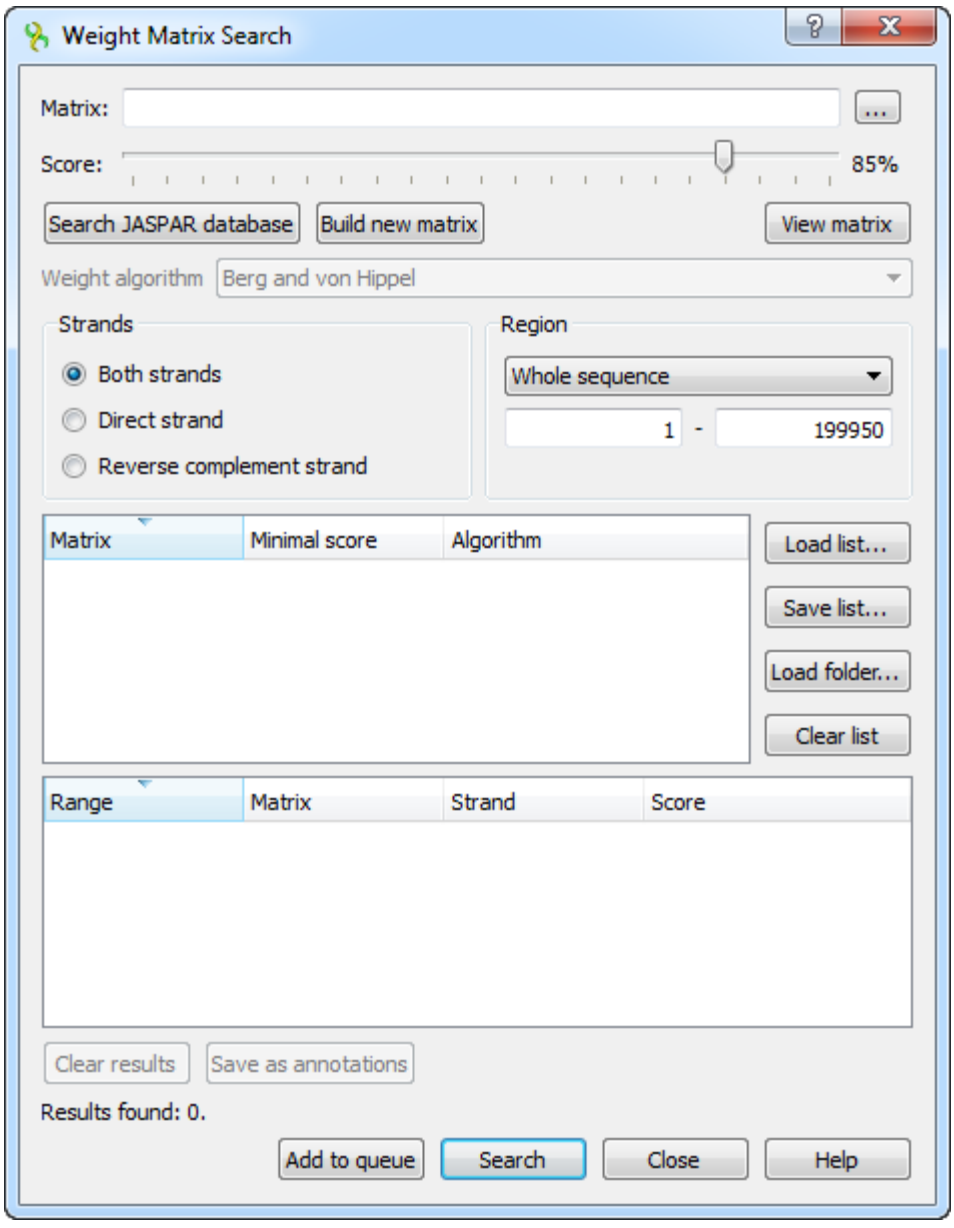
Weight Matrix

The *Weight Matrix* plugin is a tool for solving the problem of a sequence annotating. As well as for the *SITECON*, the main use case of the plugin is recognition of potential transcription factor binding sites on basis of the data about conservative conformational and physicochemical properties revealed with the binding sites sets analysis.

The *Weight Matrix* contains a lot of *position frequency matrices* (PFM 's) and *position weight matrices* (PWM 's, also known as *position specific score matrices* — PSSM 's). The matrices came from two wide-known open archives: JASPAR, which contains frequency matrices, and UniPROBE containing weight matrices.

Also the *Weight Matrix* plugin provides a tool for creating specific position frequency and weight matrices from an existing alignment or from a file with several sequences. The created matrix can be used as a profile for the search as well as the JASPAR and UNIPROBE ones.

To search for transcription factor binding sites in a DNA sequence select the *Analyze Search TFBS with matrices* context menu item. The *Weight matrix search* dialog will appear:



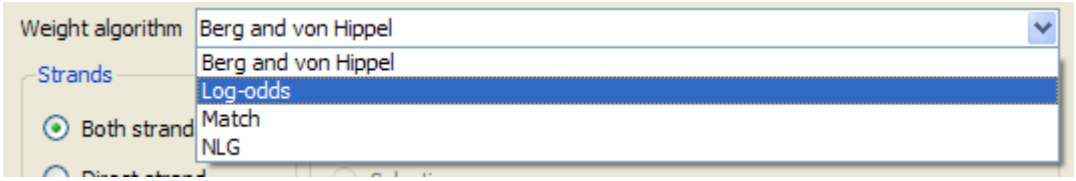
In the search dialog you must specify a file with PWM or PFM. You can do so by pressing the browse button and selecting the file.

Also you can use the *special interface to choose a JASPAR matrix* by pressing the *Search JASPAR database* button.

Alternative way to specify the position weight/frequency matrix is to create a specific one from an alignment or a file with several sequences with the *build a new matrix* tool.

After the profile (the matrix) is loaded, you can adjust the threshold value. The threshold sets the minimal identity score for a result to pass. The more the result score is, the more it is homologically related to the aligned region. By changing the threshold you can filter low- scoring results.

If the loaded matrix is a position frequency matrix, you must also specify the algorithm to build the corresponding position weight matrix which will represent the transcription factor. There are four algorithms available.

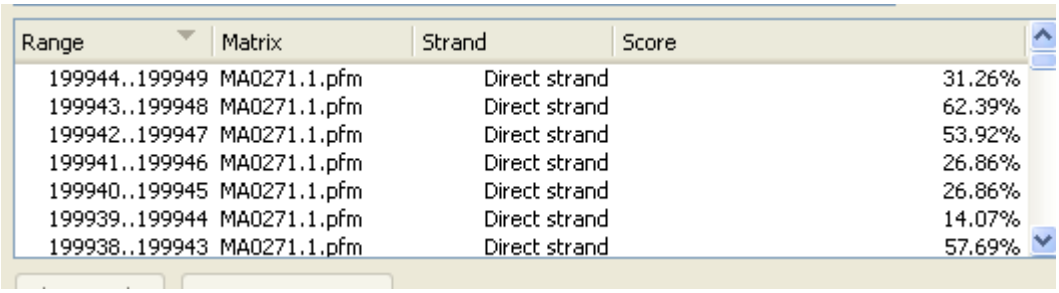


Also you can add a selected matrix with the specified *Minimal score* and the *Algorithm* to the matrices list. To do it, select the matrix and other options and press the *Add to queue* button. The plugin will search with all matrices specified in the list.

You can use the *Save list* button to export the list of matrices to a *.csv file. Later the list can be loaded from the file using the *Load list* button.

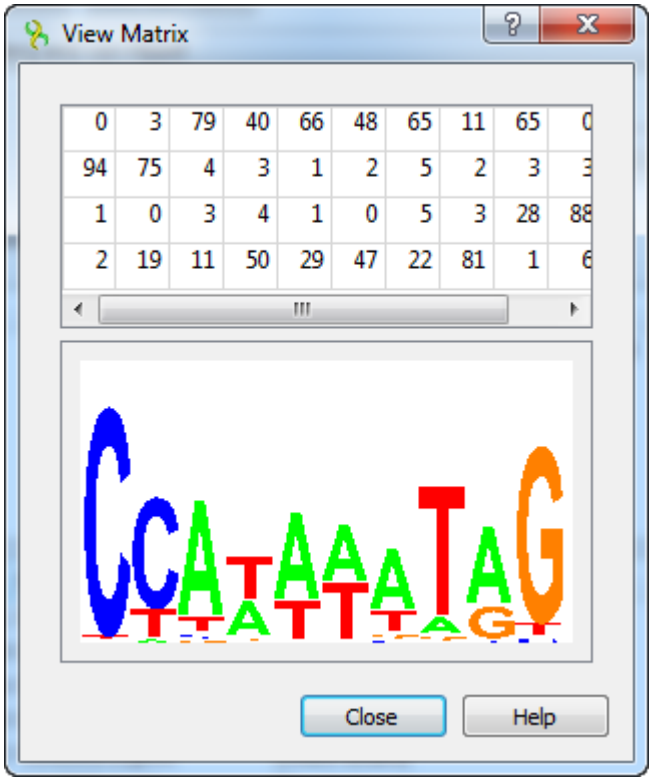
The rest options are standard sequence search options: the strand and the sequence region where to search for matches.

After specifying the necessary options press the *Search* button. The found results will appear in the dialog table. The corresponding results identity scores are in the *Score* column.



Range	Matrix	Strand	Score
199944..199949	MA0271.1.pfm	Direct strand	31.26%
199943..199948	MA0271.1.pfm	Direct strand	62.39%
199942..199947	MA0271.1.pfm	Direct strand	53.92%
199941..199946	MA0271.1.pfm	Direct strand	26.86%
199940..199945	MA0271.1.pfm	Direct strand	26.86%
199939..199944	MA0271.1.pfm	Direct strand	14.07%
199938..199943	MA0271.1.pfm	Direct strand	57.69%

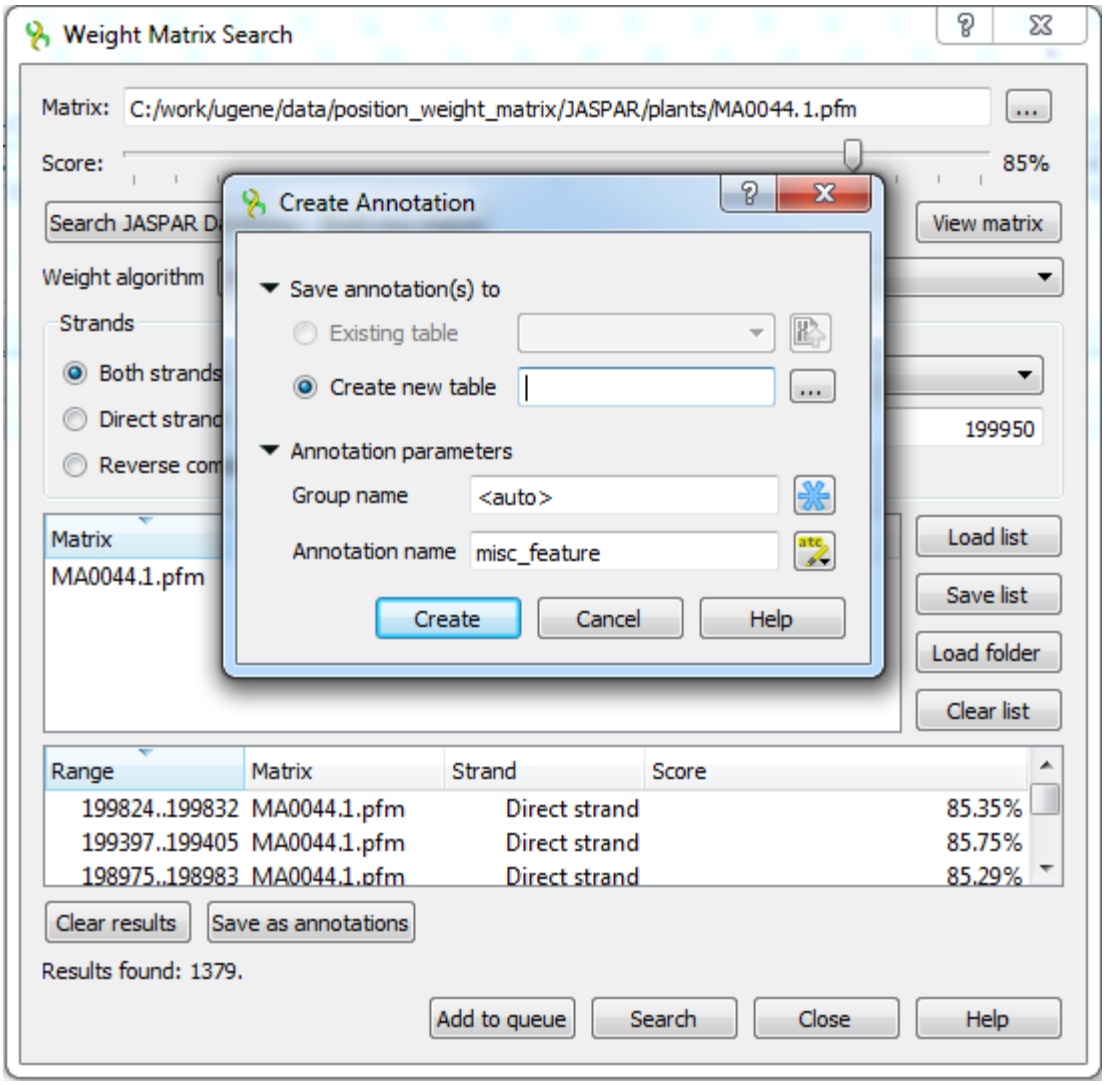
Also you can see the matrix by using the *View matrix* button:



The regions found by the weight matrix algorithm can be saved as annotations to the DNA sequence in the Genbank format by pressing the *Save as annotations* button.

After saving, the file with resulting annotations will be automatically added to the current project, and the annotations will be added to the original sequence.

Note that in case of selecting JASPAR or UNIPROBE matrix, the resulting annotations will contain the given matrix properties.



See also:

- Searching JASPAR Database
- Building New Matrix