

BacPred User's Manual

(Version 1.0)

BacPred Documentation (last updated: October 22, 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 repository: <https://github.com/suraiya14/BacPred> and save them in a directory.
- Please download and install R. Please refer to the link- <https://cran.r-project.org> for R installation.
- 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 python3. Please refer to the link- <https://jupyter.org/install> to install it (“Jupyter Notebook”) 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.
- You need to install some packages before running the notebook. So, please install rpy2 package using “pip install rpy2” command. The tzlocal package should automatically be installed with the rpy2 package; however, if not, please install tzlocal package using “pip install tzlocal” command. For any missing packages, please use “pip install” or “conda install” command to install them.
- Click on **BacPred.ipynb** to open it up in a new browser window or tab. Next, click inside a code cell and then on the “Run” button to start the notebook/script as shown in Fig 1.

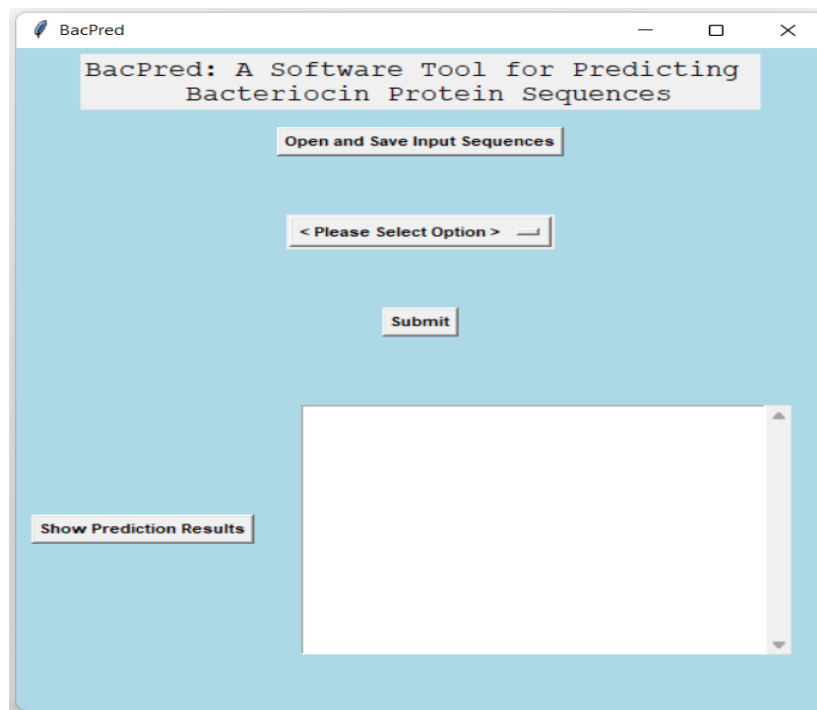


Fig 1. BacPred GUI.

Upload and Save Input FASTA Protein Sequences:

- The list of input sequences should be given as FASTA (.fasta) file format to use the file as a testing dataset or to include the sequences as additional bacteriocins or non-bacteriocins in the training dataset. A sample input file “input_seq.fasta” is supplied with this tool.
- To upload an input file from any location, you need to click on “Open and Save Input Sequences” button. You can choose any file with extension “.fasta” from any directory. A snapshot is given in Fig 2.

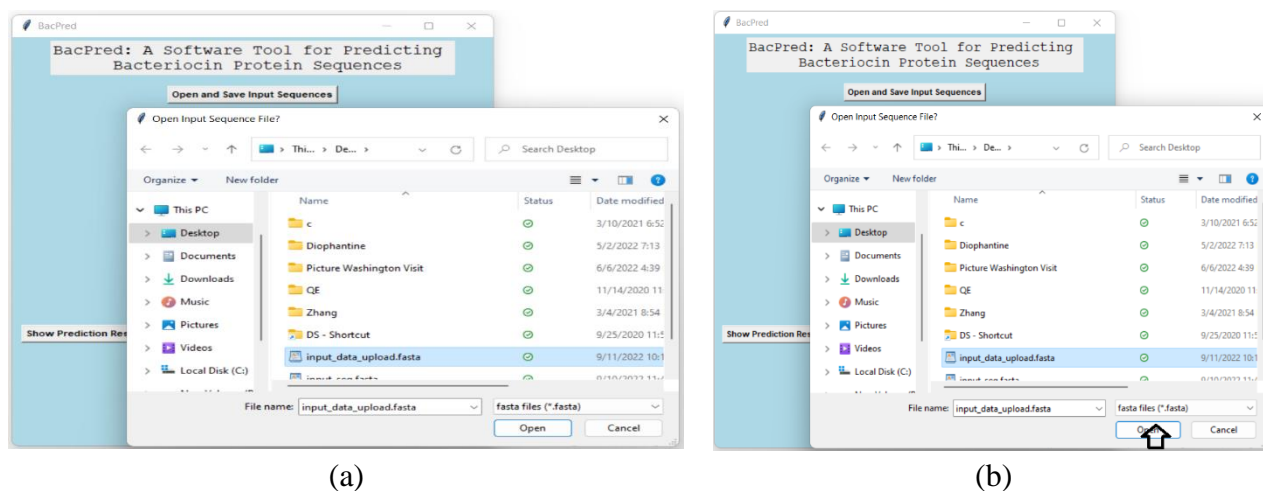


Fig 2. Uploading and saving Input file.

Service Menus:

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)). Please note that it may take a while for BacPred to complete the execution for the first time, due to the installation of all required packages and their dependencies. However, once all packages are installed, BacPred should provide the prediction results in a short time for next runs.

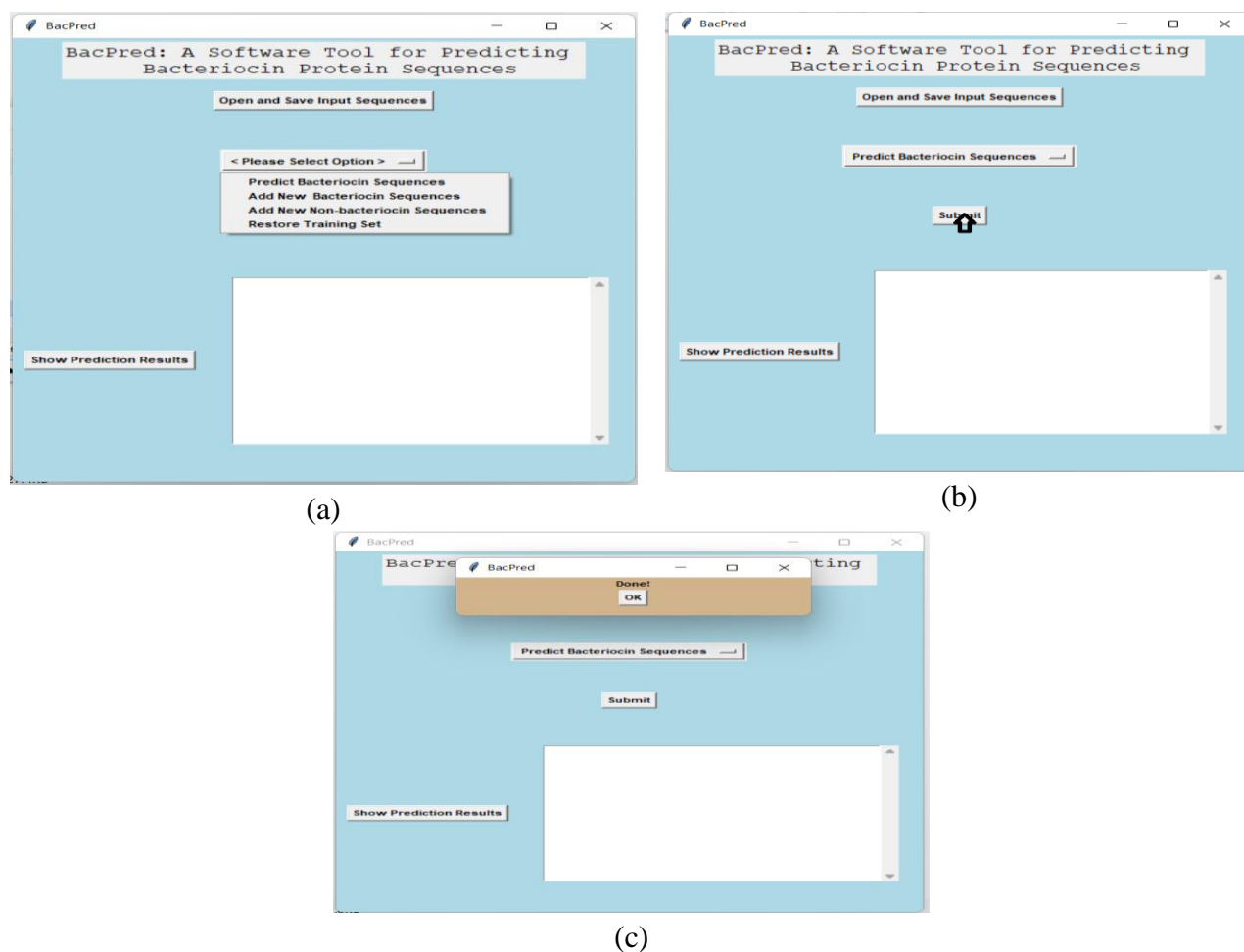


Fig 3. BacPred service options.

A brief description of the service options is given below.

- (a) **Predict Bacteriocin Sequences** for predicting bacteriocin 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 Set** 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 55 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.

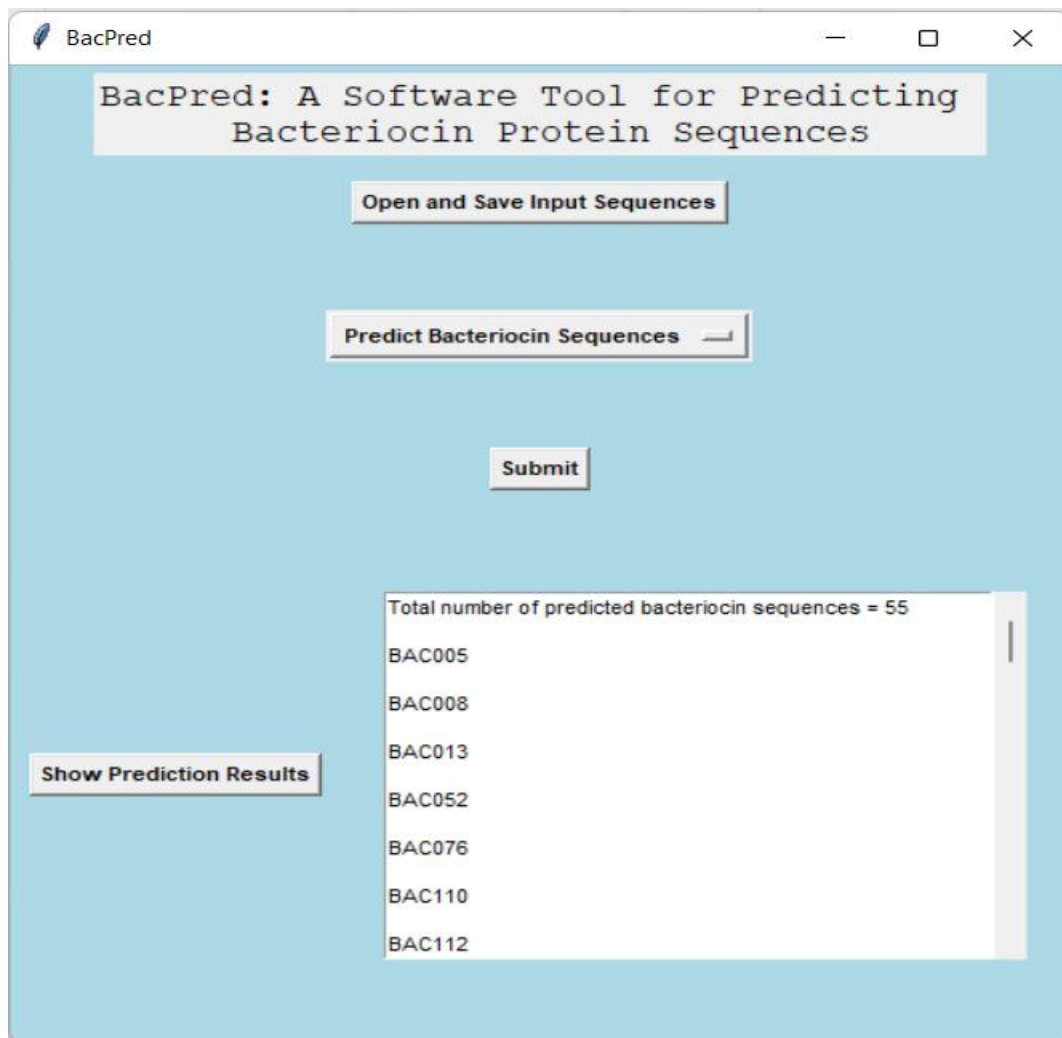


Fig 4. Sample prediction results.