GPS-PUP Manual

Prediction of Pupylation sites

Version 1.0
11/04/2011

Author: Zexian Liu, Yu Xue & Jian Ren

Contact:

Zexian Liu, lzx@mail.ustc.edu.cn

Dr. Jian Ren, renjian.sysu@gmail.com

Dr. Yu Xue, xueyu@mail.hust.edu.cn

The software is only free for academic research.

The latest version of GPS-PUP software is available from http://pup.biocuckoo.org

Index

Statement

1. **Implementation**. The softwares of the CUCKOO Workgroup are implemented in JAVA (J2SE). Usually, both of online service and local stand-alone packages will be provided.

2. **Availability**. Our softwares are freely available for academic researches. For non-profit users, you can copy, distribute and use the softwares for your scientific studies. Our softwares are not free for commercial usage.

3. **GPS**. Previously, we used the GPS to denote our Group-based Phosphorylation Scoring algorithm. Currently, we are developing an integrated computational platform for post-translational modifications (PTMs) of proteins. We re-denote the GPS as Group-based Prediction Systems. This software is an indispensable part of GPS.

4. **Usage**. Our softwares are designed in an easy-to-use manner. Also, we invite you to read the manual before using the softwares.

5. **Updation**. Our softwares will be updated routinely based on users' suggestions and advices. Thus, your feedback is greatly important for our future updation. Please do not hesitate to contact with us if you have any concerns.

6. **Citation**. Usually, the latest published articles will be shown on the software websites. We wish you could cite the article if the software has been helpful for your work.

Introduction

The Nobel Prize in Chemistry 2004 was award to Aaron Ciechanover, Avram Hershko and Irwin Rose for their discovery of ubiquitin-mediated protein degradation (1). Numerous subsequent studies showed that the selective degradation by ubiquitination provided a critical mechanism in eukaryotes to regulate the cellular processes such as cell cycle and division, immune response and inflammation and signal transduction (2). Recently, prokaryotic ubiquitin-like protein (PUP) was identified as the tagging system in prokaryotes (3), which was coupled to its targets through deamidation by dop (PUP deamidase/depupylase) and following conjugation catalyzed by PafA (PUP--protein ligase) (Fig. 1) (4). Although the detail of pup-proteasome system needs further characterization, the discovery of degradation mechanism opens the door to investigate the dynamic protein regulation in *Mycobacterium*, which could be targeted by pathogen-specific drugs (5). In this regards, experimental identification of pupylated substrates with their sites could provide fundamental insights to understanding the cellular processes in *Mycobacterium*.

After Pearce *et al.* opened up research on pupylation with the first identified site in FabD (Malonyl CoA-acyl carrier protein transacylase) (3), Festa and colleagues firstly expanded the mycobacterial "pupylome" with 50 substrates with 60 pupylation sites in *Mycobacterium tuberculosis* (6). Furthermore, proteome-wide discoveries of pupylation targets in model organism *Mycobacterium smegmatis* by Watrous *et al.* and Poulsen *et al.* provided the overall properties of PUP and the selective degradation mediated by pupylation was found to be dynamic (7-8). Since it is labour-intensive and time-consuming to experimentally identify pupylation sites, computational prediction could be an alternative and promising approach for its convenience.

In this work, we manually collected 127 experimentally indentified protein pupylation sites in 109 unique proteins from scientific literature. A previously self-developed GPS (Group-based Prediction System) algorithm was employed with great improvement. We calculated the leave-one-out validation and 4-, 6-, 8-, 10-fold cross-validations to evaluate the prediction performance and system robustness. The leave-one-out validation result is accuracy (*Ac*) of 78.85%, sensitivity (*Sn*) of 63.78%, and specificity (*Sp*) of 80.21%. Finally, both of the online service and local packages of GPS-PUP 1.0 were implemented in JAVA 1.4.2 with high speed. The GPS-PUP 1.0 could predict out potential pupylation sites for ~1,000 proteins (with an average length of ~1000aa) within three minutes. Taken together, we proposed that the GPS-PUP 1.0 will be a useful tool for experimentalists. The GPS-PUP 1.0 is freely available at: http://pup.biocuckoo.org/.
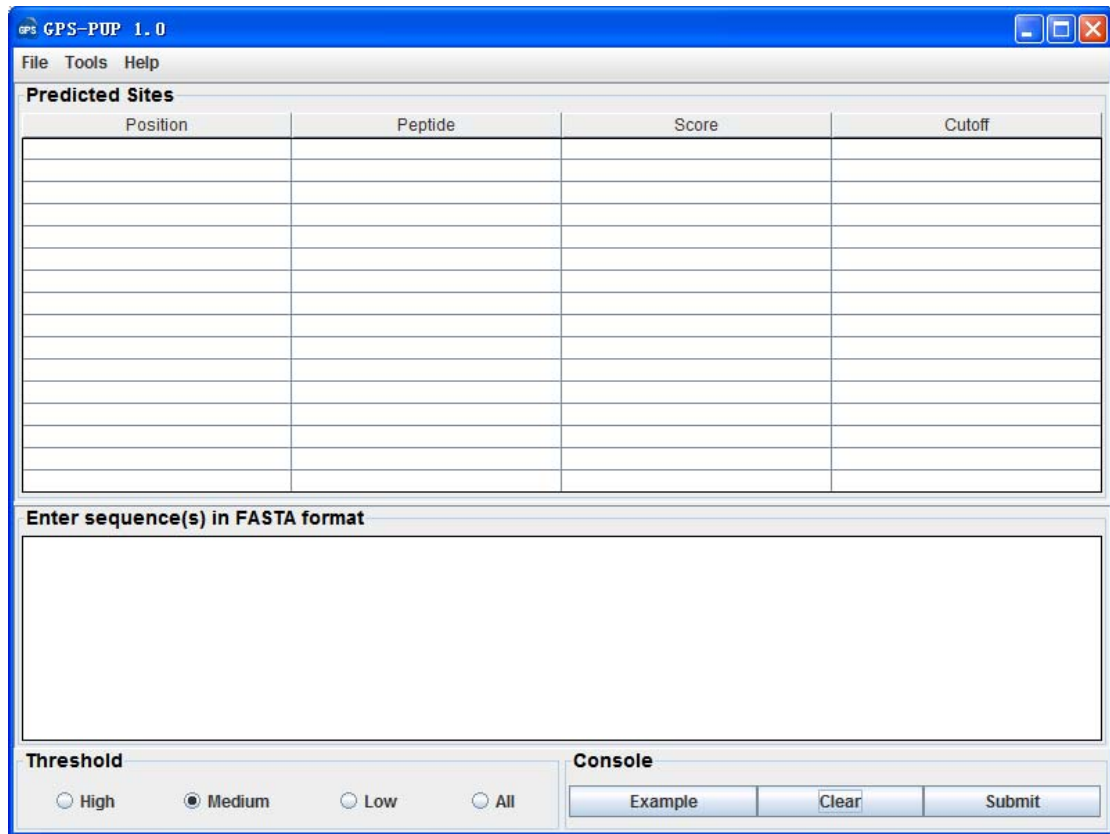
GPS-PUP 1.0 User Interface

Download & Installation

The GPS-PUP 1.0 was implemented in JAVA (J2SE), and could support three major Operating Systems (OS), including Windows, Linux/Unix or Mac OS X systems. Both of online web service and local stand-alone packages are available from: http://pup.biocuckoo.org/software.php. We recommend that users could download the latest release.

Please choose the proper package to download. After downloading, please double-click on the software package to begin installation, following the user prompts through the installation. And snapshots of the setup program for windows are shown below:
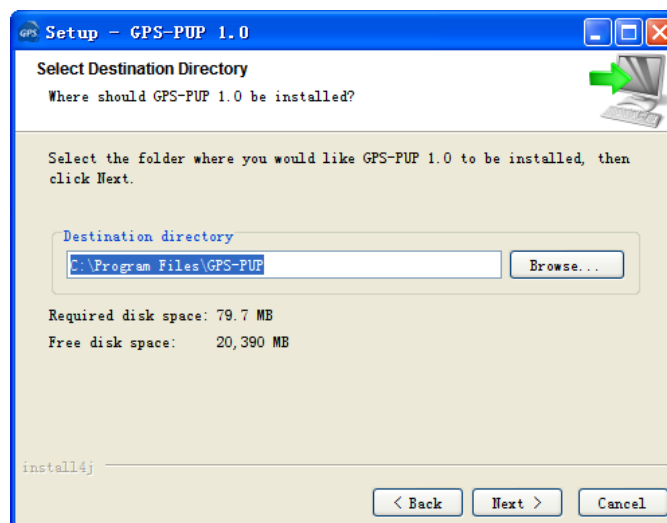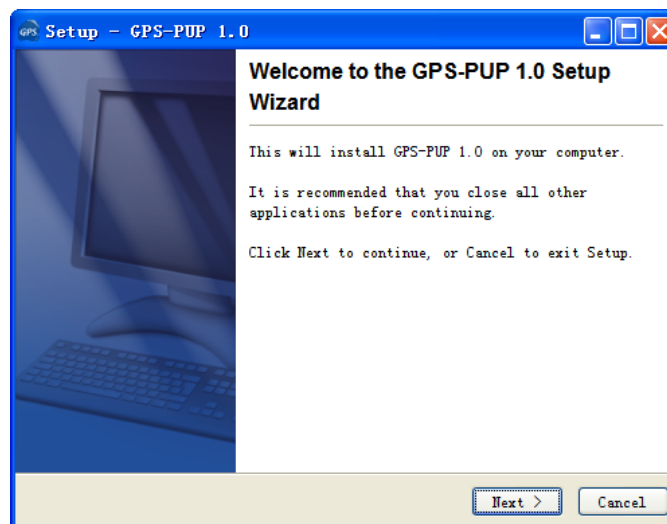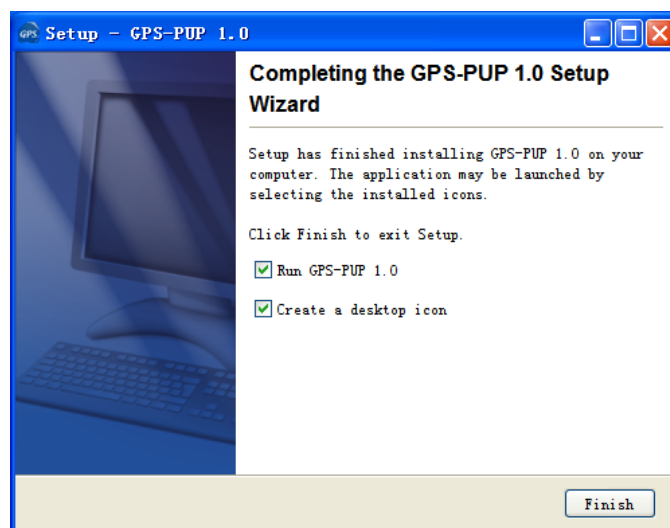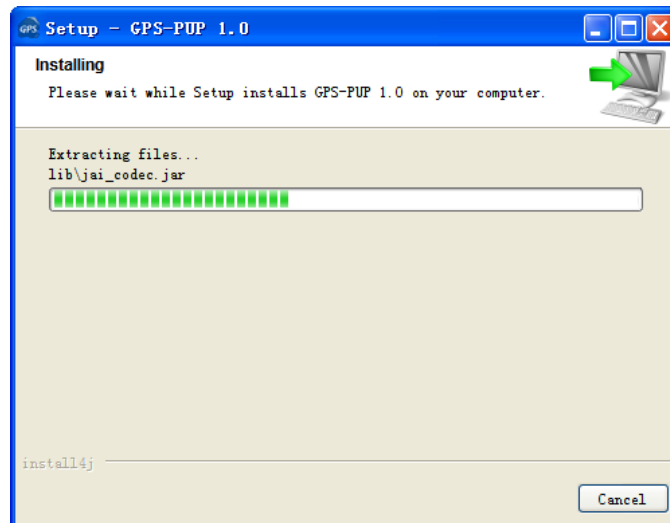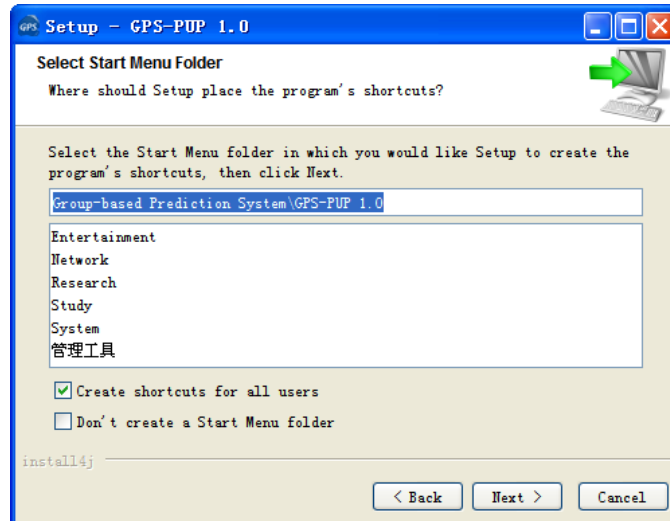
Finally, please click on the **Finish** button to complete the setup program.

Prediction of Pupylation Sites

1. A single protein sequence in FASTA format

The following steps show you how to use the GPS-PUP 1.0 to predict pupylation sites for a single protein sequence in FASTA format.

(1) Firstly, please use "Ctrl+C & Ctrl+V" (Windows & Linux/Unix) or "Command+C & Command+V" (Mac) to copy and paste your sequence into the text form of GPS-PUP 1.0

Note: for a single protein, the sequence without a name in raw format is also OK. However, for multiple sequences, the name of each protein should be presented.

(2) Choose a **Threshold** that you need, the default cut-off is **Medium**.

(3) Click on the **Submit** button, then the predicted pupylation sites will be shown.


(4) Then please click on the **RIGHT** button in the prediction form. You can use

the "**Select All**" and "**Copy Selected**" to copy the selected results into Clipboard. Then please copy the results into a file, eg., an EXCEL file for further consideration. Also, you can choose "**Export Prediction**" to export the prediction results into a tab-delimited text file.
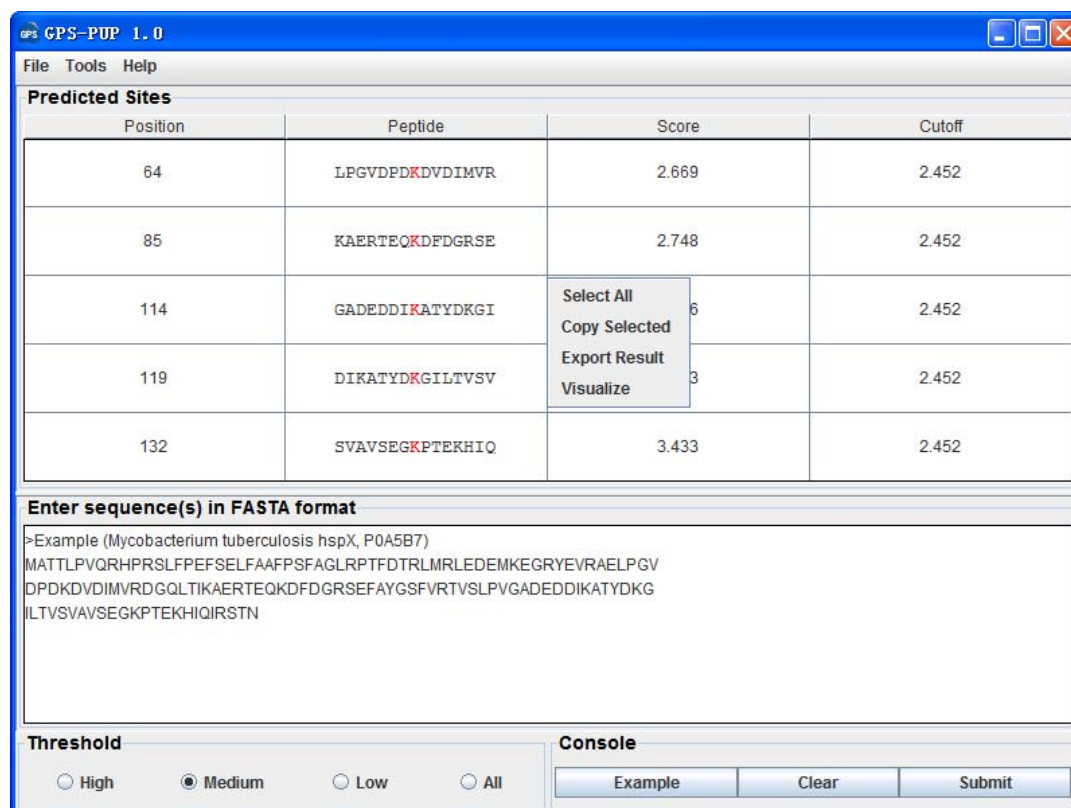
Again, you can also click the "**Export Prediction**" in **File** menu to export the results.
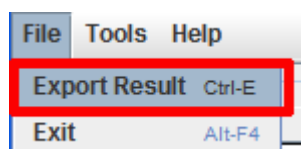

2. Multiple protein sequences in FASTA format

For multiple protein sequences, there are two ways to use the GPS-PUP 1.0.

A. Input the sequences into text form directly. (Num. of Seq ≤ 2,000)
If the number of total protein sequences is not greater than 2,000, you can just use "Ctrl+C & Ctrl+V" (Windows & Linux/Unix) or "Command+C & Command+V" (Mac) to copy and paste your sequences into the text form of GPS-PUP 1.0 for prediction.
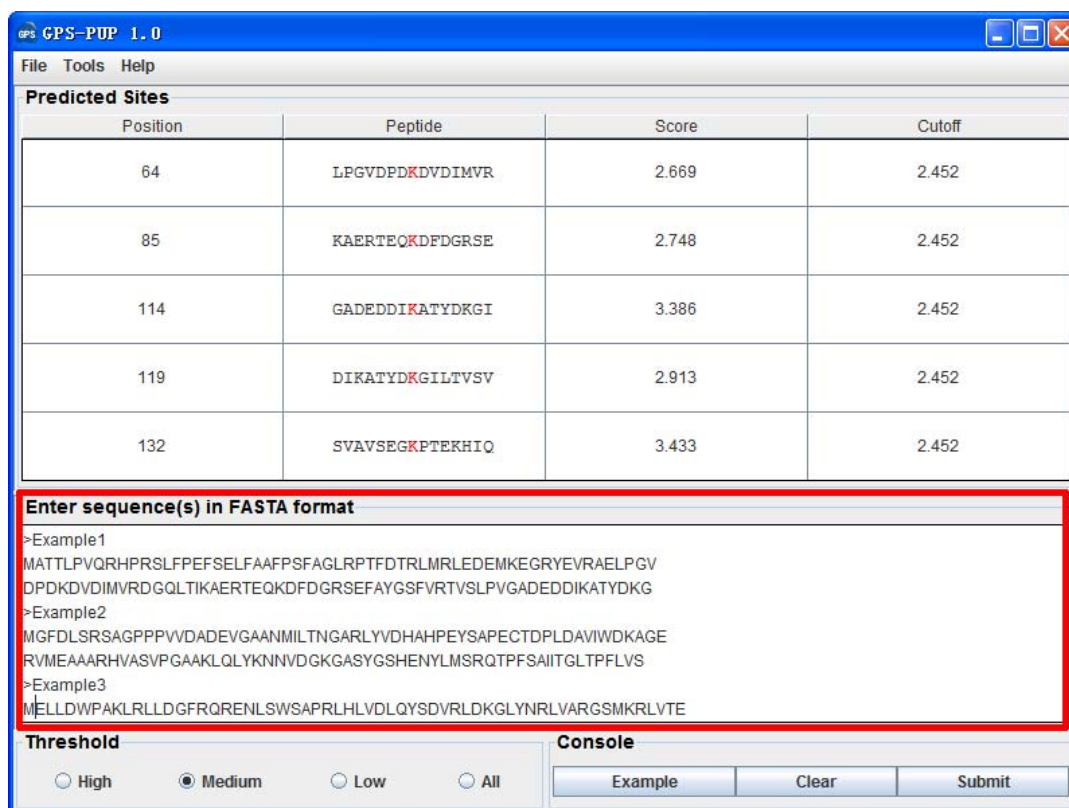
B. Use Batch Predictor tool.

If the number of protein sequences is very large, eg., yeast or human proteome, please use the **Batch Predictor**. Please click on the "**Batch Predictor**" button in the **Tools** menu.
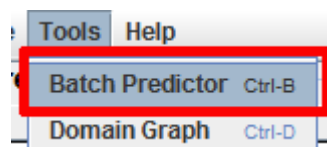

The following steps show you how to use it:

(1) Put protein sequences into one or several files (eg., SC.fas, CE.fas, and etc) with FATSA format as below:

>protein1
XXXXXXXXXXXXX
XXXXXXXX
>protein2
XXXXXXXXXXXXXXX…
>protein3
XXXXXXXXXXXX
...

Most importantly, the name of each protein should be presented.

(2) Click on the **Batch Predictor** button and then click on the **Add File** button and add one or more protein sequence files in your hard disk.
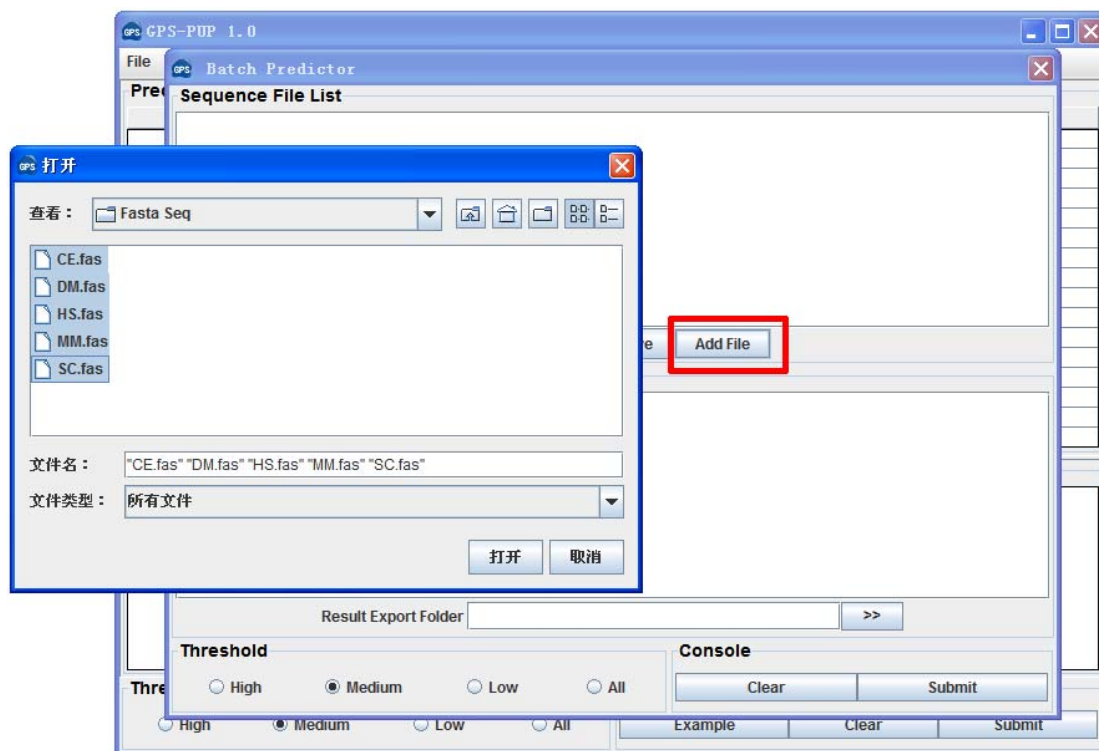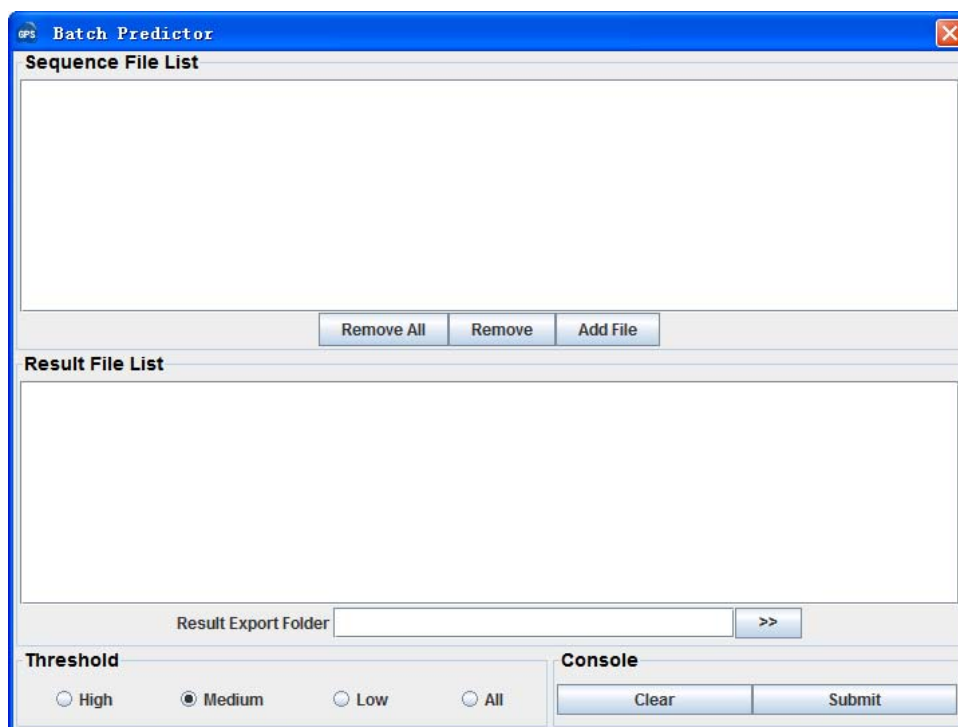
Then the names of added files will be shown in the **Sequence File List**.
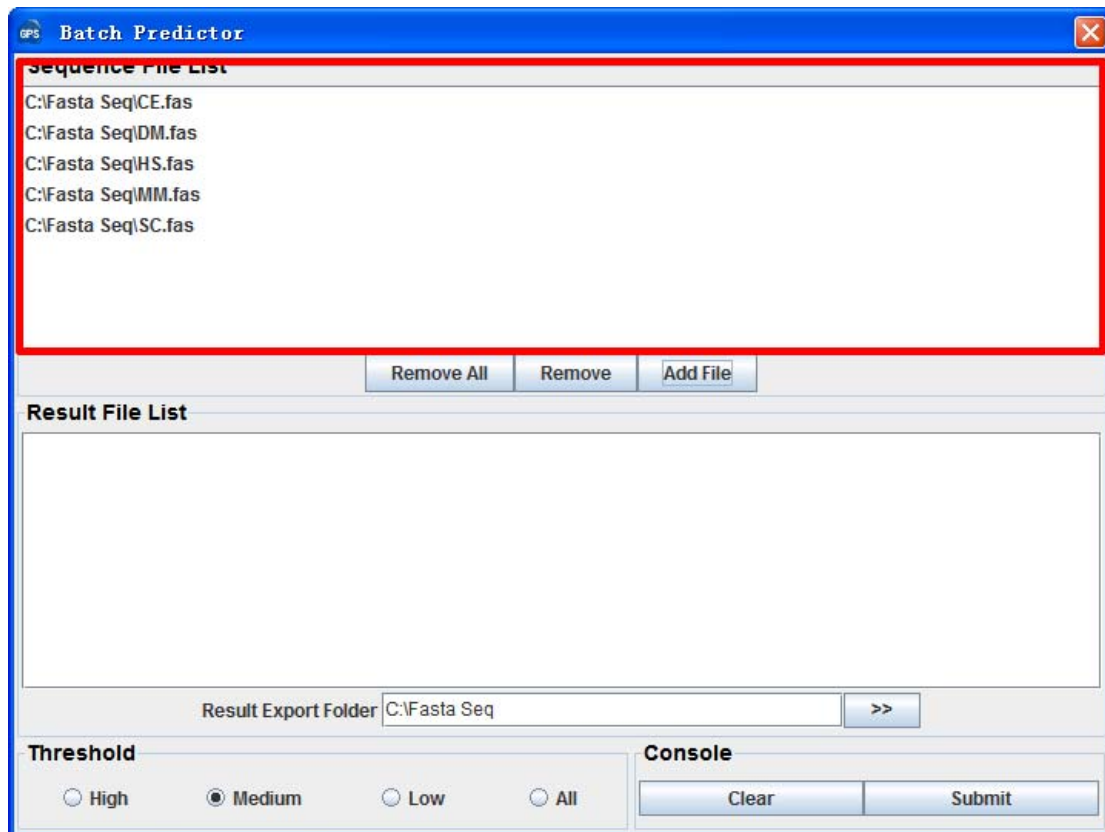
(3) The output directory of prediction results should also be defined. Please click on the **>>** button to specify the export fold.

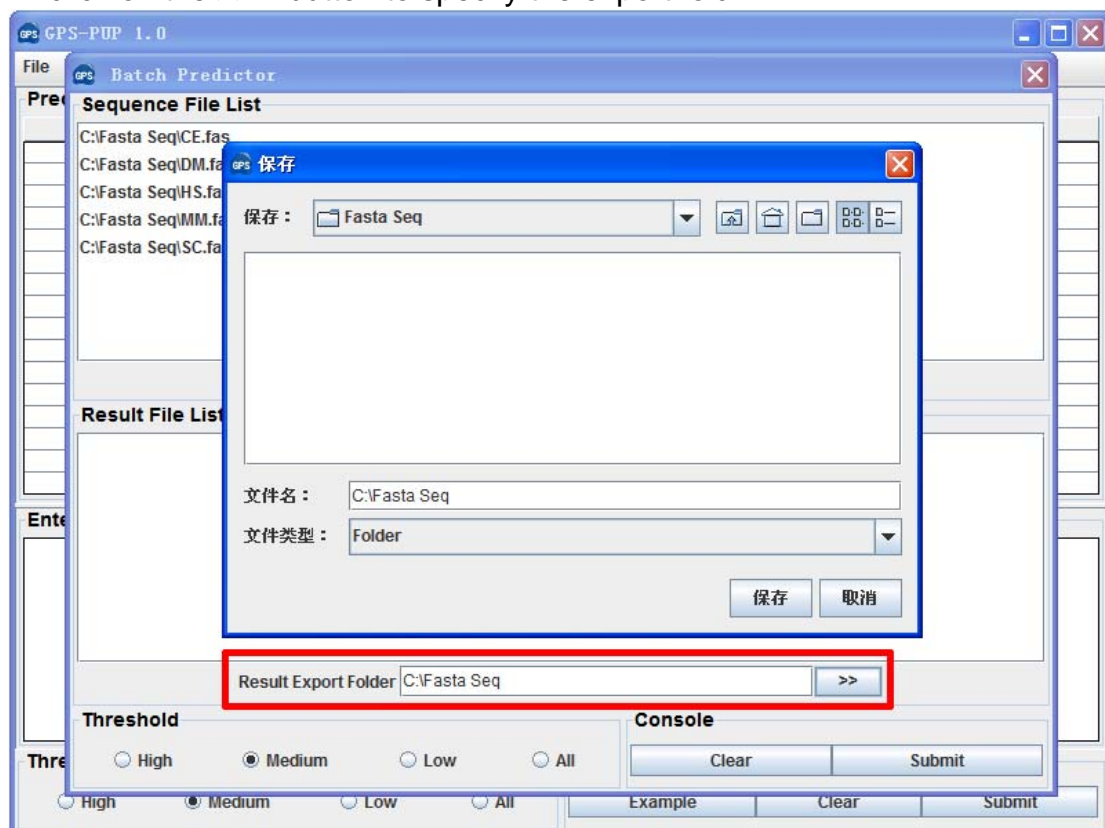(4) Please choose a proper threshold before prediction. Then please click on the **Submit** button, then the **Batch Predictor** begin to process all of the sequence files that have been added to the list. The result of prediction will be export to the **Prediction Export Fold,** and the name of result files will be shown in the **Prediction File List**.
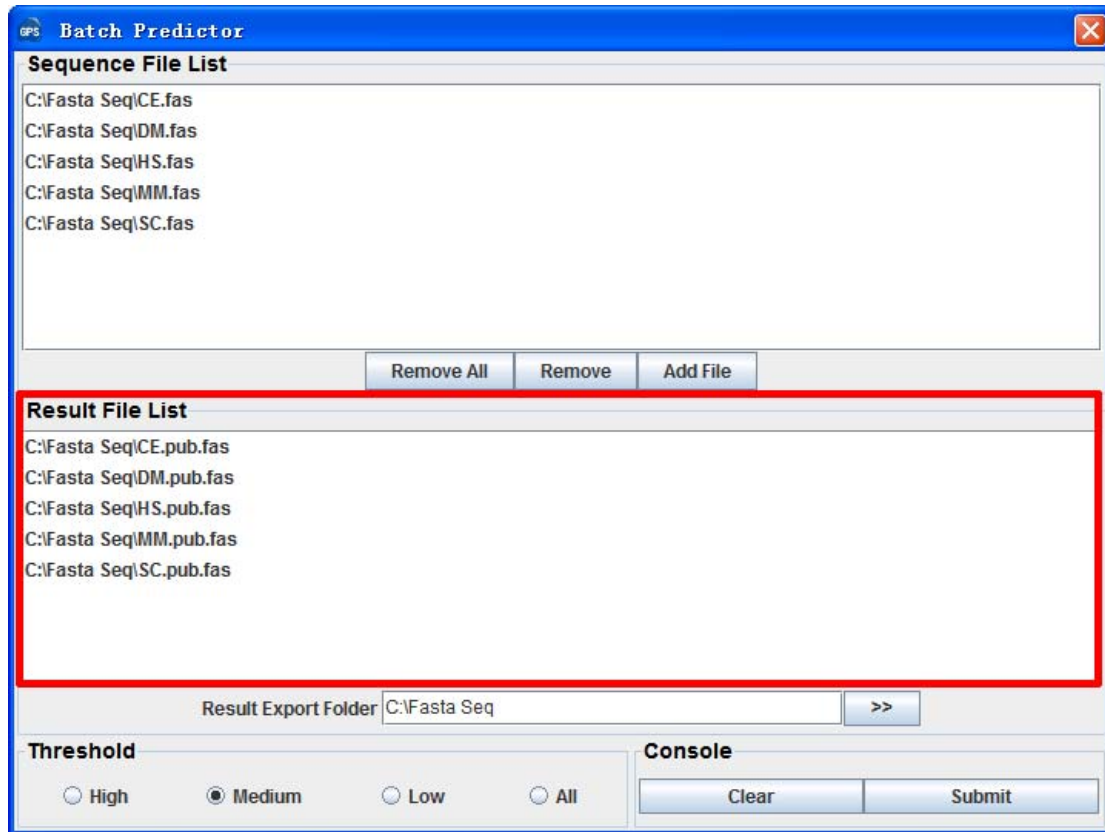
References

1.   Vogel, G. (2004) Nobel Prizes. Gold medal from cellular trash. *Science*, **306**, 400-401.

2.   R. John Mayer, Aaron J. Ciechanover and Rechsteiner, M. (eds.) (2006) *Protein Degradation: Cell Biology of the Ubiquitin-Proteasome System*. Wiley-VCH.

3.   Pearce, M.J., Mintseris, J., Ferreyra, J., Gygi, S.P. and Darwin, K.H. (2008) Ubiquitin-like protein involved in the proteasome pathway of Mycobacterium tuberculosis. *Science*, **322**, 1104-1107.

4.   Striebel, F., Imkamp, F., Sutter, M., Steiner, M., Mamedov, A. and Weber-Ban, E. (2009) Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. *Nat Struct Mol Biol*, **16**, 647-651.

5.   Salgame, P. (2008) PUPylation provides the punch as Mycobacterium tuberculosis battles the host macrophage. *Cell Host Microbe*, **4**, 415-416.

6.   Festa, R.A., McAllister, F., Pearce, M.J., Mintseris, J., Burns, K.E., Gygi, S.P. and Darwin, K.H. (2010) Prokaryotic ubiquitin-like protein (Pup) proteome of Mycobacterium tuberculosis [corrected]. *PLoS One*, **5**, e8589.

7.   Watrous, J., Burns, K., Liu, W.T., Patel, A., Hook, V., Bafna, V., Barry, C.E., 3rd, Bark, S. and Dorrestein, P.C. (2010) Expansion of the mycobacterial "PUPylome". *Mol Biosyst*, **6**, 376-385.

8.   Poulsen, C., Akhter, Y., Jeon, A.H., Schmitt-Ulms, G., Meyer, H.E., Stefanski, A., Stuhler, K., Wilmanns, M. and Song, Y.H. (2010) Proteome-wide identification of mycobacterial pupylation targets. *Mol Syst Biol*, **6**, 386.

Release Note

1. Apr. 11th, 2011, the online service and the local stand-alone packages of GPS-PUP 1.0 were released.