

Input Nucleotide FASTA of Coding Sequences

taxon1.cds

```
>Gene1.1
GCGATGCGATTGCCCGCTAAGATA
>Gene2.1
TTCTAACGTGTAACGCAGTTGAGT
```

Reference Protein Database

(default: *Arabidopsis thaliana*)

```
>AT1G06190.5
MFYLSICLLYVVYLLGNTGYTLSDS
>AT1G06620.1
MESSLPQVAALDRSTLLKAFDETKT
```

GOgetter.sh

BLASTx

taxon1.cds.blast

Rank and retain best BLAST hits based on E-value, alignment length, percent identity, or bit score

parse_best_hits.py

make_tables.py

Associate best hits with corresponding GOSlim terms

taxon1.cds.blast.besthits.tsv

Reference GOSlim Database
(default: *Arabidopsis thaliana*)

taxon1.cds.blast.besthits.tsv.freqcounts-locus.tsv
taxon1.cds.blast.besthits.tsv.freqcounts-gene.tsv
taxon1.cds.blast.besthits.tsv.rawcounts-locus.tsv
taxon1.cds.blast.besthits.tsv.rawcounts-gene.tsv

taxon1
taxon2
taxon3
....

merge_and_viz.py

Aggregated Table

GOSlim_aggregated				
GOSlimTerm	examples/DRDE_subset.fasta	examples/LYJA_subset.fasta	examples/NDUV_subset.fasta	
0 Golgi apparatus	0.0331467929401635	0.028759244042728	0.0271059216013344	
1 anatomical structure development	0.0008609556607834	0.0012325390304026	0.0033361134278565	
2 biosynthetic process	0.0004304778303917	0.000821692686935	0.000417014178482	
3 catalytic activity	0.0008609556607834	0.0	0.0	
4 cell communication	0.0004304778303917	0.0	0.0008340283569641	
5 cell wall	0.0120553792505685	0.0049301561216105	0.0079232693911593	
6 cellular protein modification process	0.002582869823504	0.0024650780608052	0.0037531276063386	
7 chloroplast	0.2651743435213090	0.2522596548890720	0.2506255212677230	
8 cytoplasm	0.069737408523461	0.0665571076417419	0.0646371976647206	
9 cytosol	0.00043047783039173	0.005341002465078	0.0133444537114261	
10 embryo development	0.0004304778303917	0.000821692686935	0.000417014178482	
11 endoplasmic reticulum	0.0004304778303917	0.0004108463434675	0.0	
12 endosome	0.0004304778303917	0.0	0.000417014178482	
13 extracellular region	0.04218682737839	0.0419063270336994	0.0271059216013344	
14 kinase activity	0.0004304778303917	0.0	0.0	
15 mitochondrion	0.078346965132957	0.0861462612982744	0.0604670558798999	
16 multicellular organism development	0.0012914334911752	0.0012325390304026	0.0029190992493744	
17 nucleus	0.005962117950925	0.005341002465078	0.006255212677231	
18 nucleoplasm	0.0004304778303917	0.0	0.0	
19 nucleotide binding	0.0030133448127421	0.0020542317173377	0.0029190992493744	
20 nucleus	0.163151097184680	0.1684470008216930	0.1655546288573810	
21 other binding	0.0043047783039173	0.0028759244042728	0.030041701417848	
22 other cellular components	0.0228153250107619	0.0238290878211175	0.0208507089241034	
23 other cellular processes	0.02038743094735256	0.0045192097781429	0.009381984987489	
24 other intracellular components	0.0473525613430908	0.0616269615201314	0.0602000417014178	
25 other membranes	0.0133448127421437	0.0164338373787017	0.0154295246029365	
26 other metabolic processes	0.0020133448127421	0.00296976170912078	0.00404871559633027	
27 peroxisome	0.002582869823504	0.0057518486085456	0.0025020850708924	
28 plasma membrane	0.0538097287989668	0.0583401807723911	0.0450275312760633	
29 plastid	0.081963096440378	0.1023007396234180	0.0946822195154295	
30 post-embryonic development	0.0008609556607834	0.001643853738701	0.0016680567129292	
31 protein binding	0.0004304778303917	0.0	0.0012510425354462	
32 reproduction	0.0008609556607834	0.0012325390304026	0.0025020850708924	
33 response to abiotic stimulus	0.0008609556607834	0.0	0.0012510425354462	
34 response to chemical	0.002582869823504	0.0012325390304026	0.0012510425354462	
35 response to endogenous stimulus	0.0004304778303917	0.0004108463434675	0.000417014178482	
36 response to external stimulus	0.0004304778303917	0.0	0.0012510425354462	
37 response to light stimulus	0.0004304778303917	0.0	0.000417014178482	
38 response to stress	0.0004304778303917	0.0	0.0008340283569641	

Data Visualizations

