

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

- Enter a FASTA file that contains **only one record**. It is not recommended to enter a file containing several records.
- 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, 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

No. of retrieved sequences

Browse Sequence FASTA file

☐ Filter without Blast

☒ Megablast

☐ Blastn

Start GAS

NCBI Taxonomy Browser

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy BioCollections

Search for as complete name ☒ lock Go Clear

Display 3 levels using filter: none

Triticum aestivum

Taxonomy ID: 4565 (for references in articles please use NCBI:txid4565)

current name

Triticum aestivum L., 1753, nom. cons. ¹⁾ in [Linnaeus C (1753c)]

homotypic synonym: ***Triticum aestivum*** subsp. ***aestivum***

Genbank common name: **bread wheat**

NCBI BLAST name: **monocots**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 1 \(Standard\)](#)

Plastid genetic code: [Translation table 11 \(Bacterial, Archaeal and Plant Plastid\)](#)

Other names:

heterotypic synonym

Triticum vulgare Vill., 1787, nom. illeg., nom. superfl. ^{2),3)} in [Villars D (1787)]

common name(s)

Canadian hard winter wheat, common wheat, wheat

Entrez records		
Database name	Subtree links	Direct links
Protein		
Structure	89	89
Genome	1	1
Popset	437	435
GEO Datasets	4,703	4,703
PubMed Central	15,693	15,693
Gene	2,285	2,285
SRA Experiments	38,131	38,026
Protein Clusters	78	78
Identical Protein Groups	136,590	136,550
Bio Project	1,125	1,122
Bio Sample	35,931	35,869
Assembly	44	43

GAS

GAS

Enter Scientific Name

Triticum aestivum

Enter Assembly Accessions

Browse the List of Assembly Accessions

Enter Your Email

No. of retrieved sequences

Browse Sequence FASTA file

☐ Filter without Blast

☒ Megablast ☐ Blastn

Start GAS

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).

Organism group Summary ▾ 20 per page ▾ Sort by Significance ▾

Plants (41)
Customize ...

Status clear

✓ Latest (41)

Latest GenBank (41)

Assembly level

Complete genome (0)

Chromosome (15)

Scaffold (17)

Contig (9)

RefSeq category

Reference (0)

Representative (1)

Exclude clear

Exclude partial (18)

✓ Exclude anomalous (0)

Customize ...

Annotation

status

Has annotation (4)

GenBank has annotation (4)

Taxonomy check

status

OK (0)

Inconclusive (0)

Failed (0)

Relation to type

material

Download Assemblies

Search results

Items: 1 to 20 of 41 Selected: 2

Filters activated: Latest, Exclude anomalous. [Clear all](#) to show 43 items

✓ [wheat_cv_fielder_v1_assembly](#)

1. Organism: Triticum aestivum (bread wheat)
Intraspecific name: Cultivar: Fielder
Submitter: Leibniz Institute of Plant Genetics and Crop Plant Research
Date: 2021/07/14
Assembly level: Chromosome
Genome representation: full
RefSeq category: representative genome
GenBank assembly accession: GCA_907166925.1 (latest)
RefSeq assembly accession: n/a
IDs: 10497421 [UID] 27857778 [GenBank]

✓ [iwgsc_refseqv1.0](#)

2. Organism: Triticum aestivum (bread wheat)
Intraspecific name: Cultivar: Chinese Spring (IWGSC RefSeq v1.0)
Submitter: IWGSC
Date: 2018/08/20
Assembly level: Chromosome
Genome representation: full
GenBank assembly accession: GCA_900519105.1 (latest)
RefSeq assembly accession: n/a

Send to: ▾ Filters: [Manage Filters](#)

Choose Destination

- ☒ File ☐ Clipboard
☐ Collections

Download 2 items.

Format

ID Table (text) ▾

Sort by

Significance ▾

Create File

Search details

txid4565[Organism
(latest[filter] AND
NOT anomalous[filter]

Search

Recent activity

txid4565[Organism
(latest[filter] AND a

GAS

Enter Scientific Name

Enter Assembly Accessions

Browse the List of Assembly Accessions

Enter Your Email

No. of retrieved sequences

Browse Sequence FASTA file

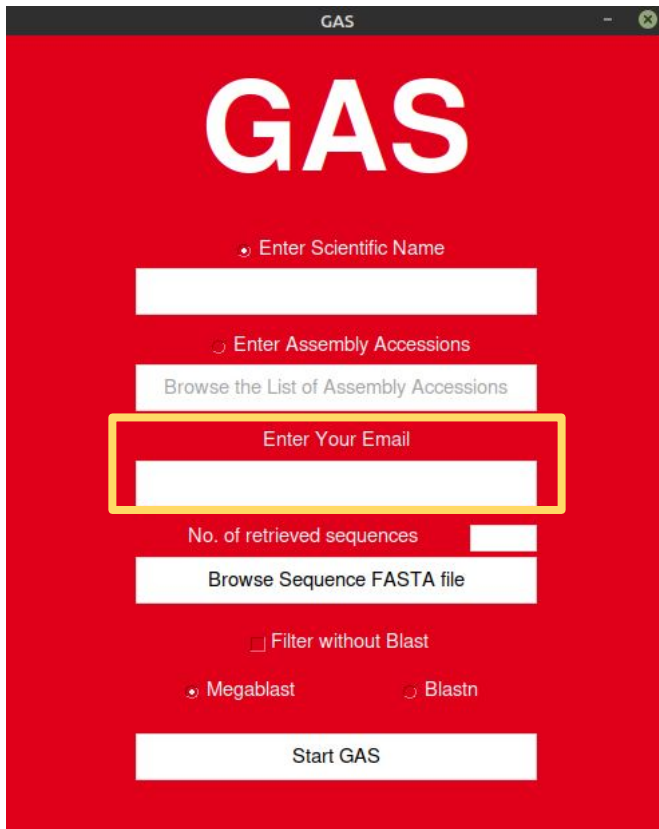
Filter without Blast

Megablast

Blastn

Start GAS

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)



The image shows a web browser window titled "GAS" with a red background. The main heading "GAS" is in large white letters. Below it, there are several input fields and buttons. The "Enter Your Email" field is highlighted with a yellow border. The "Start GAS" button is at the bottom.

GAS

☐ Enter Scientific Name

☐ Enter Assembly Accessions

No. of retrieved sequences

☐ Filter without Blast

☒ Megablast ☐ Blastn

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

GAS

GAS

☒ Enter Scientific Name

☐ Enter Assembly Accessions

[Browse the List of Assembly Accessions](#)

Enter Your Email

No. of retrieved sequences

[Browse Sequence FASTA file](#)

☐ Filter without Blast

☒ Megablast ☐ Blastn

[Start GAS](#)

This option is optional, not mandatory. If not, leave the entry blank. Normally, GAS outputs for each query sequence five results against an assembly. It means that in the case of a query sequence, one against five assemblies produces at least twenty-five results, if any. You can enter the final number of the retrieval sequences. For example, if you want the total number of blast results to be 50, type 50.

GAS

GAS

☒ Enter Scientific Name

☒ Enter Assembly Accessions

Browse the List of Assembly Accessions

Enter Your Email

No. of retrieved sequences

Browse Sequence FASTA file

☐ Filter without Blast

☒ Megablast ☐ Blastn

Start GAS

Browse a FASTA file that contains sequences

GAS

GAS

☒ Enter Scientific Name

☐ Enter Assembly Accessions

☐ Megablast

☐ Blastn

Browse the List of Assembly Accessions

Enter Your Email

Browse Sequence FASTA file

☒ Filter without Blast

Start GAS

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.

GAS

GAS

☒ Enter Scientific Name

☒ Enter Assembly Accessions

Enter Your Email

No. of retrieved sequences

☐ Filter without Blast

☒ Megablast ☐ Blastn

Choose megablast or blastn

GAS

GAS

☒ Enter Scientific Name

☒ Enter Assembly Accessions

Browse the List of Assembly Accessions

Enter Your Email

No. of retrieved sequences

Browse Sequence FASTA file

☐ Filter without Blast

☒ Megablast

☐ Blastn

Start GAS

Run GAS

```
File Edit View Search Terminal Help
ahmed@ahmed:~/GAS$ python3 GAS.py
Start 00:01:08
The number of assemblies is 5
The number of blastn searches is 1
The blastn starts at: 00:01:14
blastn search number 1 00:01:14
█
```

GAS

GAS

☒ Enter Scientific Name

☐ Enter Assembly Accessions

Browse the List of Assembly Accessions

Enter Your Email

No. of retrieved sequences

Browse Sequence FASTA file

☐ Filter without Blast

☒ Megablast ☐ Blastn

Start GAS

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