

VBRC Base By Base: Search for a sequence motif, either exact (regular expression search) or inexact (fuzzy motif search)

Nick Tang

Abstract

Viral Bioinformatic Resource Centre

- **Provide databases of viral genomic information.**
 - Please check the **Organisms** menu to see which viruses we support: we're now focusing on large DNA viruses
 - The VOCs (Virus Orthologous Clusters) database is at the heart of our system.
 - The database links directly to integrated tools for comparative analyses.
 - VOCs sorts genes into ortholog clusters (e.g. RNA polymerase) to simplify data retrieval.
- **Provide easy access to the genes, gene families, and genomes of the different virus families.**
 - via a unique series of powerful Java tools that support multiple computer platforms (see **VBRC Tools** menu).
 - design and build software to tackle specific bioinformatics/virology problems, often in collaboration with virologists.
- **Rally the research community to provide expert curation of these viral genomes by:**
 - Adding value to GenBank sequences through enhancing and updating genome annotations
 - Linking to research reviews/papers for the research community.
- **Collaborate with researchers to help on specific bioinformatics problems, e.g.**
 - Custom searches of the databases
 - Building new features into our tools
 - Help with genome annotation

Citation: Nick Tang VBRC Base By Base: Search for a sequence motif, either exact (regular expression search) or inexact (fuzzy motif search). **protocols.io**

dx.doi.org/10.17504/protocols.io.eczbax6

Published: 04 Jan 2016

Guidelines

Â

Â

Protocol

Step 1.

In BBB it is possible both to search for an exact sequence motif (regular expression) and a motif with several mismatches (fuzzy motif). Both functions are available from the *Tools* menu.

Step 2.

Suppose we wish to find a CCTGGC pattern with no mismatches.

To perform this search, first select the sequences you wish to search (click on their names in the Sequence List; hold down the **CTRL** or **Apple** key to select more than one).

Step 3.

Then, from the *Tools* menu, select **“Search”** and then click on **“Regular Expression Search”**.

In the box that appears, enter the expression (CCTGGC) that you wish to search for.

Step 4.

Now, suppose we want to find all matches to CCTGGC that are exact or almost exact — that is, that have at most one nucleotide altered.

Again, begin by selecting the sequences to be searched.

Step 5.

Then, from the *Tools* menu, select **“Search”** and then click on **“Fuzzy Motif Search”**.

Step 6.

In the box that appears, enter the expression (CCTGGC) that you wish to search for and, in the box below, set the maximum number of mismatches to 1.

Step 7.

In either case, the program will then search both the top and bottom strands of the selected DNA sequence for the given expression.

Step 8.

After the search is complete, it will return a list of matches.

You can jump to any location of any match by double clicking on it in the result list. Finally, the list of fuzzy search results can be saved to a text file by clicking on **“Save Fuzzy Results”**.