BaPreS User's Manual

(Version 1.0)

Please follow the user manual for installing the tool. If you find our software tool useful, kindly cite the following paper:

• Akhter, S. and Miller, J.H., 2023. BaPreS: a software tool for predicting bacteriocins using an optimal set of features. BMC Bioinformatics, 24(1), p.313.

If our web application for this tool (https://shiny.tricities.wsu.edu/bacteriocin-prediction/) is helpful, please cite the following paper:

• Akhter, S. and Miller, J.H., BPAGS: A web application for bacteriocin prediction via feature evaluation using alternating decision tree, genetic algorithm, and linear support vector classifier. Frontiers in Bioinformatics, 3, p.1284705.

BaPreS 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/BaPreS and save them in a directory.
- Please download and install R. Please refer to the link- https://cran.r-project.org for R installation.
- For jupyter notebook user (Windows, Linux and Mac):
    - 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 "BaPreS" 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 **BAPRES.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.
- For users who want to use executable (.exe) file (Windows machine):
    - For Windows machine, please consider using Anaconda or window command prompt as terminal. Go to the directory of your saved "BaPreS" tool using "cd" command.

- 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.
- Download the .exe file from https://emailwsu-my.sharepoint.com/:u:/g/personal/suraiya_akhter_wsu_edu/EYTh1jv8Nx1Box87eEL_ti0BIQwGRWj-xhEB4X57PnMVNw, save it to the directory where you saved "BaPreS" tool, and then run this command "BAPRES.exe". You should see the GUI as shown in Fig 1. As .exe file is very large, we couldn't upload the executable (.exe) file to the GitHub repository.

- For users who want to use python/executable file (Mac and Linux machines):
  - Open a terminal and then go to the directory of your saved "BAPRES" tool using 'cd' command.
  - 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.
  - Run command "chmod +x BAPRES.py" to make the executable file.
  - Then run command "./BAPRES.py". You should see the GUI as shown in Fig 1.
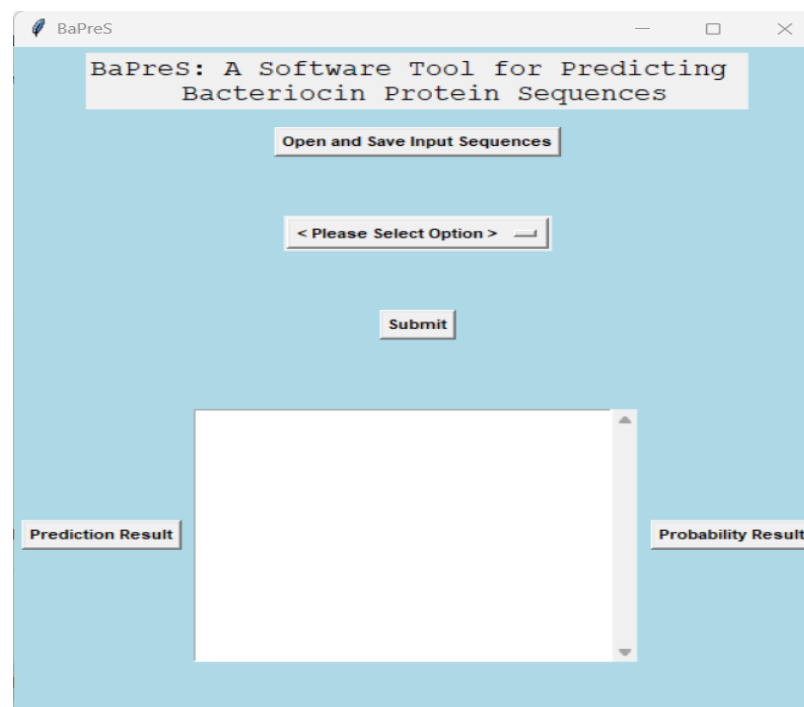
Fig 1. BaPreS GUI.

**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.

(a)   (b)
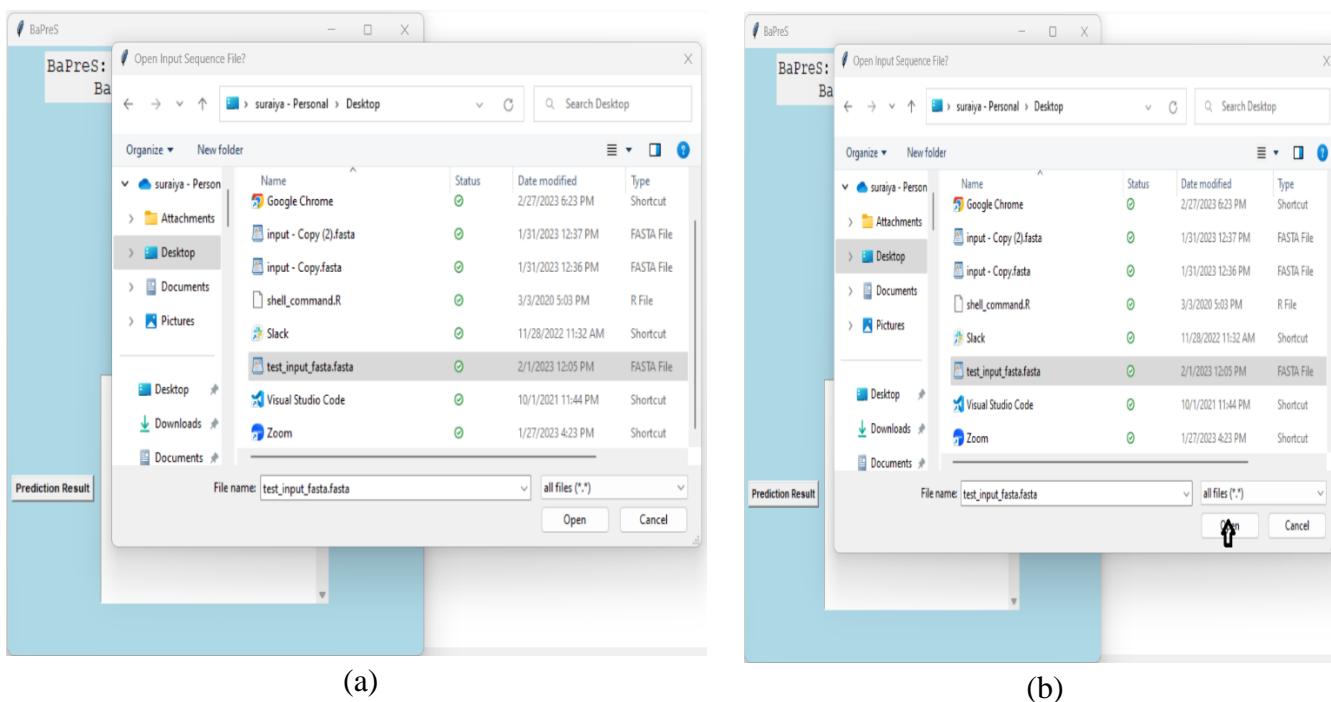
**Fig 2. Uploading and saving Input file.**

Service Menus:

Available service options of the BaPreS tool 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 BaPreS 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, BaPreS should provide the prediction results in a short time for next runs.
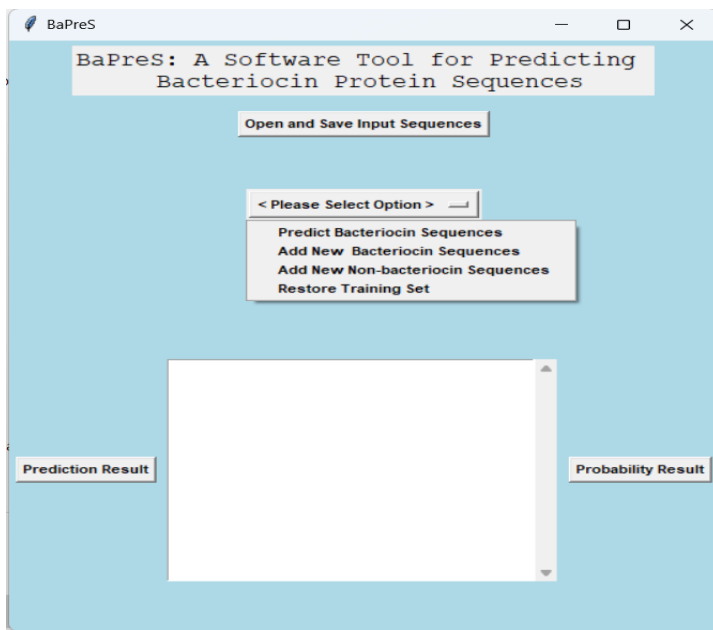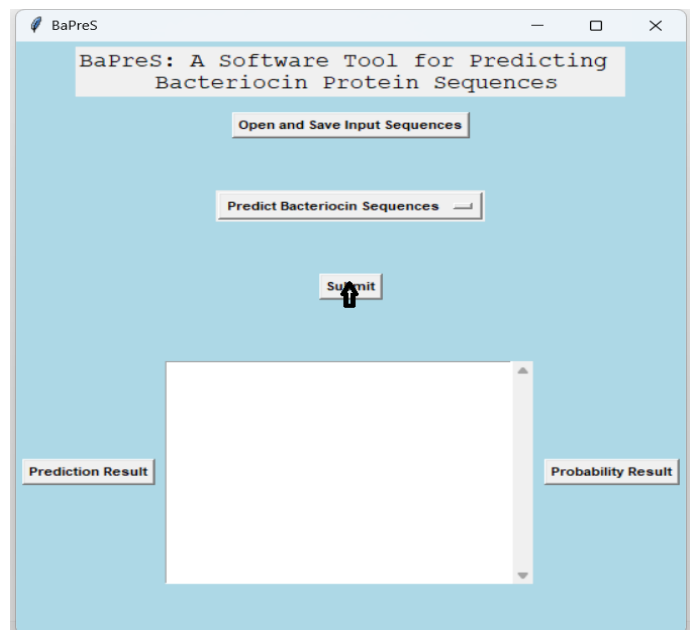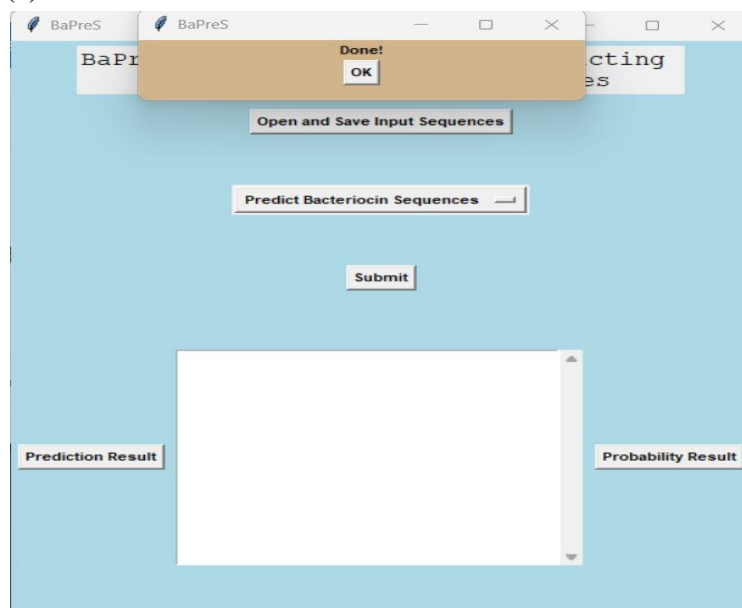
Fig 3. BaPreS service options.

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.
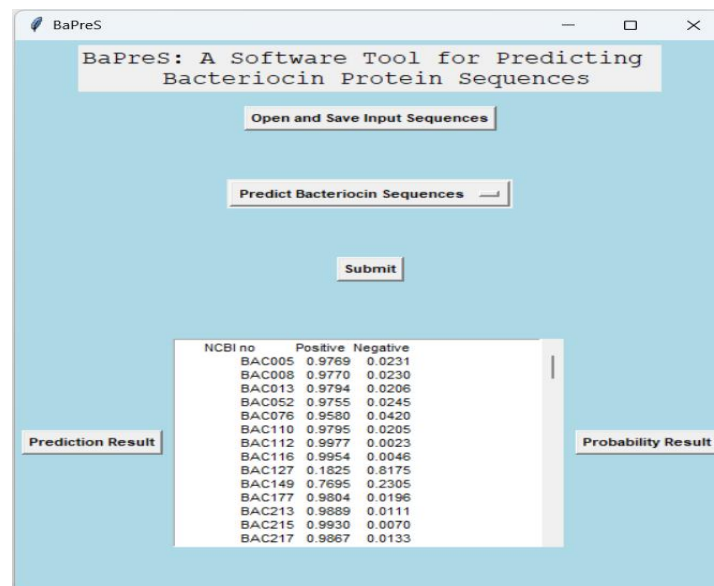
(c) **Add New Non-bacteriocin Sequences** for adding new known non-bacteriocin sequences from the input fasta file to the original training dataset.

(d) **Restore Training Set** for resetting the training set back to the original training dataset supplied with this tool.

Result Generation:

Probability outcome:
When the execution is done, you can see the probability result by clicking on "Probability Result" button. An example of predicted results is shown in the text box (refer to Fig 4) based on given "input_seq.fasta" where the total number of testing sequences is 112. Also, you can find the same prediction results in the "probability_results.csv" file located in the directory where you saved the BaPreS tool.


Fig 5. Probability results.

Prediction outcome:
Users can see the predicted bacteriocin sequences by clicking on "Prediction Result" button. An example of predicted results is shown in the text box (refer to Fig 5) where the tool detected 55 sequences as bacteriocin protein sequences from the input fasta file. Also, you can find the same prediction results in the "predicted_bacteriocin.fasta" file located in the directory where you saved the BaPreS tool.
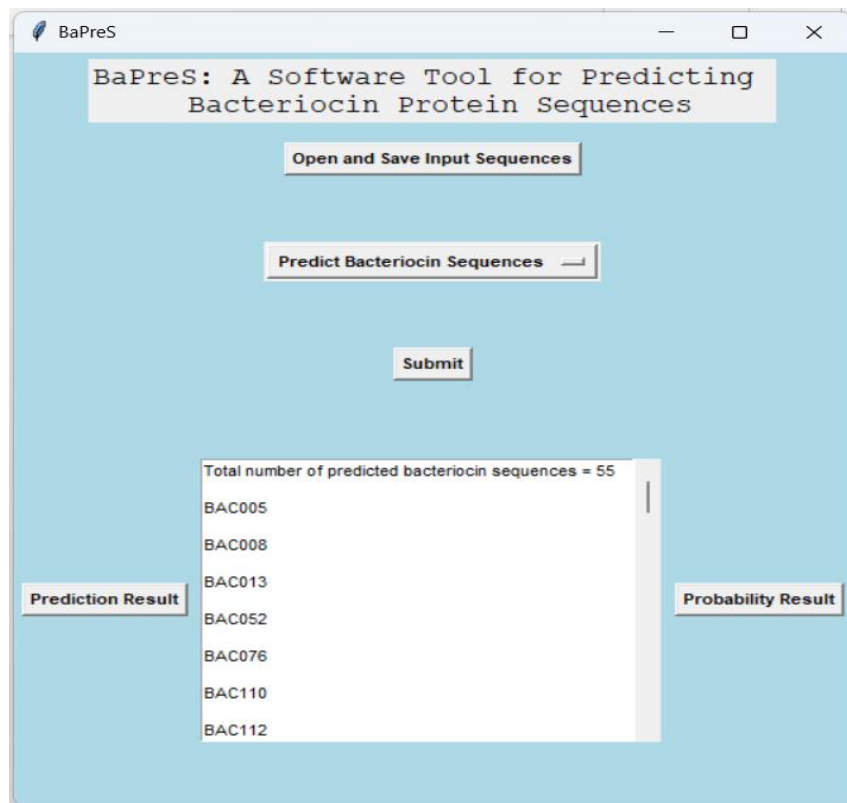
Fig 5. Sample prediction results.