GAS User Guide

Ahmed Karam

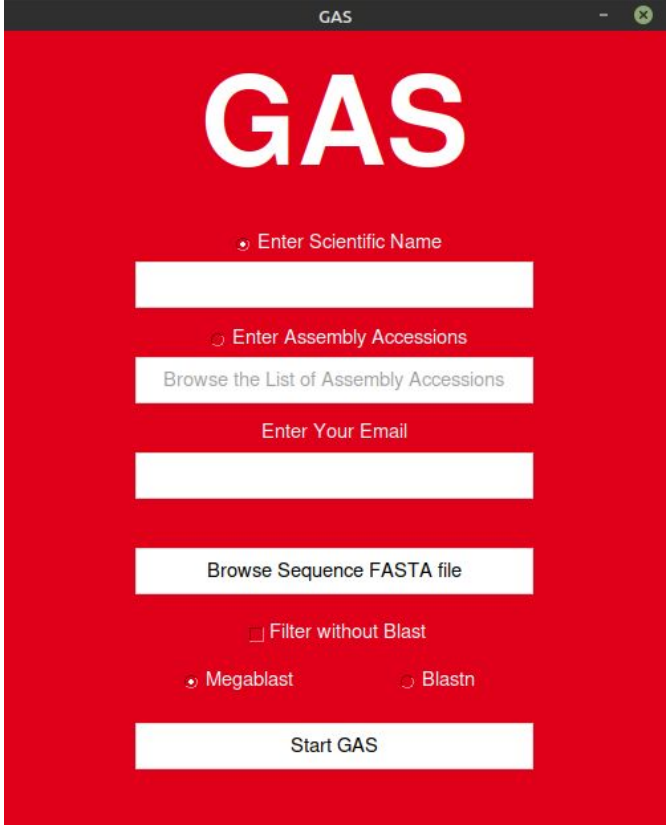
ahmedkaram.edu@gmail.com

Tips

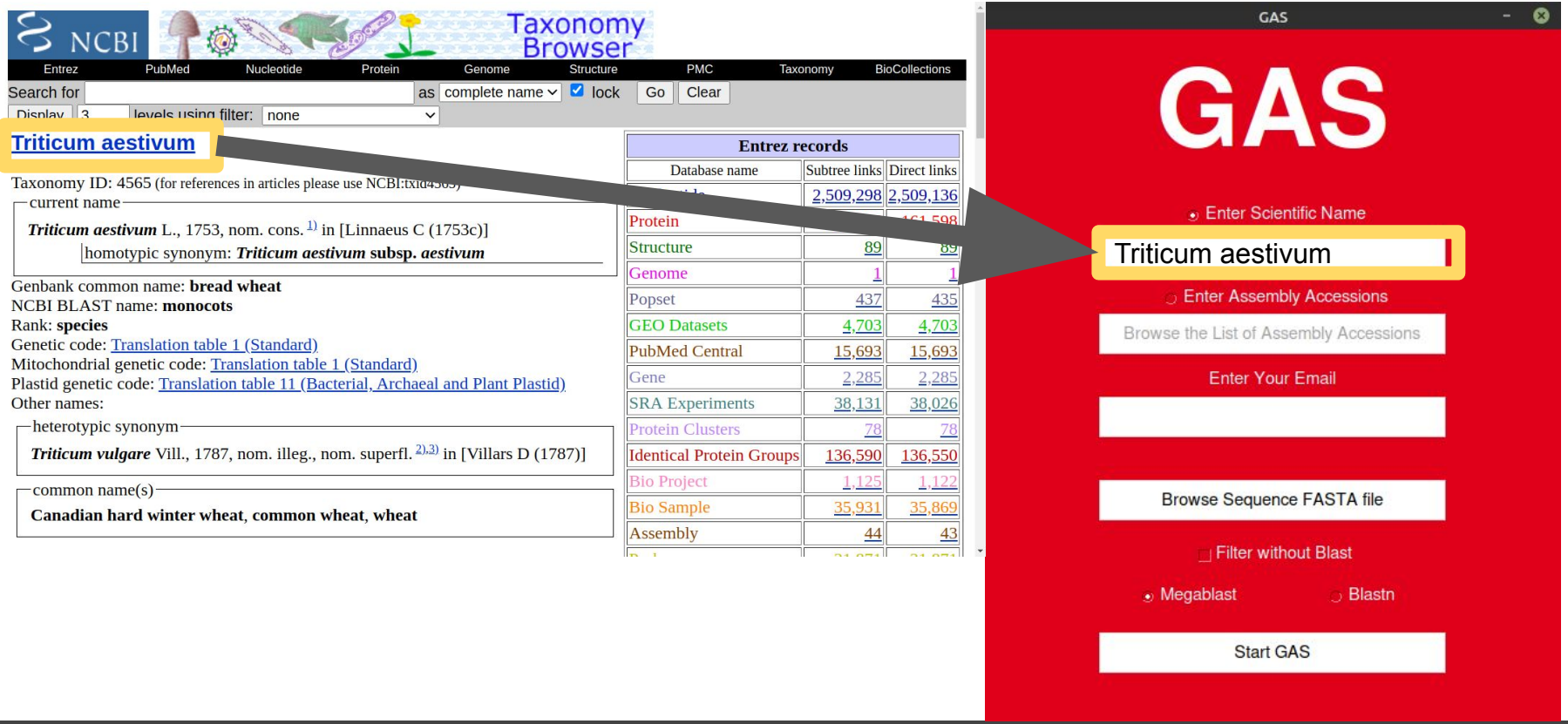
- If the first attempt was unsuccessful (i.e. the first step in the blast took a long time), try again after some time.
- GAS deals with NCBI, so in the event of any interruption or delay in the process, this is caused by NCBI and not by GAS.
- Use GAS with blast instead of NCBI blast online for large genomes and for assemblies defined by accessions. So, if the size of the genome you are working on is 12 Mb and you have 100 assemblies, the NCBI blast online can do the process without errors, and then use GAS to purify the results.
- Make sure that the "size" folder is in the same folder as GAS.py because it depends on it. That is, when you download the package, do not try to change their location.

Instructions

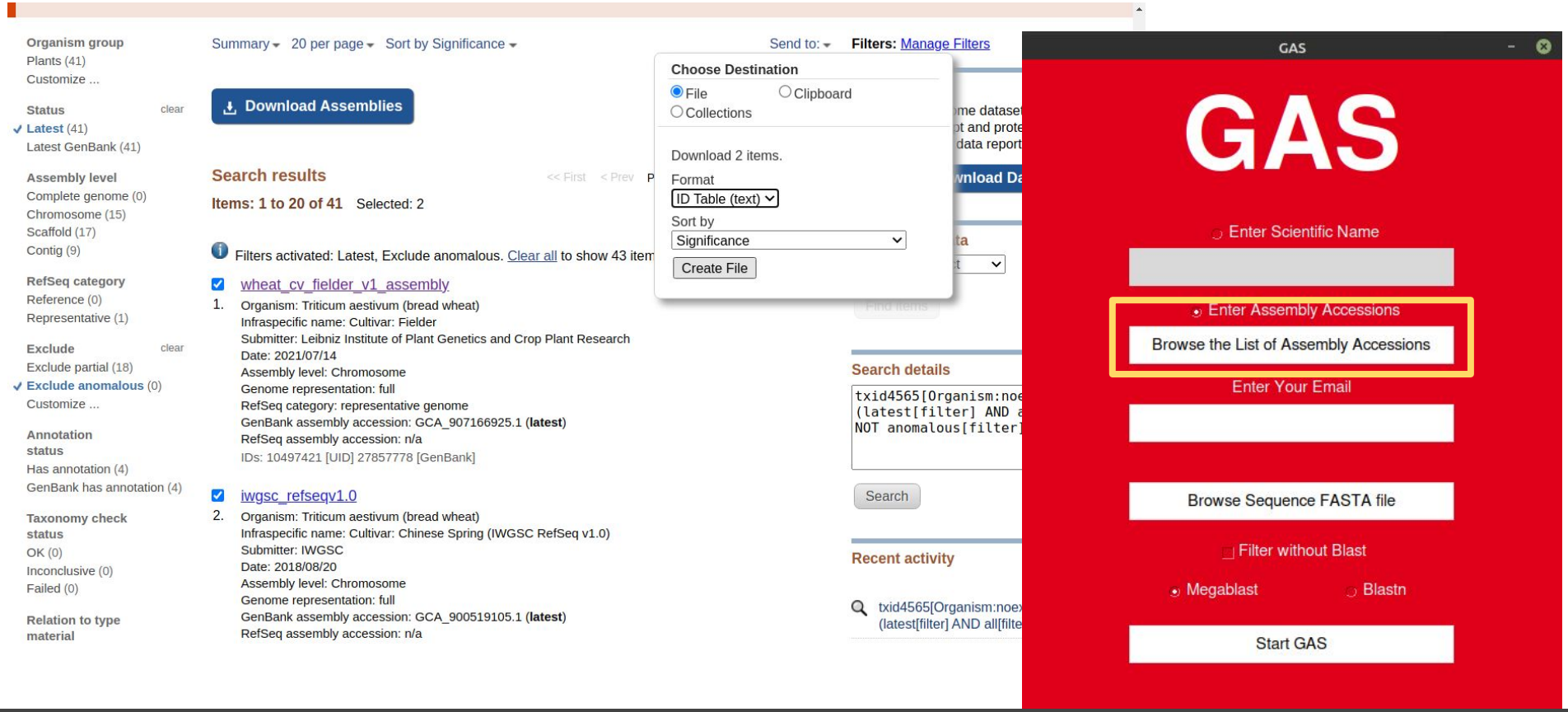
- The GUI cannot be used while interacting with NCBI and can be used again after you find the word “finish” has appeared in the terminal or CMD.
- GAS produces several files, you will find a file with the word "assemblies", which is the results of the process of retrieving the sequences, and we have distinguished the header so that it contains the query ID and the subject ID, and in this way the header is unique and will not be repeated. Thus all sequences with the same query IDs in the header can be extracted easily.
- You will find XML, table, ftp and remaining accessions files. If filtering is used without a blast action, all XML files must be located in the same folder with the query FASTA file.


GAS
GAS
Enter Scientific Name
Enter Assembly Accessions
Browse the List of Assembly Accessions
Enter Your Email
Browse Sequence FASTA file
Filter without Blast
Megablast
Blastn
Start GAS

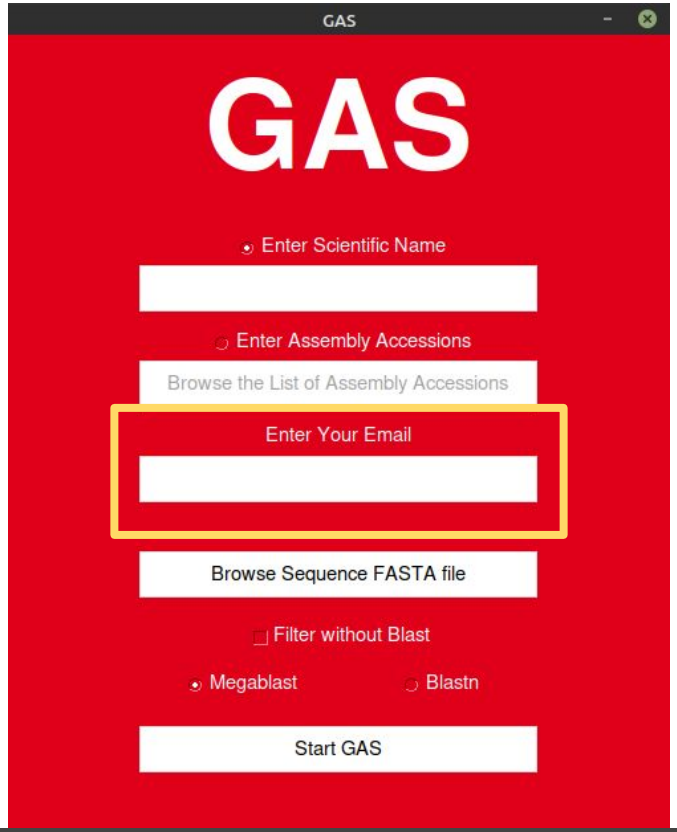
GAS GUI



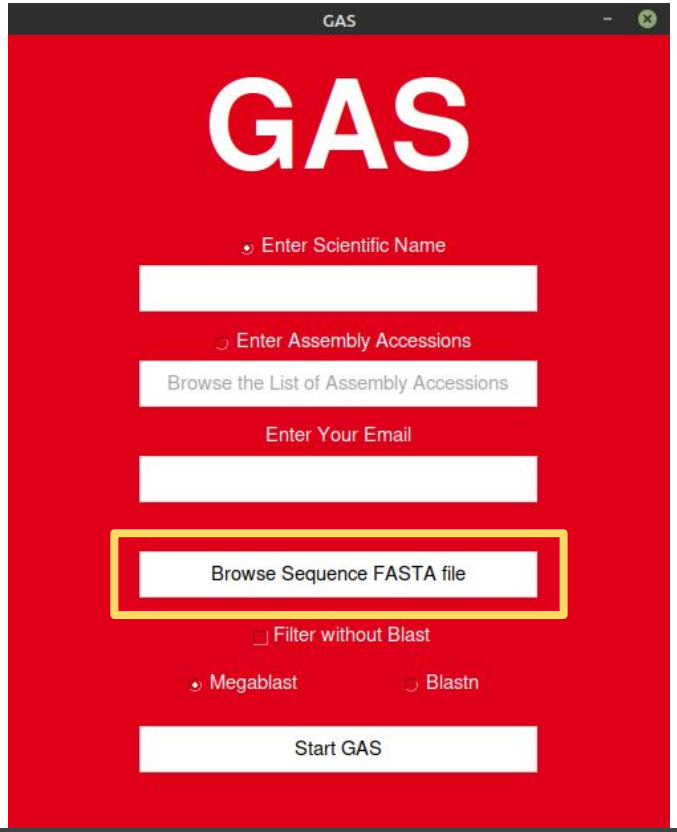
The first option to enter the data is to enter the scientific name and it must be correct. Copy correct scientific name and paste (or write it correctly).



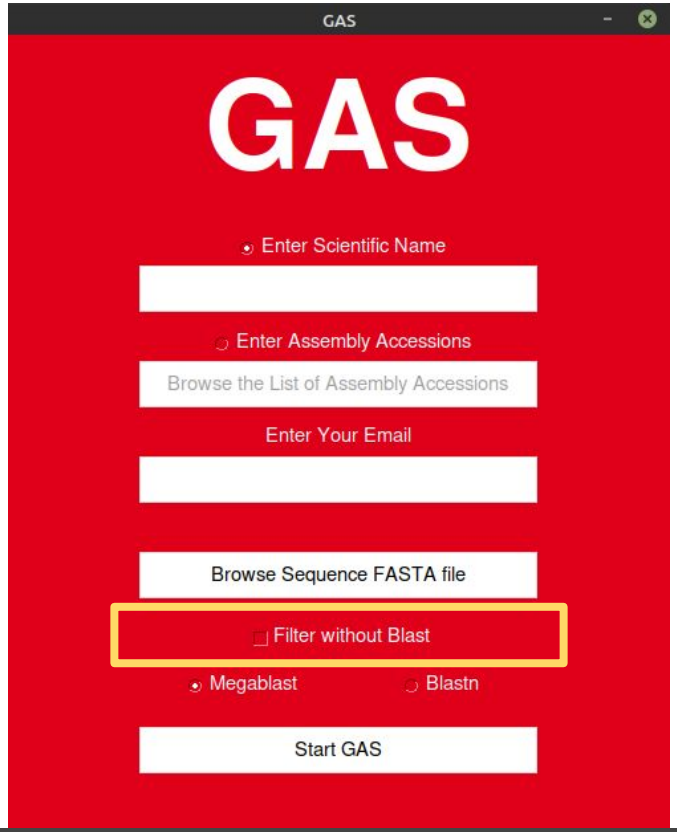
The second option for entering data is assembly accession or accessions. Search by scientific name in the assembly database, choose the assemblies, download the ID table, and browse the table through GAS. (you can merge a set of ID tables)



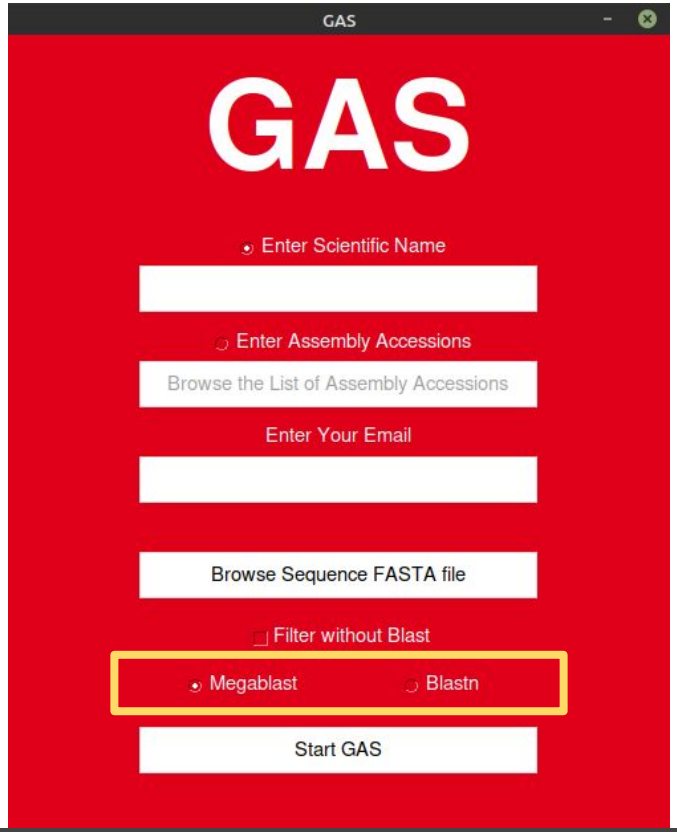
Enter your email and this is what NCBI requires to contact you if anything happens.



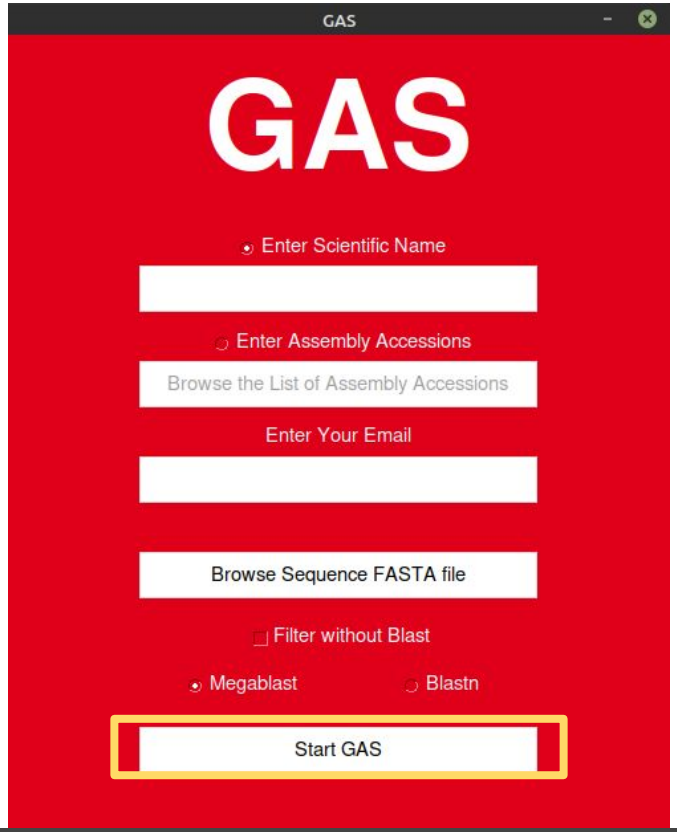
Browse a FASTA file that contains sequences



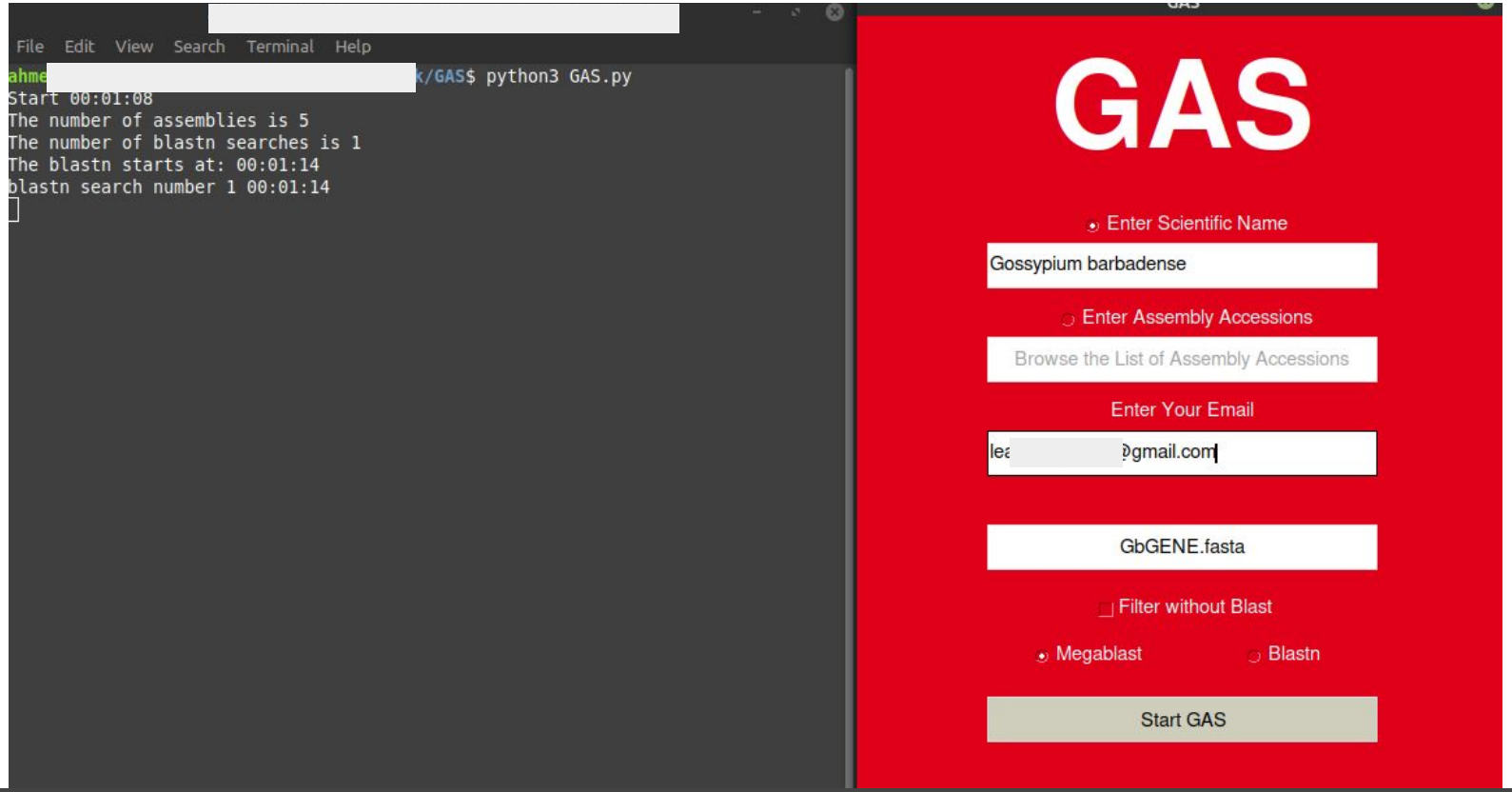
It filters all XML files that are similar to the name of the entered FASTA file that are in the same folder without blastn action.



Choose megablast or blastn


GAS
GAS
Enter Scientific Name
Enter Assembly Accessions
Browse the List of Assembly Accessions
Enter Your Email
Browse Sequence FASTA file
Filter without Blast
Megablast
Blastn
Start GAS

Run GAS



Follow the steps of the GAS through the CMD or the terminal