WebFARM: web server for finite automated restriction mapping

Tiratha Raj Singh*

Bioinformatics Centre, School of Biotechnology, DAVV, Indore, India; Tiratha Raj Singh E-mail: tiratharaj@gmail.com, *corresponding author

Abstract:
Restriction endonucleases are indispensable tools in molecular biology and biotechnology. Type II restriction endonucleases are part of restriction modification systems. DNA fragment extraction and restriction mapping are the basis for several biotechnological activities. WebFARM is a server application for identifying restriction endonuclease recognition sites and to give information regarding restriction mapping for given nucleotide sequences. WebFARM analyses given nucleotide sequence and identify restriction site for selected restriction endonucleases. It will also provide frequency of restriction for each restriction endonuclease.

Availability: http://webfarm.bioinfoindia.org/

Background:
Restriction endonucleases (REnases) are part of restriction and modification (RM) systems, ubiquitous among bacterial DNA **[1, 2].** RM systems were originally suggested to evolve as a defense mechanism against phage infection and other type of DNA invasion. Type II REnases are intensely studied enzymes from the structure-function perspective **[3, 4].** The main criterion for the classification of type II REnases is their high specificity of cleavage within or close to their recognition site and they do not require ATP hydrolysis for their nucleolytic activity **[5].** This group of enzymes constitutes one of the largest families of enzymes with the same basic function, which makes type II REnases ideal objects to manipulate biological sequences in molecular biology and biotechnology **[6].**

REnases are robust, cheap, and widely available tools for analyzing and manipulating DNA sequence. Main function of REnases is to defend their host against foreign DNA, which is achieved by cleaving incoming DNA that is recognized as foreign element at defined sites within the recognition sequence **[2, 6, 7].** Such cleavage and resulting DNA fragment extraction is an important and common activity of genetic engineering. Restriction-site mapping involves locating certain restriction sites on some sequence of DNA **[8, 9]**. There is a continuous need to generate restriction site, recognition sequence, and other related information to generate restriction map for specific DNA sequences.

Description:
Here, I develop a web server (WebFARM) that work on the principle of finite automata and recognize REnases for a given DNA sequence based on the recognition sequence of the respective REnases. It gives all possible REnases and their respective recognition sites for the given DNA sequence. Additionally it gives the position of recognition and total number of recognition sites for a particular REnase. Flowchart representing the working principle of WebFARM is shown in **Figure 1**.
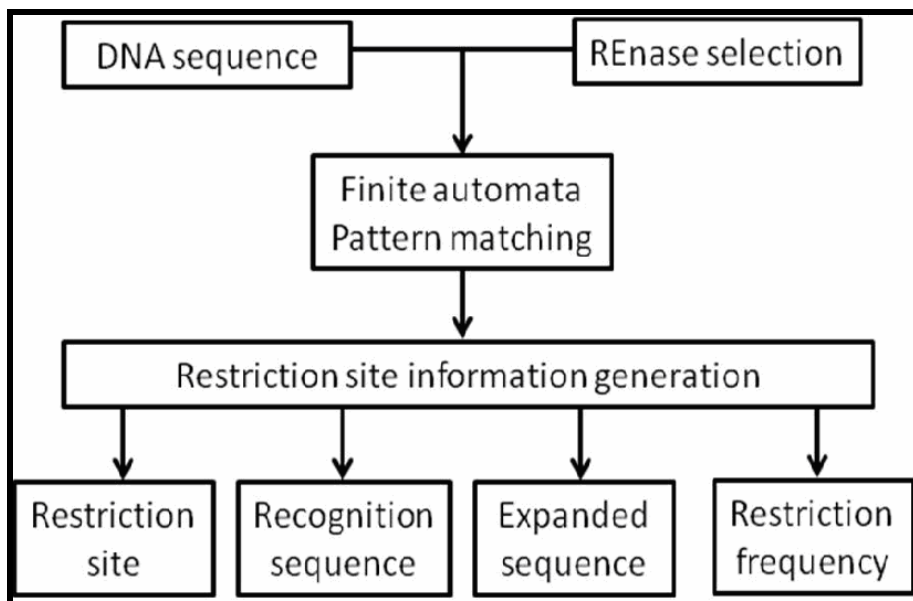

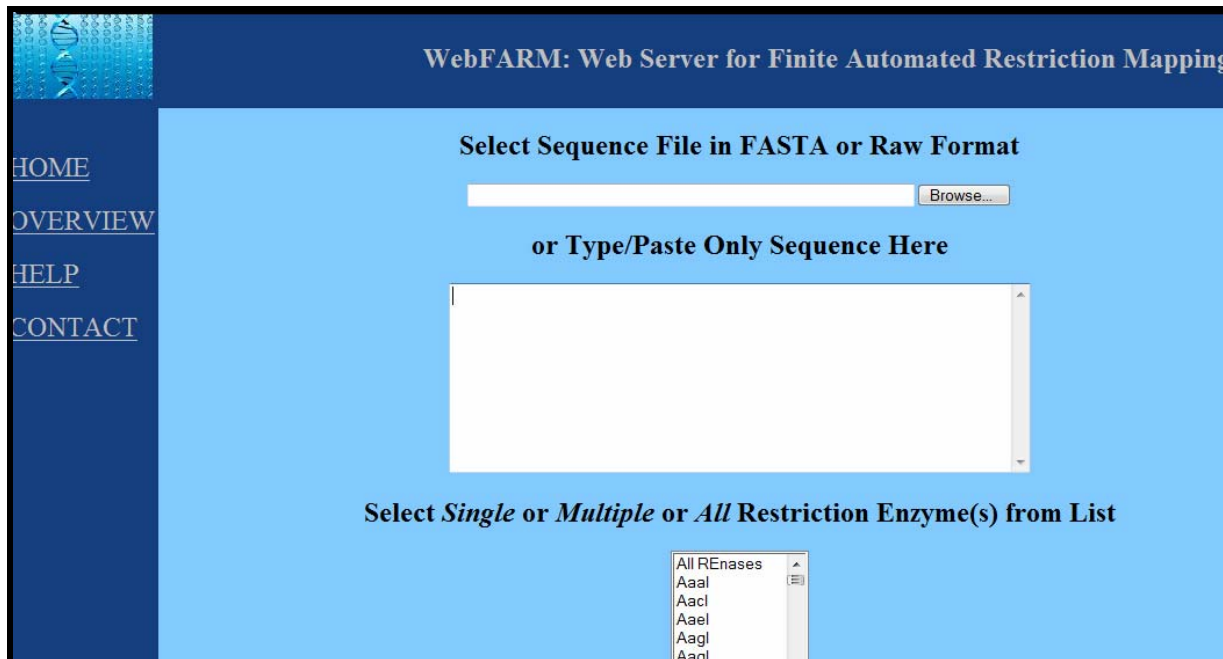**Figure 1:** Flowchart representing the working principle of WebFARM.

Figure 2: A screenshot of WebFARM with options and menu items. This page allows the user to input DNA sequence or sequence file and to select REnase(s) for restriction site information
.

The main interface of the WebFARM is a graphical display with all options and menu available in one screenshot (**Figure 2**). WebFARM accepts as input a set of DNA sequence and REnases. Only DNA sequence can be typed or pasted in the given text box or can be provided as a file input in FASTA format. Name of all REnases are given which can be selected as a single, multiple or all REnases from the given list.

WebFARM initially does a simple pattern matching job and looks for sequences which resemble with recognition sequences of the selected REnases. It will work on the principle of finite automaton matcher **[10]**, and will look for a particular pattern as REnase for the given DNA sequence as shown for Restriction Enzyme Matcher (**see supplementary material**).

It will scan REBASE **[11]** for REnases and enables the detection of respective recognition sites as well as recognition sequences for all or selected REnases. It will also provide frequency of restriction for each REnase. Final tabular output of WebFARM will provide name(s) of REnase, recognition sequence and site of restriction, expanded recognition sequence for ambiguous nucleotide characters, and the frequency of restriction for each REnase.

An online manual is provided that describes all the operations and methodology in detail. This can be accessed through help and overview options in the menu of WebFARM.

References:
[1] E Luria & ML Human, *J. Bacteriol.* (1952) **65**:557-569 [PMID: 12999684]
[2] A Pingoud & A Jeltsh, *Nucleic Acids Res.* (2001) **29/18**:3705-3727 [PMID: 11557805]
[3] K Kusano *et al.*, *Proc. Natl. Acad. Sci. USA* (1995) **92**:11095-11099 [PMID: 7479944]
[4] JM Bujnicki, *Acta Biochim. Pol.* (2001) **48/4**:935-967 [PMID: 11996004]
[5] RJ Roberts *et al.*, *Nucleic Acids Res.* (2007) **35**:D269-D270 [PMID: 17202163]
[6] A Pingound *et al.*, *Cell. Mol. Life Sci.* (2005) **62**:685-707 [PMID: 15770420]
[7] RA Kovall & B.W. Matthews, *Curr. Opin. Chem. Biol.* (1999) **3**:578-583 [PMID: 10508668]
[8] WR. Pearson, *Nucl. Acids Res.* (1982) **10**:217-227 [PMID: 6278405]
[9] L Allison & CN Yee, *CABIOS* (1988) **4/1**:97-101 [PMID: 2838139]
[10] TH Cormen *et al.*, *Introduction to Algorithms,* PHI Publications (2004) 906-932
[11] RJ Roberts *et al.*, *Nucleic Acids Res.* (2003) **31/7**:1805-1812 [PMID: 12654995]

Supplementary material:

Restriction Enzyme Matcher (S, δ, m)

$\quad\quad N \leftarrow$ length $[S]$

$\quad\quad q \leftarrow 0$

$\quad\quad$ for $i \leftarrow 1$ to n

$\quad\quad\quad\quad\quad$ do $q \leftarrow \delta(q, S[i])$

$\quad\quad\quad\quad\quad$ if $q = m$

$\quad\quad\quad\quad\quad$ then output = "Pattern P occurs with the position/shift" $i - m$

where $S[1…n]$ is the input DNA string, pattern P is of length m, the state set Q is $\{0, 1, …, m\}$, the start state is 0 and the only accepting state is m. δ is the transition function of automata.