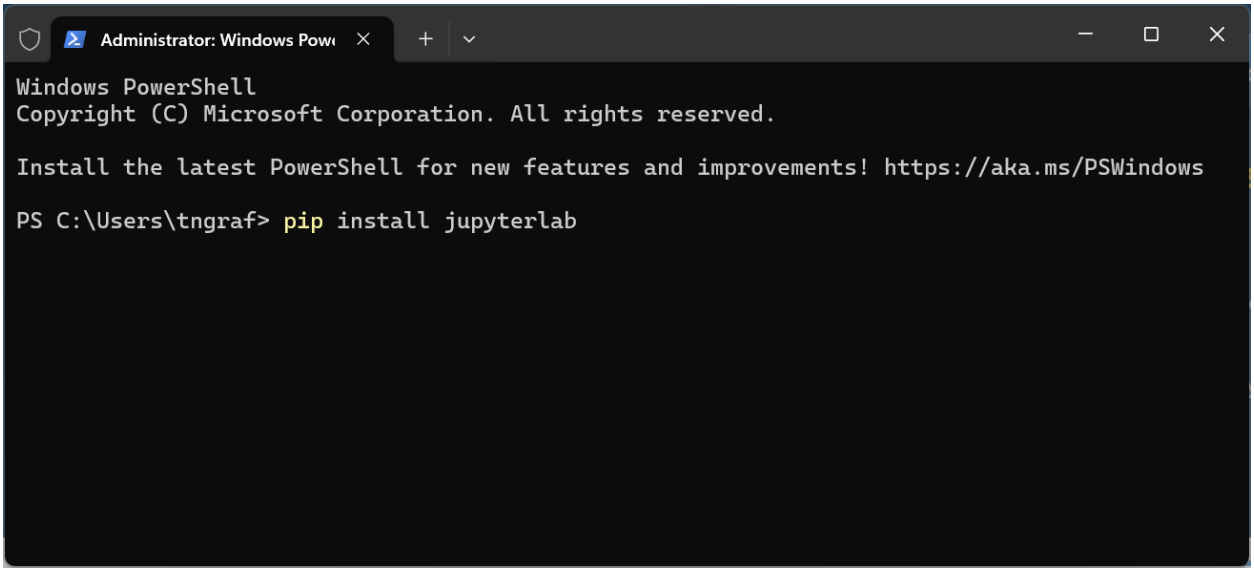
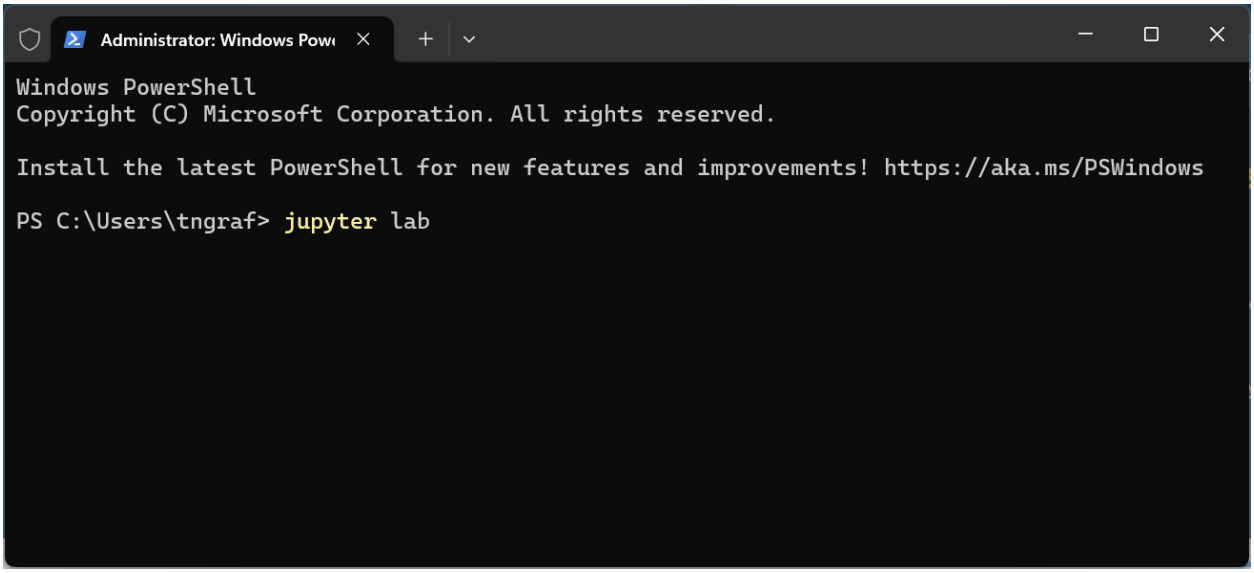
1) Install Python as an Admin, https://www.python.org/downloads/
2) In a “Terminal (Admin)” window, install Jupyter lab via a PIP install

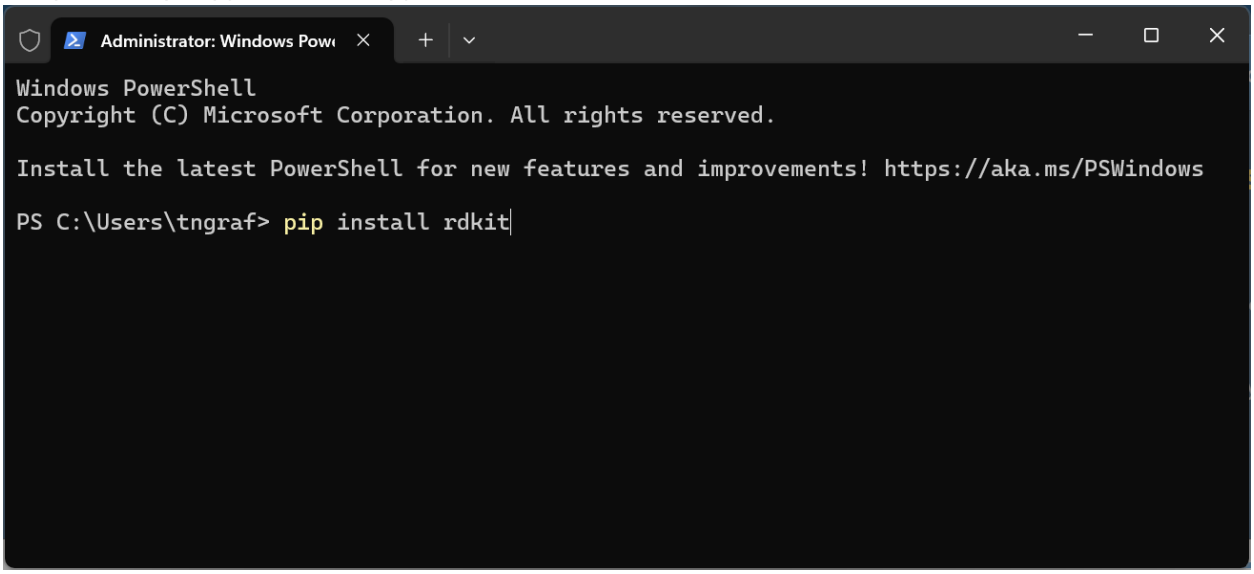


3) In a “Terminal (Admin)” window, run Jupyter lab

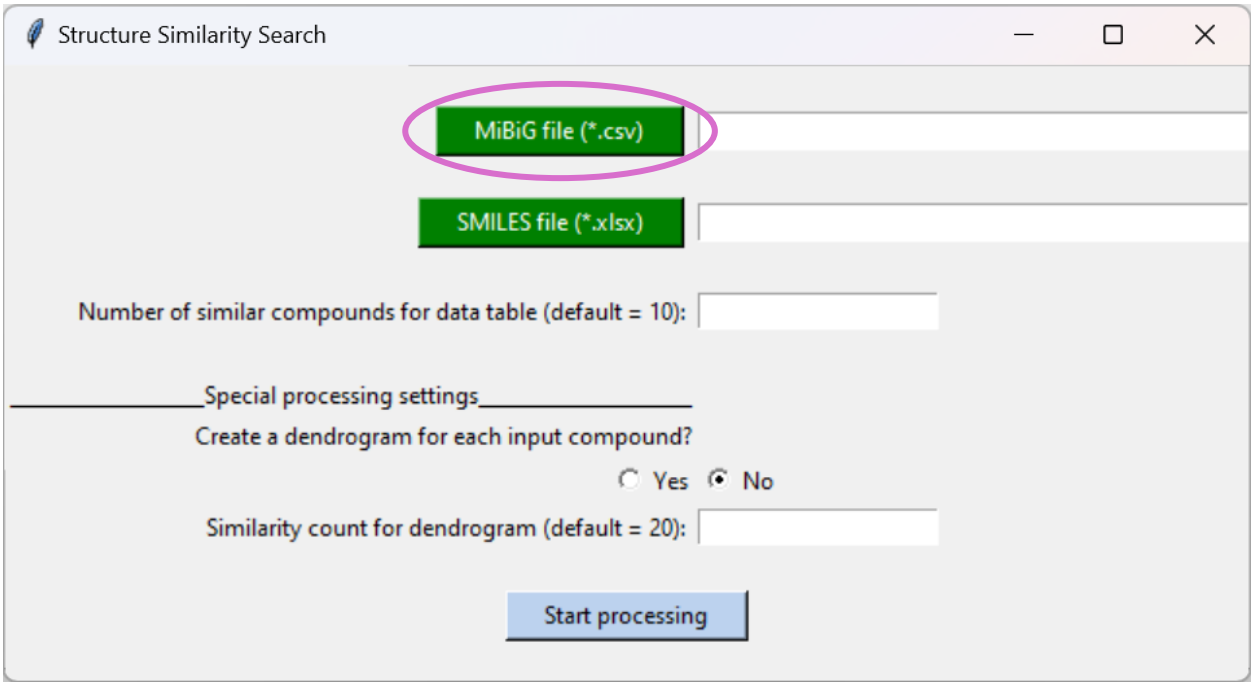


4) Open the Python script “Similarity with GUI-v2.0b.ipynb” as a notebook from within Jupyter lab.
5) In a new “Terminal (Admin)”, run a PIP install of the following required dependencies by typing “pip install ____” where the underline is one of the following: rdkit, pandas, scipy,

matplotlib, openpyxl, and numpy.

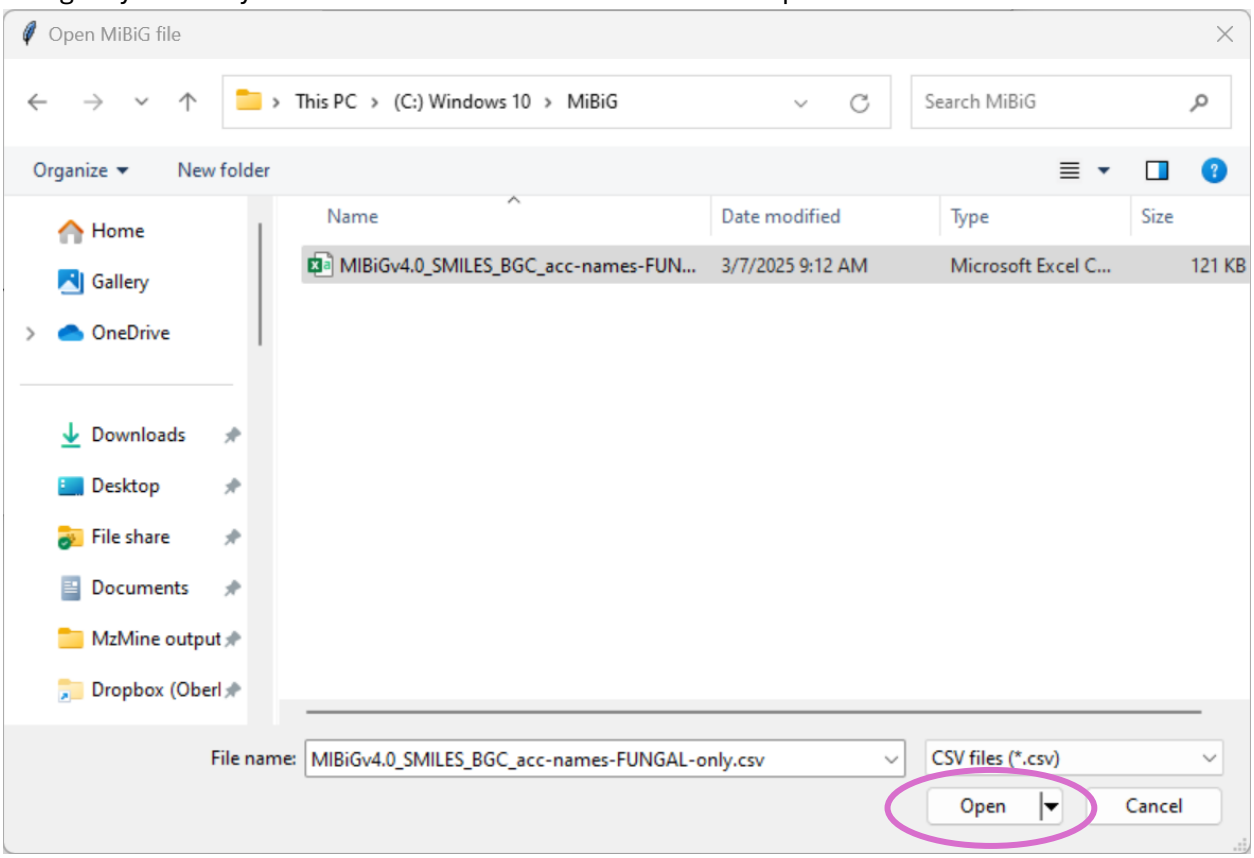


6) While selected anywhere in the python script code (within Jupyter lab), press “Ctrl + Enter” to run the code. This should bring up the Graphic user interface (GUI) for the program as a separate window/program in Windows.

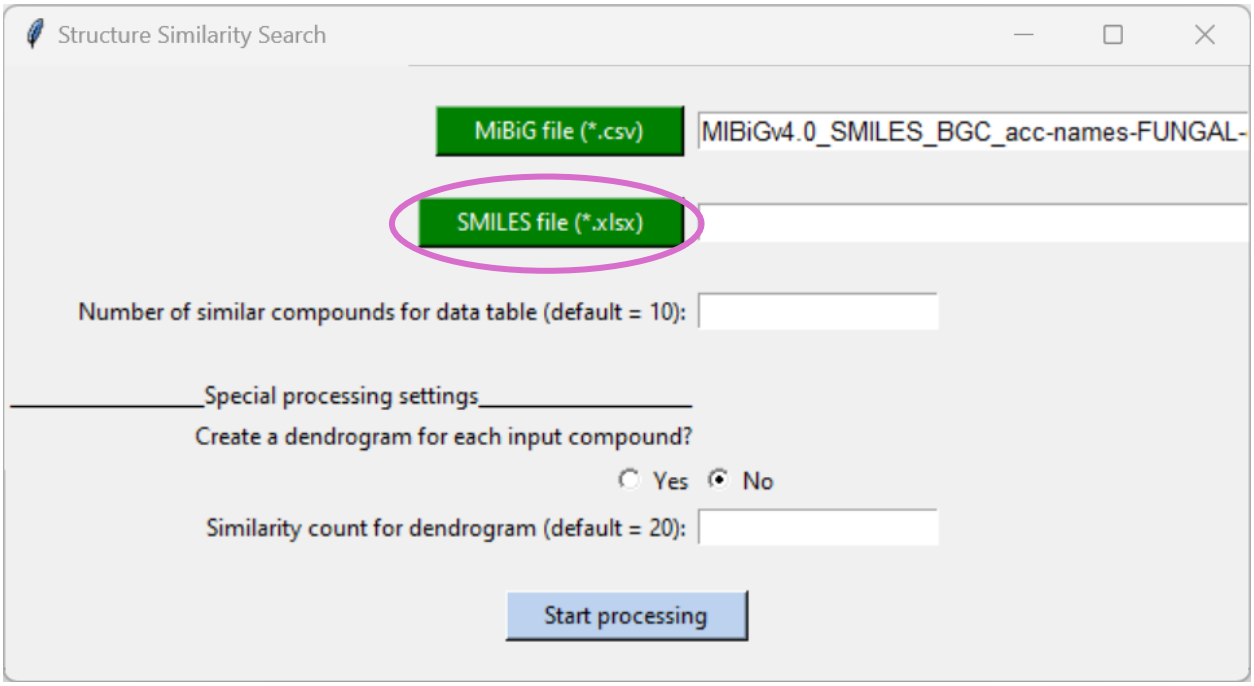


7) Click on the “MiBiG file (*.csv)” button to select the file of compounds with known BGCs.

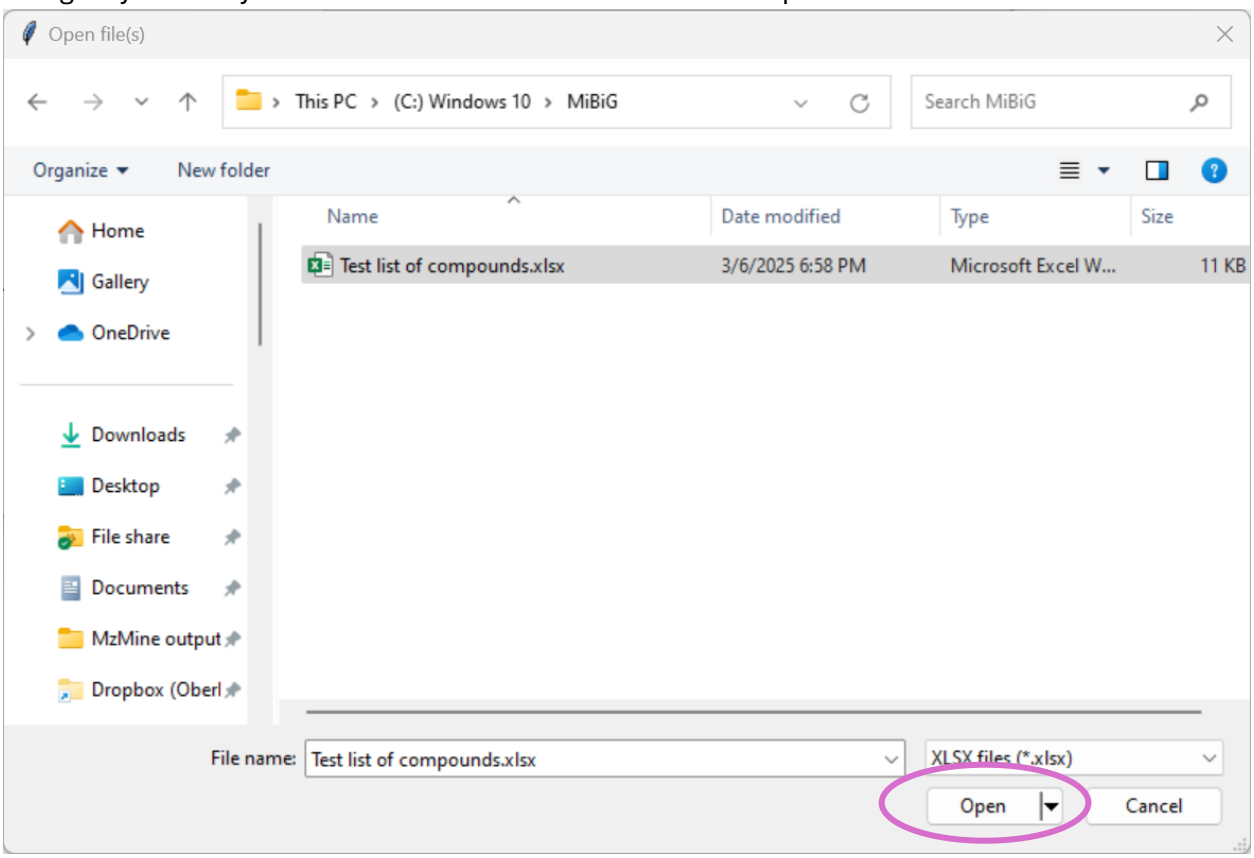
8) Navigate your file system to select the desired file and click “Open”.



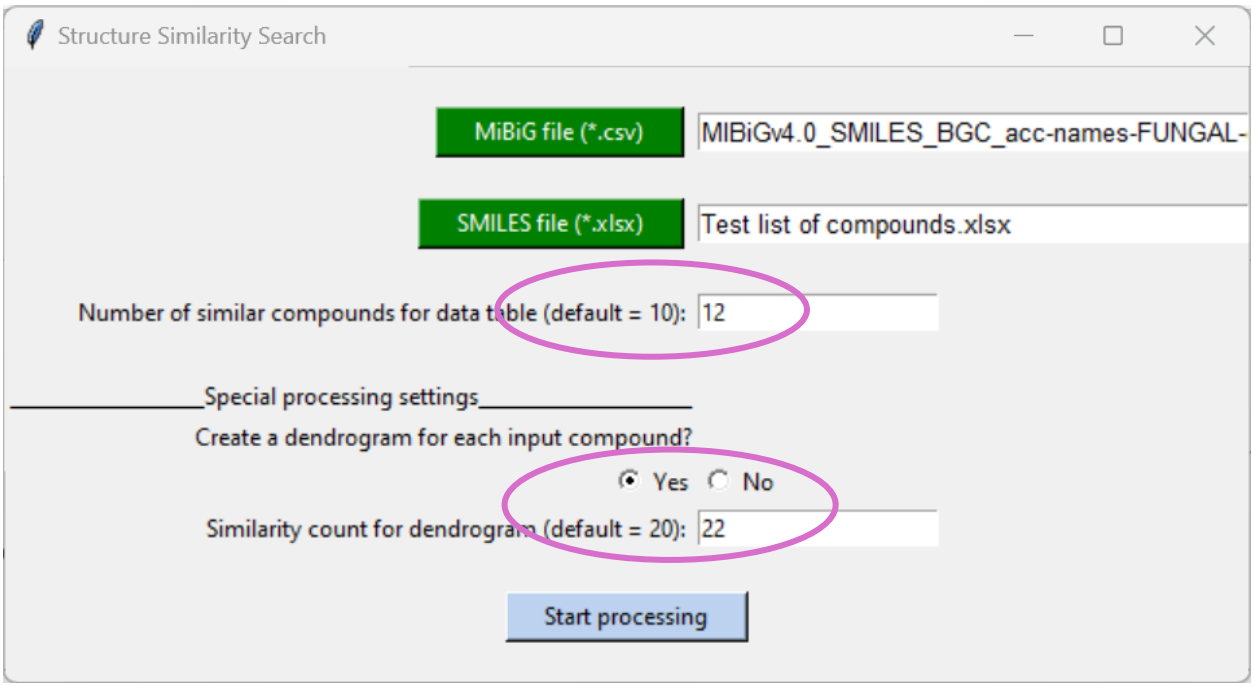
9) From the main GUI window, click on “SMILES file(*.xlsx)” button to select the Excel file containing your list of compounds you wish to search for similarities for.



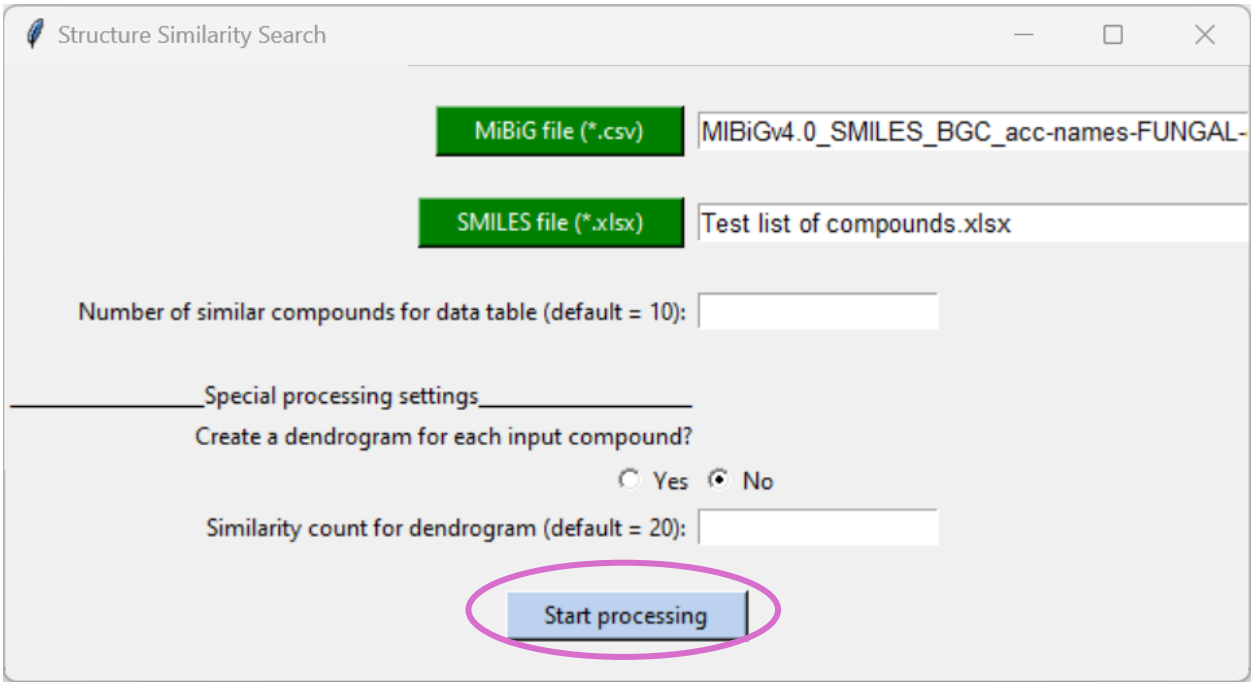
10) Navigate your file system to select the desired file and click “Open”.



11) OPTIONAL: At this point you can change some of the default settings. If you leave the boxes empty, the default value will be used. You can increase the number of similar compounds listed for each compound you input (default value is 10). You can also have similarity dendrograms created for each compound you are searching, selecting the “yes” radio button, and you can define how many similar compounds should be included in the dendrogram (default value is 20). Below is an example of changing the defaults and turning on dendrogram generation:

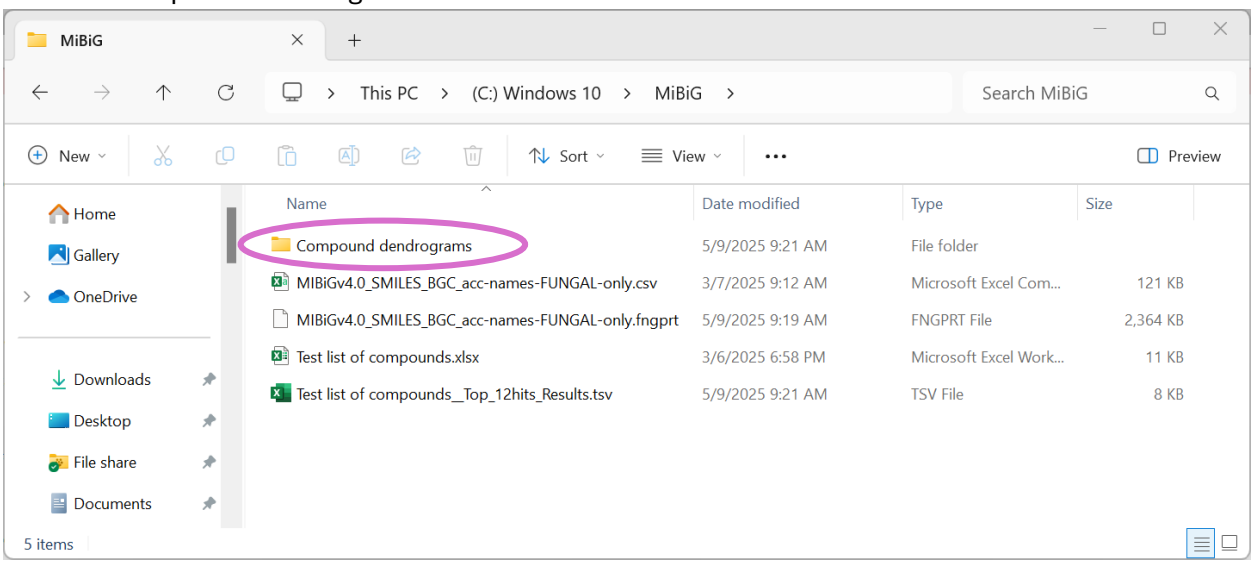


12) You can click the "Start processing" button (without making any further selections) and the Python script will start the analysis. This process will take approximately 60 seconds to complete.



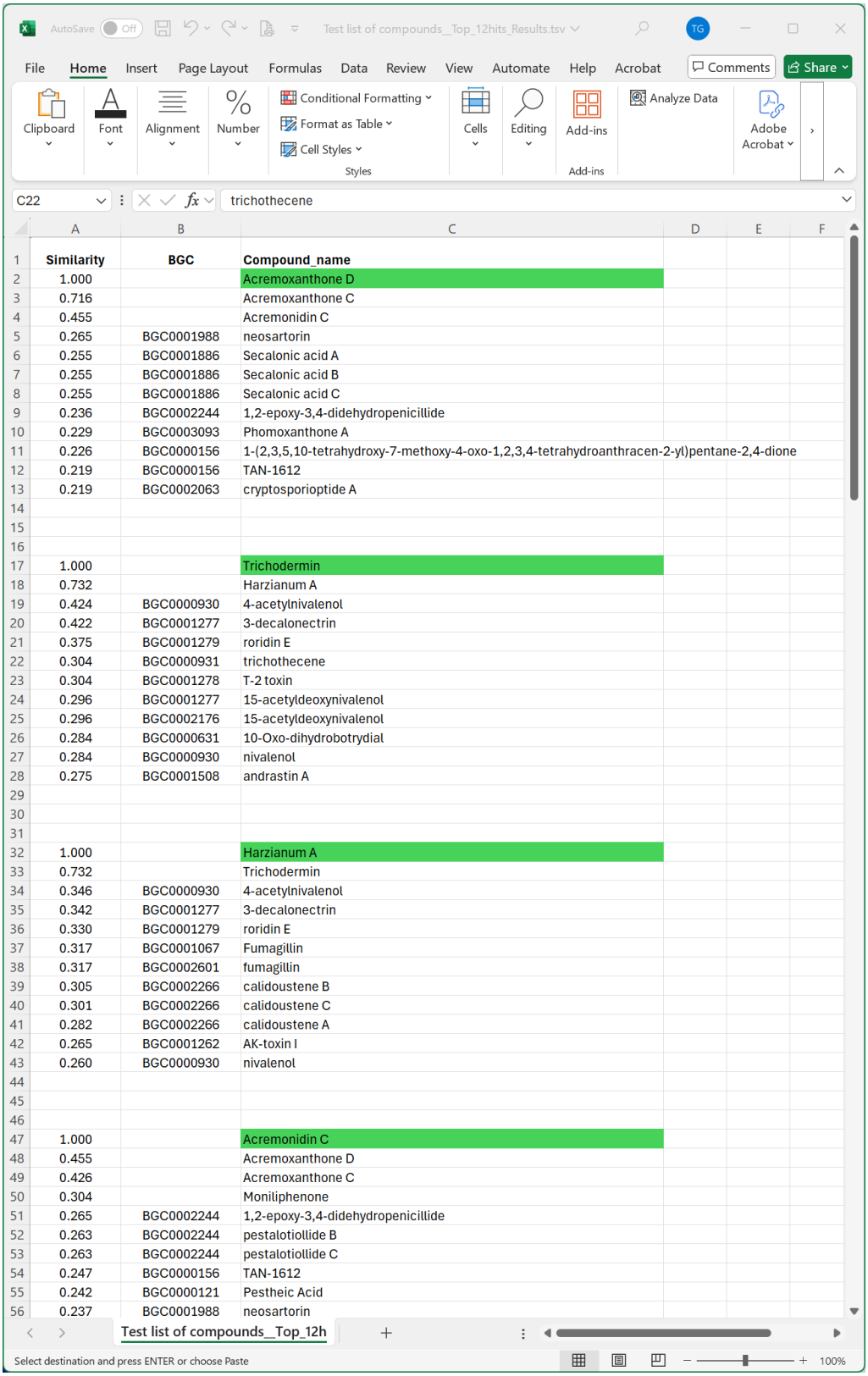
13) All output files will be placed in the folder where the list of compounds to be searched for was located. Dendrograms, if selected, will be placed as individual files in a sub-folder

named “Compound dendrograms”.



14) Files generated include “MIBiGv4.0_SMILES_BGC_acc-names-FUNGAL-only.fngprt” which has the raw fingerprints for the compounds in the MiBiG file (this is a diagnostics file and can be deleted). The “Test list of compounds__Top_12hits_Results.tsv” file has the results for the compounds searched for (in this case, the file has the top 12 hits for each). The folder containing PDF files of the dendrograms for each compound is also present (if that feature was turned on in the GUI).
15) To format the output data in an easy-to-read form, open the Excel template file “Results_format_template.xltx”, select all (Ctrl + A), copy (Ctrl + C). In the results file, right click in cell A1 and paste formatting only (paintbrush icon).

16) Note: compounds in you list to be searched for will show up as hits in the results file. They will be the entries without a BGC in column B of the Excel.



	A	B	C
1	**Similarity**	**BGC**	**Compound_name**
2	1.000		Acremoxanthone D
3	0.716		Acremoxanthone C
4	0.455		Acremonidin C
5	0.265	BGC0001988	neosartorin
6	0.255	BGC0001886	Secalonic acid A
7	0.255	BGC0001886	Secalonic acid B
8	0.255	BGC0001886	Secalonic acid C
9	0.236	BGC0002244	1,2-epoxy-3,4-didehydropenicillide
10	0.229	BGC0003093	Phomoxanthone A
11	0.226	BGC0000156	1-(2,3,5,10-tetrahydroxy-7-methoxy-4-oxo-1,2,3,4-tetrahydroanthracen-2-yl)pentane-2,4-dione
12	0.219	BGC0000156	TAN-1612
13	0.219	BGC0002063	cryptosporioptide A
14			
15			
16			
17	1.000		Trichodermin
18	0.732		Harzianum A
19	0.424	BGC0000930	4-acetylnivalenol
20	0.422	BGC0001277	3-decalonectrin
21	0.375	BGC0001279	roridin E
22	0.304	BGC0000931	trichothecene
23	0.304	BGC0001278	T-2 toxin
24	0.296	BGC0001277	15-acetyldeoxynivalenol
25	0.296	BGC0002176	15-acetyldeoxynivalenol
26	0.284	BGC0000631	10-Oxo-dihydrobotrydial
27	0.284	BGC0000930	nivalenol
28	0.275	BGC0001508	andrastin A
29			
30			
31			
32	1.000		Harzianum A
33	0.732		Trichodermin
34	0.346	BGC0000930	4-acetylnivalenol
35	0.342	BGC0001277	3-decalonectrin
36	0.330	BGC0001279	roridin E
37	0.317	BGC0001067	Fumagillin
38	0.317	BGC0002601	fumagillin
39	0.305	BGC0002266	calidoustene B
40	0.301	BGC0002266	calidoustene C
41	0.282	BGC0002266	calidoustene A
42	0.265	BGC0001262	AK-toxin I
43	0.260	BGC0000930	nivalenol
44			
45			
46			
47	1.000		Acremonidin C
48	0.455		Acremoxanthone D
49	0.426		Acremoxanthone C
50	0.304		Moniliphenone
51	0.265	BGC0002244	1,2-epoxy-3,4-didehydropenicillide
52	0.263	BGC0002244	pestalotiollide B
53	0.263	BGC0002244	pestalotiollide C
54	0.247	BGC0000156	TAN-1612
55	0.242	BGC0000121	Pestheic Acid
56	0.237	BGC0001988	neosartorin