AnnoSM: An Automated Annotation Tool for Determining the Substituent Modes on the Parent Skeleton Based on Characteristic MS/MS Fragment Ion Library

User Manual

1. Software Introduction

AnnoSM, a user-friendly interface, was built to automatically determine the SMs on the parent molecular skeleton based on characteristic MS/MS fragment ion library. This software was developed based on R version 4.1.2 and Rshiny framework

2. Software Installation

The installation process is as follows:

2.1 Download and install R with R studio.

2.2 Install packages.

Install the packages required to run the AnnoSM software.

```
# 1) install packages: readxl_1.4.2, tidyr_1.3.0, plyr_1.8.8,
dplyr_1.1.2, readr_2.1.4, openxlsx_4.2.5.2, stringr_1.5.0,
shiny_1.7.4, shinyWidgets_0.7.6, shinyFiles_0.9.3, DT_0.31,
progress_1.2.2, data.table_1.14.10
install.packages("**")

# 2) install "metaMS_1.30.0" package
if (!requireNamespace("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install("metaMS")
```

2.3 Run AnnoSM.R program.

There are two ways to access the main interface of the AnnoSM software:

2.3.1 R_GUI (graphic user's interface)

Run the AnnoSM.R program in R_GUI (graphic user's interface). (**Fig. 1**)

```
source("D:/rdata/AnnoSM.R", encoding="UTF-8")
shinyApp(ui, server)
```

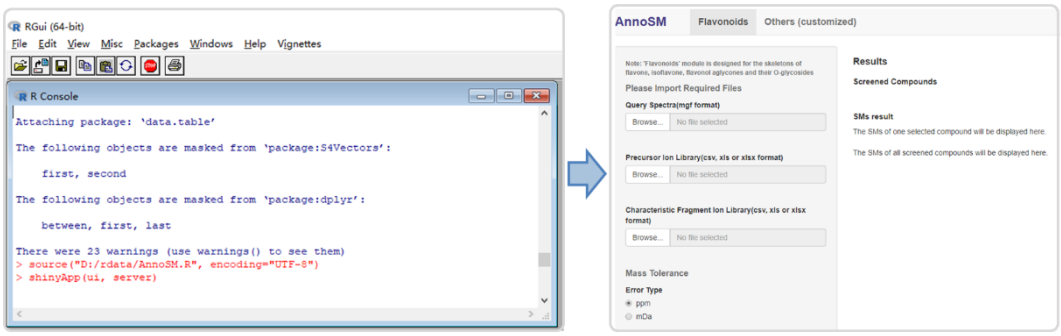

Fig. 1. Enter the main interface of AnnoSM through R_GUI.

2.3.2 Rstudio

Open the AnnoSM.R program in Rstudio, then click "Run App" to enter the main interface of the software. (**Fig. 2**)

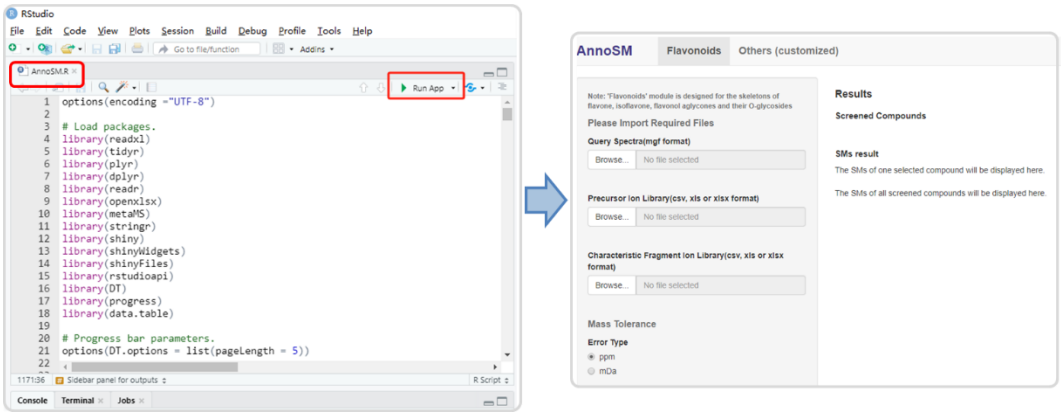

Fig. 2. Enter the main interface of AnnoSM through Rstudio.

3. Software Operation Procedures

3.1 Preparation of import files.

3.1.1 Theoretical Precursor Ions Library (TPIL).

An **Excel file (.xlsx, xls. or .csv)** contains MS1 and structural information of the target compounds is constructed as a TPIL used to annotate the total number of substituents among the matched compound candidates.

Here, the TPIL of flavonoids is used as an example (**Fig. 3**). The file contains **nine columns**:

(1) ***Compound group*****:** representing a serial of flavonoids with the same molecular formula;

(2) ***Class*****:** type of compounds;

(3) ***OHsum*****:** the total number of OH groups on the skeleton of flavonoid aglycones;

(4) ***OCH3sum*****:** the total number of OCH_3 groups on the skeleton of flavonoid aglycones;

(5) ***Sugar and phenylpropionyl units*****:** the number and type of sugar and phenylpropionyl units substituted on flavonoid aglycones;

(6) ***S&Psum*****:** the total number of sugar and phenylpropionyl units of flavonoid glycosides;

(7) ***Formulas:*** molecular formula;

(8) ***Exact Mass*****:** exact mass of compounds;

(9) ***cal.m.z.H*****:** theoretical m/z values of precursor ion ($[M+H]^+$).

	A	B	C	D	E	F	G	H	I
1	**Compound group**	**Class**	**OHsum**	**OCH3sum**	**Suger and phenylpropionyl units**	**S&Psum**	**Formulas**	**Exact Mass**	**cal.m.z.H**
2	1	Aglycone	0	0	—	0	C15 H10 O2	222.0681	223.0754
3	2	Aglycone	1	0	—	0	C15 H10 O3	238.063	239.0703
4	3	Aglycone	2	0	—	0	C15 H10 O4	254.0579	255.0652
5	4	Aglycone	3	0	—	0	C15 H10 O5	270.0528	271.0601
6	5	Aglycone	4	0	—	0	C15 H10 O6	286.0477	287.055
7	6	Aglycone	5	0	—	0	C15 H10 O7	302.0427	303.0499
8	7	Aglycone	6	0	—	0	C15 H10 O8	318.0376	319.0448
9	8	Aglycone	7	0	—	0	C15 H10 O9	334.0325	335.0398
10	9	Aglycone	8	0	—	0	C15 H10 O10	350.0274	351.0347

Fig. 3. Demonstration of Excel file for Theoretical Precursor Ions Library (TPIL).

The columns "OHsum", "OCH3sum" and "cal.m.z.H" are required for customized TPIL, remaining columns can be adjusted according to actual requires.

3.1.2 Characteristic Fragment Ions Library (CFIL).

The **Excel file (.xlsx, xls. or .csv)** of CFIL is used to annotate the substituent modes (SMs) of screened compounds. For each class of compounds, researcher require to

summarize corresponding characteristic fragmentation patterns among experimental MS/MS spectra from literatures and authentic standards.

Here, the CFIL of flavonoids is used as an example (**Fig. 4**). The file contains **eight columns**:

(1) ***Substitution modes:*** the types and positions of substituents in core structure of flavonoid aglycones.

(2) ***Class:*** subtypes of flavonoids, e.g. flavones, flavonols and isoflavones.

(3) **Types and positions of substituents:** number and position of hydroxy and methoxy substituents on the A-, B- and C-rings respectively;

(4) ***OHsum:*** the total number of OH groups on the skeleton of flavonoid aglycones;

(5) ***OCH3sum:*** the total number of OCH_3 groups on the skeleton of flavonoid aglycones;

(6) ***Formulas of Frag:*** the formulas of characteristic fragment ions (CFIs);

(7) ***Cal.m/z of Frag*****:** theoretical m/z values of CFIs;

(8) ***Annotations*****:** the bond breaking positions in MS/MS spectra, e.g. $^{1,3}A^{+}$.

The columns "Substitution modes", "OHsum", "OCH3sum" and "Cal.m/z of Frag", "Annotations" are required for customized CFIL, in which "OHsum" and "OCH3sum" columns are the same as those in TPIL.

	A	B	C	D	E	F	G	H
1	Substitution modes	Class	Types and positions of substituents	OHsum	OCH3sum	Formulas of Frag	Cal.m/z of Frag	Annotations
2	A00B00C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×0)+RingC(OH×0+OCH3×0)	0	0	C7 H5 O2	121.0284	1, 3A+
3	A00B00C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×0)+RingC(OH×0+OCH3×0)	0	0	C6 H5 O	93.0335	1, 3A+-CO/1, 4A+
4	A00B00C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×0)+RingC(OH×0+OCH3×0)	0	0	C5 H5	65.0386	1, 3A+-2×CO/1, 4A+-CO
5	A00B00C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×0)+RingC(OH×0+OCH3×0)	0	0	C6 H5	77.0386	0, 4A+/0, 2B+-CO
6	A00B00C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×0)+RingC(OH×0+OCH3×0)	0	0	C8 H7	103.0542	1, 3B+
7	A00B00C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×0)+RingC(OH×0+OCH3×0)	0	0	C9 H7 O2	147.0441	0, 4B+
8	A00B00C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×0)+RingC(OH×0+OCH3×0)	0	0	C9 H5 O	129.0335	0, 4B+-H2O
9	A00B00C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×0)+RingC(OH×0+OCH3×0)	0	0	C8 H5	101.0386	0, 4B+-H2O-CO
10	A00B00C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×0)+RingC(OH×0+OCH3×0)	0	0	C7 H5 O	105.0335	0, 2B+
11	A00B01C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×1)+RingC(OH×0+OCH3×0)	0	1	C7 H5 O2	121.0284	1, 3A+
12	A00B01C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×1)+RingC(OH×0+OCH3×0)	0	1	C6 H5 O	93.0335	1, 3A+-CO/1, 4A+
13	A00B01C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×1)+RingC(OH×0+OCH3×0)	0	1	C5 H5	65.0386	1, 3A+-2×CO/1, 4A+-CO
14	A00B01C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×1)+RingC(OH×0+OCH3×0)	0	1	C6 H5	77.0386	0, 4A+
15	A00B01C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×1)+RingC(OH×0+OCH3×0)	0	1	C9 H9 O	133.0648	1, 3B+
16	A00B01C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×1)+RingC(OH×0+OCH3×0)	0	1	C8 H9	105.0699	1, 3B+-CO
17	A00B01C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×1)+RingC(OH×0+OCH3×0)	0	1	C8 H6 O	118.0413	1, 3B+-CH3
18	A00B01C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×1)+RingC(OH×0+OCH3×0)	0	1	C7 H6	90.0464	1, 3B+-CH3-CO
19	A00B01C00	Flavones	RingA(OH×0+OCH3×0)+RingB(OH×0+OCH3×1)+RingC(OH×0+OCH3×0)	0	1	C8 H5 O	117.0335	1, 3B+-CH4

Fig. 4. Demonstration of Excel file for Characteristic Fragment Ions Library (CFIL).

3.2 Main Interface.

There are two modules on the software: ***Flavonoids*** and ***Others (customized)*** (Fig. 5). The former is specifically designed for the flavonoid aglycones and *O*-glycosides while the latter is extended to other compound classes. There are **four functional blocks** in each module.

(1) ***Import Required Files:*** Users can click the **Browse** button to select query spectra, theoretical precursor ion library (TPIL) and characteristic fragment ion library (CFIL) (Fig. 5A).

(2) ***Mass Tolerance:*** Users can select the error type (ppm or mDa) and mass error of precursor ions and fragment ions (Fig. 5B).

(3) ***Processing:*** This block is subdivided into three parts (Fig. 5C).

a. Users can tick the box in front of the **"Screening Compounds"** option and click the **RUN** button below to obtain the result of screened compounds, which are displayed in the upper part of Fig. 5D. The comprehensive results can be downloaded to local computer by clicking **Screened Compounds_result.xlsx** button.

b. After obtaining the target compounds, users can tick the box in front of the **"SMs of One Compound"** option and enter any compound ID in the screened

compounds list to generate a SM candidates list for the selected compound displayed in the middle part of **Fig. 5D**.

c. When users select the **"SMs of all Compounds"** option, the SMs of all screened compounds will be displayed in the lower part of **Fig. 5D**.

Users can download substituent results in **b.** or **c.** by clicking **SM_result.xlsx** or **SMs_results.xlsx** button respectively.

(4) ***Results:*** The results of target compounds screening, SMs of one selected compound, and SMs of all screened compounds are displayed in the corresponding part of **Fig. 5D**.

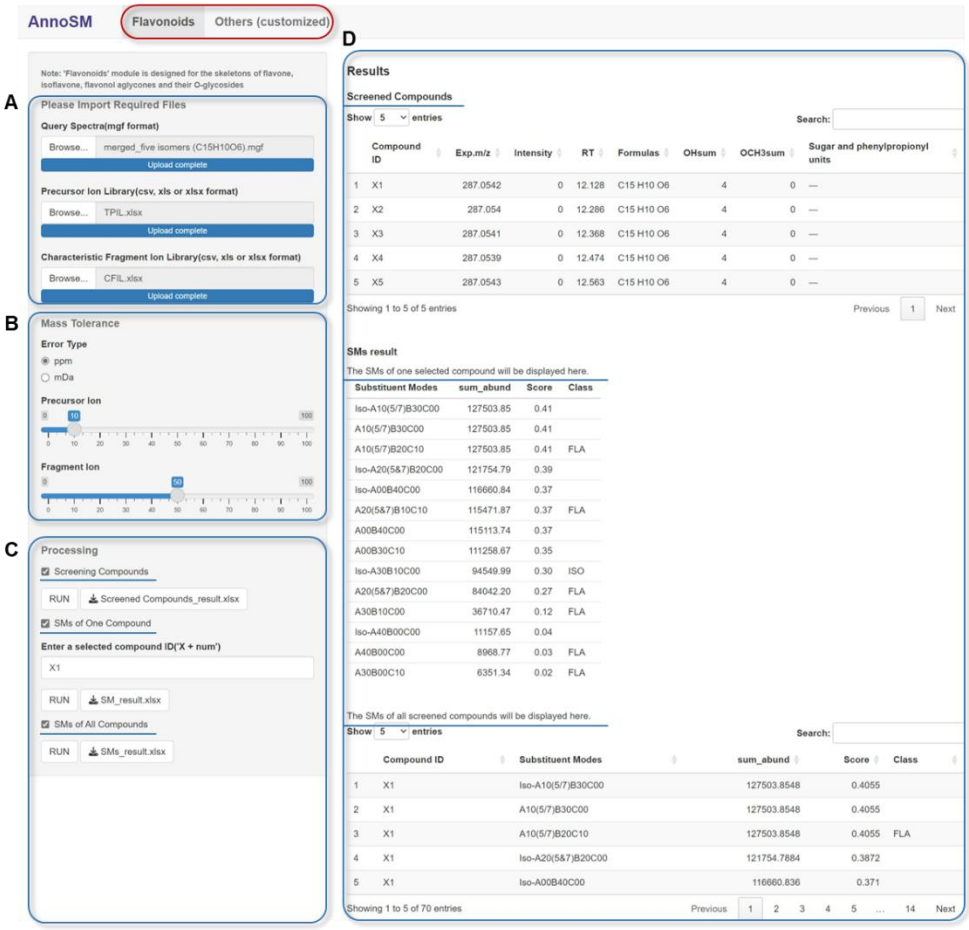

Fig. 5. Screenshot of the software tool AnnoSM (An automated annotation tool for determining the substituent modes on the parent skeleton).