PFP: Automated Protein Function

Tutorial

Cite:

Troy Hawkins, Stan Luban & Daisuke Kihara, *Enhanced automated function prediction using distantly related sequences and contextual association by PFP."* , Protein Sci, 15:1550-1556 (2006).,

Troy Hawkins, Meghana Chitale, Stan Luban & Daisuke Kihara, *PFP: Automated prediction of gene ontology functional annotations with confidence scores,* Proteins: Structure, Function, and Bioinformatics, 74: 566-582, (2009).

Related Articles:

Meghana Chitale*, Ishita K. Khan*, & Daisuke Kihara (* equal contribution), *In-depth performance evaluation of PFP and ESG sequence-based function prediction methods in CAFA 2011 experiment.* BMC Bioinformatics, 14 Suppl 3: S2 (2013),

Ishita Khan, Meghana Chitale, Catherine Rayon, & Daisuke Kihara, *Evaluation of function predictions by PFP, ESG,and PSI-BLAST for moonlighting proteins.* BMC Proceedings, 6 Suppl 7: S5 (2012),

Meghana Chitale & Daisuke Kihara , *Enhanced sequence-based function prediction methods and application to functional similairty networks.,* Protein function prediction for omics era, D. Kihara ed., Chapter 2, pp. 19-34, Springer.(2011),

Lee Sael*, Meghana Chitale*, & Daisuke Kihara (* equal contribution), *Structure- and sequence-based function prediction for non-homologous proteins.* Journal of Structural and Functional Genomics, 13: 111-123 (2012)

For any questions regarding ESG contact the administration at info@kiharalab.org

This method allows user to submit protein sequence(s) and obtain annotation prediction for the sequence(s) in terms of Gene Ontology terms. The underlying PFP algorithm can be found here . To learn more about each input you can click on the question mark next to it.

Part1: Steps for submitting job as anonymous user

1. Go to http://www.kiharalab.org/web/software.php . To start a new job click on the PFP logo highlighted in red below. This will open PFP job submission page in your browser.

2. You can enter one or more query protein sequence(s) in fasta format in the text box titled "Enter Query Sequence(s)".

Consider the following sequence that you can enter.

```
>sp|P28482|MK01_HUMAN
MAAAAAAGAGPEMVRGQVFDVGPRYTNLSYIGEGAYGMVCSAYDNVNKVRVAIKKISPFEHQTYC
QRTLREIKILLRFRHENIIGINDIIRAPTIEQMKDVYIVQDLMETDLYKLLKTQHLSNDHICYFL
YQILRGLKYIHSANVLHRDLKPSNLLLNTTCDLKICDFGLARVADPDHDHTGFLTEYVATRWYRA
PEIMLNSKGYTKSIDIWSVGCILAEMLSNRPIFPGKHYLDQLNHILGILGSPSQEDLNCIINLKA
RNYLLSLPHKNKVPWNRLFPNADSKALDLLDKMLTFNPHKRIEVEQALAHPYLEQYYDPSDEPIA
EAPFKFDMELDDLPKEKLKELIFEETARFQPGYRS
```

You can also click on "Load Sample" link to load this sequence in the text box and follow the next steps. Clicking on "Clear" link will clear the text box for sequence and load default ESG parameters in the boxes below.

Or you can click on browse button and upload a fasta file containing query sequences. Currently we support maximum 10 query sequences in one job.

To clear sequence, click on "Clear" link above the protein sequence text box

3. When you submit a job without logging in, you can bookmark the results URL and refer back to the job result. Registration to the website provides you additional features, including receiving email notifications about your jobs, viewing the progress status, and manage your jobs in a table.

4. After loading the query sequence(s) click on the Submit button. This will take you to the Your Job page and shows the status of your jobs.

When you job status shows complete, click on the jobID to view predicted Gene Ontology terms for the query protein. Or you can refresh the page to go to results. The

results page will show PFP input sequence and prediction terms for each Gene Ontology category (Molecular function, Biological Process and Cellular Component) that have confidence greater than 5% of the score of the top hit. The results page will also provide link to the results in xml format which you can download for further processing.Clicking "Visualization of Predicted GO Terms" will pop up a new window that visualize the predicted GO terms in the GO graph. The tutorial of the GO visualizer is available here . Below the redirect link, there are three links of downloading static image of visualization. Once clicked it, the server will render the SVG image for the user and display it on the web page. At the top of static image, there is a download link to download the PNG image file and user can also save the SVG by saving the static page. (After openning the page, please wait some time to allow server to generate those file.)

Clicking the plus button next to a GO term will show a list of PSI-BLAST hit sequences that contributed scores to the GO term. The percentage next to each sequence ID shows the the percentage of the score that originates from the sequence in the total score assigned to the GO term. Some of the GO terms may not have the plus sign. These GO terms are not directly retrieved from PSI-BLAST sequence hits but assigned by the GO term association matrix. To learn more about the GO term association matrix, please check the original paper

PFP Job Results

Protein Sequence:

```
MNIMDNEIDTKCINEIRMLSAELPLEAKSGHQGAPIGCAPIAHILWSYVMNYYNEDTKWINRDRF
ILSNGHASALLYTMLYLTEQGLSMEDLKSFRQFGSLTPGHPENHITKGVEVTTGPLGQGASNAVG
MATAAHNLADKYNTEEHKIFDNYVYAICGDGCMQEGVFCEAASLAGHLGLGRLILLYDDNKITID
GNTDLSFTENIEKKFEALNWEVRRVEDGNKDYKKILHEIEQGKKNLQQPTLIIVRTACGFGTKVE
GTCKSHGLALNDEDLKNAKSFFGLDPQKKFHISDEVKEFYKNVIQKKKENYIKWKNMFDDFSLKY
PQVSQEIIRRFQNDLPNNWKDALPKYTPKDAPGATRNLSGIVLNSINKIFPELIGGSADLSESNC
TSLKEENDIKKNSYGNKYIRFGVREHGMVAITNGLYAYGGFKPYCGTFLNFYTYAFGALRLAALS
NHHILCIATHDSVELGEDGPTHQPIEVLSLLRSTPNLNIIRPADGNEVSGAYLSHFSNPHTPTVI
ALCRNKVPHLNNTQPEQVLKGAYILEDFDTSNNPKVILTGSGSELHLCFEAKEILKNQHQLNVRI
VSFPSWTLFKKQPEDYQYSVMMHNHPNLPRFYIEPASTHGFDTYFNVYIGINQFGYSAPKNKIWE
HLGFTPENIVQKVLAFMKNKLK
```

Predicted GO Terms

[Download Prediction's XML](#)

[Visualization of Predicted GO Terms](#)

Download Static Visualization Figure

[MF], [BP], [CC]

List of Predicted GO Terms (the raw score can range 0 to over 20K)

Very high confidence : > 20K
High confidence : > 10K
Moderate confidence : > 500
Low confidence: >= 100 (But worthwhile to examine)
Below low confidence: < 100

Molecular Function Terms

PFP Score	Term	Description
1913.56	GO:0003824 [+]	catalytic activity
1866.10	GO:0046872 [+]	metal ion binding
1857.99	GO:0043169	cation binding
1854.75	GO:0043167	ion binding
1853.37	GO:0004802 [+]	transketolase activity
644.96	GO:0005488	binding

6 Predictions
0 Predictions > 20K; **0** Predictions > 10K; **6** Predictions > 500; **0** Predictions >= 100

Biological Process Terms

PFP Score	Term	Description
2041.19	GO:0008152 [+]	metabolic process
1583.73	GO:0006740	NADPH regeneration
1571.08	GO:0006098 [+]	pentose-phosphate shunt
1188.46	GO:0006739	NADP metabolic process
769.84	GO:0046496	nicotinamide nucleotide metabolic process
767.67	GO:0019362	pyridine nucleotide metabolic process
767.34	GO:0072524	pyridine-containing compound metabolic process
679.08	GO:0019318	hexose metabolic process
551.75	GO:0006733	oxidoreduction coenzyme metabolic process
538.97	GO:0019253 [+]	reductive pentose-phosphate cycle
488.90	GO:0006007 [+]	glucose catabolic process
452.57	GO:0019320	hexose catabolic process
441.55	GO:0019685	photosynthesis, dark reaction
437.19	GO:0015977	carbon fixation
401.66	GO:0006006 [+]	glucose metabolic process
392.75	GO:0046365	monosaccharide catabolic process
387.96	GO:0005996	monosaccharide metabolic process
368.28	GO:0071704	organic substance metabolic process
356.13	GO:0005975 [+]	carbohydrate metabolic process
332.37	GO:0046164	alcohol catabolic process
316.91	GO:0044275	cellular carbohydrate catabolic process
306.96	GO:0006066	alcohol metabolic process
270.34	GO:0016052	carbohydrate catabolic process
236.52	GO:0044262 [+]	cellular carbohydrate metabolic process
228.23	GO:0006310 [+]	DNA recombination

25 Predictions
0 Predictions > 20K; **0** Predictions > 10K; **10** Predictions > 500; **15** Predictions >= 100

Cellular Component Terms

PFP Score	Term	Description
1751.35	GO:0005737 [+]	cytoplasm
1659.59	GO:0005622 [+]	intracellular
1629.02	GO:0044424	intracellular part
1385.40	GO:0044464	cell part
1385.37	GO:0005623	cell

5 Predictions
0 Predictions > 20K; **0** Predictions > 10K; **5** Predictions > 500; **0** Predictions >= 100

5. If you close your browser and need to go back to your job results then you can use the "Job Retrieval" text box on the right side of the web page. It has been highlighted in red in the screenshot below. When you enter a valid job id and click on "Fetch Job", you will get to view the job status or results depending on if the job is complete or not.

Part2: Steps for logging in and submitting PFP job

1. Open PFP job submission page in your browser.

2. Click on the "Register" link on the right side of the page, it is highlighted in red in the screenshot above. It will take you to the account creation page shown below.

Click on the "Forgot Password" link on the right side of the page. It will take you to the password reset page. After entering the correct email, you will receive a link to reset your password.

Here you should enter your username and password information. The email address is optional; you should enter it if you wish to receive email notifications about your job status. If you don't check the "Receive Email Notification" check box you will have to log in to the server and check the status of your jobs.

3. Once you have your account set up you can log in and submit jobs for processing. Enter your username and password in the text boxes provided on the right side of the web page. This will redirect you to the page displaying all your jobs with their status and results links.

You can click on "My Jobs" and view your existing jobs and their status. The complete jobs have clickable links where you can view results of those jobs.

Your jobs

Job ID	Type	Status	Date Submitted	Batch Job
000424	ESG	Done - View or Download	2009-12-31	No
000423	PFP	Done - View or Download	2009-12-31	No
2 Jobs				

To start a new job click on "PFP Submission" link under Job Submission on the top of the right bar.

Kihara Lab Software
Thank you for using Kihara Lab Web Servers. Please refer to the tutorials below for getting started with our servers. You can register with the lab and enjoy additional features free of cost. We would be delighted to receive feedback from you.

Job Submission

PFP Submission
ESG Submission

This will direct you to the PFP job submission page.

4. You can enter one or more query protein sequence(s) in fasta format in the text box titled "Enter Query Sequence(s)".

Consider the following sequence that you can enter.

>sp|P28482|MK01_HUMAN
MAAAAAAGAGPEMVRGQVFDVGPRYTNLSYIGEGAYGMVCSAYDNVNKVRVAIKKISPFEHQTYC
QRTLREIKILLRFRHENIIGINDIIRAPTIEQMKDVYIVQDLMETDLYKLLKTQHLSNDHICYFL
YQILRGLKYIHSANVLHRDLKPSNLLLNTTCDLKICDFGLARVADPDHDHTGFLTEYVATRWYRA
PEIMLNSKGYTKSIDIWSVGCILAEMLSNRPIFPGKHYLDQLNHILGILGSPSQEDLNCIINLKA
RNYLLSLPHKNKVPWNRLFPNADSKALDLLDKMLTFNPHKRIEVEQALAHPYLEQYYDPSDEPIA
EAPFKFDMELDDLPKEKLKELIFEETARFQPGYRS

Or you can click on browse button and upload a fasta file containing query sequences. Currently we support maximum 10 query sequences in one job.

5. After loading the query sequence(s) click on submit button. This will take you to the results page and show your job waiting for processing. You can log in again later to check the status of your job and access the results when it is complete. You can also enter job id in "Job Retrieval" text box on right side of the page and click "Fetch Job" to view the status of job and results. If you have chosen to get email notifications in your account settings then you will be notified by an email when your job is complete.

6. When you job status shows complete, click on the jobID to view predicted Gene Ontology terms for the query protein. The results page will also provide link to the results in xml format which you can download for further processing.

Part3: Steps for submitting a batch job using PFP

1. Open PFP job submission page in your browser.

2. You can enter more than one query protein sequences in fasta format in the text box titled "Enter Query Sequence(s)".

Consider the following sequences that you can enter.

>sp|P56851|EP3B_HUMAN MASSLKIWGTLLALLCILCTLLVQSKEVSWREFMKQHYLSPSREFREYKCDVLMRENEAL KDKSSHMFIYISWYKIEHICTSDNWMDRFRNAYVWVQNPLKVLKCHQENSKNSYTESRSF NYIEFHCSMDGYVDSIEDLKMVEPIGN

>sp|P13811|ELBH_ECOLX MNKVKFYVLFTALLSSLCAHGAPQSITELCSEYHNTQIYTINDKILSYTESMAGKREMVI ITFKSGATFQVEVPGSQHIDSQKKAIERMKDTLRITYLTETKIDKLCVWNNKTPNSIAAI SMEN

>sp|P05928|VPR_HV1BR MEQAPEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQHIYETYGDTWAGVEAI IRILQQLLFIHFRIGCRHSRIGVTQQRRARNGASRS

Or you can click on browse button and upload a fasta file containing query sequences. Currently we support maximum 10 query sequences in one job.

3. Follow steps 3 to 5 from part 1 if you are submitting the job as anonymous or steps 5 to 6 from part 2 if you are logged in as a registered user. On the job status page in both cases the job type will be shown as batch.

Once the processing is complete the results will be displayed in the table with each sequence, its identifier and a clickable link which will take you to detailed results for that sequence.