**MFPSP**

**# 1 Description**

Protein phosphorylation plays a vital role in signal transduction pathways and diverse cellular processes. To date, a tremendous number of in silico tools have been developed for phosphorylation site identification, but few of them are suitable for the prediction of fungal phosphorylation sites.

In this paper, we present MFPSP, a machine learning method of fungal phosphorylation site identification.

The source code and datasets are accessible at https://github.com/AIforGenomics/MFPSP.

**# 2. Availability**

2.1. Datasets and source code are available at:

https://github.com/AIforGenomics/MFPSP.

2.2 Local running

2.2.1 Environment

Before running, please make sure the following packages are installed in Python environment:

pandas==1.1.3

python==3.7.3

biopython==1.7.8

numpy==1.19.2

scikit-learn==0.24.2

genism==4.2

For convenience, we strongly recommended users to install the Anaconda Python 3.7.3 (or above) in your local computer.

2.2.2 Running

Changing working dir to MFPSP-master\_source\_code. For example, if we want to predict the S site in *S. cerevisiae*, we should use the following command:

**python MFPSP.py -i testing.fasta -s sacch -t s -o prediction\_results.csv**

-i: name of input\_file in fasta format # folder “sequence” is the default file path of the input\_file

-s: fungal species name, it should be asper, crypto, fusari, manga, neuro, schizo or sacch.

# asper: *Aspergillus sp.*; crypto: *C. neoformans*; fusari: *F. graminearum*; manga: *M. oryzae*;

# neuro: *N. crassa*; sacch: *S. cerevisiae*; schizo: *S. pombe*

-t: phosphorylation\_type, it should be s, y, or t

-o name of output\_file # folder “results” is the default file path for result save.