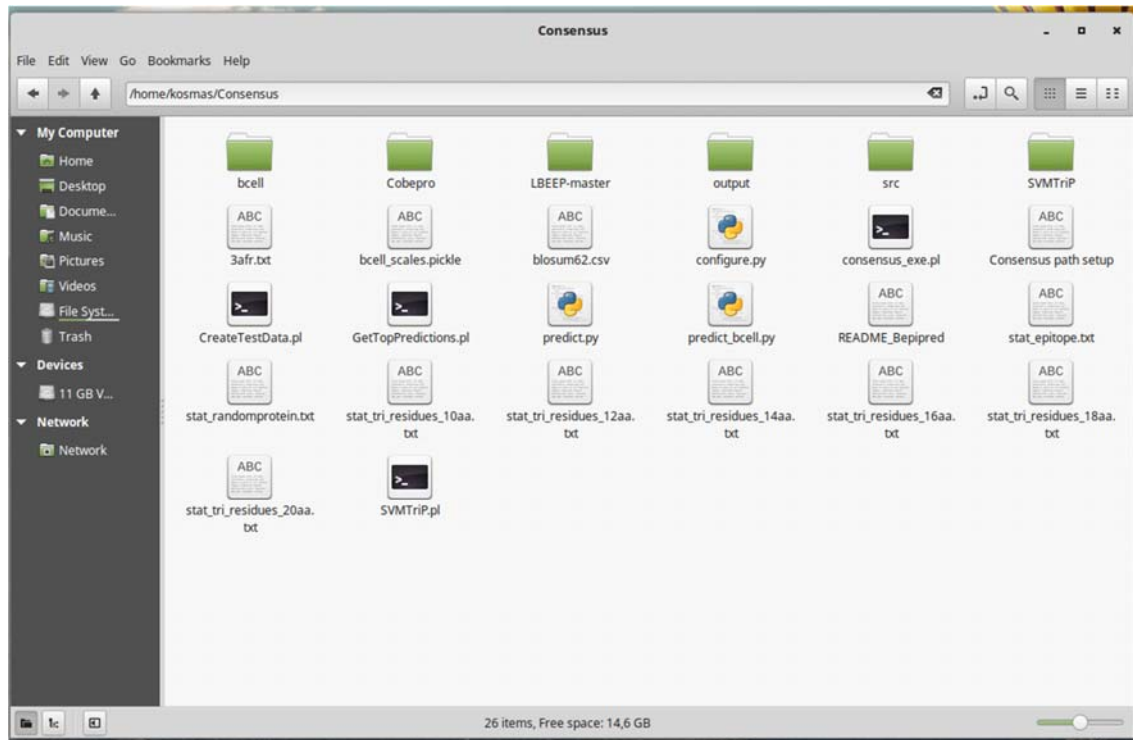


Supplementary File 1

Consensus method user guide

Introduction

The executable version of our consensus method (Supplementary Figure 1) can be obtained via this link (<http://thalis.biol.uoa.gr/BCEconsensus>), which downloads a Virtual Machine (VM) in which the package is contained. The supplied VM runs a linux distribution called “Linux Mint” (<https://www.linuxmint.com/>), inside of which one can find preinstalled all of the predictors used in this study, namely BcePred, BepiPred-1.0, ABCpred, COBEpro, SVMTriP, LBtope and LBEEP, along with our consensus algorithm that combines them. This method of software delivery was picked due to the nature of this work, in order to ensure ease of use and compatibility for any researcher that might want to use the consensus method. Most scripts used are written in the Perl programming language, while only a few of them are written in Python. It should be noted that all of these methods are freely available for academic use under Creative Commons (CC) licenses, and if you wish to use them for commercial purposes you should contact the methods’ developers and require the respective licenses for each one of them. Bepipred has a special academic license and please advice their terms before using this or the consensus method (<http://www.cbs.dtu.dk/services/BepiPred-1.0/>). The consensus method is also available under a BY-NC-SA CC license (for non-commercial use only).



Supplementary Figure 1. The Consensus directory from where you can execute the consensus method (`consensus_exe.pl`).

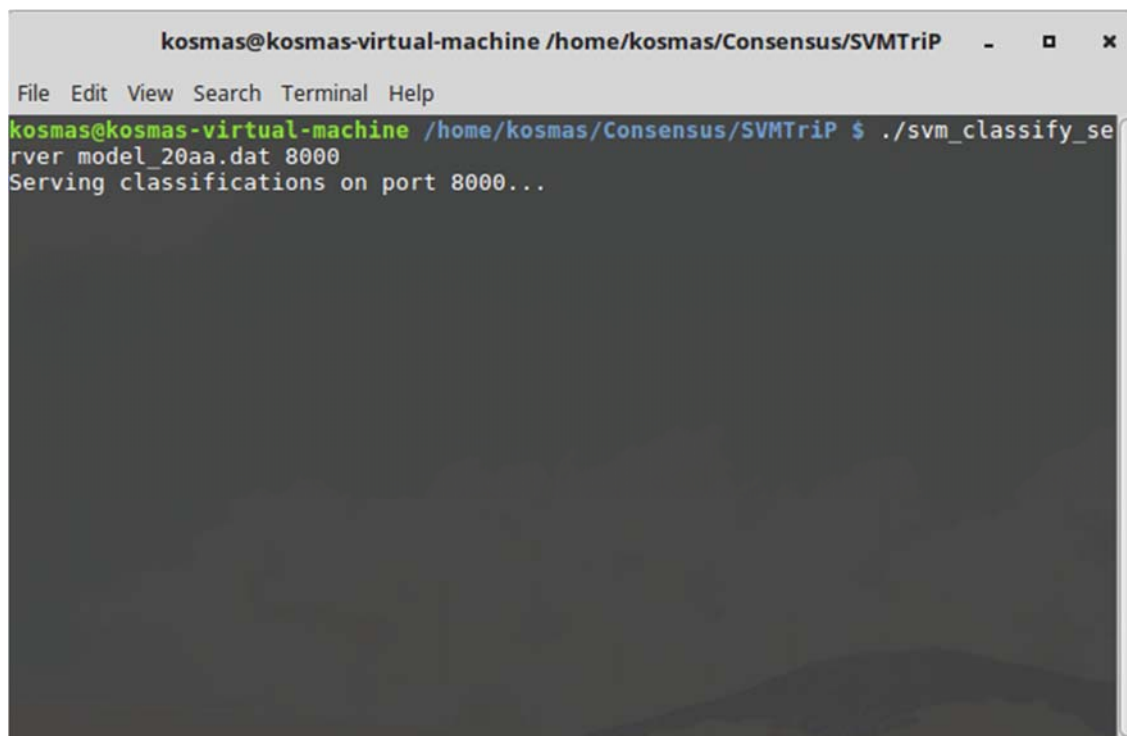
Installation/Configuration

Using the provided link the user can obtain an archive of the VM in Open Virtualization Format (.ovf), which first needs to be uncompressed. Afterwards, the files can be imported

into the user's computer using various free or commercially available Virtual Machine Monitors (https://en.wikipedia.org/wiki/Full_virtualization). We suggest a minimum of 3GB of RAM to be allocated for the imported VM. After the VM has been imported all the user has to do is execute the appropriate scripts and all the methods should run smoothly, since all scripts and dependencies for every single predictor are preconfigured. The admin password for any needs that might arise is: "consensus". The consensus script along with Bepired, COBEpro, SVMTriP and LBEEP can be found in the following directory: "/home/kosmas/Consensus". BcePred, ABCpred and LBtope can be found in: "/gpsr/standalone". If for whatever reason one needs to change the location of one or several of the software packages contained in the Consensus directory, instructions are available in the "Consensus path setup" file.

Usage

1. Load the SVMTriP model file onto the RAM using our custom "svm_classify_server" executable contained in /home/kosmas/Consensus/SVMTriP, by executing the command `./svm_classify_server model_20aa.dat 8000` inside the terminal at that specific path. Wait for a few minutes depending on your system hardware for the model to load, until the text "Serving classifications on port 8000..." appears. This only needs to be performed once every boot cycle or each time after this terminal is closed. Alternatively, with some minor modifications the "SVMTriPold.pl", contained in the SVMTriP directory can be used, which works like the original standalone package and doesn't require the first step. This comes with the drawback that for multiple sequence predictions in a row, the original software loads up the SVM model onto memory every time anew, which takes about 9 minutes per run.

A terminal window titled "kosmas@kosmas-virtual-machine /home/kosmas/Consensus/SVMTriP" with standard window controls. The terminal shows the command `./svm_classify_server model_20aa.dat 8000` being executed, followed by the output "Serving classifications on port 8000...".

```
kosmas@kosmas-virtual-machine /home/kosmas/Consensus/SVMTriP - □ ×
File Edit View Search Terminal Help
kosmas@kosmas-virtual-machine /home/kosmas/Consensus/SVMTriP $ ./svm_classify_server model_20aa.dat 8000
Serving classifications on port 8000...
```

Supplementary Figure 2. Model loading for SVMTriP

- Once the model has been loaded, we can execute the “consensus_exe.pl” script using the following command: "perl consensus_exe.pl <fasta> <outfile_name>".
 <fasta>: is the name of the file containing a single protein sequence in FASTA format.
 <outfile_name>: is the name of the file where we would like or results to be printed out to.
- For example, this is part of the results of a test run using the included test sequence “3afr.txt”, which are printed in “out.txt.”

```

1 3AFR
2 POSITION      RESIDUE ASSIGNMENT
3 1           G      -
4 2           S      -
5 3           H      -
6 4           M      E
7 5           G      E
8 6           S      E
9 7           P      E
10 8          N      E
11 9          S      E
12 10         P      E
13 11         L      E
14 12         K      -
15 13         D      E
16 14         S      E
17 15         L      E
18 16         R      E
19 17         P      E
20 18         K      E
21 19         L      E
22 20         S      E

```

Supplementary Figure 3. Results after running the consensus method on a single protein sequence

NOTES

Any errors produced in the terminal by the COBEpro package and specifically blastpgp can be safely ignored.

Inside every predictor’s directory a perl script for running batch sequences for testing purposes is included.

The included version of the “consensus_exe.pl” is modified to employ every predictor except LBEEP, but the code for its inclusion in the consensus is still there in comments (#).