

## TABLE OF CONTENTS

Understanding the FUSARIUM-ID headers.....	2
Access to the FUSARIUM-ID database.....	5
Resources available to BLAST against the FUSARIUM-ID downloadable file .....	6
BLAST parameters and understanding BLAST results.....	7
<b>BLAST using BLAST+ in Galaxy .....</b>	<b>9</b>
Creating a Galaxy account .....	9
Uploading the FASTA file .....	11
Creating the database in Galaxy .....	14
Running a BLAST search in Galaxy .....	17
<b>BLAST using SequenceServer .....</b>	<b>23</b>
Installing SequenceServer.....	24
Adding the database in SequenceServer .....	25
Running a BLAST search in SequenceServer .....	27
<b>BLAST using BLAST+ executable in command line .....</b>	<b>30</b>
Setting up BLAST+ .....	30
Adding the database using BLAST+ executable in the command line .....	32
Running a BLAST search using BLAST+ executable in the command line .....	33
<b>BLAST using Geneious Prime .....</b>	<b>35</b>
Set up Custom BLAST service.....	35
Adding the database in Geneious .....	37
Running a BLAST search in Geneious.....	40

## Understanding the FUSARIUM-ID headers

The headers include species identifiers and metadata attached to each sequence. Underscores are used instead of spaces and each metadata field is separated by a vertical line.

These fields include (see example header below):

1. Locus name
2. Isolate specific identifier
3. *Fusarium* species complex to which the sequence is assigned
4. Species name
5. Host/Substrate
6. Geographic locality
7. Additional metadata

### Example header

```
>TEF1|NRRL_20433|FOSC|'oxysporum'|Vicia_faba|Germany|fsp=fabae;gb=AF008479.1;ST=2  
---1---|---2---|---3---|---4---|---5---|---6---|---7---
```

#### **1. Locus name**

This field refers to the gene region amplified for the specific sequence (e.g., *TEF1*, *RPB1*, *RPB2*).

#### **2. Isolate specific identifier**

This includes an NRRL (ARS Culture Collection; Peoria, IL) accession number of the sequenced isolate for *Fusarium*. In cases where an NRRL accession is not available, another culture collection's identifier is used (e.g., FRC\_S2394). For taxa outside *Fusarium*, a specific identifier for the isolate is included as found in GenBank.

#### **3. *Fusarium* species complex to which the sequence is assigned**

The species complexes have been abbreviated in the header as shown below. In some clades, additional information may be included (e.g., FFSC\_American = *F. fujikuroi* Species Complex, American Clade).

Abbreviation	Assigned <i>Fusarium</i> Species Complex
FASC	<i>Fusarium albidum</i> Species Complex
FBABSC	<i>Fusarium babinda</i> Species Complex
FBSC	<i>Fusarium buharicum</i> Species Complex
FBURSC	<i>Fusarium burgessii</i> Species Complex
FBUXSC	<i>Fusarium buxicola</i> Species Complex
FCOSC	<i>Fusarium concolor</i> Species Complex

FCSC	<i>Fusarium chlamydosporum</i> Species Complex
FDECSC	<i>Fusarium decemcellulare</i> Species Complex
FDSC	<i>Fusarium dimerum</i> Species Complex
FFSC*	<i>Fusarium fujikuroi</i> Species Complex
FHSC	<i>Fusarium heterosporum</i> Species Complex
FIESC	<i>Fusarium incarnatum-equiseti</i> Species Complex
FLSC	<i>Fusarium lateritium</i> Species Complex
FNEWSC	<i>Fusarium newnesense</i> Species Complex
FNSC	<i>Fusarium nisikadoi</i> Species Complex
FOSC	<i>Fusarium oxysporum</i> Species Complex
FRSC	<i>Fusarium redolens</i> Species Complex
FSAMSC*	<i>Fusarium sambucinum</i> Species Complex
FSSC	<i>Fusarium solani</i> Species Complex
FSTASC	<i>Fusarium staphyleae</i> Species Complex
FTORSC	<i>Fusarium torreyae</i> Species Complex
FTSC	<i>Fusarium tricinctum</i> Species Complex
FVSC	<i>Fusarium ventricosum</i> Species Complex
Insertae sedis	<i>Fusarium</i> species unresolved to species complex
Outgroup	Not part of <i>Fusarium</i>

\*Subheadings: Clade designations included after the species complex name.

#### 4. Species name

This field specifies the current name for the species to which the sequence belongs (e.g., *sambucinum*). For those that are currently undescribed this space specifies the Species Complex to which the sequence belongs immediately followed by the designation “undesc” (e.g., FFSCundesc). When available, informal ad hoc nomenclature (species complex designation followed by an Arabic number) has been included for undescribed species, so that users can report identifications using this designation (e.g., FSAMSC1) rather than as *Fusarium* sp. Genus and species are included in this field for species not in the genus *Fusarium* (outgroups).

#### 5. Host/Substrate

This field specifies the substrate from which the isolate associated with the sequence was originally recovered (e.g., soil). This may include a specific host along with the location in the plant/animal where the isolate was recovered, or the environment, depending on the information available.

## 6. Geographic locality

This field includes the country for the isolate to which the sequence belongs (e.g., Costa Rica). Isolates from the United States are listed as “USA” and may be followed by the state postal abbreviation when this information is available for the isolate (e.g., USA\_PA for the state of Pennsylvania in the United States).

## 7. Additional metadata

This additional information is only available for certain sequences, depending on the availability of the data. For this field, each additional metadata identifier is separated by a semicolon. Metadata fields included may change over time as new information is added and becomes available. Currently these consist of the following:

- fsp**: forma specialis
- gb**: GenBank accession number for deposited sequence
- inf**: location of the infection (for medical/veterinary isolates only)
- race**: race of the isolate
- ST**: sequence type
- tox**: = toxin produced *in vitro*. Currently available data include trichothecenes [Laraba et al. 2021]: (3ADON = 3-acetyldeoxynivalenol, 15ADON = 15-acetyldeoxynivalenol, BUT = butenolide, CAL = calonectrin, CUL = culmorin, 15decal = 15-decalonectrin, DAS = diacetoxyscirpenol, 15OHCUL = 15-hydroxy culmorin, 37diOH = 7-hydroxy isotrichodermol, 3OH = isotrichodermol, 15ketoNX2NX3 = 15-keto NX-2 and 15-keto NX-3 – novel trichothecenes, NEO = neosolaniol, NIV = nivalenol, T2 = T-2 toxin); and the estrogenic mycotoxin ZEA = zearalenone.

## Access to the FUSARIUM-ID database

FUSARIUM-ID is publicly available and there are two main ways to access FUSARIUM-ID v.3.0 (and subsequent releases). One of them is the web-accessible FUSARIUM-ID database which will continue to be available at <http://isolate.fusariumdb.org/> (see screenshot below).

The screenshot shows the Fusarium-ID web interface. At the top, there is a navigation bar with links: Home, Introduction, Database, Search & Analysis, Guide, and CIF Gateway. Below the navigation bar, the main content area has a breadcrumb trail: MAIN » Use the database » Search by sequence. A heading "Search :: Search by sequence" is followed by a form. The form includes a dropdown menu labeled "Marker" with the option "TEF1(v.3.0)" selected. A message "New release - v.3.0 now available!" is displayed above the dropdown. Below the dropdown is a large text input field labeled "Sequence (FASTA)". To the left of this field, under the "Marker" section, are three dropdown menus: "Limit Expect Value" set to "1e-30", "Number of matching sequences" set to "10", and "Matrix" set to "BLOSUM62". Below the sequence input field are two buttons: "Okay" and "Back".

However, the online FUSARIUM-ID database server is subjected to frequent security breaches and subsequent shutdowns, with no financial resources for its maintenance. This is why, in addition to maintaining the web-accessible FUSARIUM-ID portal, we are releasing a new downloadable file that will provide users with the opportunity to set up their own FUSARIUM-ID BLAST database on a local computer. The latest FUSARIUM-ID database downloadable version is publicly available at the FUSARIUM-ID GitHub (<https://github.com/fusariumid>). All new versions and FASTA files for additional marker loci will be released at the FUSARIUM-ID GitHub. Below we provide tutorials for users that prefer to use the downloadable file to set up their BLAST searches.

## Resources available to BLAST against the FUSARIUM-ID downloadable file

This document includes general instructions to walk you through different options available to BLAST your own sequence against the FUSARIUM-ID database.

All tutorials included in this document are explained assuming you have already downloaded the corresponding **FUSARIUM-ID FASTA file** (e.g., FusariumID\_v.3.0\_TEF1), publicly available at the FUSARIUM-ID GitHub (<https://github.com/fusariumid>), and that you have a sequence at hand that you want to identify.

The tutorials are explained using the FusariumID\_v.3.0\_TEF1 file as an example, but you can apply the same steps for any of the other FUSARIUM-ID downloadable FASTA-formatted files (e.g., FusariumID\_v.3.0\_RPB2).

### Comparison of different software to BLAST against the Fusarium-ID file

Features	BLAST+ in Galaxy	Sequence Server	BLAST+ executable in command line	Geneious Prime
Free	+	+	+	
Requires account sign up	+			+
Requires use of the command line		+	+	
GUI*	+	+		+
Requires installation		+	+	+
Operating system requirements or web browser interface	Web browser interface	Mac OSX, Linux	Windows, Mac OSX, Linux	Windows, Mac OSX, Linux
Average search time for 1 sequence	<1min**	<1 min	<1 min	<1 min

\*GUI= Graphical User Interface

\*\*Internet based and may vary

## BLAST parameters and understanding BLAST results

Specific BLAST default parameters vary for each of the computing environments mentioned above. Some of these parameters can be customized depending on the software used. Default parameters and options to change them vary among applications. We suggest using megablast instead of blastn, when available, as well as a 90% sequence identity cutoff (for example, in Galaxy this option can be found under “Advanced Options”). The BLAST example below shows the result of a BLAST query using Galaxy. Note that visualization of results varies between platforms but the provided information/parameters are the same. In the example the column contains information as described below:

- **Query accession** shows the user-defined identifier for the sequence you input into BLAST
- **Subject accession** shows the header for the match in FUSARIUM-ID with the information as described under section '[Understanding the FUSARIUM-ID headers](#)'.
- **Percent identity** describes how similar the query sequence is to each match (as a percentage of how many nucleotides in each sequence are identical). The higher the percent identity the better the match.
- **Alignment length** shows the number of nucleotides that align between the query sequence and the matching sequence.
- **Mismatches** shows the number of nucleotides that do not match in your query to the matching sequence.
- **Gap openings** represents the number of spaces that are introduced into the alignment to compensate for insertions and deletions in the query sequence relative to the FUSARIUM-ID matching sequence.
- **Query start** is the first query nucleotide included in the alignment.
- **Query end** is the last query nucleotide included in the alignment.
- **Start of alignment in subject** is the first nucleotide from the matching sequence in FUSARIUM-ID that is included in the alignment.
- **End of alignment in subject** is the last nucleotide from the matching sequence in FUSARIUM-ID that is included in the alignment.
- **E-value** is a parameter that determines the number of hits expected by chance when searching against FUSARIUM-ID. The closer the E-value is to zero the more significant the match. See the [NCBI BLAST FAQ](#) for more details.
- **Bit score** is another parameter that indicates the sequence similarity independent of the length of the query sequence and the FUSARIUM-ID database size that is normalized based on the raw alignment score. It can be used to compare alignment scores from different searches.

Looking at the top match is not sufficient to assign a species identification. You will need to evaluate the range of results you are obtaining to make a decision. For more detailed information on evaluating BLAST results refer to O’Donnell et al. 2022. *DNA Sequence-based identification of Fusarium: A work in progress*.

## Example of a BLAST result conducted in Galaxy showing the top 25 matches

Query accession	Subject accession	Percent Identity	Alignment Length	Mismatches	Gap openings	Query Start	Query End	Start of Alignment in Subject	End of Alignment in Subject	E-value	Bit score
Unknown	TEF1 NRRL_26961 FOSC 'oxysporum' Dianthus_caryophyllus Australia fsp=dianthi;gb=AF246840.1;race=9;ST=88	100	666	0	0	1	666	1	666	0	1230
Unknown	TEF1 NRRL_26367 FOSC 'oxysporum' Human USA_MD gb=AY527529.1;ST=50	100	666	0	0	1	666	1	666	0	1230
Unknown	TEF1 NRRL_25609 FOSC 'oxysporum' Musa_ABB'_Bluggoe' Malawi fsp=cubense;gb=AF008490.1;ST=37	100	666	0	0	1	666	1	666	0	1230
Unknown	TEF1 NRRL_25420 FOSC 'oxysporum' Gossypium_sp USA fsp=vasinfectum;race=1;ST=28	100	666	0	0	1	666	1	666	0	1230
Unknown	TEF1 NRRL_20433 FOSC 'oxysporum' Vicia_faba Germany fsp=fabae;gb=AF008479.1;ST=2	100	666	0	0	1	666	1	666	0	1230
Unknown	TEF1 NRRL_28919 FOSC 'oxysporum' Linum_usitatissimum Argentina fsp=lini;gb=AF246882.1;ST=110	99.85	666	1	0	1	666	1	666	0	1225
Unknown	TEF1 NRRL_26964 FOSC 'oxysporum' Dianthus_caryophyllus Italy fsp=dianthi;gb=AF324328.1;race=1;ST=90	99.85	666	1	0	1	666	1	666	0	1225
Unknown	TEF1 NRRL_26962 FOSC 'oxysporum' Dianthus_caryophyllus The_Netherlands fsp=dianthi;gb=AF246841.1;race=10;ST=89	99.85	666	1	0	1	666	1	666	0	1225
Unknown	TEF1 NRRL_26437 FOSC 'oxysporum' Cucumis_sativus USA_SC fsp=cucumerinum;gb=AF362168.1;ST=71	99.85	666	1	0	1	666	1	666	0	1225
Unknown	TEF1 NRRL_26875 FOSC 'oxysporum' Spinacia_oleracea USA_AR fsp=spinaciae;gb=AF246850.1;ST=86	99.55	666	3	0	1	666	1	666	0	1214
Unknown	TEF1 NRRL_26874 FOSC 'oxysporum' Spinacia_oleracea USA_AR fsp=spinaciae;gb=AF246849.1;ST=85	99.55	666	3	0	1	666	1	666	0	1214
Unknown	TEF1 NRRL_26748 FOSC 'oxysporum' Momordica_charantia Japan fsp=momordicae;gb=AF362167.1;ST=84	99.55	666	3	0	1	666	1	666	0	1214
Unknown	TEF1 NRRL_26679 FOSC 'oxysporum' Human Australia gb=AY527526.1;inf=mouth(gum_abscess);ST=82	99.55	666	3	0	1	666	1	666	0	1214
Unknown	TEF1 NRRL_25598 FOSC 'oxysporum' Glycine_max USA_SC fsp=glycines;gb=AF008496.1;ST=35	99.55	666	3	0	1	666	1	666	0	1214
Unknown	TEF1 NRRL_28395 FOSC 'oxysporum' Lilium_sp Italy fsp=lili;gb=AF246858.1;ST=107	99.399	666	4	0	1	666	1	666	0	1208
Unknown	TEF1 NRRL_26677 FOSC 'oxysporum' Human Australia gb=AY527528.1;inf=nail(subungual_debris);ST=81	99.399	666	4	0	1	666	1	666	0	1208
Unknown	TEF1 NRRL_25375 FOSC 'oxysporum' Human South_Pacific gb=AY527521.1;ST=26	99.399	666	4	0	1	666	1	666	0	1208
Unknown	TEF1 NRRL_53156 FOSC 'oxysporum' Brassica_oleracea USA_WI fsp=conglutinans;gb=FJ985442.1;ST=255	100	649	0	0	18	666	1	649	0	1199
Unknown	TEF1 NRRL_53154 FOSC 'oxysporum' unknown Korea fsp=raphanin;gb=FJ985441.1;ST=254	100	649	0	0	18	666	1	649	0	1199
Unknown	TEF1 NRRL_38320 FOSC 'oxysporum' Solanum_melongena Israel fsp=melongenae;gb=FJ985382.1;ST=194	100	649	0	0	18	666	1	649	0	1199
Unknown	TEF1 NRRL_38309 FOSC 'oxysporum' Arachis_hypogaea_shell Australia gb=FJ985378.1;ST=190	100	649	0	0	18	666	1	649	0	1199
Unknown	TEF1 NRRL_38283 FOSC 'oxysporum' Arachis_hypogaea South_Africa gb=FJ985366.1;ST=176	100	649	0	0	18	666	1	649	0	1199
Unknown	TEF1 NRRL_37616 FOSC 'oxysporum' Pisum_sativum unknown fsp=pisi;gb=FJ985359.1;ST=169	100	649	0	0	18	666	1	649	0	1199
Unknown	TEF1 NRRL_36276 FOSC 'oxysporum' Pisum_sativum USA_ID fsp=pisi;gb=FJ985341.1;ST=151	100	649	0	0	18	666	1	649	0	1199
Unknown	TEF1 NRRL_34079 FOSC 'oxysporum' Gossypium_sp USA_LA fsp=vasinfectum;gb=FJ985323.1;ST=133	100	649	0	0	18	666	1	649	0	1199
Unknown	TEF1 NRRL_32873 FOSC 'oxysporum' Gossypium_sp USA_AR fsp=vasinfectum;gb=FJ985310.1;ST=117	100	649	0	0	18	666	1	649	0	1199
Unknown	TEF1 NRRL_26147 FOSC 'oxysporum' Dianthus_sp USA fsp=dianthi;gb=FJ985281.1;race=4;ST=42	100	649	0	0	18	666	1	649	0	1199
Unknown	TEF1 NRRL_25437 FOSC 'oxysporum' Gossypium_sp Brazil fsp=vasinfectum;gb=FJ985280.1;race=6;ST=32	100	649	0	0	18	666	1	649	0	1199
Unknown	TEF1 NRRL_22553 FOSC 'oxysporum' Raphanus_sativus Germany fsp=raphanin;ST=19	100	649	0	0	18	666	1	649	0	1199

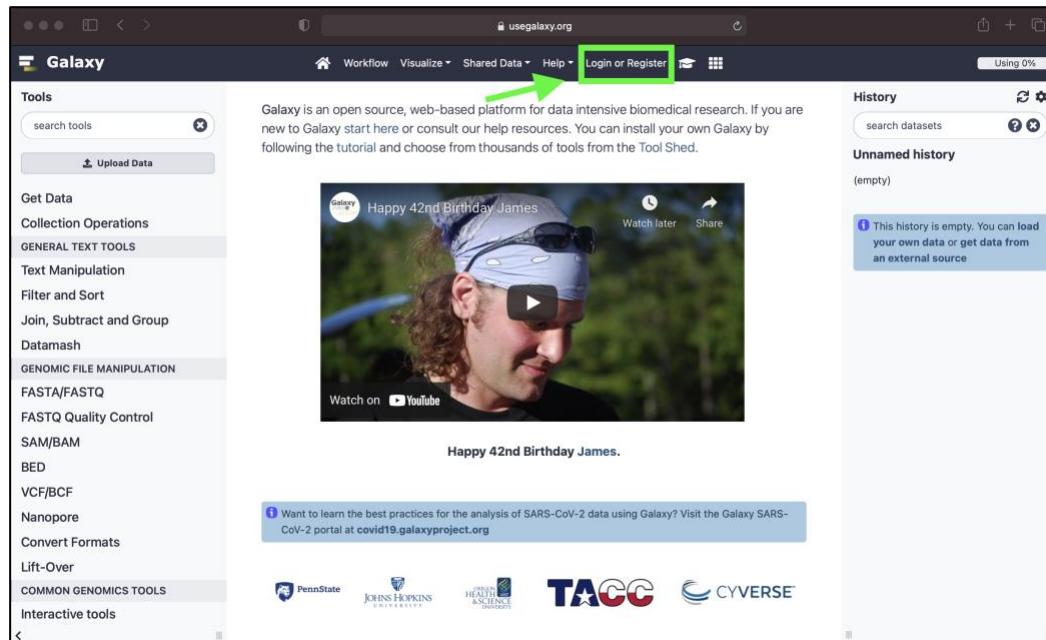
# BLAST using BLAST+ in Galaxy

Here, we provide short step-by-step instructions on how you can run a BLAST on Galaxy against the [Fusarium-ID file](#). This tutorial was generated for Galaxy version 21.05.rc1.

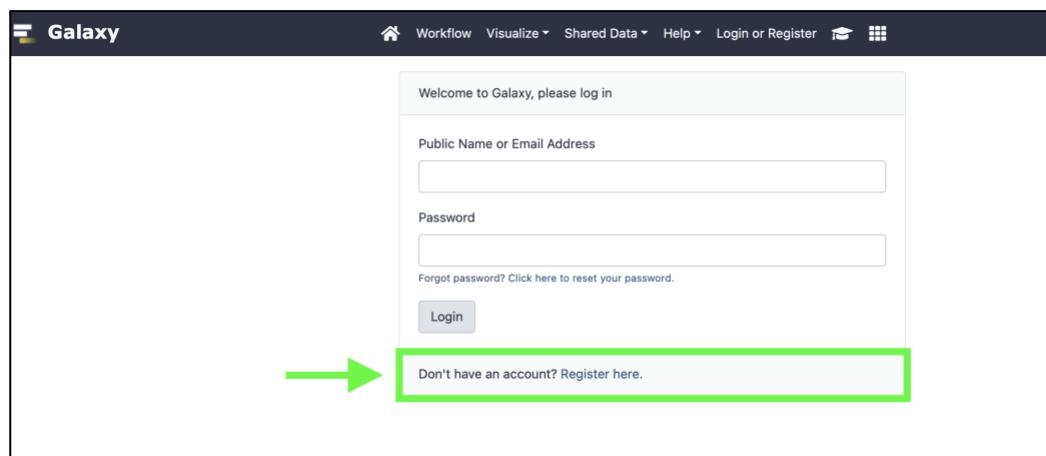
## Creating a Galaxy account

1. Go to <https://usegalaxy.org/> and register for an account

- a. Select **Login or Register** from the top menu



- b. On the next screen, click on **Register here**



- c. Fill out the required information to create an account and click **Create**

**Galaxy** Workflow Visualize ▾ Shared Data ▾ Help ▾ Login or Register

Please register only one account. The usegalaxy.org service is provided free of charge and has limited computational and data storage resources. Registration and usage of multiple accounts is tracked and such accounts are subject to termination and data deletion.

Create a Galaxy account

Email Address

Password

Confirm password

Public name

Your public name is an identifier that will be used to generate addresses for information you share publicly. Public names must be at least three characters in length and contain only lowercase letters, numbers, dots, underscores, and dashes ('.', '\_', '-').

Already have an account? Log in here.

**Galaxy Web Portal Service Agreement**

1) Use of Service. The Galaxy Web Portal is a free, public, Internet accessible resource (the "Service"). Data transfer and data storage are not encrypted. If there are restrictions on the way your research data can be stored and used, please consult your local institutional review board or the project principal investigator before uploading it to any public site, including this Service. If you have protected data, large data storage requirements, or short deadlines you are encouraged to set up your own local Galaxy instance and not use this Service. Your access to the service may be revoked at any time for reasons deemed necessary by the operators of the Service.

2) Accounts and Service Limitations. You may choose to register an account with the Service. Your registration data is primarily used so you may persistently store data on the Service and use advanced Galaxy features such as sharing and workflows. The operators of the Service will not provide your registration data to any third party unless required to do so by law. Your access to the Service is provided under the condition that you abide by any published quotas on data storage, job submissions, or any other limitations placed on the public Service. Attempts to subvert these limits by creating multiple accounts or through any other method may result in termination of all associated accounts.

3) Disclaimer. The Service is provided to you on an "AS IS" BASIS and WITHOUT WARRANTY, either express or implied, including, without limitation, the warranties of non-infringement, merchantability or fitness for a particular purpose. THE ENTIRE RISK AS TO THE QUALITY OF THE SERVICE IS WITH YOU. This attempt to subvert these limits by creating multiple accounts or through any other method may result in termination of all associated accounts.

- d. Sign into your account

**Galaxy** Workflow Visualize ▾ Shared Data ▾ Help ▾ Login or Register

Welcome to Galaxy, please log in

Public Name or Email Address

Password

Forgot password? Click here to reset your password.

Don't have an account? [Register here.](#)

## Uploading the FASTA file

Optional: If you have used Galaxy before, you may choose to “Create New History” using the plus sign next to “History” or clear any previous history in Galaxy before proceeding to the next steps.

The screenshot shows the Galaxy History panel. At the top, there's a search bar labeled "search datasets" and a row of icons including a question mark, a delete button, and a settings gear. Below this is a section titled "FusariumIDTrial" which contains the message "13 shown, 18 deleted". Underneath is a list of datasets:

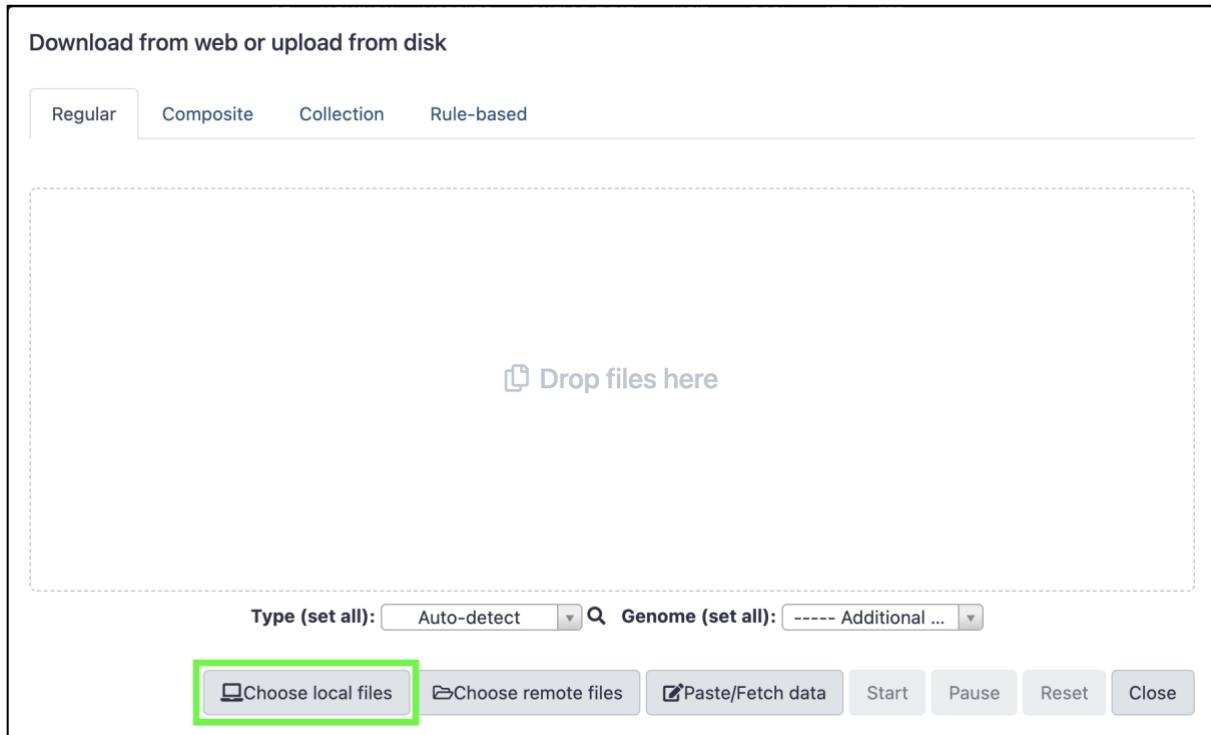
- 31: megablast OutgroupE xample.fas vs 'nucleotide BLAST database from data 28'
- 30: nucleotide BLAST dat abase from data 28
- 28: FUSARIUMID\_TEF\_for trial2.fas

2. On the left side panel, click on **Upload Data**

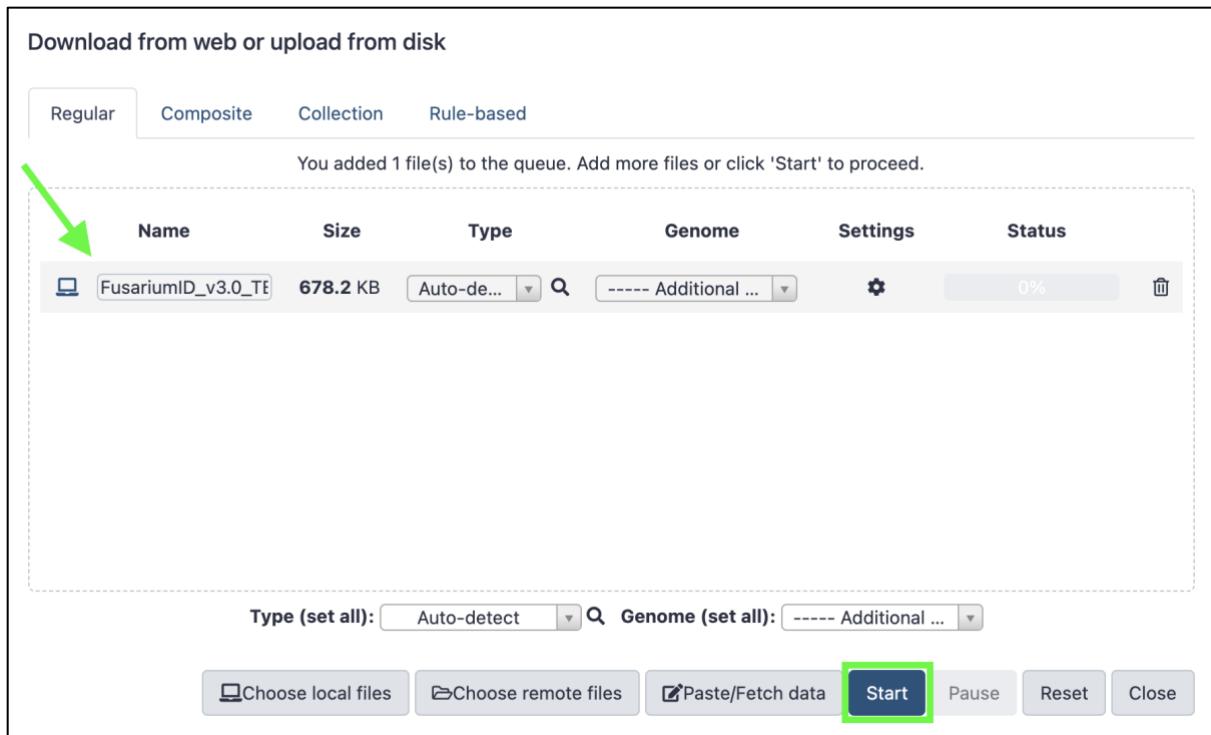
The screenshot shows the Galaxy Tools panel. At the top, there's a search bar labeled "search tools" and a "Tools" section with a star icon. Below this is a large button labeled "Upload Data" with an upward arrow icon, which is highlighted with a green rectangular box. The rest of the panel lists various tool categories:

- Get Data
- Collection Operations
- GENERAL TEXT TOOLS
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Datamash
- GENOMIC FILE MANIPULATION
- FASTA/FASTQ

3. On the pop-up screen, select **Choose local files** or drag and drop the file in this window.



4. Locate the folder in your computer that contains the downloaded [FusariumID\\_v.3.0\\_TEF1](#) FASTA file and click **Open**. On the upload screen, click on **Start**



5. Once the upload is complete (check Status bar), click **Close**

Download from web or upload from disk

Regular    Composite    Collection    Rule-based

Name	Size	Type	Genome	Settings	Status
FusariumID_v3.0_TEF1	678.2 KB	Auto-de... <input type="button" value=""/>	----- Additional ... <input type="button" value=""/>	<input type="button" value=""/>	100% <input checked="" type="checkbox"/>

Type (set all): Auto-detect  Genome (set all): ----- Additional ...

Choose local files  Choose remote files  Paste/Fetch data  Start  Pause  Reset  Close

6. Back at the main screen on the right-hand side, click on the uploaded file on the side panel to the right to display the information about the file. Make sure Galaxy recognizes it as a FASTA file.

History

search datasets

FusariumIDTrial

1 shown

678.17 KB

1: FusariumID\_v3.0\_TEF1.

fas

954 sequences    
 format: **fasta**, database: ?

uploaded fasta file

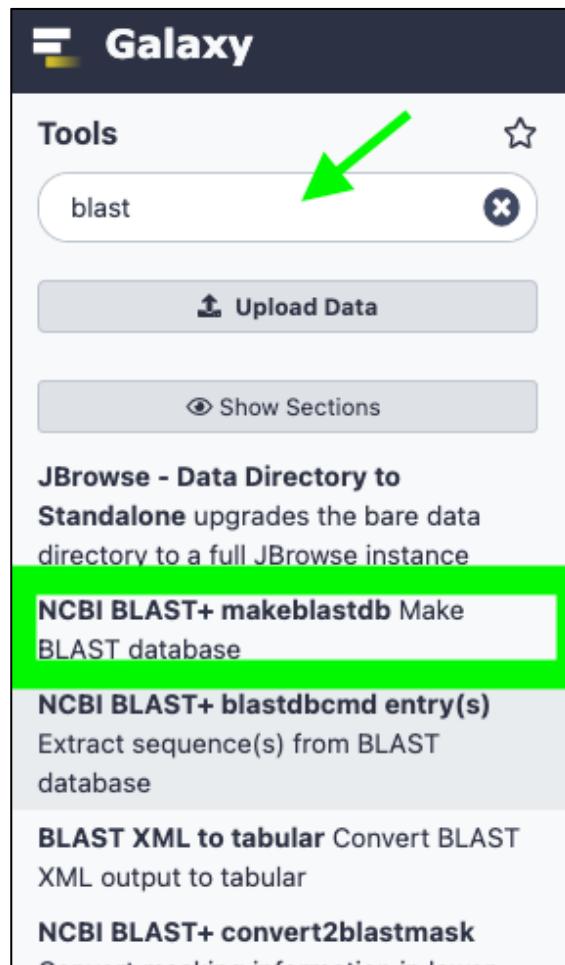
display with IGV local

```
>TEF1_NRRL_13566|FFSC_Asian|fujikuroi|Oryz  
CGTCGTATCGGCCACGTCGACTCTGGCAAGTCGACCACTGT  
GGTTCGAACGTGCTCTGTAACCCCGCTCGAGACCTAAATT  
CATCGATATTGCTTTGGAGTTGAGACTCCTCGCTANTAT  
>TEF1_NRRL_22944|FFSC_Asian|proliferatum|O
```

## Creating the database in Galaxy

7. Create a BLAST database in Galaxy using the [FusariumID\\_v.3.0\\_TEF1](#) file you just uploaded

- a. Search for “blast” in the tools search bar, on the left panel, and select **NCBI BLAST+ makeblastdb**



- b. Under “Molecule type of input” choose **nucleotide** and select the [FusariumID\\_v.3.0\\_TEF1](#) file uploaded on the previous step as the “Input FASTA file(s)”. Name the database under “Title for BLAST database”. We suggest using the name of the FUSARIUM-ID FASTA file including the version (e.g., FusariumID\_v.3.0\_TEF1 for the *TEF1* database).

NCBI BLAST+ makeblastdb Make BLAST database (Galaxy Version 2.10.1+galaxy0)

☆ Favorite    Versions    ▾ Options

**Molecule type of input**

protein  
 nucleotide

(-dbtype)

**Input FASTA files(s)**

1: FusariumID\_v3.0\_TEF1.fas

One or more FASTA files (-in)

**Title for BLAST database**

FusariumID\_v.3.0\_TEF1

This is the database name shown in BLAST search output (-title)

**Parse the sequence identifiers**

No

This is only advised if your FASTA file follows the NCBI naming conventions using pipe '|' symbols (-parse\_seqids)

**Enable the creation of sequence hash values**

Yes

These hash values can then be used to quickly determine if a given sequence data exists in this BLAST database. (-hash\_index)

- c. In that same makeblastdb screen, scroll down and click on **Execute**

Enable the creation of sequence hash values

Yes

These hash values can then be used to quickly determine if a given sequence data exists in this BLAST database. (-hash\_index)

Optional ASN.1 file(s) containing masking data

No maskinfo-asn1 or maskinfo-asn1-binary dataset available.

As produced by NCBI masking applications (e.g. dustmasker, segmasker, windowmasker) (-mask\_data)

Taxonomy options

Do not assign a Taxonomy ID to the sequences

Email notification

No

Send an email notification when the job completes.

Execute

- d. You should now see a notification that the job to create the database is on the queue

Executed NCBI BLAST+ makeblastdb and successfully added 1 job to the queue.

The tool uses this input:

- 1: FusariumID\_v3.0\_TEF1.fas

It produces this output:

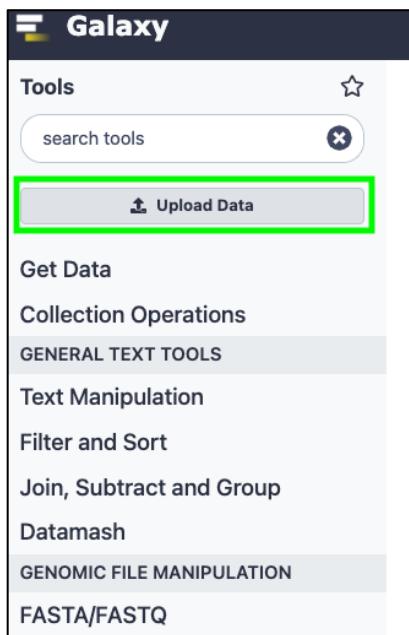
- 2: nucleotide BLAST database from data 1

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

- e. Wait for the step to be completed before running a BLAST search.

## Running a BLAST search in Galaxy

8. Upload a FASTA file with the sequence(s) you would like to identify. On the left-hand panel select **Upload Data**



9. Click on **Choose local files** and navigate to the folder with your sequence(s) or drag and drop the file with your sequence(s). If you wish to BLAST a batch of sequences at once, a concatenated FASTA file can be uploaded instead of each one individually.

This screenshot shows the 'Choose local files' dialog box. At the top, it says 'Download from web or upload from disk'. Below that are tabs for 'Regular', 'Composite', 'Collection', and 'Rule-based'. In the center is a large dashed rectangular area with the text 'Drop files here'. At the bottom, there are input fields for 'Type (set all)' (with 'Auto-detect' selected) and 'Genome (set all)' (with 'Additional ...' dropdown). A row of buttons at the bottom includes 'Choose local files' (highlighted with a green box), 'Choose remote files', 'Paste/Fetch data', 'Start', 'Pause', 'Reset', and 'Close'.

- 10.** After selecting the corresponding file, you may edit the name of the file under **Name** and click on **Start**

Download from web or upload from disk

Regular    Composite    Collection    Rule-based

You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
BLASTExample.fast	557 b	Auto-de...	----- Additional ...	⚙️	0%

Type (set all): Auto-detect Q Genome (set all): ----- Additional ...

- 11.** Once the file is uploaded, click on **Close**

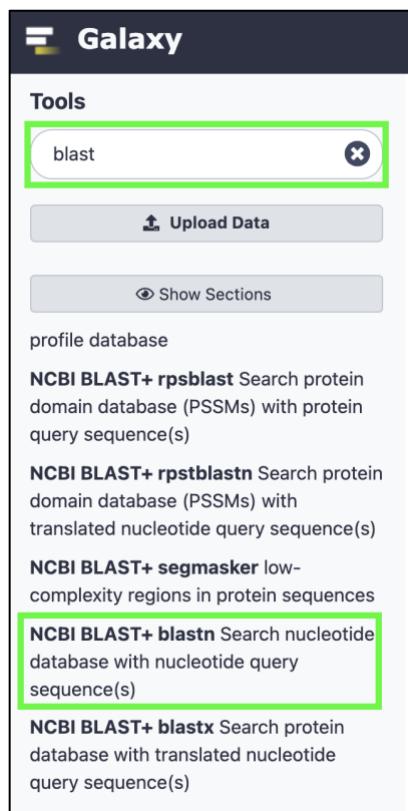
Download from web or upload from disk

Regular    Composite    Collection    Rule-based

Name	Size	Type	Genome	Settings	Status
BLASTExample.fast	557 b	Auto-de...	----- Additional ...	⚙️	100%

Type (set all): Auto-detect Q Genome (set all): ----- Additional ...

**12.** Type in “BLAST” on the tool search bar on the left panel and select **NCBI BLAST+ blastn**



**13.** Set up the BLAST search:

- a. Under **Nucleotide query sequence(s)** select the file with your sequence(s) for identification (“BLASTExample.fasta” in this example).
- b. Under **Subject database/sequences** select BLAST database from your history.
- c. Under **Nucleotide BLAST database** select the FUSARIUM-ID database you previously created.

NCBI BLAST+ blastn Search nucleotide database with nucleotide query sequence(s) (Galaxy Version 2.10.1+galaxy0)

**Nucleotide query sequence(s)**

**Subject database/sequences**

**Nucleotide BLAST database**

**Type of BLAST**  
 megablast - Traditional megablast used to find very similar (e.g., intraspecies or closely related species) sequences  blastn - Traditional BLASTN requiring an exact match of 11, for somewhat similar sequences  blastn-short - BLASTN program optimized for sequences shorter than 50 bases  dc-megablast - Discontiguous megablast used to find more distant (e.g., interspecies) sequences

**Set expectation value cutoff**

**Output format**

- d. Select the **Type of BLAST** (we suggest using megablast) and set other options for the BLAST search (we suggest filtering your results by a percent identity cutoff of 90% identity match under “Advanced Options”) and click on **Execute**

**Type of BLAST**  
 megablast - Traditional megablast used to find very similar (e.g., intraspecies or closely related species) sequences  blastn - Traditional BLASTN requiring an exact match of 11, for somewhat similar sequences  blastn-short - BLASTN program optimized for sequences shorter than 50 bases  dc-megablast - Discontiguous megablast used to find more distant (e.g., interspecies) sequences

**(-task)**

**Set expectation value cutoff**

**Output format**

**Advanced Options**

**Job Resource Parameters**

**Email notification**  
 No  
 Send an email notification when the job completes.

**Execute**

**14.** You will receive a notification that the job has been added to the queue.



Executed **NCBI BLAST+ blastn** and successfully added 1 job to the queue.

The tool uses 2 inputs:

- **13: BLASTExample.fasta**
- **10: nucleotide BLAST database from data 9**

It produces this output:

- **14: blastn BLASTExample.fasta vs 'nucleotide BLAST database from data 9'**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

**15.** Wait for your BLAST query to be completed to evaluate the results.

Optional: convert the output from tabular to CSV file format.

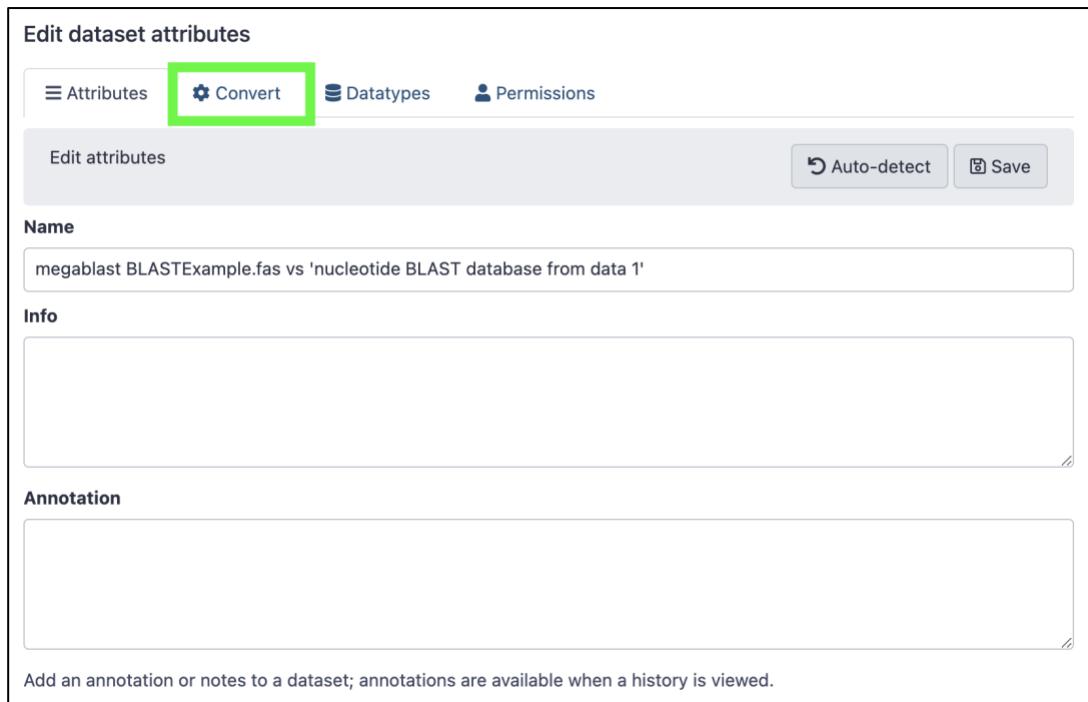
a. Go to "History" on the right-side panel and click on the  to **Edit attributes**

**4: megablast BLASTExample.fas vs 'nucleotide BLAST database from data 1'**   

**989 lines**  
**format: tabular, database: ?**

- b. Click on the **Convert** tab



The screenshot shows the 'Edit dataset attributes' interface. At the top, there are tabs: 'Attributes' (disabled), 'Convert' (highlighted with a green box), 'Datatypes' (disabled), and 'Permissions'. Below the tabs, there's a section for 'Edit attributes' with 'Auto-detect' and 'Save' buttons. The 'Name' field contains 'megablast BLASTExample.fas vs 'nucleotide BLAST database from data 1''. Under 'Info' and 'Annotation', there are large text input fields. At the bottom, a note says 'Add an annotation or notes to a dataset; annotations are available when a history is viewed.'

- c. Select **Convert tabular to CSV** and then click on **Convert datatype**



The screenshot shows the 'Edit dataset attributes' interface with the 'Convert' tab selected. A dropdown menu titled 'Convert to new format' is open, showing 'Convert tabular to CSV' (highlighted with a green arrow) and 'Convert datatype' (highlighted with a green box). The 'Name' field also has a green arrow pointing to it. A note at the bottom states: 'This will create a new dataset with the contents of this dataset converted to a new format.'

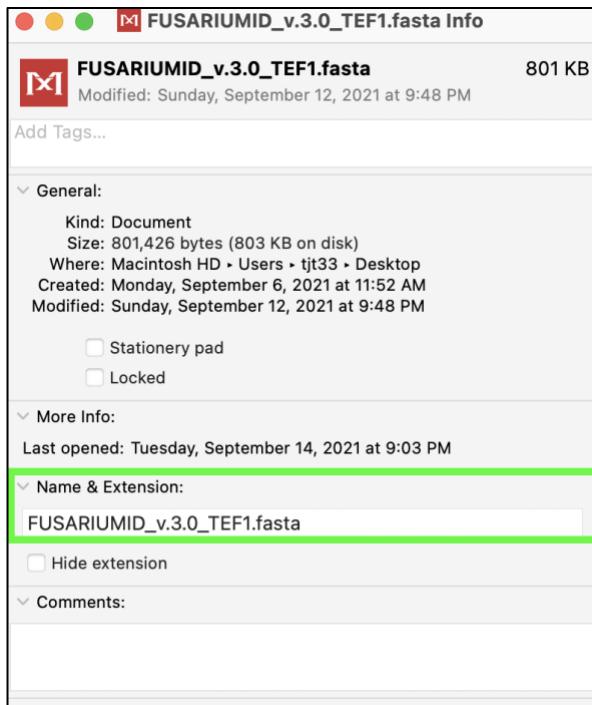
- d. On the right-side panel, click on the file and then click  to download the CSV file to your computer



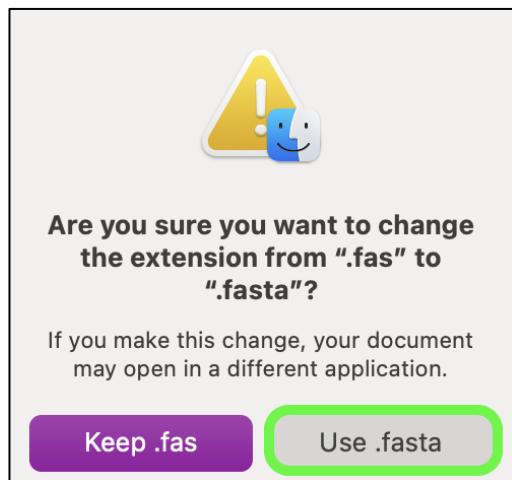
## BLAST using SequenceServer

Instructions on how to install and use SequenceServer are available on the website: <https://sequenceserver.com/>. Here we provide short detailed instructions for using SequenceServer specifically with the **FUSARIUM-ID FASTA file** using SequenceServer version 1.0.14. Installing this tool and setting up the database requires basic use of the command line.

Before starting, change the extension on the **FUSARIUM-ID FASTA file** from .fas to .fna, .fa, or .fasta (these are the formats that SequenceServer recognizes). You can do this by right clicking on the file and selecting “Get Info” to change the extension under “Name and Extension”



It will ask you if you are sure about the extension change. Click **Use .fasta**



## Installing SequenceServer (only available for Linux or Mac)

1. Open a terminal window (command line). For example, in Mac you can do this by opening the Launchpad and searching for “terminal” or by pressing ⌘ + space on your keyboard and searching for “terminal.”
2. Now you need to install Ruby, you can find several options for installation on the website <https://www.ruby-lang.org/en/documentation/installation/>

If you have a package management system, such as [Homebrew](#), this would be the easiest way to install it. But you can also do it through installers or by building Ruby from source.

You will need **administrator privileges** to install the software, since it will ask for a password. If you do not have administrator privileges on the computer, reach out to the IT department at your institution. If you have administrative privileges, enter the password when prompted and hit ENTER. It will not look on the screen like the password is being typed (no characters will show), because the terminal does not echo the keyboard’s input for security reasons. Just type the password and hit ENTER.

3. Once Ruby is installed, install SequenceServer using the following command (instructions on <https://sequenceserver.com/#installation>)

```
sudo gem install sequenceserver
```

4. You will receive a message letting you know that SequenceServer was successfully installed. Now, you can add the FUSARIUM-ID database (see next step).

```
Successfully installed json_pure-1.8.6
Fetching: ox-2.14.5.gem (100%)
Building native extensions. This could take a while...
Successfully installed ox-2.14.5
Fetching: slop-3.6.0.gem (100%)
Successfully installed slop-3.6.0
Fetching: sequenceserver-1.0.14.gem (100%)

-----
Thank you for installing SequenceServer :)!
To launch SequenceServer execute 'sequenceserver' from command line.

$ sequenceserver

Visit http://sequenceserver.com for more.
```

## Adding the database in SequenceServer

5. To add the FUSARIUM-ID BLAST database, enter the following command into the computer terminal:

```
makeblastdb -dbtype nucl -title FusariumID_v.3.0_TEF1 -in /path/to/fusariumIDfile
```

```
tjt33@D4-C02V80ACHV2J ~/Desktop
$ makeblastdb -dbtype nucl -title FusariumID_v.3.0_TEF1 -in /Users/tjt33/Desktop/FUSARIUMID_v.3.0_TEF1.fasta

Building a new DB, current time: 09/14/2021 21:36:23
New DB name: /Users/tjt33/Desktop/FUSARIUMID_v.3.0_TEF1.fasta
New DB title: FusariumID_v.3.0_TEF1
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1107 sequences in 0.036834 seconds.
```

We used a generalized file path notation ('/path/to/fusariumIDfile') in the command above. It refers to the specific path in your computer where you have the [FUSARIUM-ID FASTA file](#) downloaded. Therefore, the path will vary on each computer depending on where you save the file.

To determine the path to the file you can:

- i) navigate to the folder where the fasta file is saved in terminal and enter the command `pwd` to get the full path to the file. As an example of a specific path in the command line refer to the screenshot provided above in this tutorial.
- ii) On Mac computers, alternatively you can
  - a. open Finder and navigate to the file for which you want to copy the path
  - b. right-click (or Control+Click or two-finger click on trackpads) on the file or folder in the Finder
  - c. While on that right-click menu, hold down the **Option** key and select the “Copy as Pathname” option
  - d. It saves in your clipboard and you can paste it into your command

To avoid having to determine the path, you can either:

- i) navigate in the terminal to the folder where the **FUSARIUM-ID FASTA file** is saved before entering the command  
**or,**
- ii) move or save the **FUSARIUM-ID FASTA file** to the folder you are currently at in the terminal.

This way you do not need to enter a path to the file and can just enter the name of the file itself. For example:

```
makeblastdb -dbtype nucl -title FusariumID_v.3.0_TEF1 -in FusariumID_v.3.0_TEF1
```

```
tjt33@D4-C02V80ACHV2J ~
[$ cd Desktop/

tjt33@D4-C02V80ACHV2J ~/Desktop
[$ makeblastdb -dbtype nucl -title FusariumID_v.3.0_TEF1 -in FUSARIUMID_v.3.0_TEF1.fasta

Building a new DB, current time: 09/14/2021 21:38:42
New DB name: /Users/tjt33/Desktop/FUSARIUMID_v.3.0_TEF1.fasta
New DB title: FusariumID_v.3.0_TEF1
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 1107 sequences in 0.0531499 seconds.
```

**Note:** Additional configuration values are described on <https://sequenceserver.com/doc/>

6. Once your database is added, use the following command to open a web browser window with the BLAST search option:

```
sequenceserver -d /path/to/fusariumIDdatabase
```

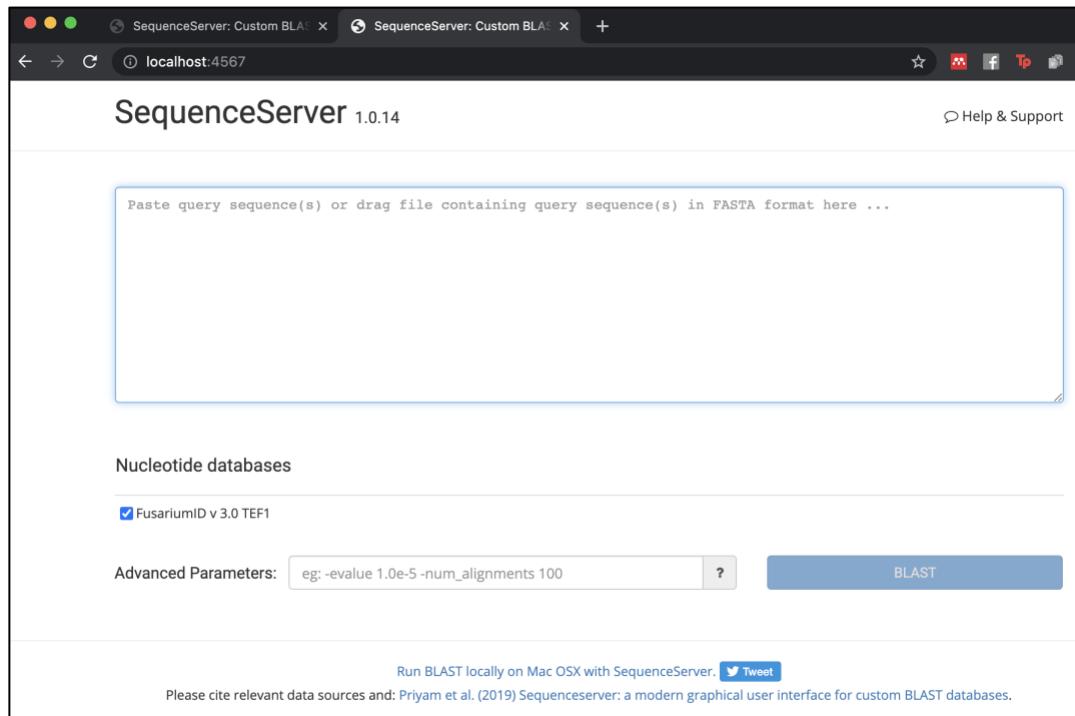
or

```
sequenceserver -d . [if you are in the directory already]
```

```
tjt33@D4-C02V80ACHV2J ~/Desktop
[$ sequenceserver -d .
[2021-09-14 21:52:49] WARN Will listen on all interfaces (0.0.0.0). Consider using 127.0.0.1 (--host option).
** SequenceServer is ready.
Go to http://localhost:4567 in your browser and start BLASTing!
Press CTRL+C to quit.
```

## Running a BLAST search in SequenceServer

7. Go to the browser window that opened after adding the database or type the address <http://localhost:4567/> on a web browser.



- a. After you have initially set up the database you can go back to this website to BLAST sequences by initializing Sequence Server on terminal using the command:

```
sequenceserver -d /path/to/fusariumIDdatabase
```

or

```
sequenceserver -d . [if you are in the directory already]
```

## 8. Set up the BLAST run

- a. Enter the potential *Fusarium* sequence(s) you are interested in identifying in the box. You may include a header or just the sequence, and you can include multiple sequences at once. You can also upload a concatenated FASTA file instead of each one individually.
- b. Select the corresponding [FUSARIUM-ID FASTA file](#) database
- c. You can adjust advanced parameters by clicking on the question mark by **Advanced Parameters**
- d. Click on **BLAST**

The screenshot shows the SequenceServer 1.0.14 interface. At the top, there's a sequence entry box containing a FASTA sequence:

```
>PotentialFusariumtoID
CTGCGGAAGGATCATTACAGTAGTCACCCAGGGTGGCGCAAGGCCTCTGGTAACCTACCACCCTTGTTATTACACTTGTGCTTGGCAAGCCTGCCCTGGG
CTGCTGGCTCCGGCCGGCGAGCGCTTCCAGAGGACCTAAACTCTGTTGTCTATACTGTCTGAGTACTATATAATAGTTAAAACTTCAACACGGATCTCTGGTT
CTGGCATCGATGAAGAACGAGCGAAATGCGATAAGTAATGTAATTGAGAATTGAGAATTTCAGTGAATCATCGAATCTTGACGCACATTGGCCCCCTGGTATTCCGGGG
GCATGCCCTGTCCGAGCGTCATTACAACCCCTCAAGCTCAGCTTGGTATTGGGCCGACCCGGGGCCCTAAAGTCAGTGGCGGTGCCGTCCGCTCCGAGCGTA
GTAATTCTTCGCTCTGGAGGTCCGTCGTGCTGCCAGCAACCCCAATTTCAGGTTGACCTCGGATCAGTAGGGATAACCGCTGAACCTAACG
```

Below the sequence entry is a section for "Nucleotide databases" with a checked checkbox for "FusariumID v 3.0 TEF1".

Under "Advanced Parameters", there is a text input field with the placeholder "eg: -evalue 1.0e-5 -num\_alignments 100" and a blue "BLASTN" button.

At the bottom, there are links for "Run BLAST locally on Linux with SequenceServer." and "Tweet". A note says "Please cite relevant data sources and: Priyam et al. (2019) Sequenceserver: a modern graphical user interface for custom BLAST databases."

**Note:** You can add multiple databases by following step #4 under '[Adding the database in SequenceServer](#)'. All added databases will then show as options on the web browser BLAST interface:

The screenshot shows the SequenceServer 1.0.14 web interface. At the top, it says "SequenceServer 1.0.14" and "Help & Support". Below that is a large input area with the placeholder text "Paste query sequence(s) or drag file containing query sequence(s) in FASTA format here ...". Underneath this is a section titled "Nucleotide databases" which is highlighted with a green border. It contains two checkboxes: " FusariumID\_v.3.0\_RPB2" and " FusariumID\_v.3.0\_TEF1". Below this section is an "Advanced Parameters:" label followed by a text input field containing "eg: -evalue 1.0e-5 -num\_alignments 100" and a question mark icon. To the right of the input field is a blue "BLAST" button.

## BLAST using BLAST+ executable in command line

Here we will briefly cover using BLAST on the command line in a simple way with the [FUSARIUM-ID FASTA file](#). For more details, we recommend reading the help information (e.g., `blastn --help`) and the NCBI BLAST Command Line Applications User Manual at <http://www.ncbi.nlm.nih.gov/books/NBK1763/>

### Setting up BLAST+

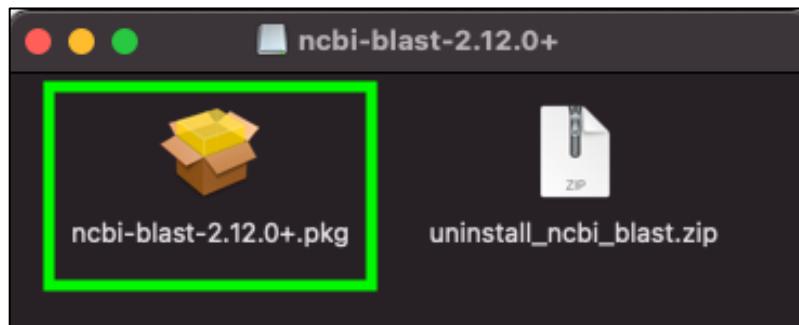
1. Install the NCBI BLAST+ tools following the respective instructions on the [NCBI BLAST Command Line Applications User Manual](#) for your operating system. Instructions are available for Windows and LINUX/UNIX.

For example, to install the tools for MacOS for users with administrator privileges:

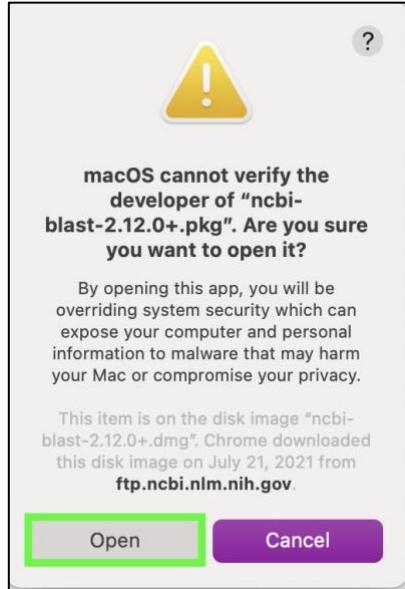
- a. Download the `ncbi-blast-version.dmg` installer (for the newest version available)

Name	Last modified	Size
<a href="#">Parent Directory</a>		-
<a href="#">ChangeLog</a>	2021-06-28 11:23	85
<a href="#">ncbi-blast-2.12.0+-1.src.rpm</a>	2021-06-28 11:23	51M
<a href="#">ncbi-blast-2.12.0+-1.src.rpm.md5</a>	2021-06-28 11:23	63
<a href="#">ncbi-blast-2.12.0+-1.x86_64.rpm</a>	2021-06-28 11:23	187M
<a href="#">ncbi-blast-2.12.0+-1.x86_64.rpm.md5</a>	2021-06-28 11:23	66
<a href="#">ncbi-blast-2.12.0+-src.tar.gz</a>	2021-06-28 11:23	56M
<a href="#">ncbi-blast-2.12.0+-src.tar.gz.md5</a>	2021-06-28 11:23	64
<a href="#">ncbi-blast-2.12.0+-src.zip</a>	2021-06-28 11:23	60M
<a href="#">ncbi-blast-2.12.0+-src.zip.md5</a>	2021-06-28 11:23	61
<a href="#">ncbi-blast-2.12.0+-win64.exe</a>	2021-06-28 11:23	94M
<a href="#">ncbi-blast-2.12.0+-win64.exe.md5</a>	2021-06-28 11:23	63
<a href="#">ncbi-blast-2.12.0+-x64-linux.tar.gz</a>	2021-06-28 11:24	236M
<a href="#">ncbi-blast-2.12.0+-x64-linux.tar.gz.md5</a>	2021-06-28 11:24	70
<a href="#">ncbi-blast-2.12.0+-x64-macosx.tar.gz</a>	2021-06-28 11:24	144M
<a href="#">ncbi-blast-2.12.0+-x64-macosx.tar.gz.md5</a>	2021-06-28 11:24	71
<a href="#">ncbi-blast-2.12.0+-x64-win64.tar.gz</a>	2021-06-28 11:24	93M
<a href="#">ncbi-blast-2.12.0+-x64-win64.tar.gz.md5</a>	2021-06-28 11:24	70
<a href="#">ncbi-blast-2.12.0+.dmg</a>	2021-06-28 11:24	147M
<a href="#">ncbi-blast-2.12.0+.dmg.md5</a>	2021-06-28 11:24	57

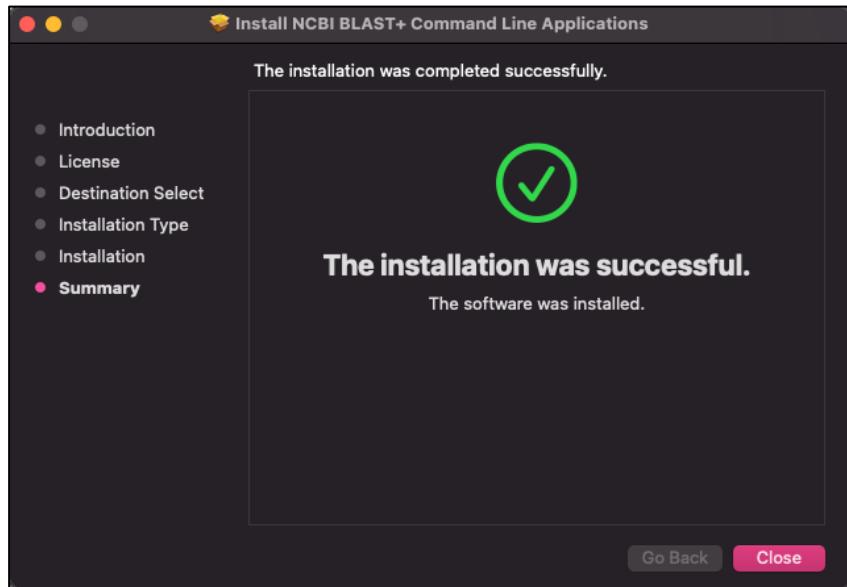
- b. Click on the downloaded file. On the pop-up screen, double click on the `ncbi-blast-version.pkg`



When you click on the downloaded file you might get a message saying it cannot be opened because it is from an unidentified developer. If this happens, right click on the file and select **Open** from the menu options. You will then get a message asking if you are sure you want to open the file. Click **Open**.



c. Follow the instructions in the installer



Refer to the manual for instructions for other operating systems or for users without administrator privileges. Alternatively, you can also install BLAST+ using other tools like Bioconda.

## Adding the database using BLAST+ executable in the command line

1. Open a terminal window (command line). For example, in Mac you can do this by opening the Launchpad and searching for “terminal” or by pressing ⌘ + space on your keyboard and searching for “terminal.”
2. Create a database using the [FUSARIUM-ID FASTA file](#) using the following command:

```
makeblastdb -in /Path/to/FusariumIDFile -out /Path/to/folder/nameofdatabase -dbtype nucl
```

```
tjt33@D4-C02V80ACHV2J ~  
$ makeblastdb -in /Users/tjt33/Desktop/FusariumID_v3.0_TEF1.fas -out /Users/tjt33/Desktop/FusariumID_v3.0_TEF1_Database -dbtype nucl  
  
Building a new DB, current time: 07/21/2021 16:09:36  
New DB name: /Users/tjt33/Desktop/FusariumID_v3.0_TEF1_Database  
New DB title: /Users/tjt33/Desktop/FusariumID_v3.0_TEF1.fas  
Sequence type: Nucleotide  
Keep MBits: T  
Maximum file size: 1000000000B  
FASTA-Reader: Title ends with at least 20 valid nucleotide characters. Was the sequence accidentally put in the title line?  
Ignoring sequence 'lcl|881' as it has no sequence data  
Adding sequences from FASTA; added 953 sequences in 0.0519571 seconds.
```

We used a generalized file path notation ('/path/to/FusariumIDFile') in the command above. It refers to the specific path in your computer where you have the [FUSARIUM-ID FASTA file](#) downloaded. Therefore, the path will vary on each computer depending on where you save the file. To determine the path to the file you can:

- i) navigate to the folder where the fasta file is saved in terminal and enter the command `pwd` to get the full path to the file. As an example of a specific path in the command line refer to the screenshot provided above in this tutorial.
- ii) On Mac computers, alternatively you can
  - a. open Finder and navigate to the file for which you want to copy the path
  - b. right-click (or Control+Click or two-finger click on trackpads) on the file or folder in the Finder
  - c. While on that right-click menu, hold down the **Option** key and select the “Copy as Pathname” option
  - d. It saves in your clipboard and you can paste it into your command

To avoid having to determine the path, you can either:

- iii) navigate in the terminal to the folder where the [FUSARIUM-ID FASTA file](#) is saved before entering the command  
**or,**
- iv) move or save the [FUSARIUM-ID FASTA file](#) to the folder you are currently at in the terminal.

This way you do not need to enter a path to the file and can just enter the name of the file itself. For example:

```
makeblastdb -in FusariumID_v.3.0_TEF1.fas -out FusariumID_v.3.0_TEF1 -dbtype nucl
```

## Running a BLAST search using BLAST+ executable in the command line

3. Run the BLAST search. You can BLAST individual sequences or a batch of sequences by using a concatenated FASTA file as input instead of each one individually.

You can decide how you want the output to look like depending on the format you prefer and even add custom fields to the output (for more information about how to do this check `blast -help`). We provide here a few examples:

- a. If you want to see the alignments on the terminal window:

```
blastn -db <database_name> -query <file_with_sequence_query>
```

The parts of the command in **bold** refer to the specific files in your computer (if you are in that directory in terminal) or the path to the folder that contains these files.

Example of this command using a path to the files:

```
blastn -db /Users/tjt33/Desktop/FusariumID_v.3.0_TEF1_Database -query /Users/tjt33/Desktop/Query.fas
```

Example of this command using the name of the files because you are currently in that directory folder:

```
blastn -db /Users/tjt33/Desktop/FusariumID_v.3.0_TEF1_Database -query /Users/tjt33/Desktop/Query.fas
```

The output will show in this format (and will be quite lengthy):

```
>TEF1_NRRL_13566|FFSC_Asian|fujikuroi|Oryza_sativa|Taiwan
Length=640

Score = 1173 bits (635), Expect = 0.0
Identities = 639/639 (100%), Gaps = 0/639 (0%)
Strand=Plus/Plus

Query  2    CGTCGTCATCGGCCACGTCGACTCTGGCAAGTCGACCACTGTGAGTACTACCCCTGACAA 61
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...
Sbjct   1    CGTCGTCATCGGCCACGTCGACTCTGGCAAGTCGACCACTGTGAGTACTACCCCTGACAA 60
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...
Query  62   TGAGCTTATCTGTCNTCGTAATCTGACCCAAAATCTGGCGGGGTATATCTCAGAACGCA 121
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...
Sbjct   61   TGAGCTTATCTGTCNTCGTAATCTGACCCAAAATCTGGCGGGGTATATCTCAGAACGCA 120
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...
Query  122   TATGCTGACTTCGTTCACAGACCGGTCACTTGATCTACAGTGCCTGGTATCGACAAG 181
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...
Sbjct   121   TATGCTGACTTCGTTCACAGACCGGTCACTTGATCTACAGTGCCTGGTATCGACAAG 180
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...
Query  182   CGAACCCATCGAGAAAGTTGAGGAAGGTTAGTCACTTCCCTTCGATGGCCCGTCCCTTGCC 241
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...
Sbjct   181   CGAACCCATCGAGAAAGTTGAGGAAGGTTAGTCACTTCCCTTCGATGGCCCGTCCCTTGCC 240
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...
Query  242   CACCGATTCCCTTACGGTTGAAACGTGCCTGCTACCCCGCTCGAGACCTAAATTTTG 301
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...
Sbjct   241   CACCGATTCCCTTACGGTTGAAACGTGCCTGCTACCCCGCTCGAGACCTAAATTTTG 300
|||...|||...|||...|||...|||...|||...|||...|||...
Query  302   CGATATGACCGTAATTttttttttttttttttttttttttttttttttttttttttttttttt 361
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...
Sbjct   301   CGATATGACCGTAATTttttttttttttttttttttttttttttttttttttttttttttt 360
|||...|||...|||...|||...|||...|||...|||...|||...
Query  362   TTTTCCCCCTTCCGTCCACAACCTCAATGAGCGCAATGTCACGTCTCAAACATAACATT 421
|||...|||...|||...|||...|||...|||...|||...|||...
Sbjct   361   TTTTCCCCCTTCCGTCCACAACCTCAATGAGCGCAATGTCACGTCTCAAACATAACATT 420
|||...|||...|||...|||...
Query  422   CGACAAATAGGAAGCCGCTGAGCTCGGTAAAGGGTTCTTCAGTACGCCCTGGTTCTTGAC 481
|||...|||...|||...|||...
Sbjct   421   CGACAAATAGGAAGCCGCTGAGCTCGGTAAAGGGTTCTTCAGTACGCCCTGGTTCTTGAC 480
```

- b. If you prefer to get a tabular output in the terminal window, you can use this command

```
blastn -db /Users/tjt33/Desktop/FusariumID_v.3.0_TEF1_Database -
query /Users/tjt33/Desktop/Query.fas -outfmt 7
```

The output will look like this:

blastn -db /Users/tjt33/Desktop/FusariumID_v.3.0_TEF1_Database -query /Users/tjt33/Desktop/Query.fas -outfmt 7												
# BLASTN 2.12.0+												
# Query: Query												
# Database: /Users/tjt33/Desktop/FusariumID_v.3.0_TEF1_Database												
# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score												
# 512 hits found												
Query TEF1_NRRL_13566 FFSC_Asian fujikuroi Oryza_sativa Taiwan 100.000 639 0 0 2 640 1 639 0.0 1173												
Query TEF1_NRRL_44887 FFSC_Asian fujikuroi Human USA inf=eye 97.809 639 14 0 2 640 1 639 0.0 1107												
Query TEF1_NRRL_31418 FFSC_Asian species? tomato_soil USA_FL 97.496 639 16 0 2 640 1 639 0.0 1096												
Query TEF1_NRRL_26131 FFSC_Asian globosum Zea_mays South Africa 94.892 646 25 3 2 640 1 645 0.0 1007												
Query TEF1_NRRL_46612 FFSC_Asian proliferatum Solanum_lycopericicum Italy_(Sardinia) 94.427 646 28 3 2 640 1 645 0.0 990												
Query TEF1_NRRL_26794 FFSC_Asian fractiflexum Cymbidium_sp Japan 94.272 646 27 6 2 640 1 643 0.0 983												
Query TEF1_NRRL_28852 FFSC_Asian fractiflexum Cymbidium_sp Japan 94.118 646 28 6 2 640 1 643 0.0 977												
Query TEF1_NRRL_22944 FFSC_Asian proliferatum Cattleya_hybrid Germany 93.498 646 34 3 2 640 1 645 0.0 957												
Query TEF1_NRRL_31259 FFSC_Clade mangiferae Ficus_carica Africa 92.558 645 32 8 2 640 1 635 0.0 915												
Query TEF1_NRRL_25226 FFSC_Asian mangiferae Mangiferae_indica India gb=AF160281.1 92.272 647 33 9 1 640 1 637 0.0 905												
Query TEF1_NRRL_26427 FFSC_Asian FFSCundesc rain_forest_soil Papua_New Guinea 92.248 645 33 9 2 640 1 634 0.0 902												
Query TEF1_NRRL_13164 FFSC_African diamini corn_soil South Africa 91.705 651 42 6 2 640 1 651 0.0 896												
Query TEF1_NRRL_22949 FFSC_African udum unknown unknown 91.680 649 42 8 2 640 1 647 0.0 893												
Query TEF1_NRRL_31852 FOSC foetens Begonia_elatior_hybrid The_Netherlands ST=29 92.126 635 35 10 18 639 1 633 0.0 885												
Query TEF1_NRRL_26061 FFSC_African FFSCundesc Striga_heemonthica Madagascar 91.512 648 44 6 2 640 1 646 0.0 885												
Query TEF1_NRRL_31631 FFSC_African pseudocircinatum unknown South Africa 91.538 650 38 11 2 640 1 644 0.0 883												
Query TEF1_NRRL_25309 FFSC_Asian concentricum Triticum_aestivum Japan 91.641 646 37 9 2 640 1 636 0.0 881												
Query TEF1_NRRL_25303 FFSC_Asian concentricum Oryza_sativa Japan 91.641 646 37 9 2 640 1 636 0.0 881												
Query TEF1_NRRL_25181 FFSC_Asian concentricum Musa_sapientum Costa Rica 91.641 646 36 10 2 640 1 635 0.0 881												
Query TEF1_NRRL_13448 FFSC_African nygamai Sorghum_bicolor Australia 91.411 652 38 14 2 640 1 647 0.0 881												
Query TEF1_NRRL_38302 FOSC foetens Pinus_radiata_seedling Chile ST=186 91.969 635 35 11 18 639 1 632 0.0 880												
Query TEF1_NRRL_13617 FFSC_African phyllophilum Dracaena_deremensis Italy 91.358 648 45 8 2 640 1 646 0.0 880												
Query TEF1_NRRL_28366 FOSC oxysporum [Liliums_sp USA gb=AF246857.1;ST=102 91.244 651 43 11 1 639 1 649 0.0 878												
Query TEF1_NRRL_13308 FFSC_African acutatum soil Malaysia 91.231 650 43 8 2 640 1 647 0.0 876												
Query TEF1_NRRL_26757 FFSC_American FFSCundesc reed South Africa 91.358 648 39 10 2 640 1 640 0.0 874												
Query TEF1_NRRL_26756 FFSC_American FFSCundesc grass South Africa 91.358 648 39 10 2 640 1 640 0.0 874												
Query TEF1_NRRL_28372 FOSC oxysporum [Asparagus_officinalis USA fsp=aspargragi;gb=AF246866.1;ST=106 91.104 652 43 11 1 639 1 650 0.0 872												
Query TEF1_NRRL_26622 FOSC oxysporum [Phoenix_sp Morocco fspa=albedinis;gb=D0837688.1;ST=80 91.104 652 43 11 1 639 650 0.0 872												
Query TEF1_NRRL_22545 FOSC oxysporum [Matthiola_incana Germany fsp=matthiolis;gb=D0837682.1;ST=13 91.104 652 43 11 1 639 1 650 0.0 872												
Query TEF1_NRRL_26793 FFSC_American sudanense Striga_heemonthica Sudan 91.231 650 40 11 2 640 1 644 0.0 872												
Query TEF1_NRRL_26437 FOSC oxysporum [Cucumis_sativus USA_SC fsp=cucumerinum;gb=AF362168.1;ST=71 91.091 651 43 12 1 639 1 648 0.0 870												
Query TEF1_NRRL_22533 FOSC oxysporum [Aechmea_fasciata Germany fsp=achmeae;gb=AY527622.1;ST=5 91.091 651 43 11 1 639 1 648 0.0 870												
Query TEF1_NRRL_31633 FFSC_African lactis unknown South Africa 91.104 652 40 14 2 640 1 647 0.0 870												
Query TEF1_NRRL_31632 FFSC_African lactis unknown South Africa 91.104 652 40 14 2 640 1 647 0.0 870												
Query TEF1_NRRL_31630 FFSC_African lactis unknown South Africa 91.091 651 41 11 2 640 1 646 0.0 869												
Query TEF1_NRRL_31649 FFSC_Asian sacchari Musa_AAB Kenya 91.036 647 48 7 2 640 1 645 0.0 869												

If you would like to save the output, you can do so by adding the following to the previous command:

```
-out results.txt
```

tjt33@D4-C02V80ACHV2J ~/Desktop												
\$ blastn -db /Users/tjt33/Desktop/FusariumID_v.3.0_TEF1_Database -query /Users/tjt33/Desktop/Query.fas -outfmt 7 -out results.txt												

This will generate a txt file with the output that you can either visualize on the terminal or open separately on your computer. You can even open this tab-delimited file using Excel for easier visualization.

# BLAST using Geneious Prime

Geneious Prime is a bioinformatics program that provides tools for sequence analysis with a Graphical User Interface (GUI). There are different types of subscriptions (pricing details: <https://www.geneious.com/pricing/>) and a 14-day free trial option is available.

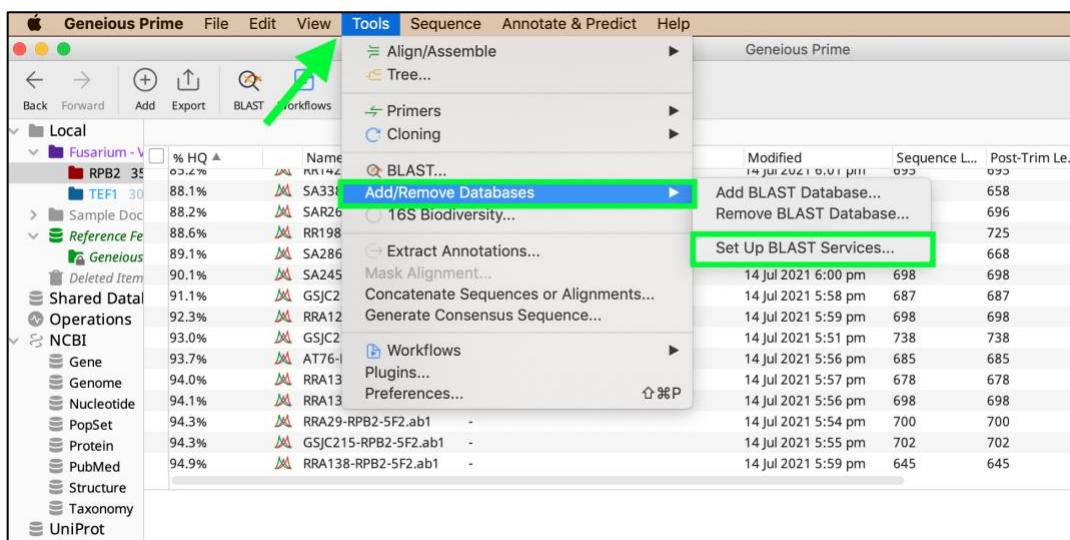
Geneious provides general instructions to BLAST against a local database in this link under “Custom BLAST”: <https://www.geneious.com/tutorials/sequence-searching/>. Here, we provide short step-by-step instructions on how you can run a BLAST on Geneious against the **FUSARIUM-ID FASTA file**. This tutorial was generated for Geneious Prime version 2021.1.1.

## 1. Open Geneious Prime on your computer

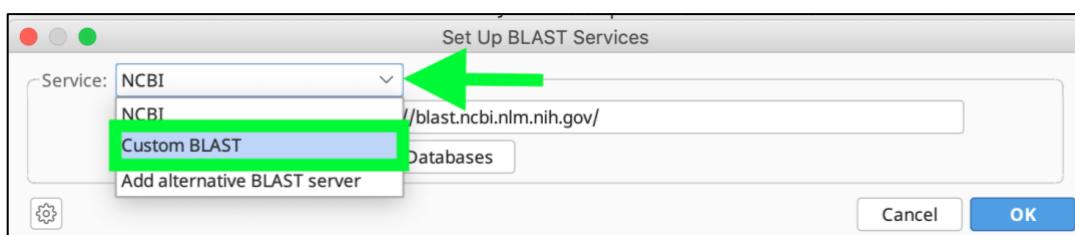
### ***Set up Custom BLAST service***

You will need to do this step only if you have not already set this feature before. If you have already set up Custom BLAST Service, go to the next step ('[Adding the database in Geneious](#)').

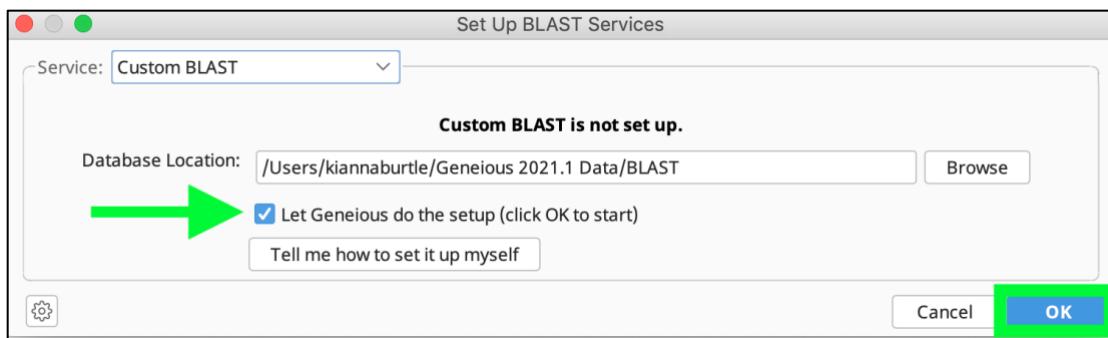
## 2. Go to Tools, select Add/Remove Databases and then click on Set Up BLAST Services



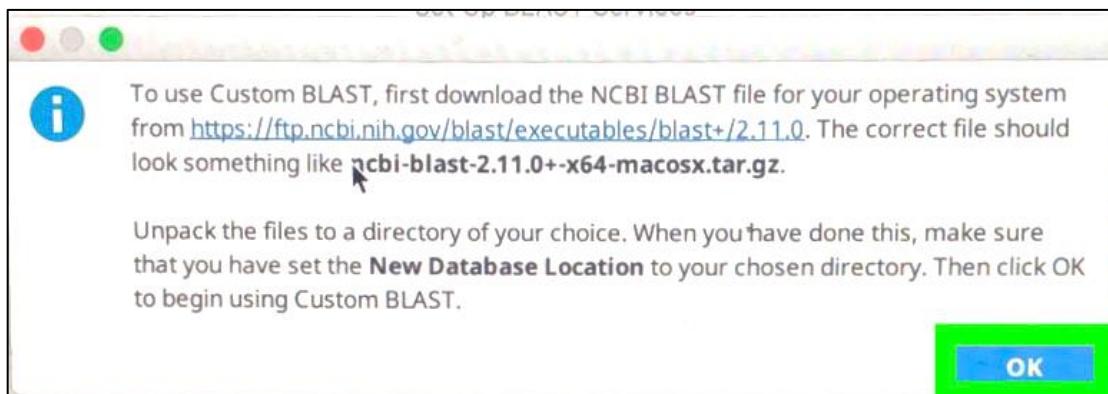
## 3. In the pop-up screen, select **Custom BLAST** using the dropdown menu



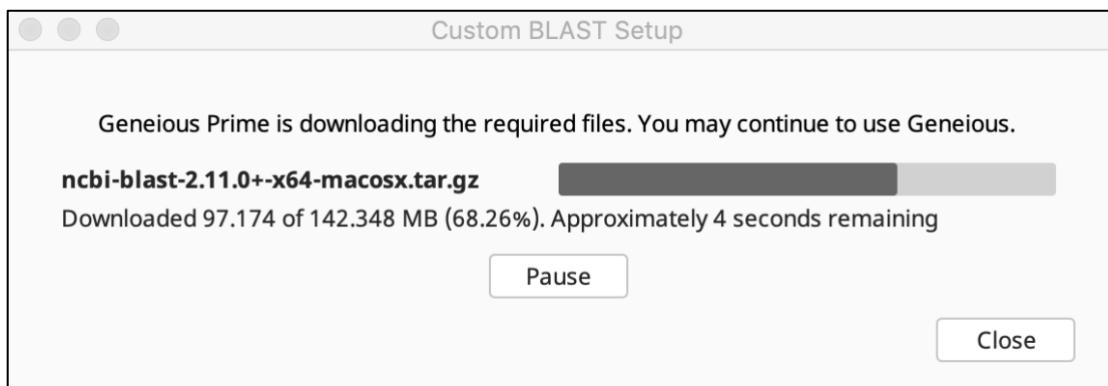
- On the 'Set Up BLAST Services' screen, make sure to check the option to **Let Geneious do the setup** and then click on **Ok**.



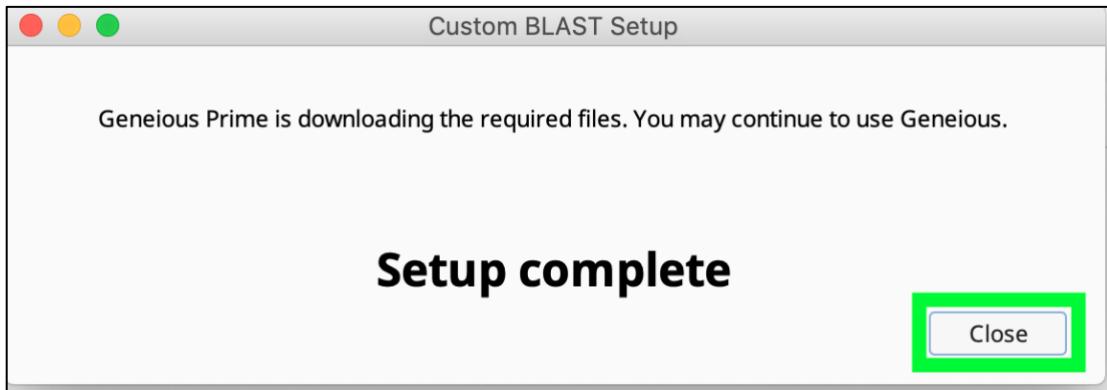
- You will see a pop-up window with instructions on how to manually download the NCBI BLAST file. To do the process automatically in Geneious, just click **Ok**.



- You will see a screen telling you that Geneious Prime is downloading the required files.



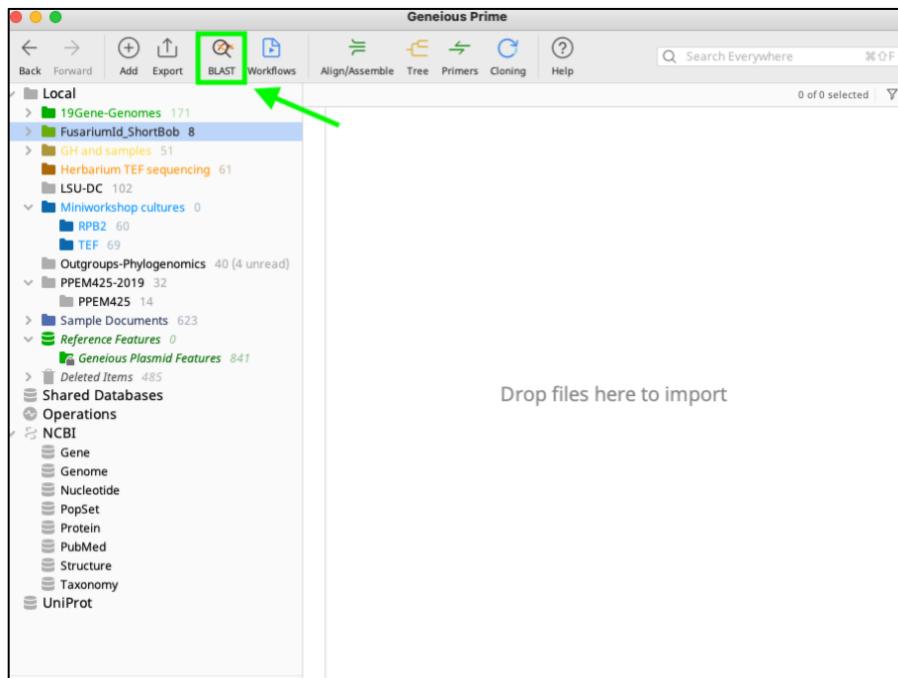
7. When the download is complete, you will see a screen saying that the Setup is complete. Click on **Close**. Now you can add the database.



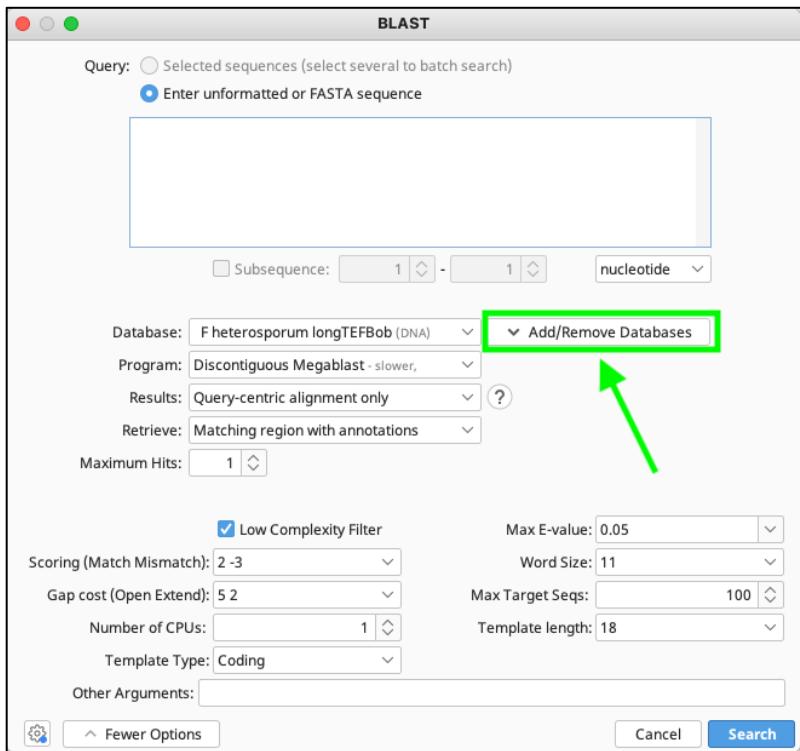
### Adding the database in Geneious

You only need to do this step once. Once the database is added you can access it anytime you need to BLAST a sequence. If there is a new version of the [FUSARIUM-ID FASTA file](#) or if you want to set up a database for each locus [FUSARIUM-ID FASTA file](#), you will need to do this step again to create a new database.

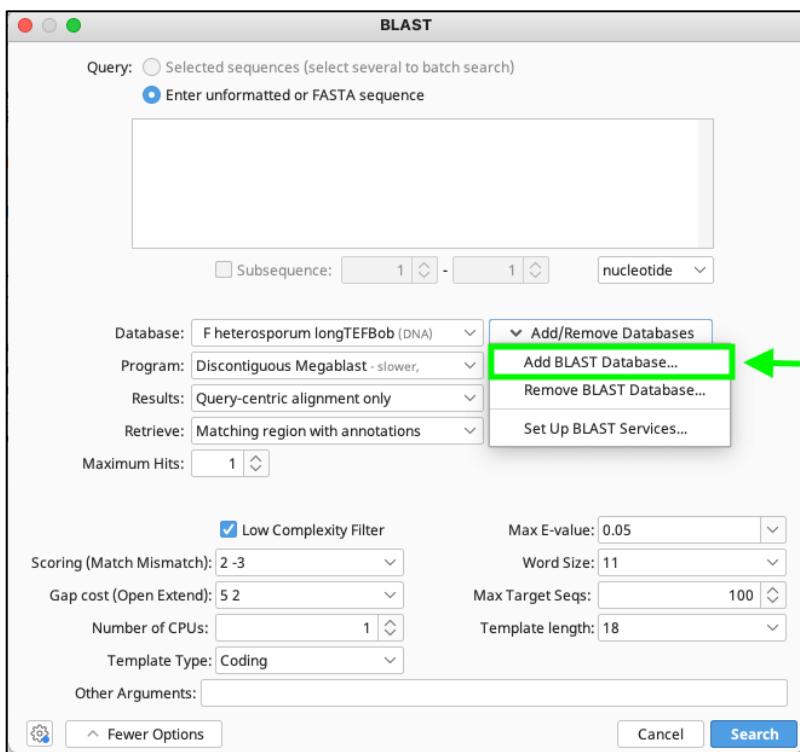
8. On the top menu, click on **BLAST**



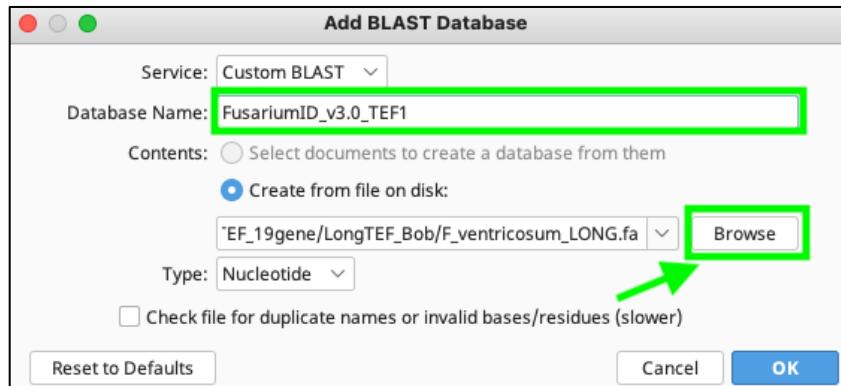
**9. On the pop-up window, click on Add/Remove Databases**



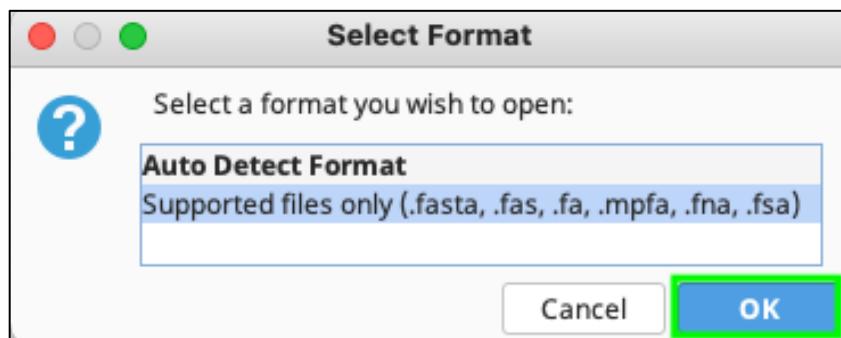
**10. Select Add BLAST Database**



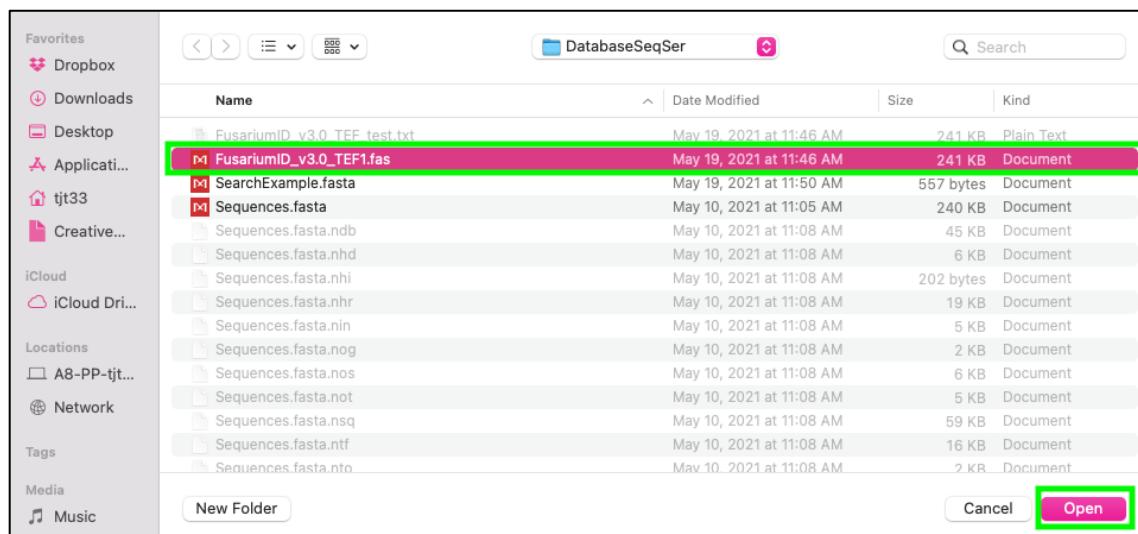
- 11.** Type in a name for the database in the corresponding text box (“Database Name”). We suggest using the name of the [FUSARIUM-ID FASTA file](#) including the version (e.g. **FusariumID\_v.3.0\_TEF1**). Then, click on Browse to select the file from where you have it saved on your computer.



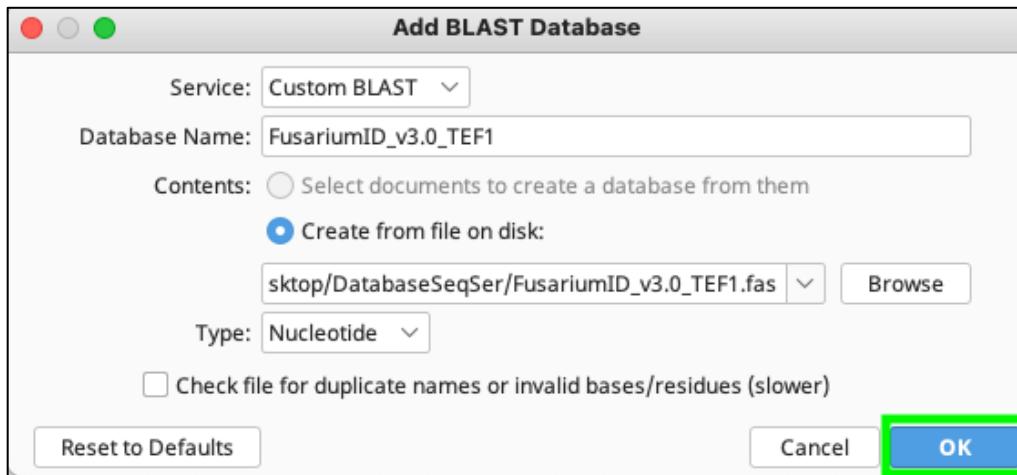
- 12.** You will receive a message indicating the formats recognized by Geneious. Click on **Ok**



- 13.** Search in your computer, select the corresponding [FUSARIUM-ID FASTA file](#) by clicking on it, and then click on **Open**

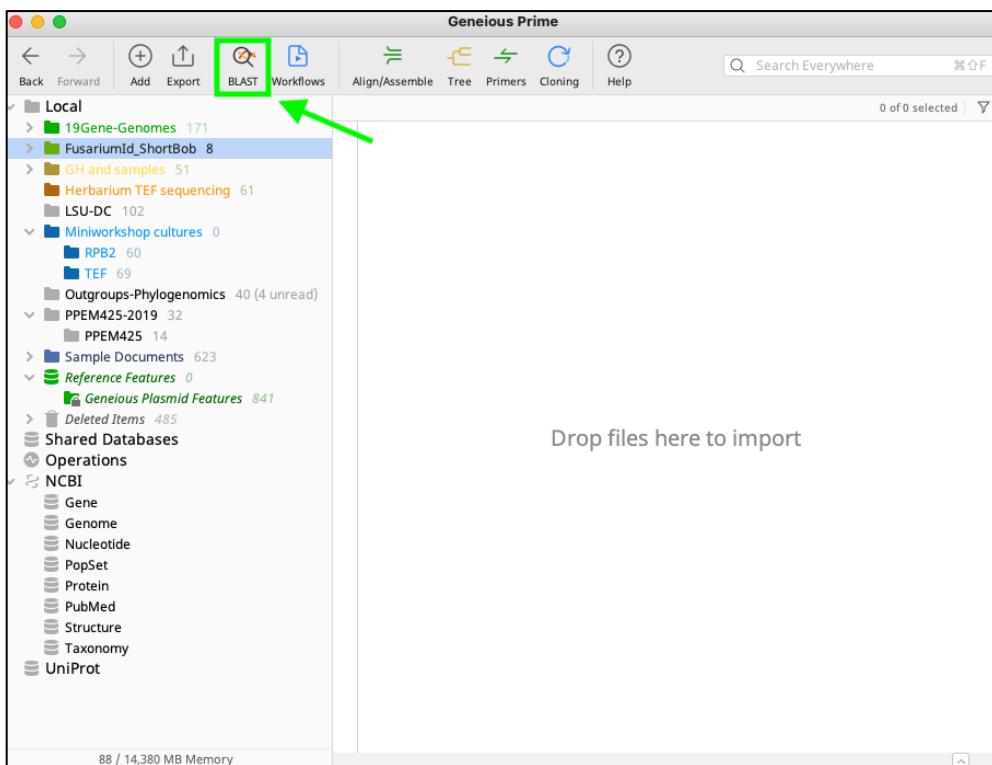


**14.** On the “Add BLAST Database” screen, click on **Ok**



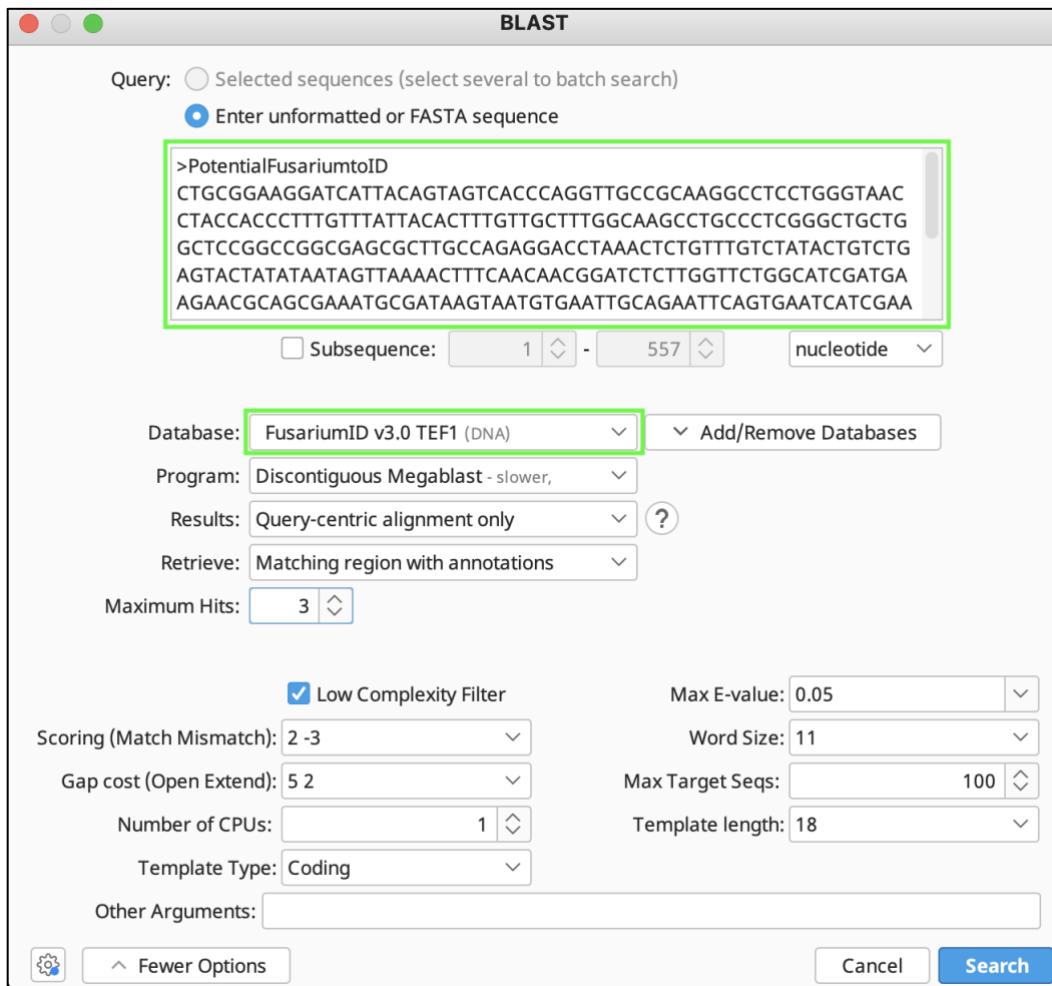
### Running a BLAST search in Geneious

**15.** On the top menu, click on **BLAST**



**16.** Paste the sequence(s) you are trying to identify in the corresponding box under **Query** and select the respective database to BLAST against from the dropdown menu under **Database**.

- a. If you wish to BLAST a batch of sequences at once you can paste them one after the other in FASTA format in the respective box.



**Note:** there are options for Geneious to perform different types of BLAST searches (blastn, megablast, discontiguous megablast, etc) and options to display the results (hit table, query-centric alignment, bin into has hit vs. no hit).

This tutorial document was created by Terry J. Torres-Cruz with valuable input from Briana Whitaker, Hye-Seon Kim, Kerry O'Donnell, Imane Laraba, Chyanna McGee, Tania Estrada-Rodríguez, Emma Wallace and Robert Proctor. We thank Kiana Burtle for providing suggestions and screenshots used in this document.