BacPred User’s Manual

(Version 1.0)

BacPred Documentation (last updated: August 17, 2022)

Washington State University, Tri-cities, Richland, WA 99354, USA

For more information, please contact Suraiya Akhter (Email: [suraiya.akhter@wsu.edu](mailto:suraiya.akhter@wsu.edu))

To predict bacteriocin sequences, **BacPred** is implemented that is developed by python3. It is standalone software tool and most of scripts are written in R and main scrip is called by **BacPred.ipynb**.

**Installation steps:**

* Download all from Github link-https://github.com/suraiya14/ML\_bacteriocins/tree/main/BacPred and save it in a directory.
* You will need to install “Jupyter Notebooks” if you want to use the notebook version of this program. For installation, visit the link: https://www.anaconda.com/download/ and install anaconda for python 3. To install it from a terminal window, please visit the link: <https://jupyter.org/install>.
* Next, use the ‘cd’ command to go to the directory where the “BacPred” has been saved and then type the command “jupyter notebook”.
* The list of input sequences should be given fasta file format named “input\_seq.fasta” to include sequences you want to test.

After a couple of seconds, a new page for Jupyter Notebook will appear in your default browser. Click on **BacPred.ipynb**. BacPred will open up in a new browser window or tab. A screenshot is shown below.

**Before running the notebook, please install the rpy2 package using the command “conda install -c r rpy2” from the command prompt. Also, please install the tzlocal package using command “conda install -c conda-forge tzlocal” if needed. For details, please visit** [**https://anaconda.org/r/rpy2**](https://anaconda.org/r/rpy2) **and** [**https://anaconda.org/conda-forge/tzlocal**](https://anaconda.org/conda-forge/tzlocal) **for the rpy2 and tzlocal packages, respectively. If your system cannot locate the e1071 package, please install the e1071 package using the command “conda install -c r r-e1071”. For details, please visit** [**https://anaconda.org/r/r-e1071**](https://anaconda.org/r/r-e1071) **for the e1071 packages. All of the remaining packages should come with Jupyter Notebook by default. In the worst case, if any package is missing, please use the ‘conda install’ command to install the required packages. Alternatively, users can use the ‘pip install’ command to install the necessary packages rather than using the ‘conda install’ command.**

Next, click inside a code cell and then click on the “Run” button.

If you want to stop running the notebook, please click on “Interrupt the kernel” (i.e., the stop button). If an error occurs, you can restart the kernel. Click on the “restart the kernel” button.

After starting the notebook or script, a GUI should appear as in the following screenshot.

**Graphical user interface

Description automatically generated with medium confidence**

**Upload Input Sequences:**

To upload input file (file named should be “input\_seq.fasta”) from any location, you need to click on “Open” button. You can choose “input\_seq.fasta” from any directory. Then you have to click on “Save” button. Input file upload and saving are showed in below figure 2.

|  |  |
| --- | --- |
| Graphical user interface, text, application  Description automatically generated  (a) | Graphical user interface  Description automatically generated with low confidence  (b) |

Figure 2: Input file upload and save

**Input Sequences Processes:**

Please click on the option menu to see all options available in the tool as shown in the following screenshot.

A picture containing graphical user interface

Description automatically generated

A brief description of the options is given below.

(a) Predict bacteriocin sequences: This option is for predicting bacteriocin peptide sequences from the input fasta sequences in the default “input\_seq.fasta” file.

(b) Include new bacteriocin sequences: These options are given for a user who wants to include new known AVP sequences to the original training data that comes with this tool. Again the sequences that a user wants to add need to be placed in the “input\_seq.fasta” file. Please note that the training sequences are in the “selected\_training\_merged\_file.csv” file.

(c) Include new non-bacteriocin sequences: These options are given for a user who wants to include new known non-bacteriocin sequences to the training data that comes with this tool. Again the sequences that a user wants to add need to be placed in the “input\_seq.fasta” file. Please note that the training sequences are in the “selected\_training\_merged\_file.csv” file.

(d) Restore training sets: This option allows a user to reset all training datasets back to the original versions that came with this tool.

After successful execution, the tool will show a message- “The operation is successful!”.

**Result generation:**

You can see the predicted bacteriocin sequences by clicking on “Show Predicted Results” button. The predicted results are showed in middle text box. The example is given in figure 3. Where the tool can detect 3 sequences are bacteriocin protein sequences from “input\_seq.fasta” file for instanc.

After that you need to click on “Delete Text Result” button in order that previous predicted results can remove and show every time new results. Similarly, text box will be clear after clicking on that button.

Graphical user interface

Description automatically generated