Instruction for conservation analysis tool

Changzhi Wang

This tool is coded by Changzhi Wang for Dr. Yue Chen's lab in University of Minnesota, Twin City. And this tool based on database from PhosphositePlus (www.phosphosite.org/), Eggnog (http://eggnogdb.embl.de/#/app/home), WebGestalt (http://www.webgestalt.org/option.php) also including database from Uniprot (http://www.uniprot.org/). This tool is designed based on the format of those database, so that potential error might exist without using these data format.

The programming language of this tool is python 3.5. Other python version might cause some unpredictable error.

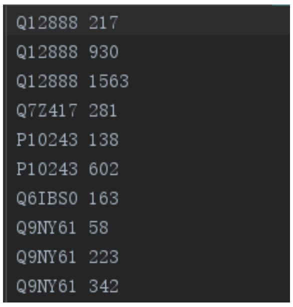
There are 12 steps in 2 python coded documents. The first document includes step1 and step2, and the second document includes from step 3 to step12. The output of the first document is to be uploaded on eggnog mapper for OG identification, the second document is to further analyze the output from eggnog, and the output of document 2 is the result include the conservation information on each site with p value and percentage.

STEP 1 – Information screening on original database from PhosPhosite plus.

PhosPhositePlus provides database for different biological mechanisms for downloading on their website (https://www.phosphosite.org/staticDownloads.action). After extract the database, those databases will input our first python document.

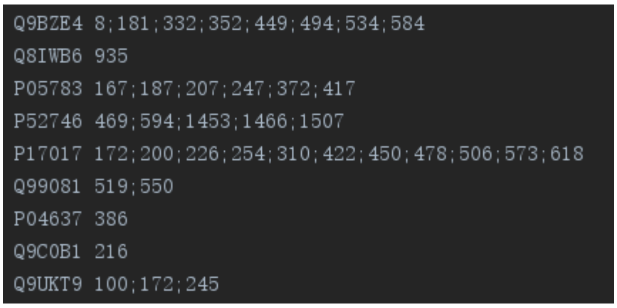
Once we run, the coding will ask for the name of original dataset, and all dataset should follow the format of database from PhosphositePlus.org. And the dataset should be put in the same folder of the python coding. After typing the name of database, user can choose the specie they want to analyze and the specific amino acid (for all amino acid in this mechanism, input 'all') , then press enter for running.

The running time should be short, and there will be 2 output files named: 1-sites.txt and 1-set.txt. And sites.txt represent all human sites in this database with Uniprot Id and sites position. set.txt means the cluster for each protein with uniprot ID and all its corresponded sites positions from the database, separate by semicolon.



```
Q12888 217
Q12888 930
Q12888 1563
Q7Z417 281
P10243 138
P10243 602
Q6IBS0 163
Q9NY61 58
Q9NY61 223
Q9NY61 342
```

Example output for 1-sites.txt



```
Q9BZE4 8;181;332;352;449;494;534;584
Q8IWB6 935
P05783 167;187;207;247;372;417
P52746 469;594;1453;1466;1507
P17017 172;200;226;254;310;422;450;478;506;573;618
Q99081 519;550
P04637 386
Q9C0B1 216
Q9UKT9 100;172;245
```

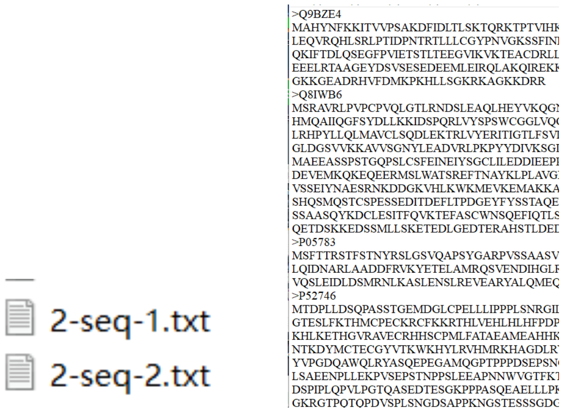
Example output for 1-set.txt

STEP 2-Protein sequence download and instruction for eggnog mapper

Eggnog mapper (http://eggnogdb.embl.de/#/app/emapper) is a powerful tool for the identification of orthology group. By searching and aligning the sequences we upload, it can provide the output about best identified orthology group of each protein. And in our analysis, we need eggnog mapper to tell us which is the orthology group that our protein belong to, as well as the raw alignment for the protein with all other proteins in the same OG, it will be useful for us to determine the conservation of each modified sites.

The eggnog mapper has a number limitation of uploaded proteins, which can not up to 5000, so that to identify the OG for our proteins, we need all their amino acid sequences and upload them to eggnog mapper. And 2seqsrequest.py is coded for sequences from uniprot.org, and separate our data into files containing proteins less than 4500 for eggnog mapper.

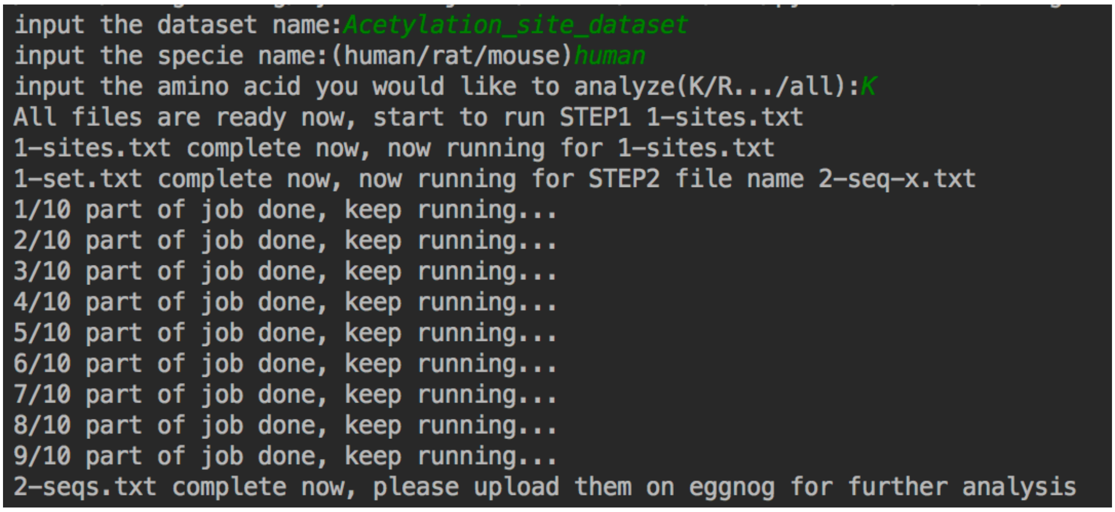
Step2 will automatically process by importing the file's name from step1 (1-set.txt), the coding will connect to the uniprot.org for amino sequences and the output files will be listed.



2-seq-1.txt
2-seq-2.txt

```
>Q9BZE4
MAHYNFKKITVVPSAKDFIDLTLSKTQRKTPTVIH
LEQVRQHLSRLPTIDPNTRTLLLCGYPNVGKSSFIN
QKIFTDLQSEGFPVIETSTLTEEGVIKVKTEACDRL
EEELRTAAGEYDSVSESEDEEMLEIRQLAKQIREK
GKKGEADRHVFDMKPKHLLSGKRKAGKKDRR
>Q8IWB6
MSRAVRLPVPCPVQLGTLRNDSLEAQLHEYVKQG
HMQAIIQGFSYDLLKKIDSPQRLVYSPSWCGGLVQ
LRHPYLLQLMAVCLSQDLEKTRLVYERITIGTLFSV
GLDGSVVKKAVVSGNYLEADVRLPKPYYDIVKSG
MAEEASSPSTGQPSLCSFEINEIYSGCLILEDDIEEP
DEVEMKQKEQEERMSLWATSREFTNAYKLPLAVG
VSSEIYNAESRNKDDGKVHLKWKMEVKEMAKKA
SHQSMQSTCSPESSEDITDEFLTPDGEYFYSSTAQE
SSAASQYKDCLESITFQVKTEFASCWNSQEFIQTLS
QETDSKKEDSSMLLSKETEDLGEDTERAHSTLDE
>P05783
MSFTTRSTFSTNYRSLGSVQAPSYGARPVSSAASV
LQIDNARLAADDFRVKYETELAMRQSVENDIHGL
VQSLEIDLDSMRNLKASLENSLREVEARYALQME
>P52746
MTDPLLDSQPASSTGEMDGLCPELLLIPPPLSNRGI
GTESLFKTHMCPECKRCFKKRTHLVEHLHLHFPDP
KHLKETHGVRAVECRHHSCPMLFATAEAMEAHH
NTKDYMCTECGYVTKWKHYLRVHMRKHAGDLR
YVPGDQAWQLRYASQEPEGAMQGPTPPPDSEPSN
LSAEENPLLEKPVSEPSTNPPSLEEAPNNWVGTFK
DSPIPLQPVLPGTQASEDTESGKPPPASQEAELLLP
GKRGTPQTQPDVSPLSNGDSAPPKNGSTESSSGD
```

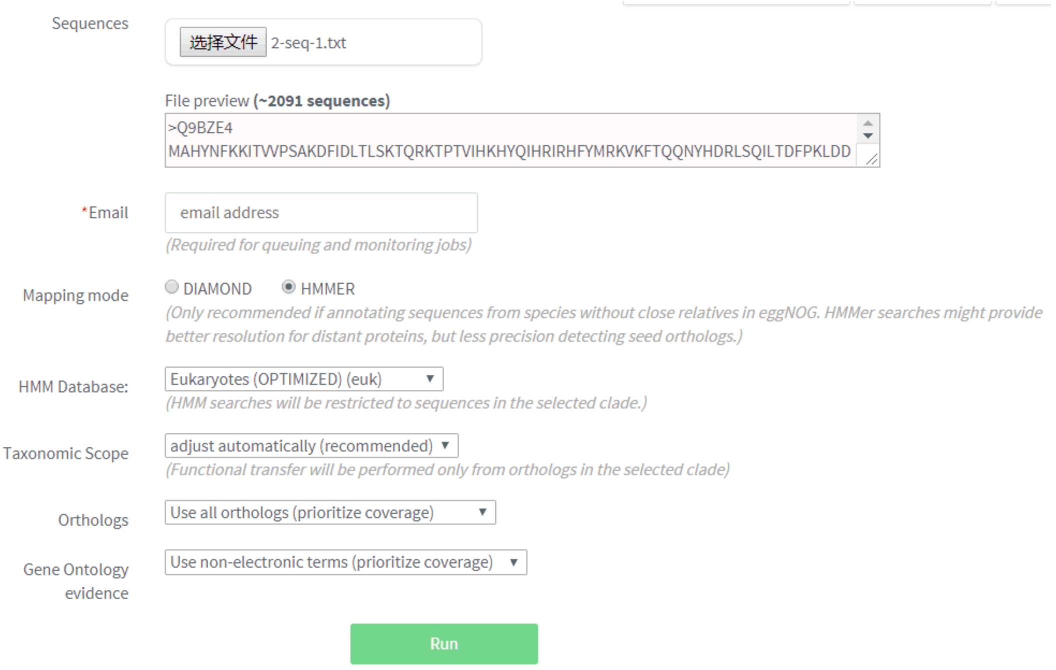
Output sample files and contents



```
input the dataset name:Acetylation_site_dataset
input the specie name:(human/rat/mouse)human
input the amino acid you would like to analyze(K/R.../all):K
All files are ready now, start to run STEP1 1-sites.txt
1-sites.txt complete now, now running for 1-sites.txt
1-set.txt complete now, now running for STEP2 file name 2-seq-x.txt
1/10 part of job done, keep running...
2/10 part of job done, keep running...
3/10 part of job done, keep running...
4/10 part of job done, keep running...
5/10 part of job done, keep running...
6/10 part of job done, keep running...
7/10 part of job done, keep running...
8/10 part of job done, keep running...
9/10 part of job done, keep running...
2-seqs.txt complete now, please upload them on eggnog for further analysis
```

Running example of step1 and step2

Then we should upload the outputs singly to eggnog mapper as follow:



Sequences: 选择文件 2-seq-1.txt

File preview **(~2091 sequences)**

>Q9BZE4
MAHYNFKKITVVPSAKDFIDLTLSKTQRKTPTVIHKHYQIHRIRHFYMRKVKFTQQNYHDRLSQILTDFPKLDD

*Email: email address

(Required for queuing and monitoring jobs)

Mapping mode: ○ DIAMOND ● HMMER

(Only recommended if annotating sequences from species without close relatives in eggNOG. HMMer searches might provide better resolution for distant proteins, but less precision detecting seed orthologs.)

HMM Database: Eukaryotes (OPTIMIZED) (euk)

(HMM searches will be restricted to sequences in the selected clade.)

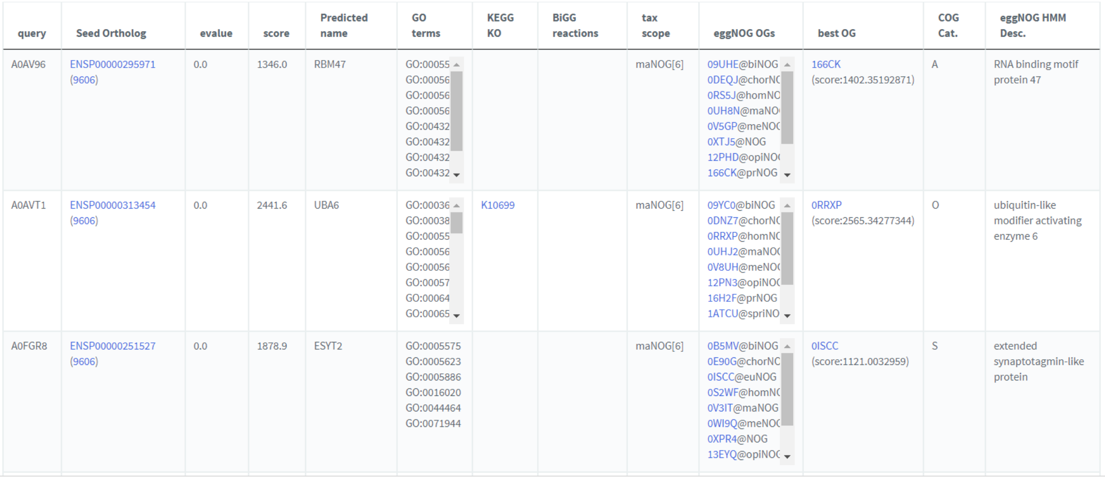
Taxonomic Scope: adjust automatically (recommended)

(Functional transfer will be performed only from orthologs in the selected clade)

Orthologs: Use all orthologs (prioritize coverage)

Gene Ontology evidence: Use non-electronic terms (prioritize coverage)

Once the eggnog mapper finished the job, we can see the web page as follow:



query	Seed Ortholog	evalue	score	Predicted name	GO terms	KEGG KO	BiGG reactions	tax scope	eggNOG OGs	best OG	COG Cat.	eggNOG HMM Desc.
A0AV96	ENSP00000295971 (9606)	0.0	1346.0	RBM47	GO:00055 GO:00056 GO:00056 GO:00056 GO:00432 GO:00432 GO:00432 GO:00432			maNOG[6]	09UHE@biNOG 0DEQJ@chorNC 0RS5J@homNC 0UH8N@maNO 0V5GP@meNOC 0XTJ5@NOG 12PHD@opiNO 166CK@prNOG	166CK (score:1402.35192871)	A	RNA binding motif protein 47
A0AVT1	ENSP00000313454 (9606)	0.0	2441.6	UBA6	GO:00036 GO:00038 GO:00055 GO:00056 GO:00056 GO:00057 GO:00064 GO:00065	K10699		maNOG[6]	09YC0@biNOG 0DNZ7@chorNC 0RRXP@homNC 0UHJ2@maNOC 0V8UH@meNOC 12PN3@opiNOC 16H2F@prNOG 1ATCU@spriNO	0RRXP (score:2565.34277344)	O	ubiquitin-like modifier activating enzyme 6
A0FGR8	ENSP00000251527 (9606)	0.0	1878.9	ESYT2	GO:0005575 GO:0005623 GO:0005886 GO:0016020 GO:0044464 GO:0071944			maNOG[6]	0B5MV@biNOG 0E90G@chorNC 0ISCC@euNOG 0S2WF@homNC 0V3IT@maNOG 0WI9Q@meNOC 0XPR4@NOG 13EYQ@opiNOC	0ISCC (score:1121.0032959)	S	extended synaptotagmin-like protein

And we need to download these outputs for further analysis.

Sumoseq-1.txt.emapper.annotations

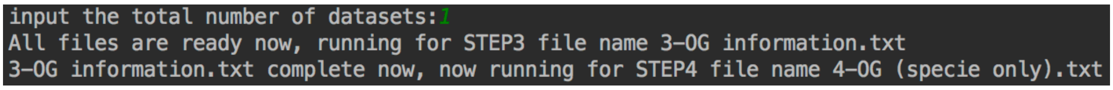
Sumoseq-2.txt.emapper.annotations

Sumoseq-3.txt.emapper.annotations

Output sample files

STEP 3-OG files modification (information filter)

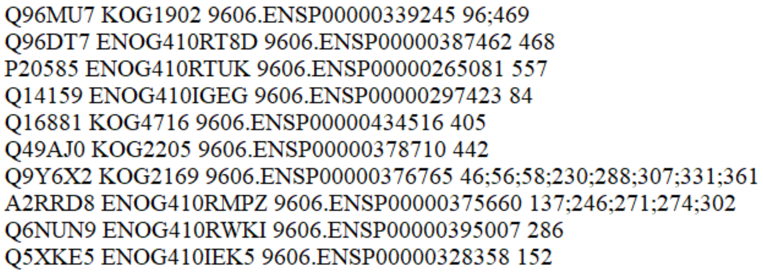
After we download the annotations datasets, we will start to run python document 2 for further analysis. And the step3 is to screen the information from eggnog output. By inputting the total number of eggnog output, those annotation files would collect together into one file 3-OG information.txt.



```
input the total number of datasets:1
All files are ready now, running for STEP3 file name 3-OG information.txt
3-OG information.txt complete now, now running for STEP4 file name 4-OG (specie only).txt
```

Running Example for step3

And the output file 3-OG information.txt is the output we need, which contains the ID, OG and sites.

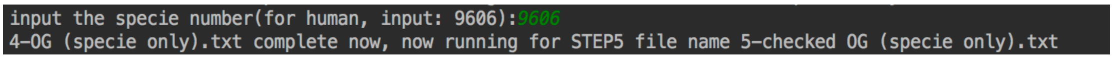


```
Q96MU7 KOG1902 9606.ENSP00000339245 96;469
Q96DT7 ENOG410RT8D 9606.ENSP00000387462 468
P20585 ENOG410RTUK 9606.ENSP00000265081 557
Q14159 ENOG410IGEG 9606.ENSP00000297423 84
Q16881 KOG4716 9606.ENSP00000434516 405
Q49AJ0 KOG2205 9606.ENSP00000378710 442
Q9Y6X2 KOG2169 9606.ENSP00000376765 46;56;58;230;288;307;331;361
A2RRD8 ENOG410RMPZ 9606.ENSP00000375660 137;246;271;274;302
Q6NUN9 ENOG410RWKI 9606.ENSP00000395007 286
Q5XKE5 ENOG410IEK5 9606.ENSP00000328358 152
```

Contents for sample

STEP 4-Specie filter

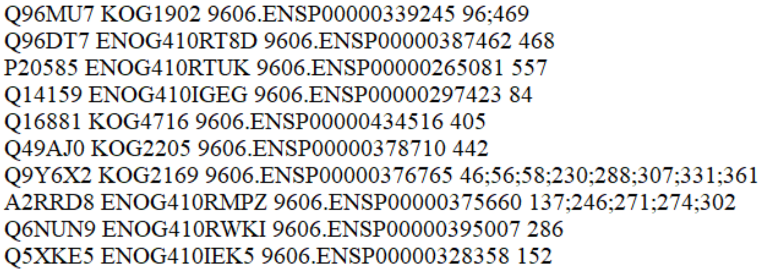
And from the last step, we get the file with annotation, name ID and sites, although we have filter the analysis with a specific specie protein only, but we need to make sure that these proteins are actually identified in the right specie. So that step 4 is to filter again those proteins identified in the same specific specie. By inputting the specie number (e.g. human 9606), the result will be cleared with human protein left only.



```
input the specie number(for human, input: 9606):9606
4-OG (specie only).txt complete now, now running for STEP5 file name 5-checked OG (specie only).txt
```

Running Example for step4

And this step should complete extremely quick.

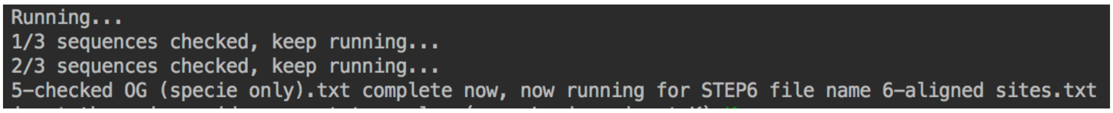


```
Q96MU7 KOG1902 9606.ENSP00000339245 96;469
Q96DT7 ENOG410RT8D 9606.ENSP00000387462 468
P20585 ENOG410RTUK 9606.ENSP00000265081 557
Q14159 ENOG410IGEG 9606.ENSP00000297423 84
Q16881 KOG4716 9606.ENSP00000434516 405
Q49AJ0 KOG2205 9606.ENSP00000378710 442
Q9Y6X2 KOG2169 9606.ENSP00000376765 46;56;58;230;288;307;331;361
A2RRD8 ENOG410RMPZ 9606.ENSP00000375660 137;246;271;274;302
Q6NUN9 ENOG410RWKI 9606.ENSP00000395007 286
Q5XKE5 ENOG410IEK5 9606.ENSP00000328358 152
```

Contents for sample

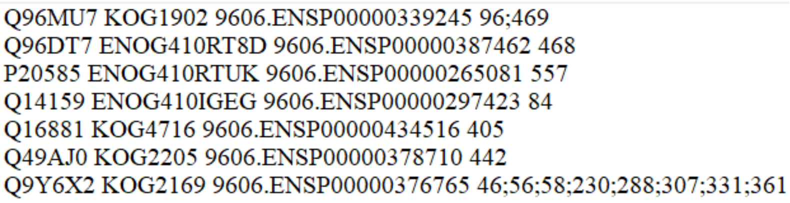
STEP 5-Sequence check

This step is to check if the sequence in OG and sequence by uniprot ID are 100% same. Because the same sequence is important to identify the sites in OG for conservation analysis. And by automatically keep running, we can get the output file 5-checked OG (specie only).txt which contain the proteins which are same sequence in uniprot ID and assemble ID.



```
Running...
1/3 sequences checked, keep running...
2/3 sequences checked, keep running...
5-checked OG (specie only).txt complete now, now running for STEP6 file name 6-aligned sites.txt
```

Running Example for step5



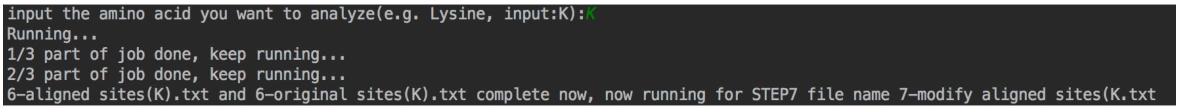
```
Q96MU7 KOG1902 9606.ENSP00000339245 96;469
Q96DT7 ENOG410RT8D 9606.ENSP00000387462 468
P20585 ENOG410RTUK 9606.ENSP00000265081 557
Q14159 ENOG410IGEG 9606.ENSP00000297423 84
Q16881 KOG4716 9606.ENSP00000434516 405
Q49AJ0 KOG2205 9606.ENSP00000378710 442
Q9Y6X2 KOG2169 9606.ENSP00000376765 46;56;58;230;288;307;331;361
```

Contents for sample

STEP 6-aligned sites position searching

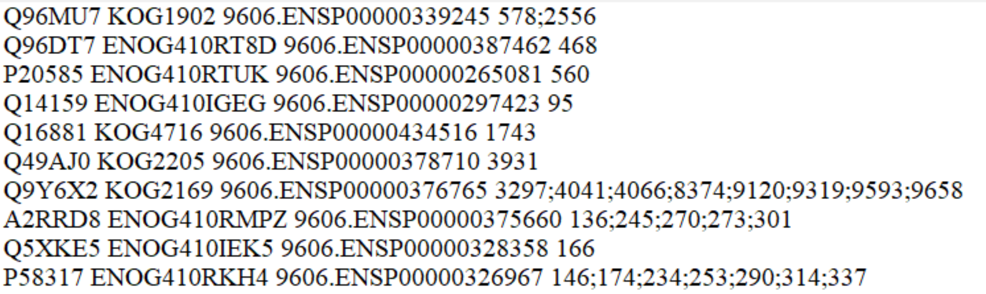
This Eggnog database provide raw_alignment dataset for each OG, and that will be crucial to identify the position of our modification sites' position after the raw alignment. So step6 will search for the aligned position for each modification sites.

In step6, type in the specific amino acid we want (upper letter). For instance, when we type in K for sumoylation database, all modified lysine sites will be select with the aligned position number in the output named 6-aligned sites(K).txt.



```
input the amino acid you want to analyze(e.g. Lysine, input:K):K
Running...
1/3 part of job done, keep running...
2/3 part of job done, keep running...
6-aligned sites(K).txt and 6-original sites(K).txt complete now, now running for STEP7 file name 7-modify aligned sites(K.txt
```

Running Example for STEP6

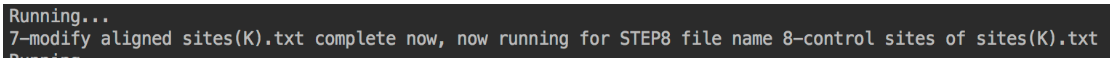


```
Q96MU7 KOG1902 9606.ENSP00000339245 578;2556
Q96DT7 ENOG410RT8D 9606.ENSP00000387462 468
P20585 ENOG410RTUK 9606.ENSP00000265081 560
Q14159 ENOG410IGEG 9606.ENSP00000297423 95
Q16881 KOG4716 9606.ENSP00000434516 1743
Q49AJ0 KOG2205 9606.ENSP00000378710 3931
Q9Y6X2 KOG2169 9606.ENSP00000376765 3297;4041;4066;8374;9120;9319;9593;9658
A2RRD8 ENOG410RMPZ 9606.ENSP00000375660 136;245;270;273;301
Q5XKE5 ENOG410IEK5 9606.ENSP00000328358 166
P58317 ENOG410RKH4 9606.ENSP00000326967 146;174;234;253;290;314;337
```

Contents for output sample

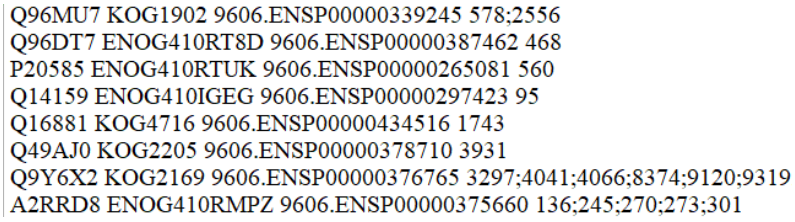
STEP 7-aligned sites position check and control sites screening

After we get the file contain the aligned sites, we need to screen it because we want to delete all proteins which do not have any target amino acid as modification sites. The second python document will automatically enter this step, it will output the file 7-control sites of sites(K).txt. Then by input the target again, it will search for all target amino acids in each protein which have any modified target amino acid site.



```
Running...
7-modify aligned sites(K).txt complete now, now running for STEP8 file name 8-control sites of sites(K).txt
```

Running Example for step 7

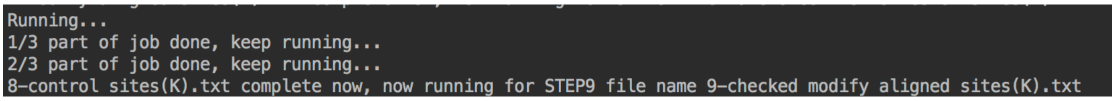


```
Q96MU7 KOG1902 9606.ENSP00000339245 578;2556
Q96DT7 ENOG410RT8D 9606.ENSP00000387462 468
P20585 ENOG410RTUK 9606.ENSP00000265081 560
Q14159 ENOG410IGEG 9606.ENSP00000297423 95
Q16881 KOG4716 9606.ENSP00000434516 1743
Q49AJ0 KOG2205 9606.ENSP00000378710 3931
Q9Y6X2 KOG2169 9606.ENSP00000376765 3297;4041;4066;8374;9120;9319
A2RRD8 ENOG410RMPZ 9606.ENSP00000375660 136;245;270;273;301
```

Contents for 7-control sites of sites(K).txt

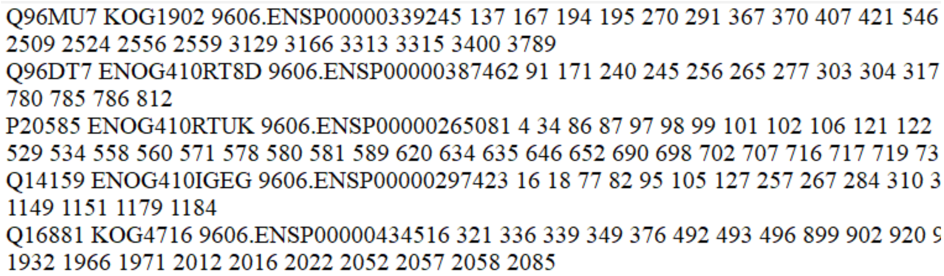
STEP 8-control sites position searching

Then we need to find out all positions of that specific amino acids as our control to compare with the percentage of functional amino acids for conservative analysis in further steps. And this step takes quite a long time and the output file will be ID, OG, assemble name and amino acid position separated by space. And this step is still automatically processed without any input from user.



```
Running...
1/3 part of job done, keep running...
2/3 part of job done, keep running...
8-control sites(K).txt complete now, now running for STEP9 file name 9-checked modify aligned sites(K).txt
```

Running Example for step 8



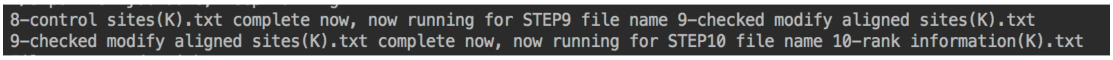
```
Q96MU7 KOG1902 9606.ENSP00000339245 137 167 194 195 270 291 367 370 407 421 546
2509 2524 2556 2559 3129 3166 3313 3315 3400 3789
Q96DT7 ENOG410RT8D 9606.ENSP00000387462 91 171 240 245 256 265 277 303 304 317
780 785 786 812
P20585 ENOG410RTUK 9606.ENSP00000265081 4 34 86 87 97 98 99 101 102 106 121 122
529 534 558 560 571 578 580 581 589 620 634 635 646 652 690 698 702 707 716 717 719 73
Q14159 ENOG410IGEG 9606.ENSP00000297423 16 18 77 82 95 105 127 257 267 284 310 3
1149 1151 1179 1184
Q16881 KOG4716 9606.ENSP00000434516 321 336 339 349 376 492 493 496 899 902 920 9
1932 1966 1971 2012 2016 2022 2052 2057 2058 2085
```

Contents for control sites of step 8

STEP 9-format transfer for modify aligned sites file

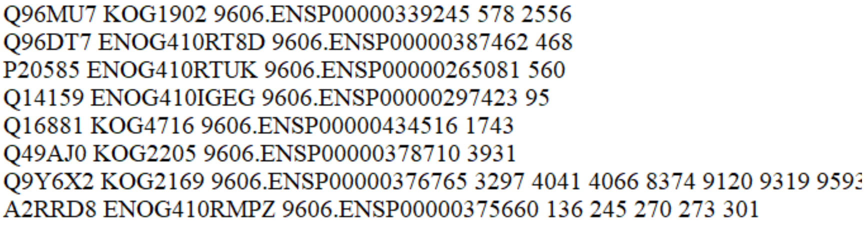
Because of some private reason, we need to transfer the format of 8-modifyalignedxx.txt, the semicolon between each position number will change to blank. And the output name is checked8-modifyaligned sitesxxx.txt.

This step is still automatically processed with extremely quick in time.



```
8-control sites(K).txt complete now, now running for STEP9 file name 9-checked modify aligned sites(K).txt
9-checked modify aligned sites(K).txt complete now, now running for STEP10 file name 10-rank information(K).txt
```

Running Example for step 9

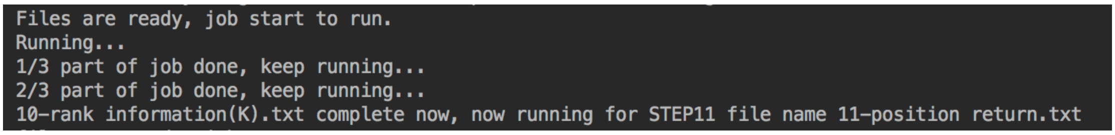


```
Q96MU7 KOG1902 9606.ENSP00000339245 578 2556
Q96DT7 ENOG410RT8D 9606.ENSP00000387462 468
P20585 ENOG410RTUK 9606.ENSP00000265081 560
Q14159 ENOG410IGEG 9606.ENSP00000297423 95
Q16881 KOG4716 9606.ENSP00000434516 1743
Q49AJ0 KOG2205 9606.ENSP00000378710 3931
Q9Y6X2 KOG2169 9606.ENSP00000376765 3297 4041 4066 8374 9120 9319 959
A2RRD8 ENOG410RMPZ 9606.ENSP00000375660 136 245 270 273 301
```

Contents for output sample

STEP 10-conserve analysis part1-percentage calculation

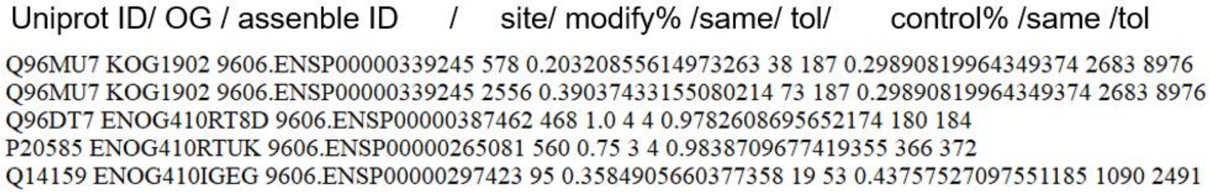
Because of The analysis for conservation depend on the percentage of modified sites with controlled sites in each OG. By input modified sites and control sites, the code will search each position of all proteins in each OG to detect if they have the same amino acid with target sites.



```
Files are ready, job start to run.
Running...
1/3 part of job done, keep running...
2/3 part of job done, keep running...
10-rank information(K).txt complete now, now running for STEP11 file name 11-position return.txt
```

Running Example for step 10

The output file will indicate each correspond sites information. The information listed as: uniprot ID, OG, assemble ID, aligned sites number in OG, percentage in modification site, number of same amino acid for modification site, total number in the position for modification site, percentage of control sites, number of same amino acid for control, total number of control sites.



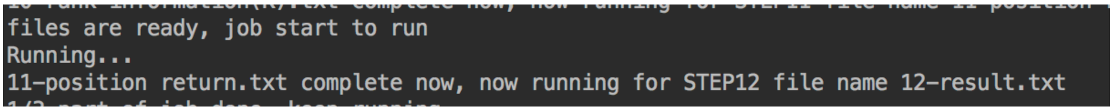
```
Uniprot ID/ OG / assenble ID      /     site/ modify% /same/ tol/       control% /same /tol
Q96MU7 KOG1902 9606.ENSP00000339245 578 0.20320855614973263 38 187 0.29890819964349374 2683 8976
Q96MU7 KOG1902 9606.ENSP00000339245 2556 0.39037433155080214 73 187 0.29890819964349374 2683 8976
Q96DT7 ENOG410RT8D 9606.ENSP00000387462 468 1.0 4 4 0.9782608695652174 180 184
P20585 ENOG410RTUK 9606.ENSP00000265081 560 0.75 3 4 0.9838709677419355 366 372
Q14159 ENOG410IGEG 9606.ENSP00000297423 95 0.3584905660377358 19 53 0.43757527097551185 1090 2491
```

Contents for output sample

STEP 11-aligned sites position return to original position

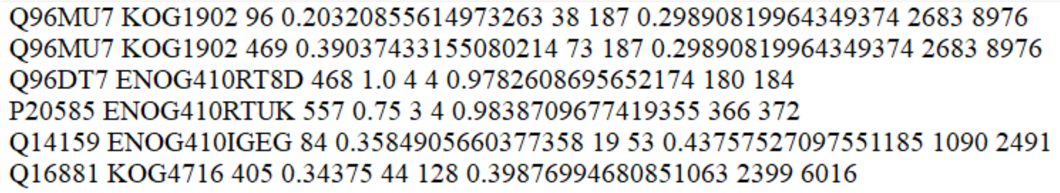
Before we get access to peptide for each site, we need to return those sites from aligned position back to original position. And this step requires an internet connection for eggnog OG data. Step11 will output the file with each ID correspond to original sites.

This step also process automatically.



```
files are ready, job start to run
Running...
11-position return.txt complete now, now running for STEP12 file name 12-result.txt
```

Running Example for step 11



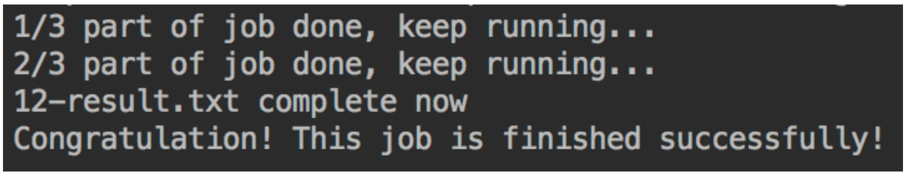
```
Q96MU7 KOG1902 96 0.20320855614973263 38 187 0.29890819964349374 2683 8976
Q96MU7 KOG1902 469 0.39037433155080214 73 187 0.29890819964349374 2683 8976
Q96DT7 ENOG410RT8D 468 1.0 4 4 0.9782608695652174 180 184
P20585 ENOG410RTUK 557 0.75 3 4 0.9838709677419355 366 372
Q14159 ENOG410IGEG 84 0.3584905660377358 19 53 0.43757527097551185 1090 2491
Q16881 KOG4716 405 0.34375 44 128 0.39876994680851063 2399 6016
```

Contents for output sample

STEP 12-peptide searching and p value calculation.

The last step is to get the peptide for each site. The peptide will start from 15 amino acids in front of the sites and end by 15 amino acids behind. And the calculation of p value depends on the 4 numbers about modification sites and control sites. By inputting the file from last step, step12 will output the file named result.txt including uniprot ID, peptide, OG, identified numbers with percentage and p value. The final result can open through excel for further analysis.

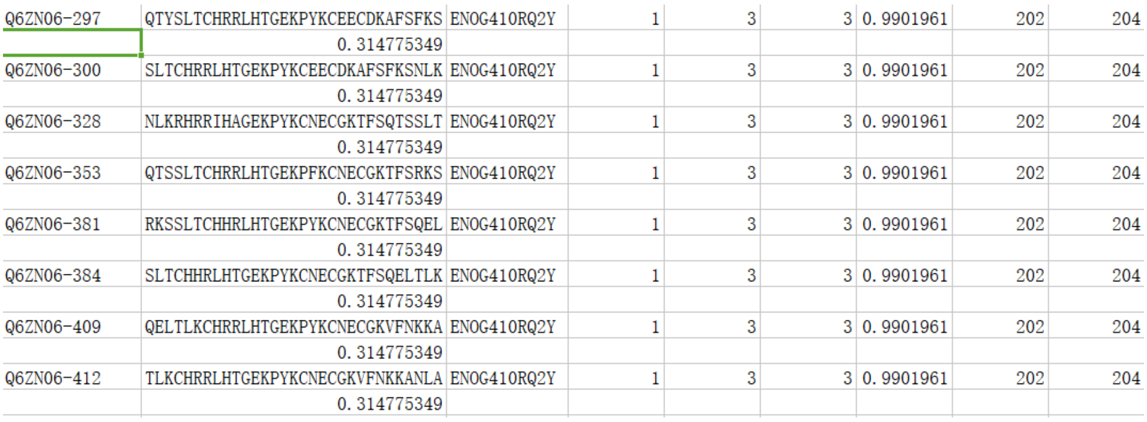
And this step process automatically but it takes a long time in p value calculation, basically it takes days even several weeks depend on the feature of computer and the size of database.



```
1/3 part of job done, keep running...
2/3 part of job done, keep running...
12-result.txt complete now
Congratulation! This job is finished successfully!
```

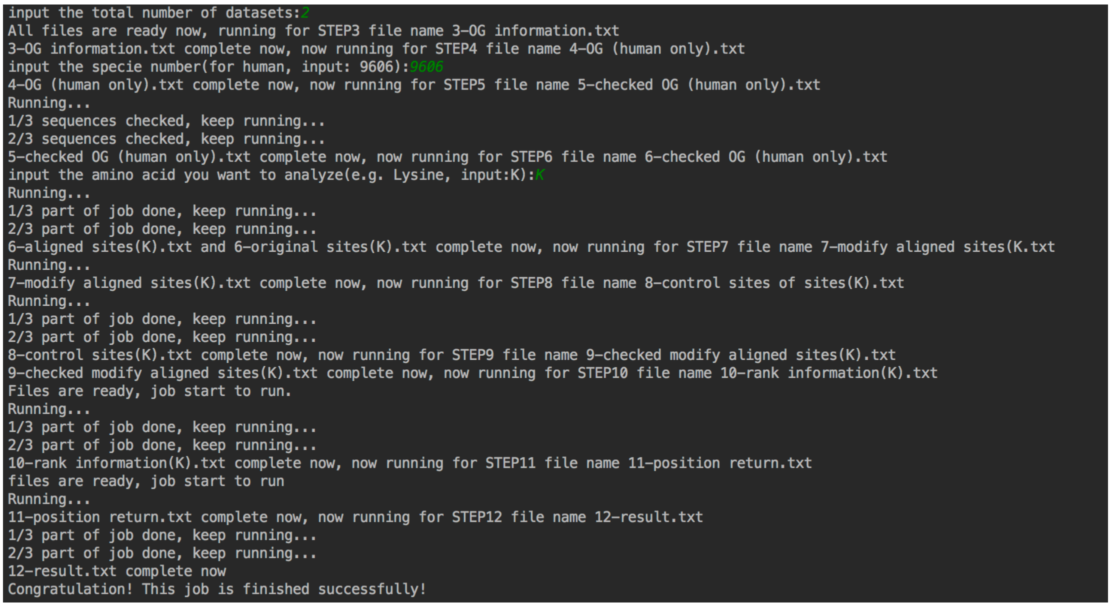
Running Example for step 12

The output of this step can open through excel.



Q6ZN06-297	QTYSLTCHRRLHTGEKPYKCEECDKAFSFKS	ENOG410RQ2Y	1	3	3	0.9901961	202	204
	0.314775349							
Q6ZN06-300	SLTCHRRLHTGEKPYKCEECDKAFSFKSNLK	ENOG410RQ2Y	1	3	3	0.9901961	202	204
	0.314775349							
Q6ZN06-328	NLKRHRRIHAGEKPYKCNECGKTFSQTSSLT	ENOG410RQ2Y	1	3	3	0.9901961	202	204
	0.314775349							
Q6ZN06-353	QTSSLTCHRRLHTGEKPFKCNECGKTFSRKS	ENOG410RQ2Y	1	3	3	0.9901961	202	204
	0.314775349							
Q6ZN06-381	RKSSLTCHHRLHTGEKPYKCNECGKTFSQEL	ENOG410RQ2Y	1	3	3	0.9901961	202	204
	0.314775349							
Q6ZN06-384	SLTCHHRLHTGEKPYKCNECGKTFSQELTLK	ENOG410RQ2Y	1	3	3	0.9901961	202	204
	0.314775349							
Q6ZN06-409	QELTLKCHRRLHTGEKPYKCNECGKVFNKKA	ENOG410RQ2Y	1	3	3	0.9901961	202	204
	0.314775349							
Q6ZN06-412	TLKCHRRLHTGEKPYKCNECGKVFNKKANLA	ENOG410RQ2Y	1	3	3	0.9901961	202	204
	0.314775349							

Contents for output sample



```
input the total number of datasets:2
All files are ready now, running for STEP3 file name 3-OG information.txt
3-OG information.txt complete now, now running for STEP4 file name 4-OG (human only).txt
input the specie number(for human, input: 9606):9606
4-OG (human only).txt complete now, now running for STEP5 file name 5-checked OG (human only).txt
Running...
1/3 sequences checked, keep running...
2/3 sequences checked, keep running...
5-checked OG (human only).txt complete now, now running for STEP6 file name 6-checked OG (human only).txt
input the amino acid you want to analyze(e.g. Lysine, input:K):K
Running...
1/3 part of job done, keep running...
2/3 part of job done, keep running...
6-aligned sites(K).txt and 6-original sites(K).txt complete now, now running for STEP7 file name 7-modify aligned sites(K.txt
Running...
7-modify aligned sites(K).txt complete now, now running for STEP8 file name 8-control sites of sites(K).txt
Running...
1/3 part of job done, keep running...
2/3 part of job done, keep running...
8-control sites(K).txt complete now, now running for STEP9 file name 9-checked modify aligned sites(K).txt
9-checked modify aligned sites(K).txt complete now, now running for STEP10 file name 10-rank information(K).txt
Files are ready, job start to run.
Running...
1/3 part of job done, keep running...
2/3 part of job done, keep running...
10-rank information(K).txt complete now, now running for STEP11 file name 11-position return.txt
files are ready, job start to run
Running...
11-position return.txt complete now, now running for STEP12 file name 12-result.txt
1/3 part of job done, keep running...
2/3 part of job done, keep running...
12-result.txt complete now
Congratulation! This job is finished successfully!
```

Complete process of the second python file. (example in figure is sumoylation database from phosphositePlus.org)