from REBASE and it marks all sites that are potentially affected by DNA methylation. Many options exist to choose the enzymes used for digestion, including all known specificities, subsets of those that are commercially available or sets of enzymes that produce compatible termini.

Keywords: Web Server, DNA etc.

Introduction

More than 3000 type II restriction endonucleases have been discovered. They recognize short, usually palindromic, sequences of 4–8 bp and, in the presence of Mg²⁺, cleave the DNA within or in close proximity to the recognition sequence (1,2). REBASE contains comprehensive information about restriction enzymes, DNA methyltransferases and related proteins such as nicking enzymes, specificity subunits and control proteins (3). In the 1970s they were used extensively to provide physical maps of small DNAs and, in the 1980s, were used to map large DNAs. As methods for the determination of DNA sequence improved, it became convenient to search those sequences for potential restriction enzyme cleavage sites since that would permit the facile further manipulation of specific fragments. In fact, until the advent of the polymerase chain reaction (PCR) (4), they provided the most convenient way to manipulate individual genes and move them from one vector to another. For a while, it seemed that the ability of PCR to permit precise amplification of individual stretches of DNA might render the use of restriction enzymes obsolete. However, they merely found new utility by then serving as diagnostic reagents to show that DNA constructs had been made correctly. They still provide one of the cheapest and most convenient ways to characterize DNA constructs. They have also found use in analyzing the genomes of higher organisms using restriction fragment length polymorphisms (RFLPs) as physical markers (5) or by directly detecting the presence of single nucleotide polymorphisms (SNPs) (6). Since most DNA constructs are now quickly sequenced, tools to

locate restriction enzyme sites within these constructs are especially valuable. Every commercial software package to manipulate DNA sequences always includes one or more modules to detect restriction enzyme recognition sites. However, most rely on data files of recognition sites that are out-of-date, do not have links into REBASE or are cumbersome to use.

Materials and methods

Availability and requirements Lists the following:

Project name: SEQcutter

Project home page: Web Based

Operating system(s): Platform independent

Scripting language: HTML, CSS, Javascript PHP

Detailed Design

Algorithm for linear display

```
$str = $_POST['sequence'];  
$str = preg_replace('/\s*/', '', $str);  
$sequence = str_split($str, 75);  
$replacement = array(  
    'AseI' => 'attaat',  
    'VspI' => 'attaat',  
    'BspHI' => 'tcatga',  
    'AlwNI' => 'ctggt',  
    'EcoRI' => 'GAATTC',  
    'EcoRII' => 'CCWGG',  
    'BamHI' => 'GGATCC',  
    'HindIII' => 'AAGCTT',  
    'TaqI' => 'TCGA',
```

```

'Smal' => 'CCCGGG',
'HaellI' => 'GGCC',
'SpeI' => 'ACTAGT',
'SphI' => 'AGGCCT',
'Sau3A' => 'GATC',
'PvuII' => 'CCTAGG',
'AvrBII' => 'CAGCTG',
'AvrII' => 'CCTAGG',
'AvrIII' => 'TGCGCA',
'AsuNHI' => 'GCTAGC',
'AspLEI*' => 'GCGC',
'EcoRII' => 'CCWGG',
'BamHI' => 'GGATCC',
'HindIII' => 'AAGCTT',
'TaqI' => 'TCGA',
'Smal' => 'CCCGGG',
'HaellI' => 'GGCC',
'SpeI' => 'ACTAGT',
'SphI' => 'AGGCCT',
'Sau3A' => 'GATC',
'PvuII' => 'CCTAGG',
'AvrBII' => 'CAGCTG',
'AvrII' => 'CCTAGG',
'AvrIII' => 'TGCGCA',
'AsuNHI' => 'GCTAGC',
'AspLEI*' => 'GCGC',
);

```

Algorithm for graphical display

```

$INFO = array();
$counter = 1;
foreach($sequence as $pattern) {
    $pattern = strtolower($pattern);
    $complimentary = str_replace('a', 'z', $pattern);
    $complimentary = str_replace('g', 'y', $complimentary);
    $complimentary = str_replace('t', 'a', $complimentary);
    $complimentary = str_replace('c', 'g', $complimentary);
    $complimentary = str_replace('z', 't', $complimentary);
    $complimentary = str_replace('y', 'c', $complimentary);

    foreach($replacement as $key => $value) {
        $position = strpos($pattern, $value);
        if ($position !== false) {
            $INFO[$key][] = $position + $counter + 1;
            if ($isLinear) echo "<br /><span style='visibility:hidden>" . str_ireplace($value, "</span><span style='color:red>" . $key . "</span><span style='visibility:hidden>", $pattern) . "</span>";
        }
    }
    if ($isLinear) {
        echo "<br />" . $pattern;
        echo "<br />" . $complimentary;
    }
    foreach($replacement as $key => $value) {
        $position = strpos($complimentary, $value);

```

```

        if ($position !== false) {
            $INFO[$key][] = $position + $counter + 1;
            if ($isLinear) echo "<br /><span style='visibility:hidden>" . str_ireplace($value, "</span><span style='color:red>" . $key . "</span><span style='visibility:hidden>", $complimentary) . "</span>";
        }
    }
    if ($isLinear) {
        echo "<br />";
        echo "<br />";
    }
    $counter += 75;
}
if ($isLinear) {
    echo '<table>
<tr>
<th>Enzyme Name</th><th>Number of
Cuts</th><th>Positions of sites</th>
</tr>';
    foreach($INFO as $enzyme => $total) {
        $count = count($total);
        $positions = implode(", ", $total);
        echo
"<tr><td>$enzyme</td><td>$count</td><td>$positions<
/td></tr>";
    }
    echo "</table>";
} else {
    $data = "[['Name', 'Position of sites', 'Number of cuts',
'Region']";
    foreach($INFO as $enzyme => $total) {
        $i = 1;
        foreach($total as $num => $position) {
            $data .= ",['$enzyme', $position, $i, '$enzyme']";
            $i++;
        }
    }
}

```

Using SEQcutter

SEQcutter accepts an input sequence, which can be pasted in, picked up from a local file. The program calculates the positions of restriction enzyme sites where they are going to cut the sequence. It then displays sequence in linear form and graphical form.

The initial display also shows the enzymes that can be used in a complete digest to excise each ORF that is displayed. Figure 2 shows linear and Figure 3 graphical displays are offered. The particular enzymes are highlighted and moving the mouse over an enzyme name will produce a box with the recognition sequence noted and the sequence itself becomes underlined in the display.



Figure 1: Home page of SEQcutter

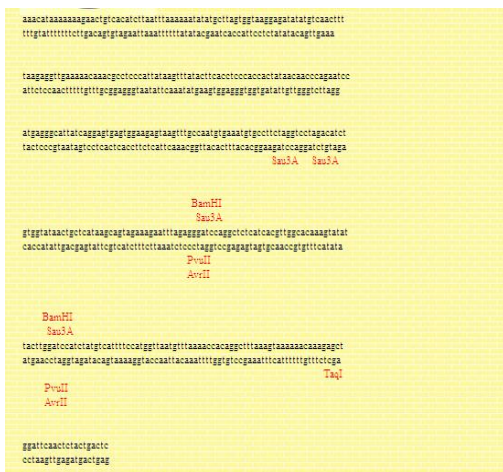


Figure 2: The linear display of a SEQcutter which shows restriction enzymes in highlighted form



Figure 3: The linear display of a SEQcutter which shows restriction enzymes in graphical form



Figure 4: Contact us page

Source Code

SEQcutter Source code can be downloaded from this link: <https://www.dropbox.com/sh/ow2m5pw29kg20nf/AADAo5MY0LuDMj42QcMsHFMqa?dl=0>

Results

SEQcutter shows the restriction mapping of nucleotide sequences and analysis of nucleotide sequences is that of searching for restriction sites.

Acknowledgement

I would like to thank Miss Nadia Rasheed for her support and to give idea for this project.

References

- [1]. Pingoud, A. and Jeltsch, A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.*, 29, 3705–3727.
- [2]. Roberts, R.J. and Halford, S.E. (1993) Type II restriction enzymes. In Linn, S.M., Lloyd, R.S. and Roberts, R.J. (eds). *Nucleases*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 35–88.
- [3]. Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2003) REBASE—restriction enzymes and ethyltransferases. *Nucleic Acids Res.*, 31, 418–420.
- [4]. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239, 487–491.
- [5]. Lander, E.S. and Botstein, D. (1986) Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. *Proc. Natl Acad. Sci. USA*, 83, 7353–7357.
- [6]. Hao, K., Niu, T., Sangokoya, C., Li, J. and Xu, X. (2002) SNPkit: an efficient approach to systematic valuation of candidate single nucleotide polymorphisms in public databases. *Biotechniques*, 33, 822–828.