BLAST EXT

V1.1

OPERATIONS MANUAL

Table of Contents

Introduction

BLASTEXT is a web browser extension developed exclusively for use within the Del Castillo Lab for facilitating a user in the recording of results from NCBI BLAST analyses by scraping and organizing information for export to a spreadsheet. BLASTEXT is purely a temporary intermediate databank between NCBI BLAST and a spreadsheet. It is not intended to perform analyses or store user data permanently. Likewise, it cannot create or record data other than what is contained within the NCBI BLAST Webpage and NCBI Entrez API.

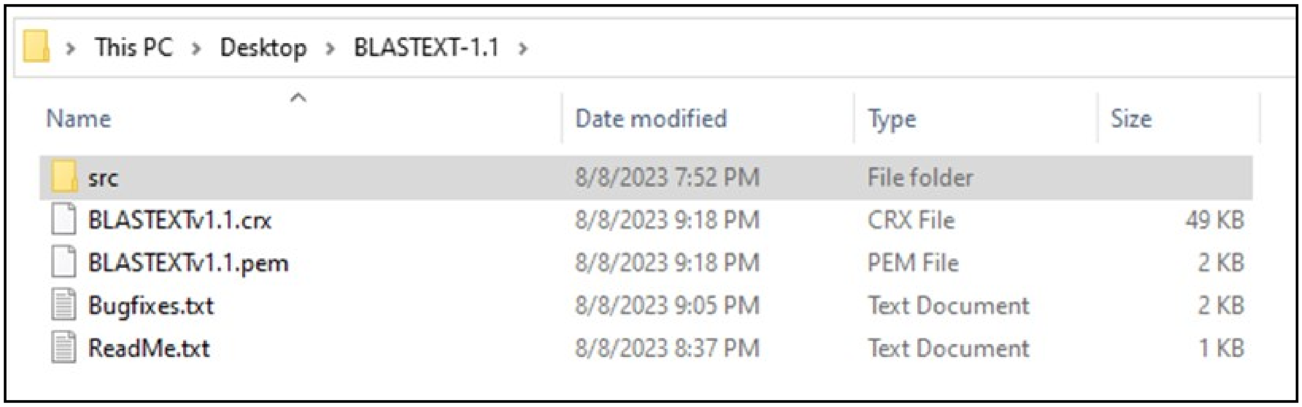
This documentation is for BLASTEXT V1.1. Last updated 8/9/23.

Installation

BLASTEXT is confirmed to work with Google Chrome and Microsoft Edge web browsers. Compatibility with Firefox is unverified at this moment. Due to security restrictions in Chromium-based browsers, BLASTEXT may only be installed and distributed as an unpacked folder.

Installation- Google Chrome

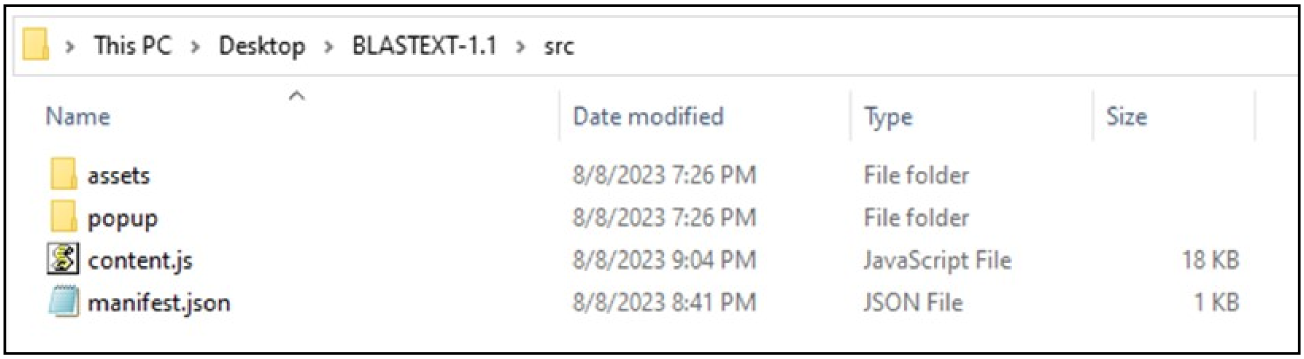
1. Download the zip folder containing the current version of BLASTEXT.
2. Right click the zip folder and extract the contents to an area within the local disk such as the desktop.
3. Open the extracted folder to confirm its contents. The folders should look similar to figure 1.1 and 1.2. Make note of the location of these folders.



› This PC › Desktop › BLASTEXT-1.1 ›

Name	Date modified	Type	Size
src	8/8/2023 7:52 PM	File folder	
BLASTEXTv1.1.crx	8/8/2023 9:18 PM	CRX File	49 KB
BLASTEXTv1.1.pem	8/8/2023 9:18 PM	PEM File	2 KB
Bugfixes.txt	8/8/2023 9:05 PM	Text Document	2 KB
ReadMe.txt	8/8/2023 8:37 PM	Text Document	1 KB

Figure 1.1: BLASTEXT root folder (as of version 1.1)



› This PC › Desktop › BLASTEXT-1.1 › src

Name	Date modified	Type	Size
assets	8/8/2023 7:26 PM	File folder	
popup	8/8/2023 7:26 PM	File folder	
content.js	8/8/2023 9:04 PM	JavaScript File	18 KB
manifest.json	8/8/2023 8:41 PM	JSON File	1 KB

Figure 1.2: BLASTEXT src folder (as of version 1.1)

4. Open Google Chrome and click the three-dot menu in the upper-right corner. Hover over **Extensions**, then select **Extension Manager** (Figure 1.3).

   Alternatively, typing:
   chrome://extensions/
   into the address bar will also bring you to the extension manager.

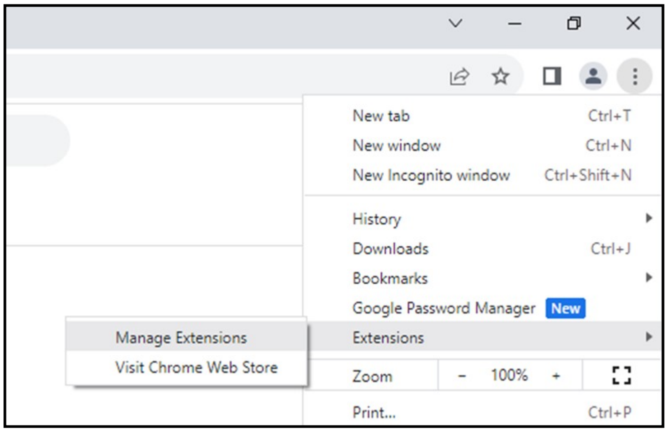


Figure 1.3: Opening Chrome extension manager

5. Click the toggle in the upper right corner of the extension manager labelled **Developer Mode** to enable it (Figure 1.4). A new option for Load unpacked should appear.

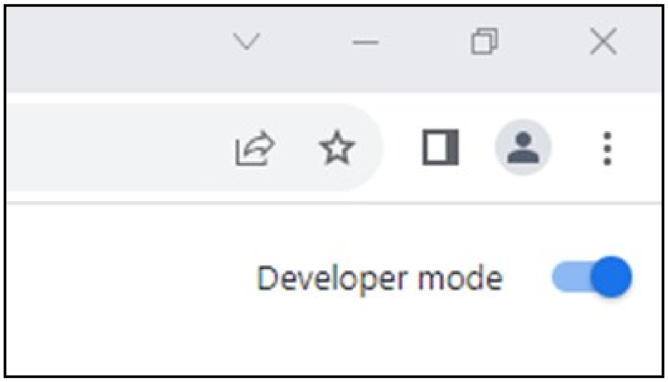


Figure 1.4: Enabled Developer mode toggle.

6. Select the Load Unpacked button (Figure 1.5).

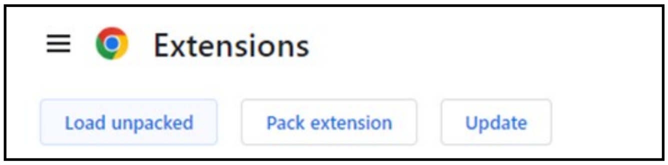
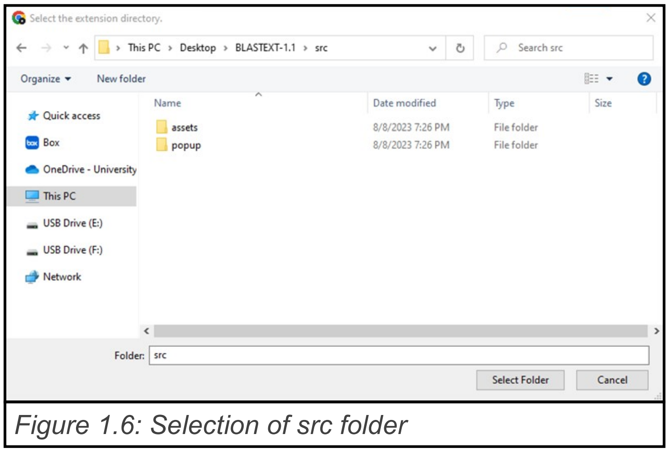


Figure 1.5: Load unpacked button

7. Navigate to the extracted BLASTEXT folder you opened earlier in step 3 click the folder titled src to open it. The folder prompt should now read **src** (Figure 1.6).

8. Click the **Select Folder** button. The extension should load successfully.

   If loading fails, repeat step 6 and 7 and confirm that the folder you selected was the **src** folder containing the file **manifest. json.**

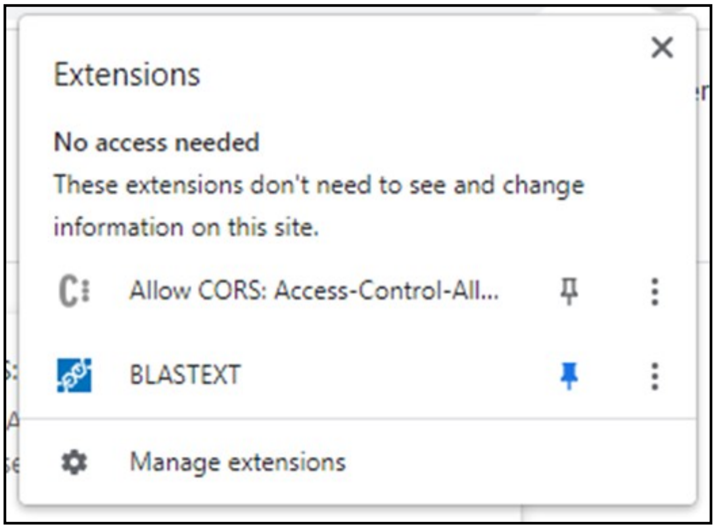


Name	Date modified	Type	Size
assets	8/8/2023 7:26 PM	File folder	
popup	8/8/2023 7:26 PM	File folder	

Folder: src

Figure 1.6: Selection of src folder

9. Pinning the extension to the chrome extension bar is recommended. To do this, click the extension icon (shaped like a puzzle piece) in the upper right corner. In the menu click the pin next to BLASTEXT to pin it.



Extensions

No access needed

These extensions don't need to see and change information on this site.

Figure 1.7: Pinning the extension

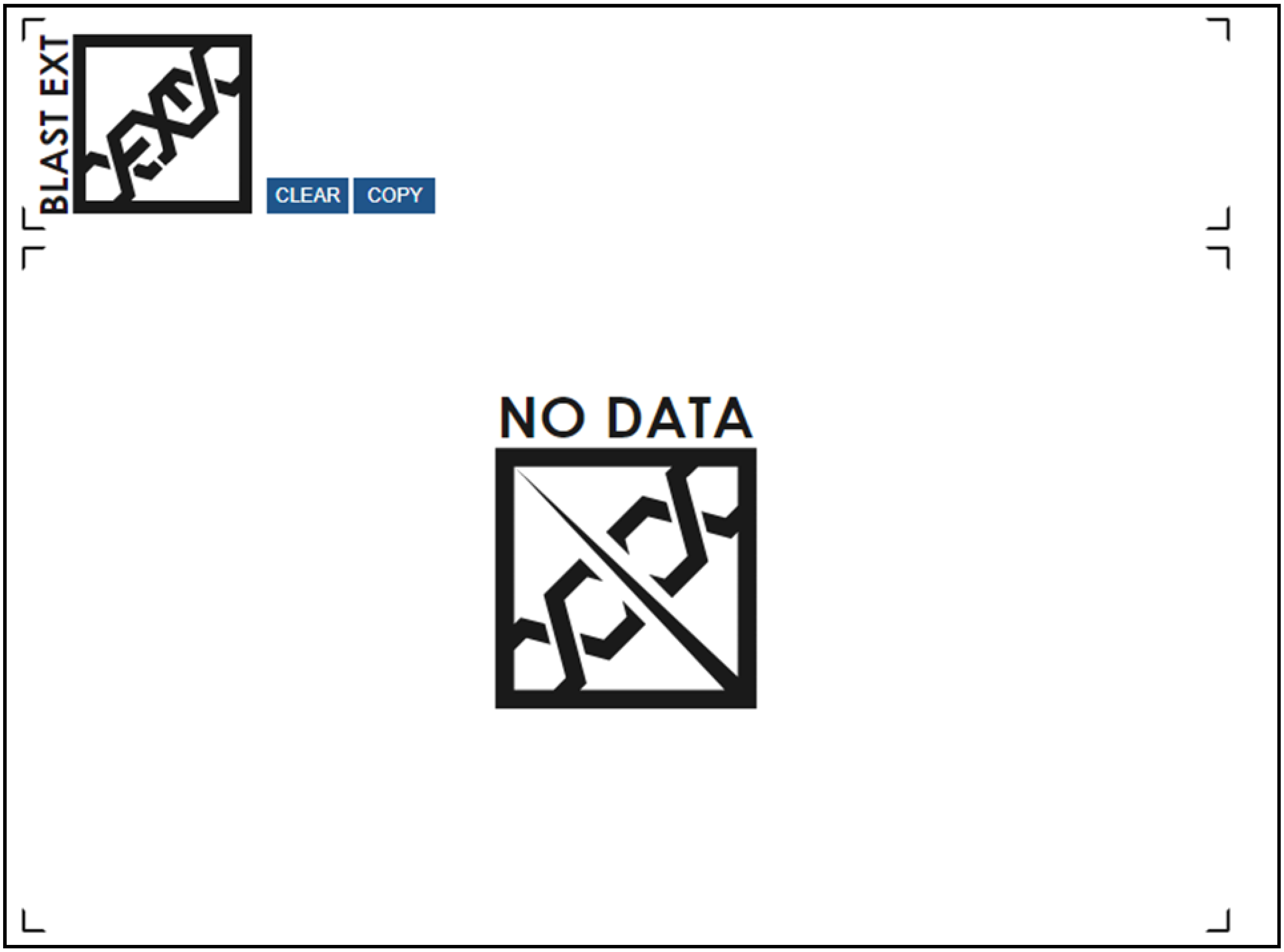


Figure 1.8: BLASTEXT UI

Workflow

BLASTEXT runs automatically in the background any time an NCBI BLAST associated webpage is open and active. This section will use a set of sample data included in the extension files to demonstrate how to operate the extension.

Workflow- Sample Data

1. Navigate to the NCBI BLAST website (https://blast.ncbi.nlm.nih.gov/Blast.cgi)
2. Select Nucleotide BLAST
3. Open the **sample files** folder in the BLASTEXT directory
4. Click and drag the **BLASTEXT_SampleFasta.fa** file from the folder into the NCBI BLAST webpage. Drop the file ON TOP of the **Choose File** button.
5. Press the **BLAST** button on the bottom of the webpage. Wait for the analysis to conclude.

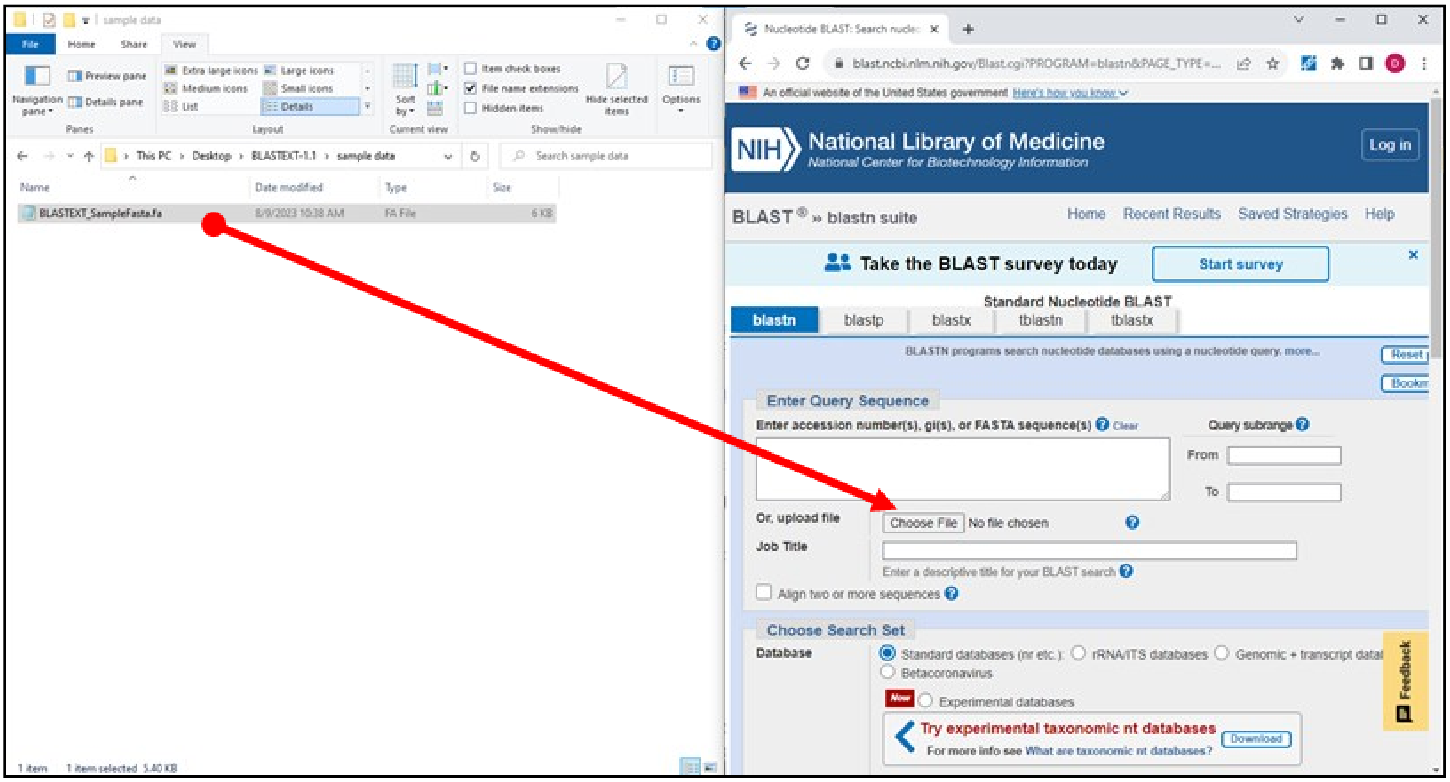
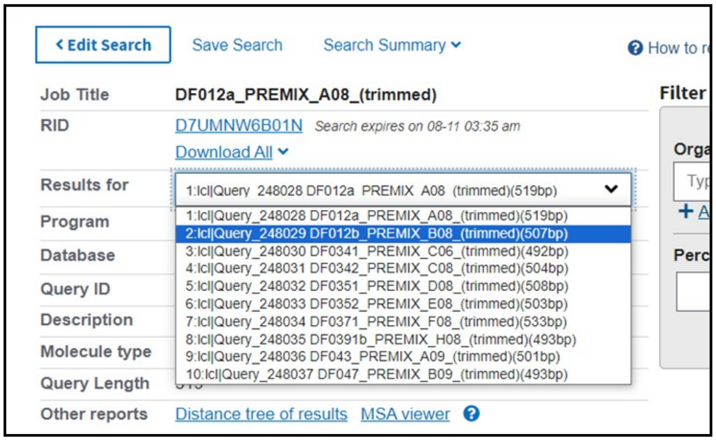

sample data
File
Home
Share
View
This PC › Desktop › BLASTEXT-1.1 › sample data
Name
Date modified
Type
Size
BLASTEXT_SampleFasta.fa
8/9/2023 10:38 AM
FA File
6 KB
1 item 1 item selected 5.40 KB
Nucleotide BLAST: Search nucle
blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=...
An official website of the United States government Here's how you know
NIH National Library of Medicine
National Center for Biotechnology Information
Log in
BLAST® » blastn suite
Home
Recent Results
Saved Strategies
Help
Take the BLAST survey today
Start survey
Standard Nucleotide BLAST
blastn
blastp
blastx
tblastn
tblastx
BLASTN programs search nucleotide databases using a nucleotide query. more...
Enter Query Sequence
Enter accession number(s), gi(s), or FASTA sequence(s) Clear
Query subrange
From
To
Or, upload file
Choose File No file chosen
Job Title
Enter a descriptive title for your BLAST search
Align two or more sequences
Choose Search Set
Database
Standard databases (nr etc.)
rRNA/ITS databases
Genomic + transcript datal
Betacoronavirus
New
Experimental databases
Try experimental taxonomic nt databases
For more info see What are taxonomic nt databases?
Download
Feedback

Figure 2.1: Loading a FASTA into the NCBI BLAST Interface. Ensure the FASTA file being loaded is dropped directly on top of the Choose File Button.

6. Click on the BLASTEXT extension icon. A list of all the samples should be present in the popup that appears, each with [Null] listed for details.

7. Select the first sample from the result drop down.



Job Title DF012a_PREMIX_A08_(trimmed)

RID D7UMNW6B01N *Search expires on 08-11 03:35 am*

Results for 1:lcl|Query 248028 DF012a PREMIX A08 (trimmed)(519bp)

1:lcl|Query_248028 DF012a_PREMIX_A08_(trimmed)(519bp)
2:lcl|Query_248029 DF012b_PREMIX_B08_(trimmed)(507bp)
3:lcl|Query_248030 DF0341_PREMIX_C06_(trimmed)(492bp)
4:lcl|Query_248031 DF0342_PREMIX_C08_(trimmed)(504bp)
5:lcl|Query_248032 DF0351_PREMIX_D08_(trimmed)(508bp)
6:lcl|Query_248033 DF0352_PREMIX_E08_(trimmed)(503bp)
7:lcl|Query_248034 DF0371_PREMIX_F08_(trimmed)(533bp)
8:lcl|Query_248035 DF0391b_PREMIX_H08_(trimmed)(493bp)
9:lcl|Query_248036 DF043_PREMIX_A09_(trimmed)(501bp)
10:lcl|Query_248037 DF047_PREMIX_B09_(trimmed)(493bp)

Program

Database

Query ID

Description

Molecule type

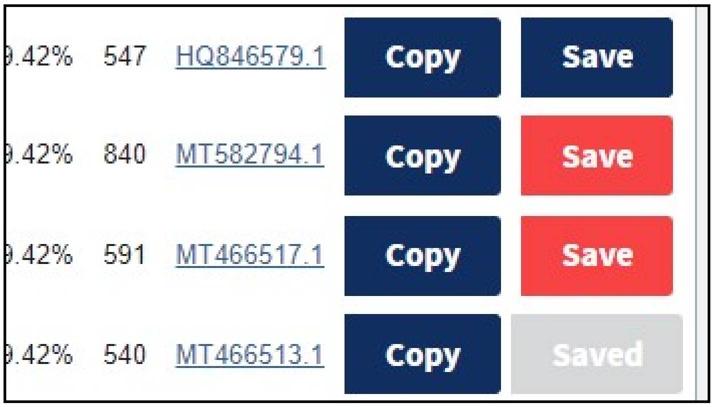
Query Length

Other reports

Figure 2.2: NCBI BLAST sample dropdown list.

8. Select the **SAVE** button next to the result you would like to save. Note that after a short delay, some of the **SAVE** buttons will turn red. This indicates that the associated result is NOT from a published article. **SAVE** buttons that remain blue are from published articles (Figure 2.3). Note that BLASTEXT will **ONLY CHECK THE FIRST 50 RESULTS**. This number can be adjusted by modifying the `pubThreshold` constant in content.js.

   Since this feature is dependent on the NCBI Entrez API, it's performance may be unreliable. Bacterial samples have notable lag. In some cases, it may be better to check each entry manually.



.42%	547	HQ846579.1	Copy	Save
.42%	840	MT582794.1	Copy	Save
.42%	591	MT466517.1	Copy	Save
.42%	540	MT466513.1	Copy	Saved

Figure 2.3: Various save buttons, top to bottom: Published and not saved entry, two unpublished and not saved entries, saved entry.

9. Click on the BLASTEXT extension icon, confirm the details have been recorded correctly (Figure 2.4).

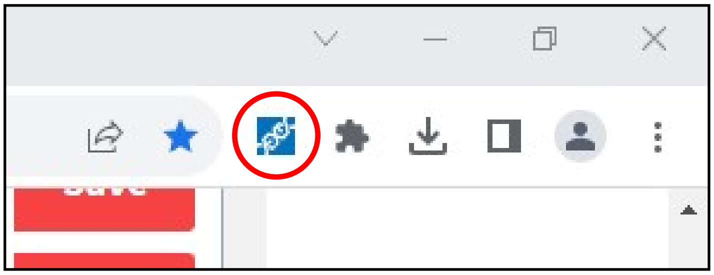


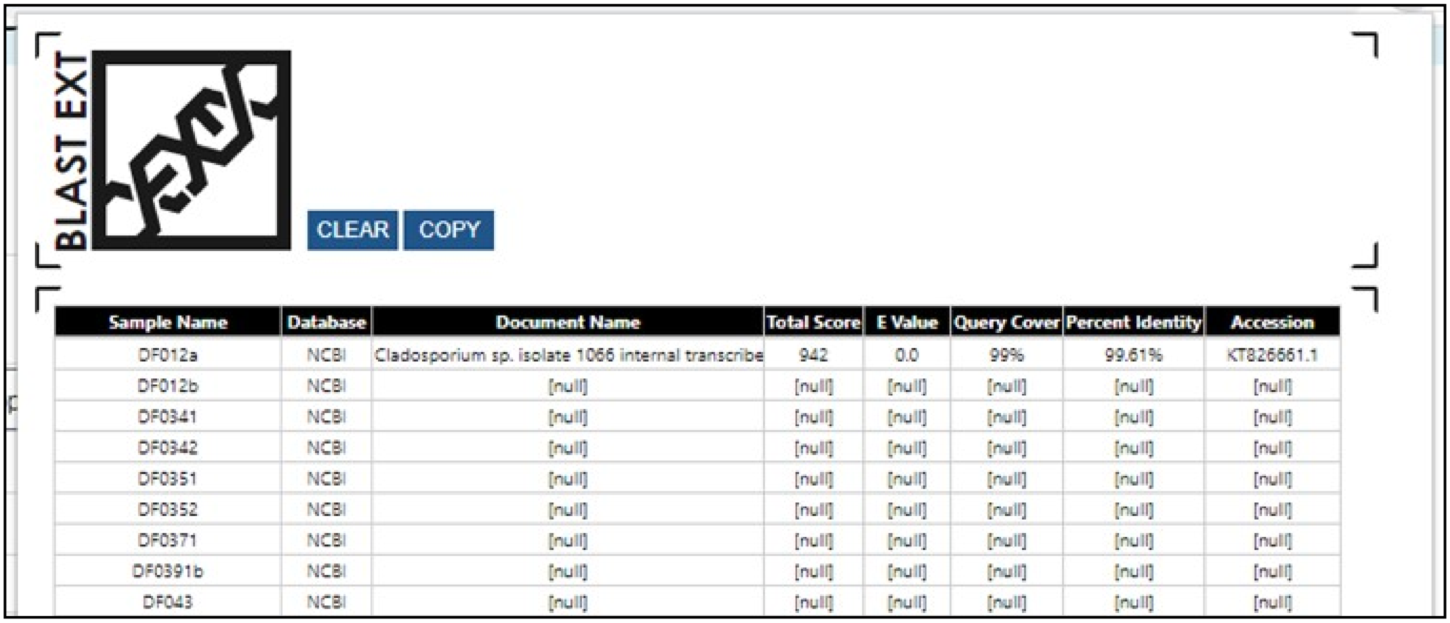
Figure 2.4: BLASTEXT extension icon.

10. Repeat steps 7-9 for the remaining samples.

11. Once a result for each sample has been saved, click on the BLASTEXT icon again.

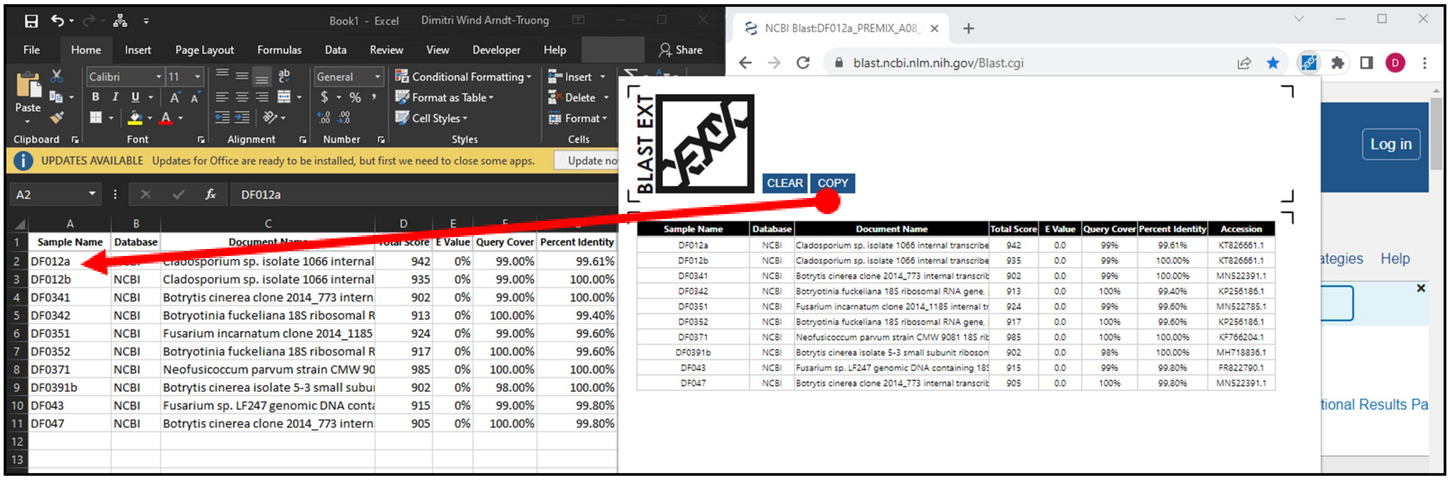
12. Click the **Copy** button.

13. Open the relevant sequence analysis log and paste (Control-V) the finished results into the sheet (Figure 2.6).



Sample Name	Database	Document Name	Total Score	E Value	Query Cover	Percent Identity	Accession
DF012a	NCBI	Cladosporium sp. isolate 1066 internal transcribe	942	0.0	99%	99.61%	KT826661.1
DF012b	NCBI	[null]	[null]	[null]	[null]	[null]	[null]
DF0341	NCBI	[null]	[null]	[null]	[null]	[null]	[null]
DF0342	NCBI	[null]	[null]	[null]	[null]	[null]	[null]
DF0351	NCBI	[null]	[null]	[null]	[null]	[null]	[null]
DF0352	NCBI	[null]	[null]	[null]	[null]	[null]	[null]
DF0371	NCBI	[null]	[null]	[null]	[null]	[null]	[null]
DF0391b	NCBI	[null]	[null]	[null]	[null]	[null]	[null]
DF043	NCBI	[null]	[null]	[null]	[null]	[null]	[null]

Figure 2.5: BLASTEXT UI with an entry recorded.

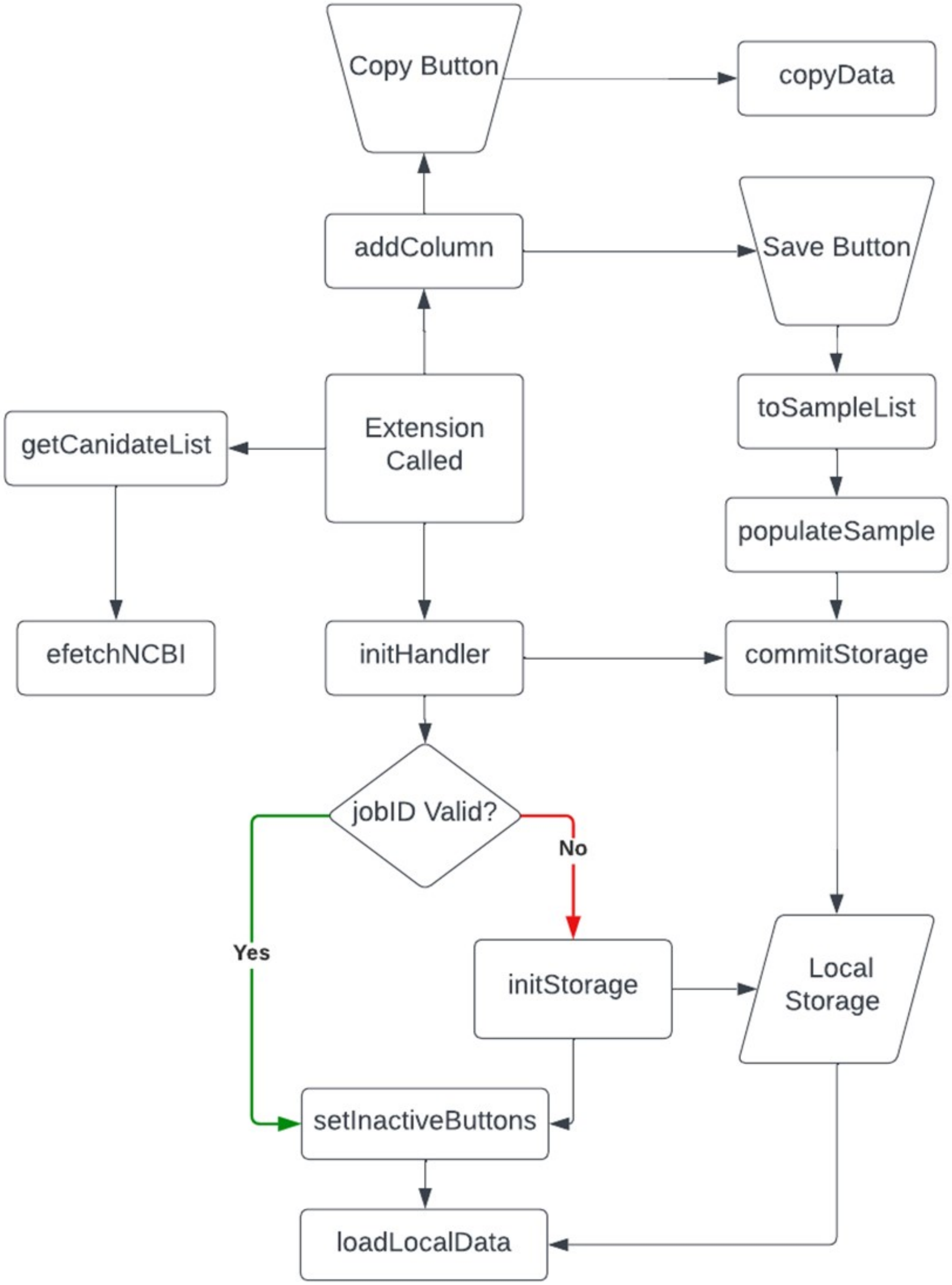


	A	B	C	D	E	F	G
1	Sample Name	Database	Document Name	Total Score	E Value	Query Cover	Percent Identity
2	DF012a	NCBI	Cladosporium sp. isolate 1066 internal	942	0%	99.00%	99.61%
3	DF012b	NCBI	Cladosporium sp. isolate 1066 internal	935	0%	99.00%	100.00%
4	DF0341	NCBI	Botrytis cinerea clone 2014_773 intern	902	0%	99.00%	100.00%
5	DF0342	NCBI	Botryotinia fuckeliana 18S ribosomal R	913	0%	100.00%	99.40%
6	DF0351	NCBI	Fusarium incarnatum clone 2014_1185	924	0%	99.00%	99.60%
7	DF0352	NCBI	Botryotinia fuckeliana 18S ribosomal R	917	0%	100.00%	99.60%
8	DF0371	NCBI	Neofusicoccum parvum strain CMW 90	985	0%	100.00%	100.00%
9	DF0391b	NCBI	Botrytis cinerea isolate 5-3 small subu	902	0%	98.00%	100.00%
10	DF043	NCBI	Fusarium sp. LF247 genomic DNA conta	915	0%	99.00%	99.80%
11	DF047	NCBI	Botrytis cinerea clone 2014_773 intern	905	0%	100.00%	99.80%
12							
13							

Sample Name	Database	Document Name	Total Score	E Value	Query Cover	Percent Identity	Accession
DF012a	NCBI	Cladosporium sp. isolate 1066 internal transcribe	942	0.0	99%	99.61%	KT826661.1
DF012b	NCBI	Cladosporium sp. isolate 1066 internal transcribe	935	0.0	99%	100.00%	KT826661.1
DF0341	NCBI	Botrytis cinerea clone 2014_773 internal transcrib	902	0.0	99%	100.00%	MN522391.1
DF0342	NCBI	Botryotinia fuckeliana 18S ribosomal RNA gene,	913	0.0	100%	99.40%	KP256186.1
DF0351	NCBI	Fusarium incarnatum clone 2014_1185 internal tr	924	0.0	99%	99.60%	MN522785.1
DF0352	NCBI	Botryotinia fuckeliana 18S ribosomal RNA gene,	917	0.0	100%	99.60%	KP256186.1
DF0371	NCBI	Neofusicoccum parvum strain CMW 9081 18S rib	985	0.0	100%	100.00%	KF766204.1
DF0391b	NCBI	Botrytis cinerea isolate 5-3 small subunit riboson	902	0.0	98%	100.00%	MH718836.1
DF043	NCBI	Fusarium sp. LF247 genomic DNA containing 18S	915	0.0	99%	99.80%	FR822790.1
DF047	NCBI	Botrytis cinerea clone 2014_773 internal transcrib	905	0.0	100%	99.80%	MN522391.1

Figure 2.6: Exporting data from BLASTEXT to Excel using the copy button

Code Schematic



Bugs and Potential Improvements

Identical JobIDs

BLASTEXT tells apart old and new requests by comparing a combination of the sample list length and the job name of the BLAST request being run. If this happens to be identical between two separate runs, as may be the case if the query is modified but not renamed, BLASTEXT will not identify a change in the JobID and will not refresh it's internal sample array to reflect the changes. In the event of this bug, the sample array can be manually refreshed via the clear button in the BLASTEXT UI. As this bug is relatively difficult to trigger no action has yet been taken against it, however in future updates it is recommended to have the jobID getter access the search's RID instead of the job title

Premature truncation of sample names

BLASTEXT assigns sample names by removing the latter detail portion of the sequence name that is generated by Eurofins during processing. Since this portion is consistently preceded by a single underscore, BLASTEXT will simply detect the first underscore in the sample name and truncate everything after it. This strategy poses an issue if spaces or underscores were in the sample names submitted to Eurofins as this will result in Eurofins sending back result files containing additional underscores before the underscore that BLASTEXT relies on. While this is easily avoided by removing spaces and underscores from sample names before sending to Eurofins, one could also search for the underscore starting from the end of the string instead of the start of the string, thereby bypassing any variability in the sample name itself.

Slow NCBI Entrez API Response

BLASTEXT relies on the NCBI Entrez API to detect if a given result has been published or not based on its accession number, however the response time of this sever can vary drastically from seconds to several minutes. Slow-down is most noticeable with bacterial accessions which can take up to 15 minutes to respond. Current attempts to mitigate this issue by limiting the number of entries requested per sample have been mostly unsuccessful, with a few accessions taking roughly the same time as all one hundred at once. Further testing into request phrasing and perhaps smarter requests (i.e. only sending one representative accession if many seem to be from the same paper, as indicated by very similar accession numbers) may help alleviate this issue.

General Coding Practices

BLASTEXT was produced by an amateur programmer and did not have an opportunity to undergo critical analysis of it's programming practices and code structure. As such, there are likely many portions of code that are poorly written, redundant, or just unconventional. While the program is not resource intensive enough to justify heavy optimization, a review of the code by a programmer well-versed in web development is needed.

Notes