

BacPred User's Manual

(Version 1.0)

BacPred Documentation (last updated: September 25, 2022)

Washington State University, Richland, WA 99354, USA

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

BacPred is a standalone software tool that is implemented using R and python3 for predicting bacteriocin sequences. This tool can run on Windows, Linux, and macOS operating systems.

Installation steps:

- Please download all files from the Github link: https://github.com/suraiya14/ML_bacteriocins/tree/main/BacPred and save them in a directory.
- Please install “Jupyter Notebook” to use the notebook version of the script. Please refer to the link- <https://www.anaconda.com/download/> to install anaconda for python 3. Please refer to the link- <https://jupyter.org/install> to install it from a terminal. For Windows machine, please consider using Anaconda command prompt as terminal.
- Go to the directory of your saved “BacPred” tool using ‘cd’ command and then use the command “jupyter notebook” to open a new page for Jupyter Notebook in your machine’s browser.
- The list of input sequences should be given as FASTA file format like “input_seq.fasta” to include sequences that you want to test.
- Click on **BacPred.ipynb** to open it up in a new browser window or tab.

You need to install a couple of packages before running the notebook. So, please install the rpy2 and tzlocal packages using “pip install rpy2” and “conda install -c conda-forge tzlocal” commands, respectively. For any missing package, please use the “conda install” or “pip install” command to install them.

Next, click inside a code cell and then on the “Run” button to start the notebook/script as shown in Fig 1.

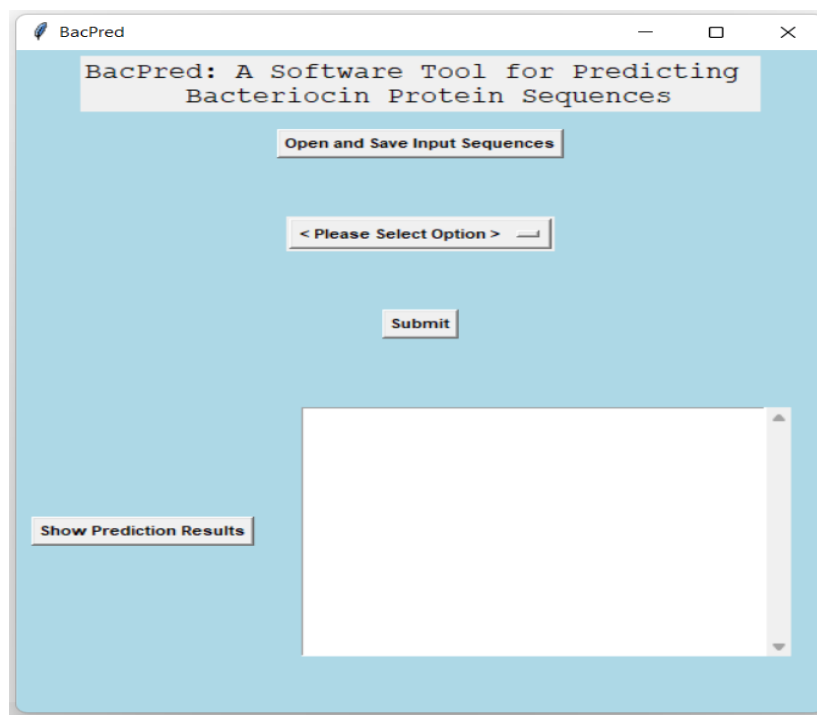


Fig 1. BacPred GUI.

Upload and Save Input FASTA Protein Sequences:

To upload input file with extension of the file “.fasta” from any location, you need to click on “Open and Save Input Sequences” button. You can choose any file with extension “.fasta” like “input_seq.fasta” from any directory. A snapshot is given in Fig 2.

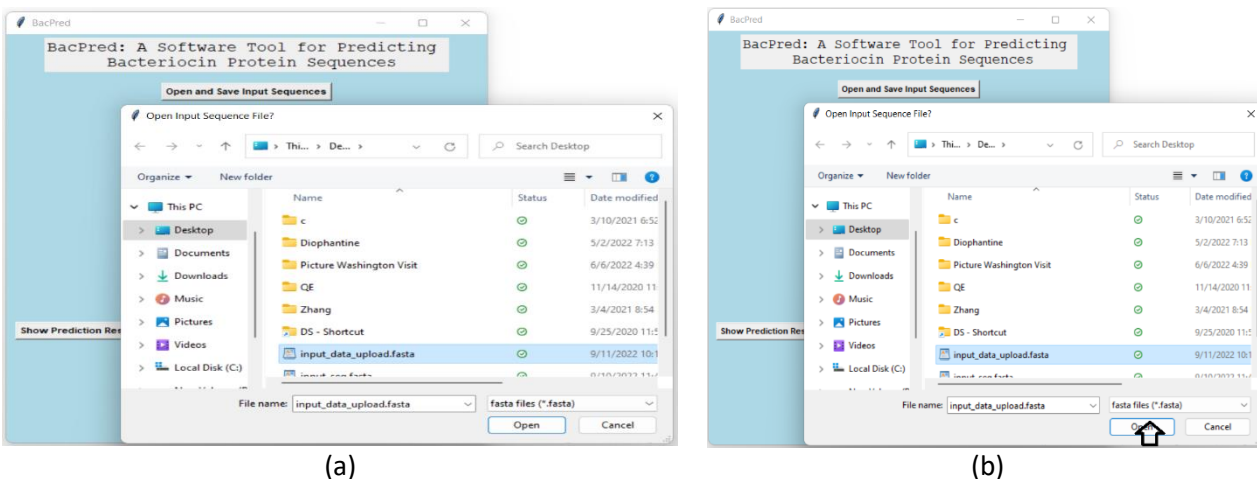
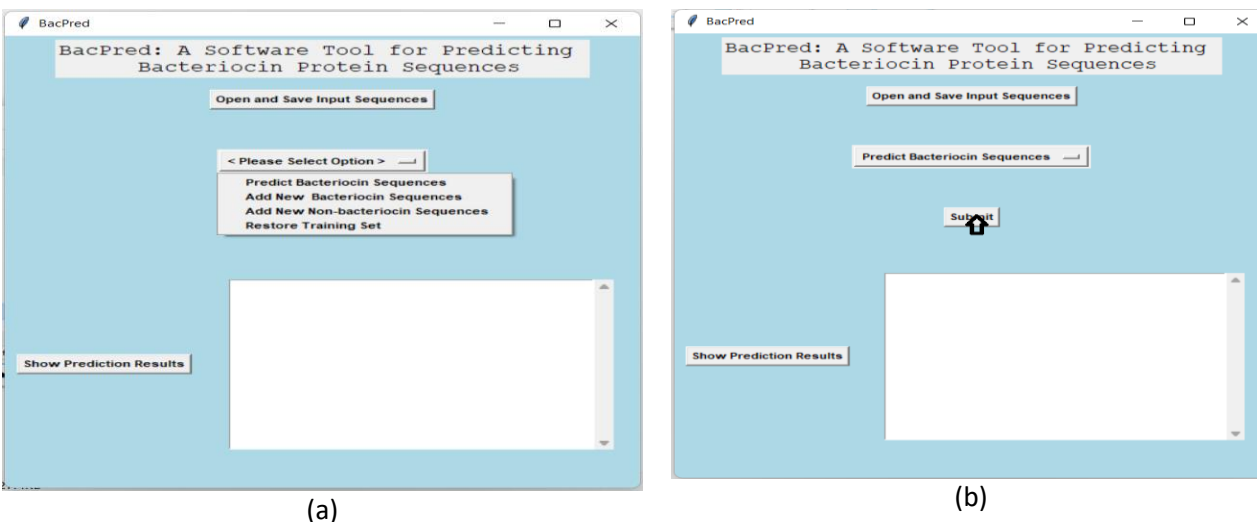
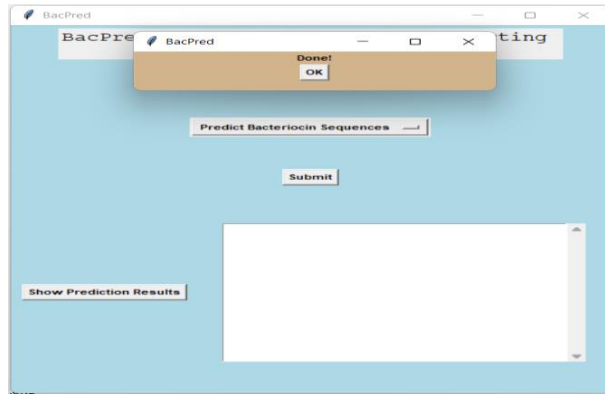


Fig 2. Uploading and saving Input file.

Service Menu:

Available service options of the BacPred are shown in Fig 3. Select appropriate service (Fig 3(a)) and then click on “submit” button (Fig 3(b)). When the execution is done, you will get a message- “Done!” (Fig 3(c)).





(c)

Fig 3. BacPred service options.

A brief description of the service options is given below.

- (a) **Predict bacteriocin sequences** for predicting bacteriocin peptide sequences from the input fasta file.
- (b) **Add new bacteriocin sequences** for adding new known bacteriocin sequences from the input fasta file to the original training dataset.
- (c) **Add new non-bacteriocin sequences** for adding new known non-bacteriocin sequences from the input fasta file to the original training dataset.
- (d) **Restore training sets** for resetting the training set back to the original training dataset supplied with this tool.

Result generation:

When the execution is done, you can see the predicted bacteriocin sequences by clicking on “Show Prediction Results” button. An example of predicted results is shown in the text box (refer to Fig 4) where the tool detected 6 sequences as bacteriocin protein sequences from the input fasta file. Also, you can see the prediction results in the “predicted_bacteriocin.fasta” file located where you saved the BacPred tool.

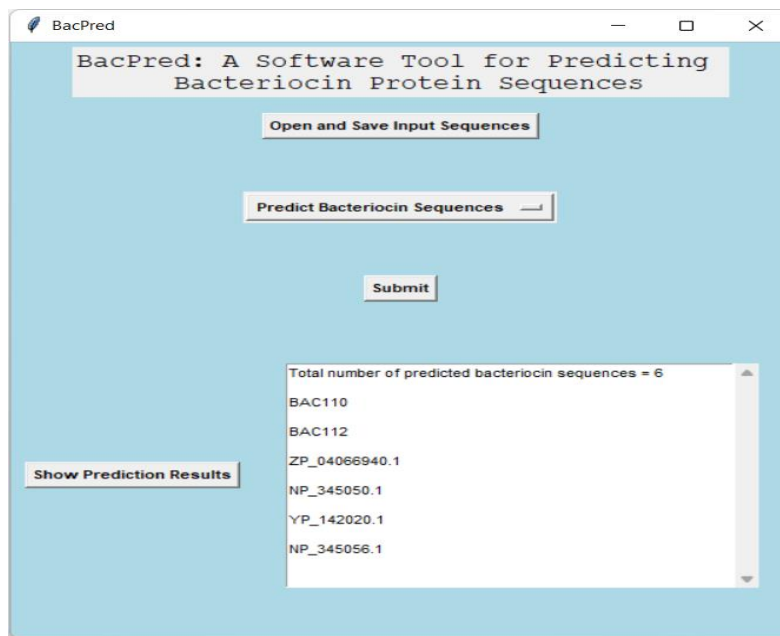


Fig 4. Sample prediction results.