

ESG: Extended Similarity Group Job Submission

Cite:

[Meghana Chitale, Troy Hawkins, Changsoon Park, & Daisuke Kihara](#) *ESG: Extended similarity group method for automated protein function prediction*, *Bioinformatics*, 25: 1739-1745(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 similarity 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)

Tutorial

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 ESG 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 ESG logo highlighted in red below. This will open ESG job submission page in your browser.



ESG is a sequence similarity-based protein function prediction server. It employ PSI-BLAST iteratively and essentially selects GO term annotations that appear consistently in the searches. [Visit the server](#) to submit a sequence or [read the paper](#). ESG was among top in the 1st CAFA function prediction assessment.



PFP is a sequence similarity-based protein function prediction server designed to predict GO annotations for a query sequence beyond what can be found by conventional database search such as BLAST. It takes into account weakly similar sequences as well as GO term associations observed in known annotations. [Visit the server](#) to submit a sequence or [read the paper](#). PFP had the highest total score in a function prediction contest held at AFP-SIG'05 (ISMB2005), and also was the best in the function prediction category at CASP7.

ESG: Extended Similarity Group Job Submission

Gene Ontology Prediction by Iterative PSI-BLAST Search

Enter Query Sequence(s)

Enter your protein sequence here: [?](#) [Clear](#) [Load Sample](#)
Limit 10 sequences

or

Upload your FASTA File: [?](#)

 No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#)

Enter the number of stages [?](#)

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

2. You can enter one or more query protein sequence(s) in [fasta](#) format in the text box titled "Enter Query Sequence(s)". Example of an input sequence:

```
>sp|P28482|MK01_HUMAN  
MAAAAAAGAGPEMVRGQVFDVGPRTYNTLSYIGEGAYGMVCSAYDENVNKKVRAIKKISPFHQTYC  
QRTLREIKILLRFRHENIIGINDIIRAPTIEQMKDYYIVQDLMETDLYKLLKTQHLSNDHICYFL
```

YQILRGLKYIHSANVLHRDLKPSNLLLNTTCDLKICDFGLARVADPDHDHTGFLTEYVATR WYRA
PEIMLNSKGYTKSIDIWSVGCILAEMLSNRPIFPKGKHYLDQLNHILGILGSPSQEDLNCIINLKA
RNYLLSLPHKNKVPWNRLFPNADSKALDLLDKMLTFNPHKRIEVEQALAHPLYEQYYDPSDEPIA
EAPFKFDMELDDLPEKELKELIFEETARFQPGYRS

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.

ESG: Extended Similarity Group Job Submission
Gene Ontology Prediction by Iterative PSI-BLAST Search

Enter Query Sequence(s)
Enter your protein sequence here: [?](#) [Clear](#) [Load Sample](#)
Limit 10 sequences
>sp|P56851|EP3B_HUMAN
MASSLKIWGTLLALLCILCTLLVQSKVSWREFMKQHYLSPREFREYKCDVLMRENEAL
KDKSSHMFYISWYKIEHICTSDNWMDFRNAYVWVQNLKVLKCHQENSKNSYTESRSF
NYIEFHCSMDGYVDSIEDLKMVEPIGN

or

Upload your FASTA File: [?](#)
 No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#)

Enter the number of stages [?](#)

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

Or you can click on browse/choose file 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. Choose ESG parameters

"Number of hits" indicates number of PSI-BLAST hits to be considered at each level of ESG. Default value of this parameter is 10. "Number of stages" indicates the levels of neighborhood around the query protein that will be considered by ESG. Default value for this parameter is 2. Due to computational constraints currently we support parameters where $\{ \text{Number of hits}^{\text{Number of stages}-1} \}$ does not exceed 100. Details about the parameters can be found in the [original paper](#).

ESG: Extended Similarity Group Job Submission
Gene Ontology Prediction by Iterative PSI-BLAST Search

Enter Query Sequence(s)
Enter your protein sequence here: [?](#) [Clear](#) [Load Sample](#)
Limit 10 sequences
>sp|P56851|EP3B_HUMAN
MASSLKIWGTLLALLCILCTLLVQSKVSWREFMKQHYLSPREFREYKCDVLMRENEAL
KDKSSHMFYISWYKIEHICTSDNWMDFRNAYVWVQNLKVLKCHQENSKNSYTESRSF
NYIEFHCSMDGYVDSIEDLKMVEPIGN

or

Upload your FASTA File: [?](#)
 No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#) → Number of Hits

Enter the number of stages [?](#) → Number of Stages

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

4. 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.

5. 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.

Job Status

Job ID	Type	Status	Date Submitted
000425	ESG	Waiting for CPU time	2009-12-31

You may bookmark this page to view your results at a later date or use the Job Retrieval form with your Job ID#.

When your job is complete, predicted Gene Ontology terms for the query protein will be displayed. Or you can refresh the page to go to results. The results page will show ESG input parameters, query sequence and prediction terms for each Gene Ontology category (Molecular function, Biological Process and Cellular Component) that have probability greater than 5%. The results page will also provide link to the results in xml format which you can download for further processing. Clicking the plus button next to a GO term will show a list of 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.

If you click "Download Prediction's XML", an XML file that have all information shown in the result page can be obtained. Clicking "Visualization of Predicted GO Terms" will pop up a new window that visualizes 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 opening the page, please wait some time to allow server to generate those file.)

ESG Job Results

ESG Parameters

Hits per stage: Number of stages:

Protein Sequence:

```

MQSWSRVYCSLAKRGHFNRI SHGLQGLSAVPLRKYADQPIDADVTVIGSGPGGYVAAIKAAQLGF
KTVCIEKNETLGGTCLNVGCIPSKALLNNSHYHMAHGKDFASRGIEMSEVRLNLDKMMEQKSTA
VKALTGGIAHLFKQNKVVHNGYKKTGKNQVTATKADGGTQVIDTKNILIATGSEVTPFFGITI
DEDTIIVSSTGALSLLKVPKMMVIGAGVIGVELGSVWQRLGADVTAVEFLGHVGGVGDMEISKN
FQRILQKQGFKLNKVTGATKSDGKIDVSIEAASGGKAEVITCDVLLVCIGRRPPTKRLGLE
ELGIELDFRGRIPVNTRFQTKIPNIYAIGDVVAGPMLAHKAEDEGIICVEGMAGGAVHIDYNCVP
SVIYTHPEVAWVGKSEEQLKEEGIEYKVGKFFAANSRAKTNADTGMVKILGQKSTDRVLGAHI
LPGGAGEMVNEAALALEYGASCEDIARVCHAHPTLSEAFREANLAASFGKSINF
        
```

Predicted GO Terms

[Download Prediction's XML](#)

[Visualization of Predicted GO Terms](#)

Download Static Visualization Figure

[\[ME\]](#), [\[BP\]](#), [\[CC\]](#)

List of Predicted GO Terms

Very high confidence : > 70%

High confidence : > 40%

Moderate confidence : >= 30%

Below moderate confidence : < 30%

Molecular Function Terms

Probability	Term	Description
79.2%	GO:0004148 [+]	dihydropolyl dehydrogenase activity
79.2%	GO:0050660 [+]	flavin adenine dinucleotide binding
48.6%	GO:0016491 [+]	oxidoreductase activity
48.6%	GO:0016668 [+]	oxidoreductase activity, acting on a sulfur group of donors, NAD(P) as acceptor

4 Predictions
2 Predictions > 70%; 2 Predictions > 40%; 0 Predictions >= 30%

Biological Process Terms

Probability	Term	Description
79.2%	GO:0045454 [+]	cell redox homeostasis
48.6%	GO:0055114 [+]	oxidation-reduction process
10.3%	GO:0006090 [+]	pyruvate metabolic process
10.3%	GO:0006099 [+]	tricarboxylic acid cycle
10.3%	GO:0006120 [+]	mitochondrial electron transport, NADH to ubiquinone
10.3%	GO:0006508 [+]	proteolysis
10.3%	GO:0006554 [+]	lysine catabolic process
10.3%	GO:0007369 [+]	gastrulation
10.3%	GO:0009083 [+]	branched-chain amino acid catabolic process
10.3%	GO:0010510 [+]	regulation of acetyl-CoA biosynthetic process from pyruvate
10.3%	GO:0034641 [+]	cellular nitrogen compound metabolic process
10.3%	GO:0042391 [+]	regulation of membrane potential
10.3%	GO:0044237 [+]	cellular metabolic process
10.3%	GO:0044281 [+]	small molecule metabolic process
10.3%	GO:0048240 [+]	sperm capacitation
9.0%	GO:0006629 [+]	lipid metabolic process
9.0%	GO:0006631 [+]	fatty acid metabolic process
9.0%	GO:0006633 [+]	fatty acid biosynthetic process

18 Predictions
1 Predictions > 70%; 1 Predictions > 40%; 0 Predictions >= 30%

Cellular Component Terms

Probability	Term	Description
79.2%	GO:0005759 [+]	mitochondrial matrix
58.9%	GO:0005739 [+]	mitochondrion
10.3%	GO:0005634 [+]	nucleus
10.3%	GO:0005730 [+]	nucleolus
10.3%	GO:0005929 [+]	cilium
10.3%	GO:0043159 [+]	acrosomal matrix

6 Predictions
1 Predictions > 70%; 1 Predictions > 40%; 0 Predictions >= 30%

6. 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 see the job status or results page depending on if the job is complete or not.

Member Portal Access

Username:

Password:

[Forgot your password?](#)

[Register](#)

Job Retrieval

Job ID#:

Part2: Steps for logging in and submitting ESG job

1. Open ESG job submission page in your browser.

ESG
EXTENDED SIMILARITY GROUP

ESG: Extended Similarity Group
Gene Ontology Prediction by Iterative PSI-BLAST Search

Enter Query Sequence(s)
Enter your protein sequence here: [?](#) [Clear](#) [Load Sample](#)
Limit 10 sequences

or

Upload your FASTA File: [?](#)
 No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#)

Enter the number of stages [?](#)

Email Notifications

To receive email notifications you must first login or create a new account

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

[PPP Submission](#)
[ESG Submission](#)

Tutorials

[ESG Tutorial](#)
[PPP Tutorial](#)

New [PPP and ESG Features](#)

Member Portal Access

Username:

Password:

[Register](#) [Forgot Password?](#)

Job Retrieval

Job ID#:

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.

New Account Creation

Account Settings

Username: *

Email Address:

Receive Email Notifications

Password: *

Confirm Password: *

* denotes required field

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.

Reset Password

Enter Email:

Email Address:

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 and click login. This will redirect you to the software page with links to your jobs appears on right side frame on the page.

Member Portal

[Change Settings](#)
[My Jobs](#)
[Logout](#)

Job Retrieval

Job ID#:

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 ESG job submission page shown in step 1.

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

```
>sp|P28482|MK01_HUMAN
MAAAAAGAGPEMVRGQVDFVGPRTYTNLSYIGEGAYGMVCSAYDNVNVKVRVAIKKISPFQHTYC
QRTLREIKILLRFRHENIIGINDIIRAPTIEQMKDVYIVQDLMETDLYKLLKTQHLSNDHICYFL
YQILRGLKYIHSANVLRDLKPSNLLLNTTCDLKICDFGLARVADPDHDHTGFLTEYVATRKYRA
PEIMLNSKGYTKSIDIWSVGCILAEMLSNRPIFPGKHYLDQLNHILGILGSPSQEDLNCIINLKA
RNYLLSLPHKNKVPWNRLFPNADSKALDLLDKMLTFNPHKRIEVEQALAHPLYEQYYDPSDEPIA
EAPFKFDMELDDLPEKELKELIFEETARFQPGYRS
```

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.

ESG: Extended Similarity Group Job Submission

Gene Ontology Prediction by Iterative PSI-BLAST Search

Enter Query Sequence(s)

Enter your protein sequence here: [?](#) [Clear](#) [Load Sample](#)
Limit 10 sequences

>sp|P56851|EP3B_HUMAN
MASSLKIWGTLLALLCILCTLLVQSKEVSWREFMKQHYLSPSREFREYKCDVLMRENEAL
KDKSSHMFYISWYKIEHICTSDNWMDFRNAYVWVQNLKVLKCHQENSKNSYTESRSF
NYIEFHCSMDGYVDSIEDLKMVEPIGN

or

Upload your FASTA File: [?](#)
 No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#)

Enter the number of stages [?](#)

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

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

5. Choose ESG parameters

"Number of hits" indicates number of PSI-BLAST hits to be considered at each level of ESG. Default value of this parameter is 10. "Number of stages" indicates the levels of neighborhood around the query protein that will be considered by ESG. Default value for this parameter is 2. Due to computational constraints currently we support parameters where $\{ \text{Number of hits}^{\text{Number of stages}-1} \}$ does not exceed 100. Details about the parameters can be found in the [original paper](#).

6. After loading the query sequence(s) and parameters 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.

7. When your job status shows complete, click on the job ID 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 ESG

1. Open ESG job submission page in your browser.

ESG: Extended Similarity Group Job Submission
Gene Ontology Prediction by Iterative PSI-BLAST Search

Enter Query Sequence(s)
Enter your protein sequence here: [?](#) [Clear](#) [Load Sample](#)
Limit 10 sequences

or

Upload your FASTA File: [?](#)
 No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#)

Enter the number of stages [?](#)

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

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
MASSLKIWGTLLALLCILCTLLVQSKEVSWREFMKQHLYLSPSREFREYKCDVLMRENEAL
KDKSSHMFYISWYKIEHICTSDNWMDRFRNAYVWVQNPLKVLKCHQENSKNSYTESRSF
NYIEFHCSMDGYVDSIEDLKMVEPIGN
>sp|P13811|ELBH_ECOLX
MNVKVFYVLFALLSSLCAHGAPQSITELCSEYHNTQIYTINDKILSYTESMAGKREMVI
ITFKSGATFQVEVPGSQHIDSQKKAIERMKDTRLRITYLTETKIDKLCVWNNKTPNSIAAI
SMEN
>sp|P05928|VPR_HV1BR
MEQAPEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQHIYETYGDTWAGVEAI
IRILQQLLFIHFRIGCRHSRIGVTQRRRANGASRS
```

ESG: Extended Similarity Group Job Submission

Gene Ontology Prediction by Iterative PSI-BLAST Search

Enter Query Sequence(s)

Enter your protein sequence here: [?](#) [Clear](#) [Load Sample](#)
Limit 10 sequences

```
>sp|P56851|EP3B_HUMAN
MASSLKIWGTLALLCILCTLLVQSKEVSWREFMKQHYLSPSREFREYKCDVLMRENEAL
KDKSSHMFYISWYKIEHICTSDNWMDFRNAYVWVQNPLKVLKCHOENSKNSYTESRSF
NYIEFHCSMDGYVDSIEDLKMVEPIGN
>sp|P13811|ELBH_ECOLX
MNKVKFYVLFALLSSLCAHGAPOSITELCSEYHNTQIYTINDKILSYTESMAGKREMVI
ITFKSGATEQVEVPGSQHIDSQKKAIERMKDILRITYLTETKIDKLCVWNNKTPNSIAAI
SMEN
```

or

Upload your FASTA File: [?](#)
 No file chosen

Choose ESG Parameters

Enter the number of hits per stage [?](#)

Enter the number of stages [?](#)

Email Notifications

To receive email notifications you must first login or create a new account

Submit

This website is free and open to all users and there is no login requirement.

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 6 from part 1 if you are submitting the job as anonymous or steps 5 to 7 from part 2 if you are logged in as a registered user. On the job status page the job type will be shown as batch.

Your jobs

Job ID	Type	Status	Date Submitted	Batch Job
000401	ESG	Job Processing	2009-12-19	Yes
000400	ESG	Done - Results	2009-12-19	No

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.

©2014 Kihara Bioinformatics Laboratory - For technical issues please contact the [administrator](#)