1. About

MS2Compound v1.0.0 is a user friendly Graphical User Interface (GUI) for the identification of the compounds from MS/MS metabolomics data. The tool features use of custom database for the identification of compounds. The current version is compatible with the customized database prepared using the fragment prediction using CFM-id tool.

The current version of the tool is compatible with Mascot Generic Format (.mgf) files as input.

The tool has been developed and maintained by Center for Systems Biology and Molecular Medicine (http://csbmm.yenepoya.res.in/), Yenepoya Research Centre, Yenepoya (Deemed to be University).

2. Installing MS2Compound

2.1 System requirements

2.1.1 Operating system

Windows 7, 8 and 10

2.1.2 Minimum system configuration

2GB or more RAM

Intel i3 64-bit CPU

2.2 Installation

2.2.1 Dependencies

The tool is independent from all the pre-requisite dependencies.

2.2.2 MS2Compound installation

- Download the application "MS2Compound_Setup.exe" from https://github.com/beherasan/MS2Compound or https://sourceforge.net/projects/ms2compound/
- Click on "MS2Compound_Setup.exe" to install. The default installation directory will be "C:\MS2Compound", user can select the installation directory of choice, however avoid spaces (" ") in the directory path name.
- Please select to create a shortcut of the MS2Compound in desktop for easy access.
- A successful installation will create a directory C:\MS2Compound. This directory will contain the executable "MS2Compound.exe".

3. Start MS2Compound

Click on the shortcut "MS2Compound", created in desktop after successful completion of installation. If the user has not opted for desktop shortcut during installation, the executable can be start by clicking the "MS2Compound.exe" in "C:\MS2Compound" directory.

4. Availability of test dataset

The test dataset for MS2Compound are available at https://github.com/beherasan/MS2Compound/tree/master/TestData

The current version of MS2Compound majorly performs the following five tasks (Link for the test data for each function are provided in parentheses).

a. Create custom MS/MS database using CFM-id prediction models. Please cite CFM-id (https://cfmid.wishartlab.com/) if the user has used this module in MS2Compound. (https://github.com/beherasan/MS2Compound/tree/master/TestData/FragmentPrediction)
b. Identification of compounds from LC-MS data using in-built database. (https://github.com/beherasan/MS2Compound/tree/master/TestData/MS_InBuiltDB)
c. Identification of compounds from LC-MS data using custom (user-defined) database. (https://github.com/beherasan/MS2Compound/tree/master/TestData/MS_CustomDB)
d. Identification of compounds from LC-MS/MS data using in-built database. (https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_InBuiltDB)
e. Identification of compounds from LC-MS/MS data using custom (user-defined) database. (https://github.com/beherasan/MS2Compound/tree/master/TestData/MSMS_CustomDB)

4. Run MS2Compound

Clicking the "MS2Compound.exe" will start the GUI as shown in figure 1.

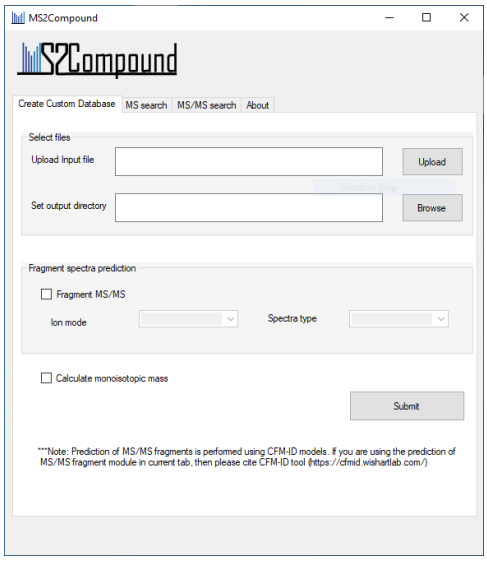


Figure 1. Start page of MS2Compound

Create Custom Database

This module will help the user to create their custom database from a list of compounds (given the SMILES id of the listed compounds)

The page for creating custom database is the default start page for MS2Compound.

The input file as given the test dataset "inputSMILES.txt" is a tab separated file containing two columns; first column is compound id or any unique identifier, and second column containing the SMILES id for the same (Figure 2).

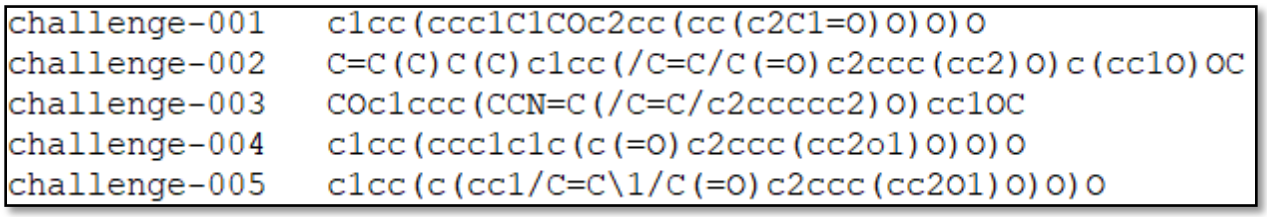


Figure 2. Sample input file for creating custom database for MS/MS search.

Select the Output directory: Please avoid spaces in name of the directory or the directory path

User can select both the option "Fragment MS/MS" and "Calculate monoisotopic mass" or any one according to the need.

After selecting the "Fragment MS/MS", please select the desired "Ion mode" and "Spectra type" for the prediction.

Click "Submit after selecting the desired options". After submit it will start a cmd.exe and the main program cannot be assessed until the current work completed.

This program will generate a directory "CFM-ID_Predicted" inside the selected output directory for the MS/MS fragments. If the directory "CFM-ID_Predicted" already exists in selected output directory, program will be halted with an error. Please rename the existing file or delete it before running the command. After successful completion of the prediction it will create separate log files for each compounds in "CFM-ID_Predicted" directory, this directory can be used as the custom database in the MS/MS search using custom Database module in the tool.

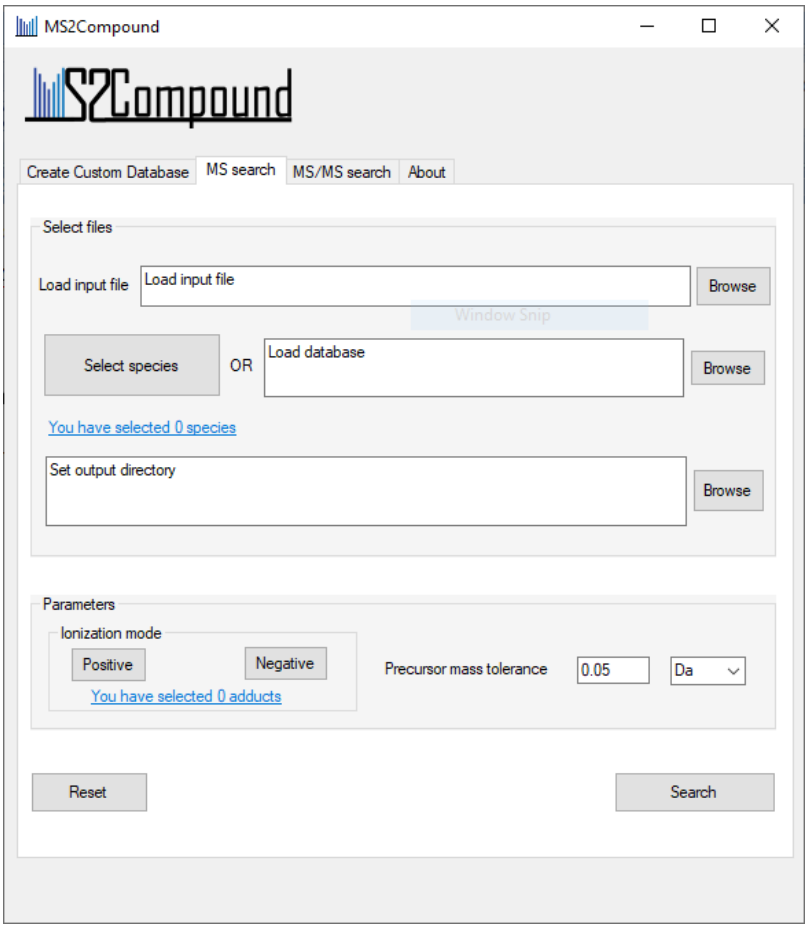
Alongside the program will create a output file named "MonoisotopicMass.txt" for the calculated monoisotopic masses.

The sample outputs from this module using the given test data, and "Positive" ionization mode and "ESI-MS/MS (CE)" spectra type is provided in https://github.com/beherasan/MS2Compound/tree/master/TestData/FragmentPrediction/OUTPUT .

********* *The CFM-id prediction will take some time based on the number of the compounds in input file. The current monoisotopic mass calculation cannot be used for the isotopes of compounds.*

MS-Search using in-built Database

Click on MS search tab. This will open a window as following:

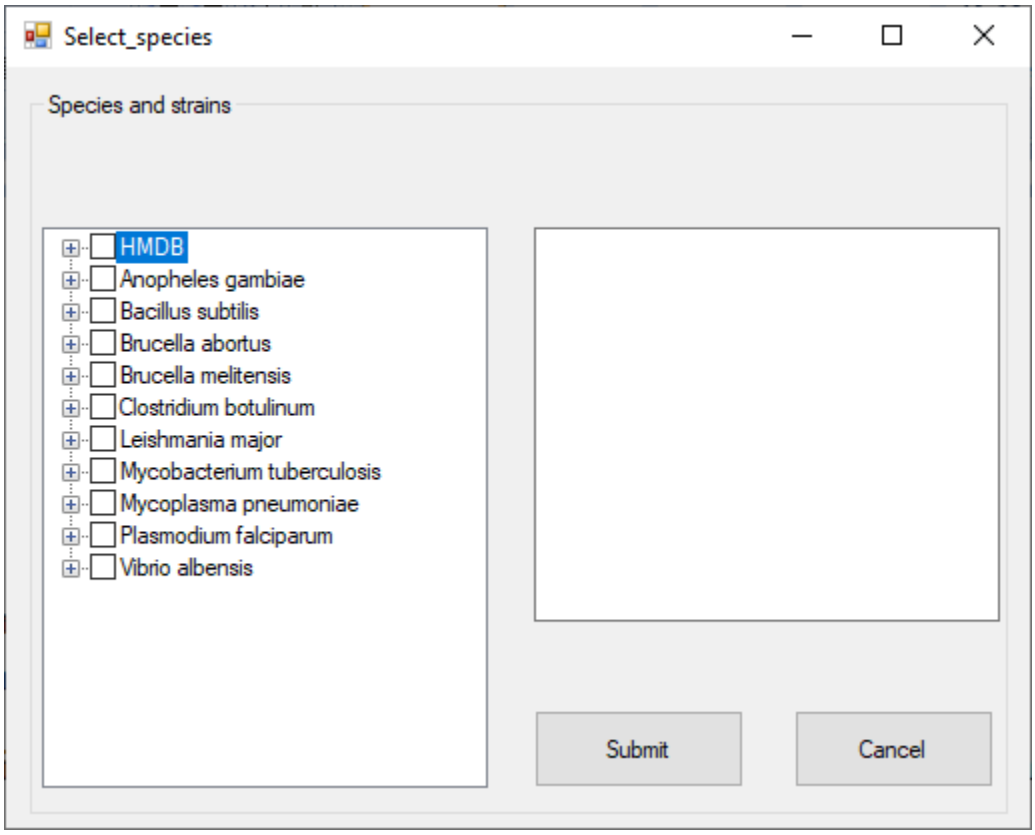


The input file should be a tab separated file with three columns; first column containing unique identifier, second containing the precursor m/z value, and the third column with charge information. The format of the charge should be "a numeric number followed by + or - sign", for example "1+". A sample file has been shown below:

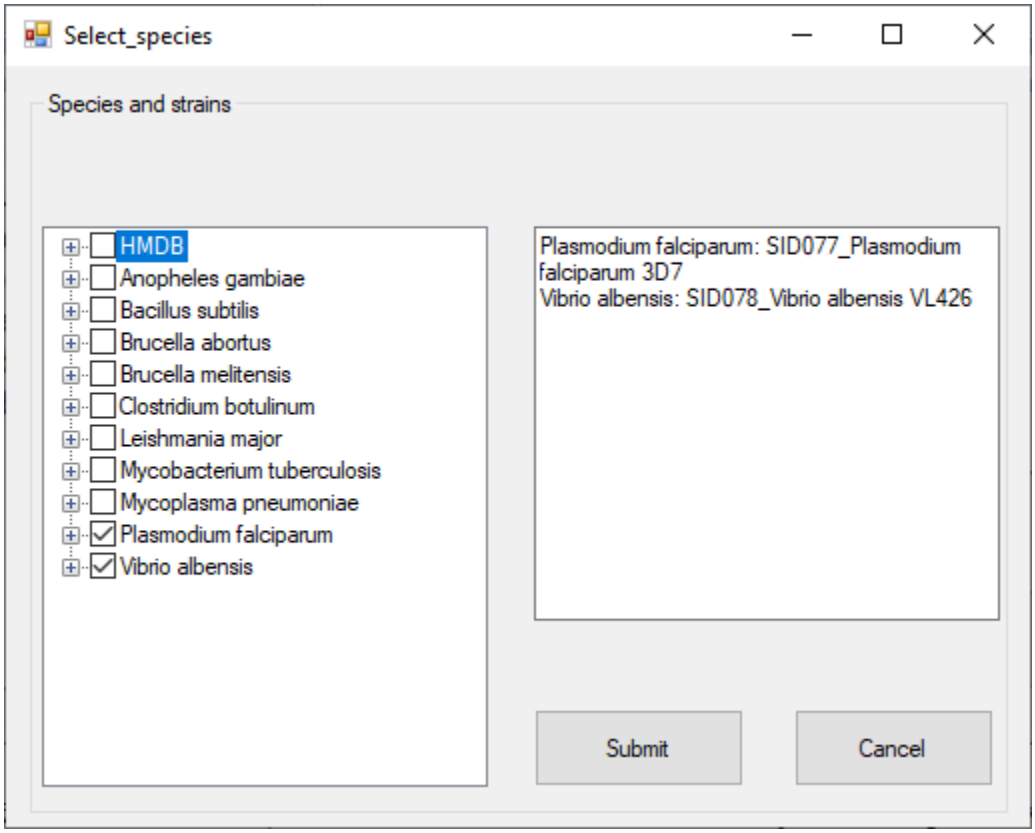
1	145.54	1+
2	530.3015	1+
3	758.2237	1+
4	336.2166	2+
5	747.3701	1+
6	345.2206	2+
7	445.2429	2+
8	536.1644	
9	610.1874	
10	182.9848	1+

If the charge information is provided the match will consider particular adducts for the given charge, else it will search for all adducts selected.

After uploading the input file, select the "Select species" button. This will open a window as following:



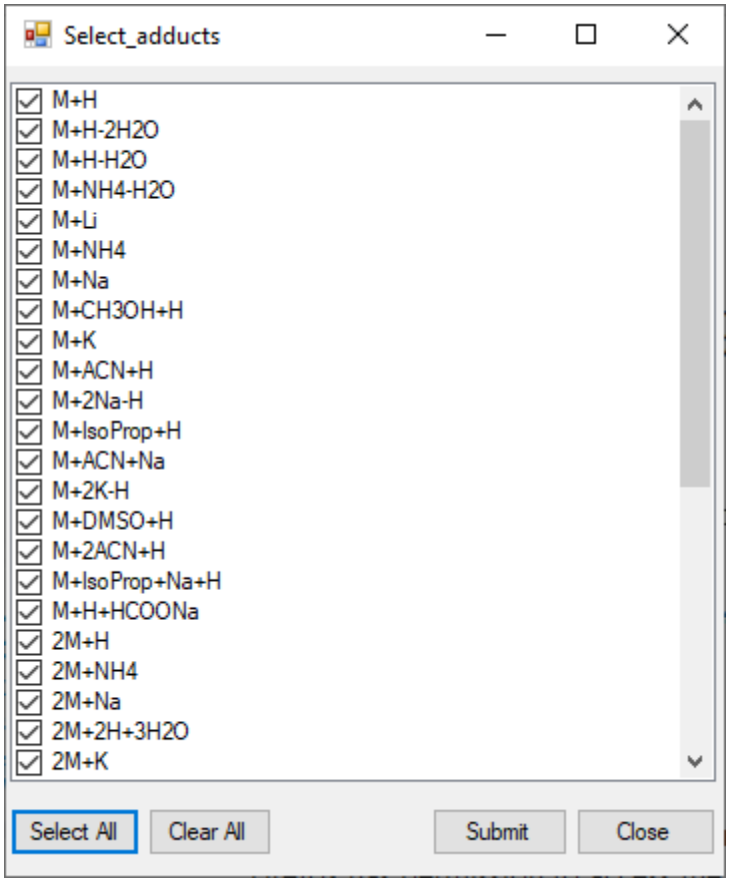
User can select any species of interest, or the search can be strain specific by clicking the "+" sign of species, and then select the desired strain. The species selection is shown as following:



Click “Submit” after selecting the species.

Select the “Output directory”

Select the “Ionization mode” by clicking either “Positive” or “Negative” button. This will open a new window for adducts.



If adducts are not known please select all adducts.

Select the “Precursor mass tolerance”, and the unit.

Click “Search”

This generates three output files in selected output directory.

1. Assigned.txt

2. Assigned.html

3. Unassigned.txt

The sample output files can be found at “https://github.com/beherasan/MS2Compound/tree/master/TestData/MS_InBuiltDB/OUTPUT”

MS-Search using Custom Database

This module allows the user to search MS data against a user defined monoisotopic mass file for a selected list of compounds.

The input file will be similar as that of the input for "**MS-Search using in-built Database**".

The monoisotopic mass file should be a tab separated file with two columns; one for unique identifier, and the second one for corresponding monoisotopic mass.

Remaining parameters and the output will be similar as that of "**MS-Search using in-built Database**".