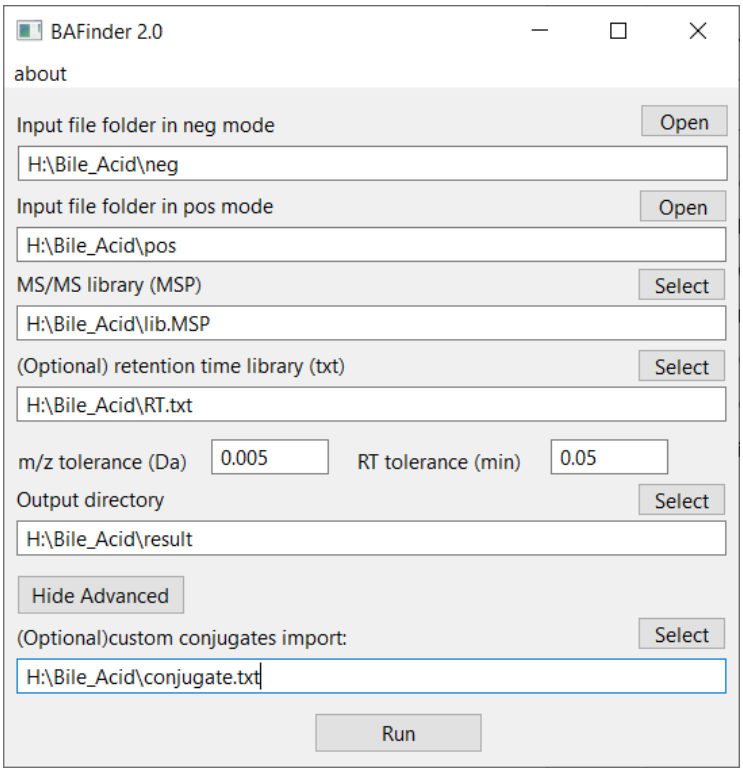
BAFinder 2.0 User Manual

https://bafinder.github.io/

By Yan Ma, updated on July 27, 2023



1. Introduction

BAFinder is a program developed for the unknown bile acid identification from LC-MS/MS data in both positive and negative modes. It was developed in Java (jdk 15.0.1) with a graphical user interface (GUI) using the Open Source IDE Eclipse. BAFinder takes the alignment and peak list generated from XCMS (csv) and MS/MS spectra (MGF), searches them against a MS/MS reference library (MSP), a build-in characteristic feature (e.g. product ion or neutral loss) query list and an optional retention time library(txt) within user-defined *m/z* and RT tolerance, and exported the annotation summary (csv), processing details (csv) and representative MS/MS spectra (MGF) into an output folder (see the work-flow chart in page 3).

The following bile acids are covered by BAFinder 1.0: free bile acids (non-oxidized & oxidized), glycine-conjugated bile acids (GBA), taurine-conjugated bile acids (TBA), bile acid sulfates (BA-S), bile acid glucuronides (BA-GlcA) and double-conjugated bile acids (GBA-S, TBA-S, GBA-GlcA, TBA-GlcA).

The updated BAFinder 2.0 adds 18 common amino acid-conjugated bile acids (alanine, arginine, asparagine, aspartate, glutamine, glutamate, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine and valine) as well as user-defined amino acid conjugated bile acid.

2. Start BAFinder

BAFinder 2.0 is provided as a package bundled with java run time environment (jre) therefore no java installation is required.

For Windows64 system:
Open folder “BAFinder2_windows64”, and double click “BAFinder2_win64.exe”.

For macOS system:
Open folder “BAFinder2_macosx”, in the sub-directory “Contents/MacOS”, find BAFinder2_macosx and double click it.

The bundled packages were generated using Packr (https://github.com/libgdx/packr). They have been tested on Windows 10 and macOS Monterey 12 system. In case the packaged program doesn’t work, BAFinder 2.0 could be started from the .jar file (java installation required). Please refer to the User Manual for BAFinder 1.0 (https://bafinder.github.io/) for more information.

2. Preparation of input files

2.0 Prerequisite

To be processed by BAFinder, the LC-MS/MS data in positive and negative mode must be generated from the same set of samples. The file names of positive and negative data need to be matched (e.g. pos_sample1 & neg_sample1) for BAFinder to correctly recognize their corresponding relationships.

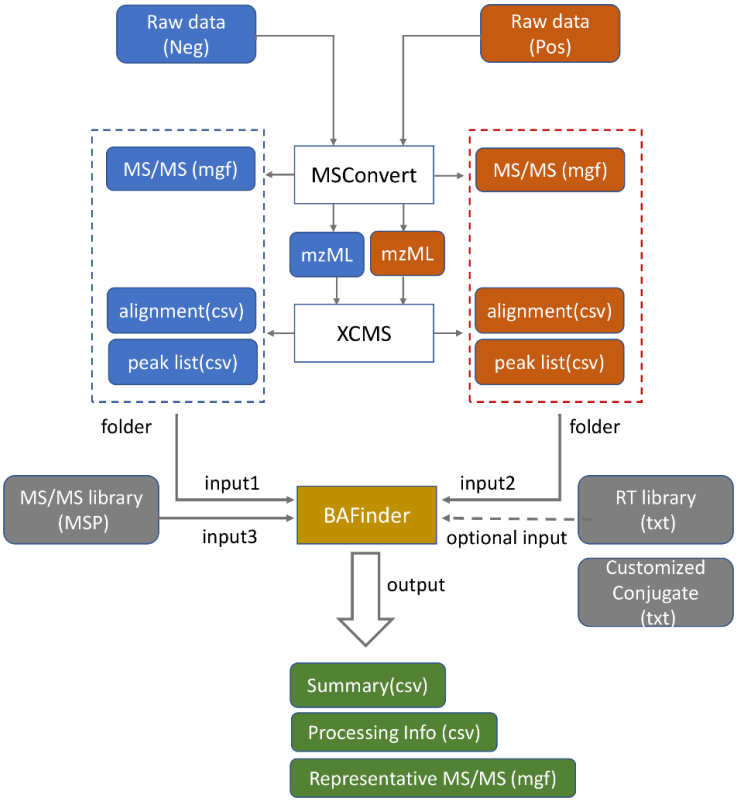


Figure 1. input and output of BAFinder

2.1 Raw data conversion

Raw data files are converted to mzML and MGF format using ProteoWizard MSConvert software.

Note: 1.For profile data, pick-picking must be selected as the first filter in MSConvert, since XCMS centWave algorithm can only process data in centroid mode. 2. For Thermo Orbitrap data, RawConverter (http://fields.scripps.edu/rawconv/) is recommended for MGF format conversion for its accurate precursor selection.

2.2 Data preprocessing

Data files in mzML format from positive and negative modes are put into two separate folders and processed using R package XCMS respectively. Below are the example R codes for positive mode:

```
library(xcms)
data=list.files("H:/Bile_Acid/mzML_pos", full.names=TRUE)
raw_data <- readMSData(files=data,mode = "onDisk")
cwp <- CentWaveParam(ppm=10,peakwidth=c(2.5,8),mzdiff=0.005)
xpeak <- findChromPeaks(raw_data, param = cwp)
# number of sample in this folder
sampleg=rep(1,9)
pdp <- PeakDensityParam(sampleGroups = sampleg, bw = 1, minFraction=0.1, minSamples=1,
binSize=0.005)
xgroup <- groupChromPeaks(xpeak, param = pdp)
xfeature <- featureDefinitions(xgroup)
xvalue <- featureValues(xgroup, value = "into")
xreport <- merge(xfeature,xvalue, by="row.names")
xreport$peakidx <- sapply(xreport$peakidx, paste, collapse=";", sep=";")
write.csv (xreport,file="H:/Bile_Acid/alignment_pos.csv")
write.csv (chromPeaks(xpeak) , file="H:/Bile_Acid/peak_pos.csv")
```

The parameters in bold need to be adjusted according to the user's data. The lines in green are essential to generate csv file format that can be recognized by BAFinder, and therefore cannot be changed.

After the data-preprocessing is done, put the resulted two csv files and the MGF files for each mode into a folder.

2.3 MS/MS library

A MS/MS library generated from bile acid reference standards can be downloaded on BAFinder website. The updated library for BAFinder 2.0 includes in-silico MS/MS spectra of common amino acid conjugated bile acids. Also feel free to expand this library or use your own library.

2.4 Retention time library (optional)

RT library is a text file that contains the information of retention time of reference standard. It can be easily written by users, as long as the following format is followed:

Class RT Name

Class: type of bile acids, defined by two parts:1) number of hydroxyl groups and keto groups (double bond in ring is also counted as keto); 2) conjugate type. The two parts were linked by underline (_).
E.g. 1OH_BA; 3OH_GBA-S; 1O1OH_TBA-GlcA

Part 1
1OH
1O
2OH
1O1OH
2O
3OH
1O2OH
2O1OH
3O
4OH

Part 2	
BA	Ala-BA
GBA	Arg-BA
TBA	Asn-BA
BA-S	Asp-BA
BA-GlcA	Gln-BA
BA-Hex	Glu-BA
GBA-S	His-BA
TBA-S	Leu/Ile-BA
GBA-GlcA	Lys-BA
TBA-GlcA	Met-BA
GBA-Hex	Phe-BA
TBA-Hex	Pro-BA
	Ser-BA
	Thr-BA
	Trp-BA
	Tyr-BA
	Val-BA

RT: retention time in mins
Name: name of reference bile acid

Please note:

1.The separator between Class, RT and Name is tab. It is recommended to create the list in Excel and then save as txt or copy to txt editors.

2.Optional: separate retention time tolerance for each bile acid can be defined by adding 2 numbers (the left and right boundaries of RT error) after the bile acid name. This will overwrite the universal threshold in GUI.
In the following example, if the RT tolerance is 0.06 min in GUI, the RT range for the first line is from 13.5-0.05 to 13.5+0.05, while the second line is from 13.5-0.1 to 13.5+0.05.

```
2OH_BA    13.5    CDCA
2OH_BA    13.5    CDCA    -0.1    0.05
```

2.5 *m/z* window and retention time tolerance

m/z window: set according to the mass accuracy of MS instrument (actual performance)

RT tolerance: set according to the retention time shift during the batch. It is recommended to check various conjugation types in QC samples or internal standards, as they may shift differently.

2.6 Customized conjugate list (optional)

BAFinder 2.0 allows user-defined bile acid conjugate list in the following format:

Name Mass

Example: Alanine 89.048

Please note: Customized conjugate is annotated based on the fragmentation rules of amino acid-conjugated bile acids. Therefore, it is recommended to pick customized conjugate with similar structure to amino acids.

3. Output Files

BAFinder exported three files in user-selected folders: a final summary report in csv format, a detailed processing report in csv format and a set of representative MS/MS spectra in MGF format.

3.1 Summary

RT(min)	the retention time of the most abundant adduct in negative mode
annotation	hydroxyls, ketos and conjugation of bile acids
confidence	L1 to L4: highest to lowest annotation confidence (please see the paper for details) empty: result is matched only by RT
std	hit of RT library search
neg MS2 hit	characteristic MS/MS library hit in neg mode (dot product>500)
fragment	characteristic product ion ratio, i.e.145/159>1 in m/z 355 (+) indicate 12-OH
comments	characteristic product ion intensity, i.e. intensity of NL 176 indicate position of GlcA
annotation details	from [neg_MS2_feature] to [top_neg_adduct], summary of processing details
peak intensity	the last columns, with the data file name as headings

3.2 Processing details

For advanced users to further validate the annotation result. Most headings are self-explanatory.

XCMSfeature	feature number in XCMS alignment result
pass MS1 filters	empty: MS1 filter passed, proceeding to MS2 annotation neg only: top pos/neg intensity>2, pos features are removed
MS2 feature	1: characteristic product ion or neutral loss is detected in MS/MS
modified lib hit	hit after remove mz 76 or 126 in GBA/TBA pos MS/MS

3.3 Representative MS/MS spectra

For advanced users to further validate the annotation result.

4. Contact us

To report bugs, ask questions and offer suggestions, please contact us by email: mayan@nibs.ac.cn

Thanks for using BAFinder!