Exercise 2: Using BLAST2GO to generate GO annotation for 25 nucleotide sequences.

For this exercise, we will use an old version of BLAST2GO (v2.8). You will run BLAST2GO on a BioHPC Linux computer. (Here is the link to download BLAST2GO v2.8. If you want to run BLAST2GO on your laptop, here is the link to download v2.8 https://www.blast2go.com/start-blast2go-2-8).

1. Download the fasta file for this exercise (https://cornell.box.com/s/ffrxai61n6u64oewc82c). Then use FileZilla to upload the file to the directory /workdir /myUserName of your assigned computer (replace myUserName with your BioHPC ID).

2. Run BLASTX against SWISSPROT database. BLASTX is a command line software. To run the following commands, connect to your assigned computer using putty (Windows) or terminal (Mac)

```
cd /workdir/myUserName

cp /shared_data/genome_db/BLAST_NCBI/nr.* ./

cp /shared_data/genome_db/BLAST_NCBI/swissprot* ./

blastx -num_threads 8 -query annot_exercise.fasta  -db swissprot -out blastresults.xml
-max_target_seqs 20 -evalue 1e-5 -outfmt 5  -culling_limit 10 >& logfile &
```

Note:
"nr" is theThe NCBI non-redundant protein database. NCBI also provides multiple masking files if you want to BLAST against a subset of nr database. Swissprot is a subset of nr. At this step, first you copy the NCBI non-redundant BLAST database nr to the working directory. Then copy the swissprot masking files to the working directory. Run blastx (nucleotide queries against the protein database). The output file blastresults.xml is in XML format (-outfmt 5).

3. Run BLAST2GO on BioHPC computers with local BLAST2GO database server at Cornell. (Our local BLAST2GO database server is only accessible from BioHPC lab computers)
   a. As BLAST2GO has a graphic user interface, you will need VNC to access the graphic interface. To initiate VNC, go to http://cbsu.tc.cornell.edu/, login with your BioHPC user name and password, click User:xxxxx->"My Reservations". You will see a table with the information of your reserved computer. Click the "Connect VNC" link, you will see a message include this part: "Typically machine name and port number are used together: cbsuxxxxxxx.tc.cornell.edu:5901". Keep this string, you will need it in the next step. (As we have multiple people on the same computer for this workshop, there could be a problem when more than 2 people initiating VNC simultaneously. If you has a problem, wait for a moment, and try it again.)
   b. Windows users can use VNC Viewer (http://www.realvnc.com/download/viewer/), Mac users can use Chicken-of-VNC (http://sourceforge.net/projects/cotvnc/). Launch the Viewer.

In VNC server field, enter the **serve name you constructed in the last step**. (In VNC window, sometimes you would see a popup Window prompt for authorization, just dismiss it by clicking the "X" at the upper right corner)

c. Start BLAST2GO by clicking the "BLAST2GO" icon on the desktop of the VNC window. It is normal that you do not see anything on the screen for about a minute, it takes some time for the software to start.

d. Change the database setting of BLAST2GO to point to Cornell server. From the menu, click "File"->"DataAccess setting", check the "Own database" checkbox.

Fill out the following information (copy-paste does not work in VNC, you will have to type)

DB Name: b2gdb

DB Host: cbsuss06.tc.cornell.edu

DB User: blast2go

DB Password: blast4it

e. Load sequences in FASTA file (the FASTA file that you have transferred to the BioHPC computer home directory)

From menu, click "File" -> "Load sequences"

f. Load BLAST results (the BLAST xml file that you have transferred to the BioHPC computer home directory)

"File" ->"Import" -> "Import blast results" -> "One xml file", click the triangle start button to start importing.

g. Run mapping. "Mapping"->"Run Mapping". This step might take a long time. You can close you laptop computer, and come back later by clicking "My Reservations" -> "Connect VNC" link on the cbsu.tc.cornell.edu web site.

h. Run annotation. "Annotation"-> "Run Annotation". This step might take a long time. You can close your computer and come back later.

i. (Optional) Run interproscan. "Annotation " -> "Interproscan" -> "Run interproscan", followed by merging annotation: "Annotation " -> "Interproscan" -> "Merge interproscan GO to annotation".

j. After the annotation is finished, you can export the annotations to a file, "File"->"Export annotations" ->"Export annotations(.annot)". Then, you can transfer the ".annot" file to your local computer, and continue the work by using the "BLAST2GO" installed on your own computer ("File"->"Load annotations").