



浙江中醫藥大學
ZHEJIANG CHINESE MEDICAL UNIVERSITY

Seminar

Lichuang Huang

Supervisor: Cao Gang

Zhejiang Chinese Medical University

2022-09-21

1 Identification of MASS spectra

2 The principle of MCnebula

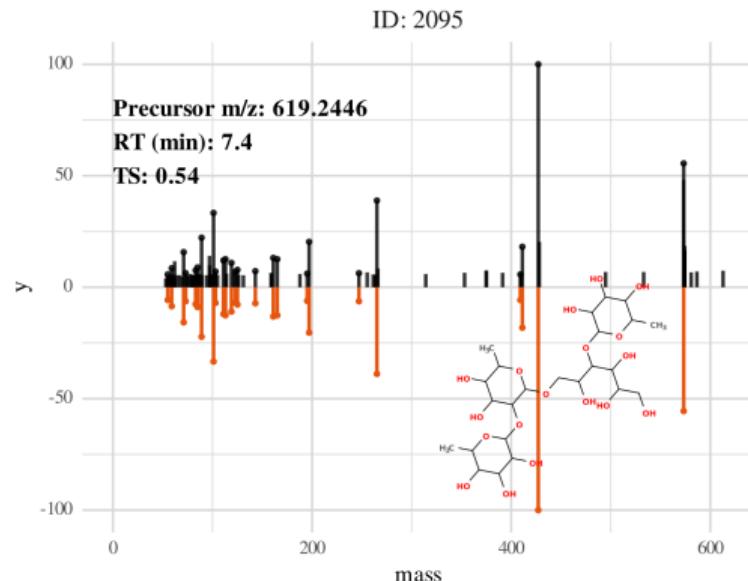
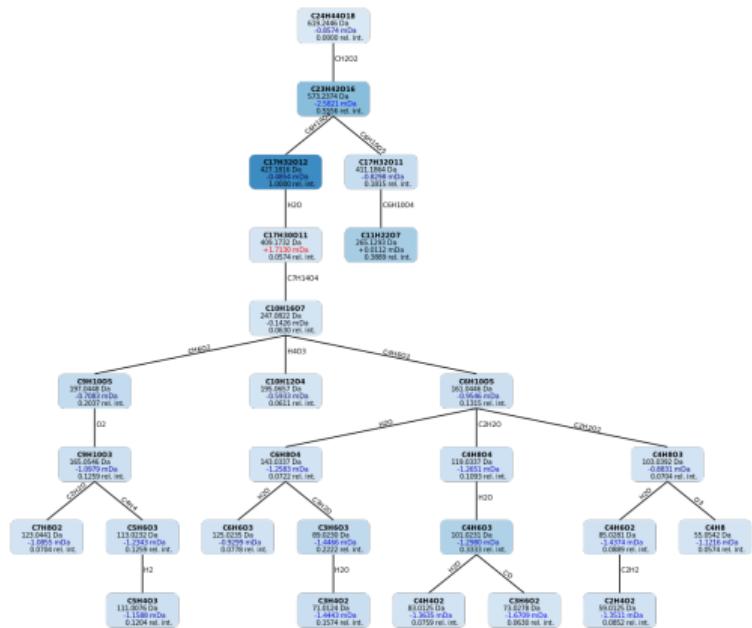
3 Progress

4 Next Schedule

Identification of MASS spectra

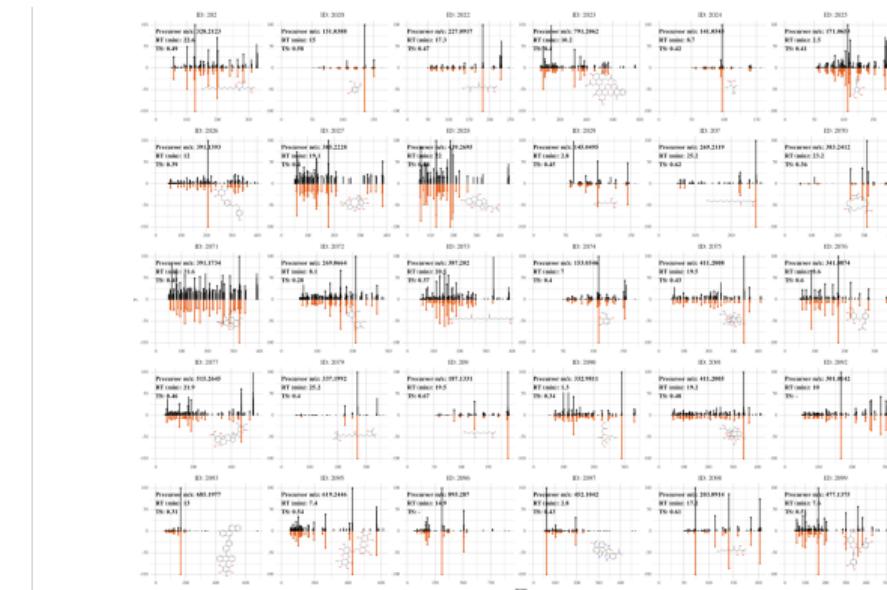
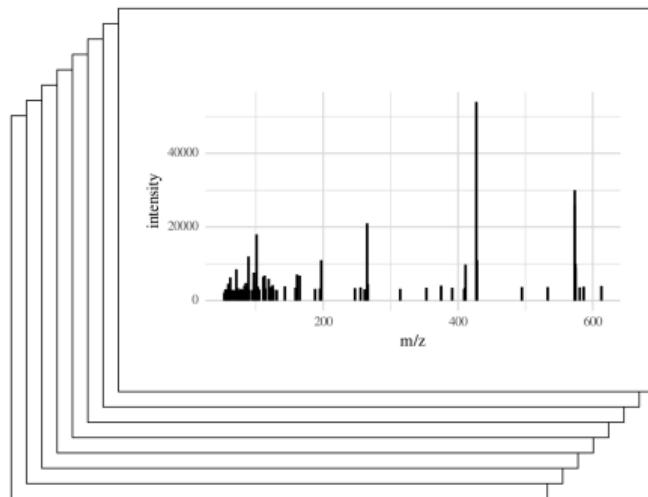
From MS/MS spectrum to compound

Compounds smashed by energy



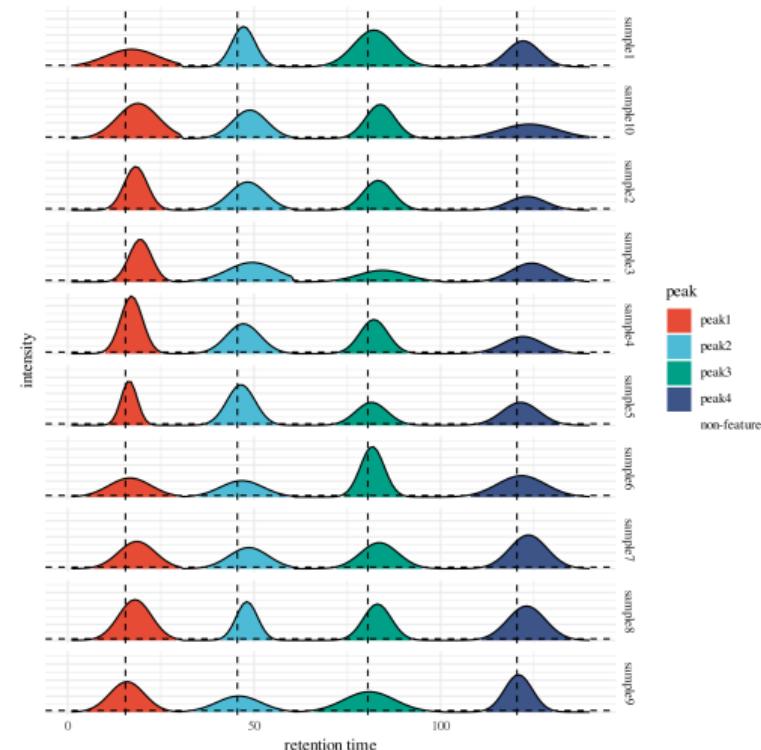
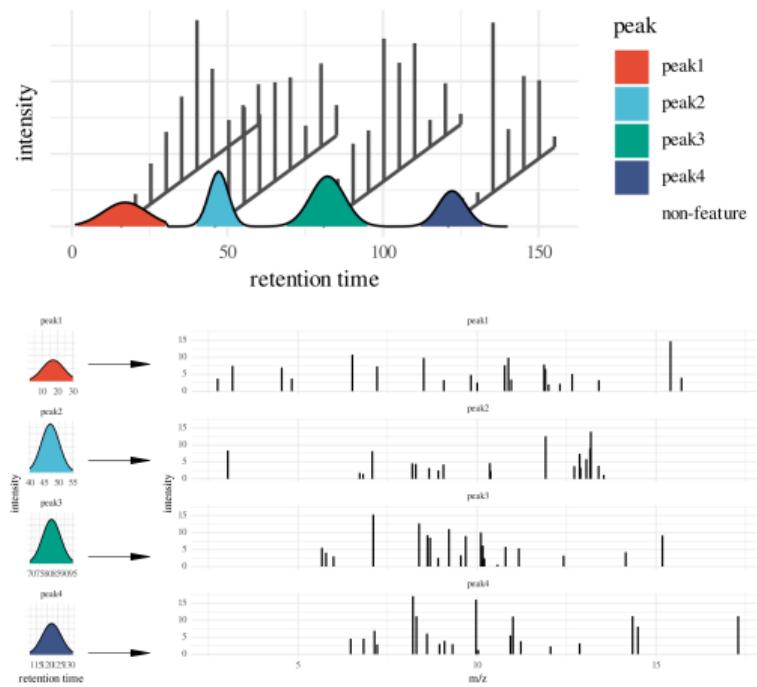
Massive MS/MS spectra

A metabolitic research need lots of identification.



Why so many spectra ?

—LC-MS/MS alignment



From manpower to machine

The speculation of MS/MS spectra.

1 MS¹ ⇒ chemical formula

- ion mass
- isotopes
- ...

2 MS² ⇒ chemical structure

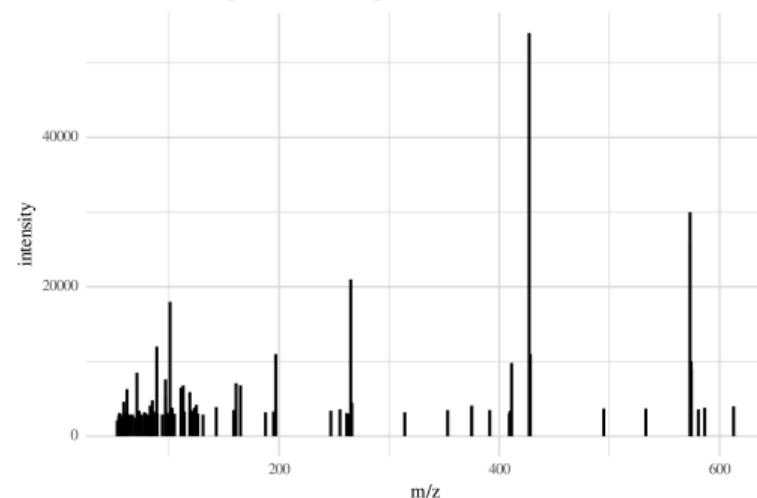
- fragmentation m/z
- relative intensity
- ...

3 structure ⇒ chemical classes

- dominant structure
- sub-structure

Ion mass: 619.2446

Ion type: [M + ?]-



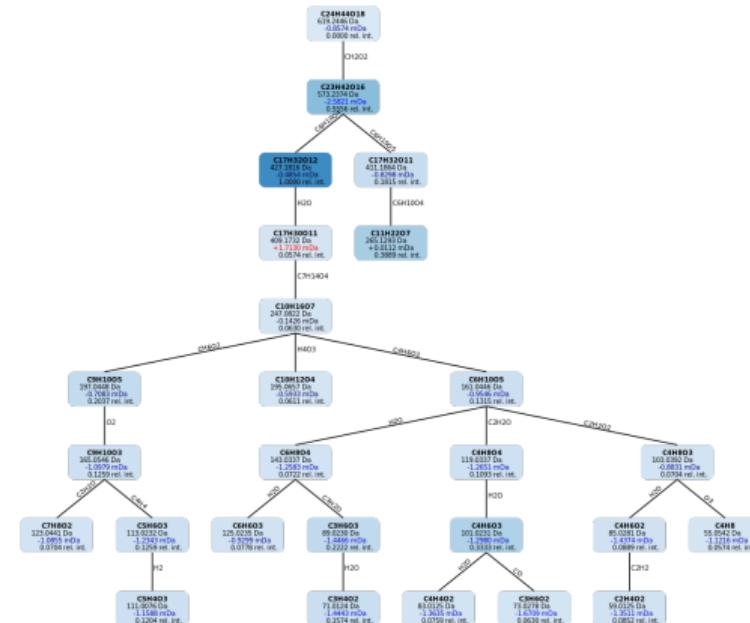
From manpower to machine Extra

Another speculation mode of MS/MS spectra.

- MS¹ and MS² ⇒ formula
 - ion mass
 - isotopes
 - fragmentation m/z
 - relative intensity

- 2 Database search ⇒ structure
– fragmentation tree

- 3 structure \Rightarrow classes
 - dominant structure
 - sub-structure



Identification

 instance: ID:2095

Identification candidates

1 formula

rank	molecularFormula	adduct
1	C31H40O13	[M - H]-
2	C24H44O18	[M - H]-
3	C20H40N6O16	[M - H]-
4	C23H44N2O15S	[M - H]-
5	C25H40N4O14	[M - H]-
6	C32H36N4O9	[M - H]-
7	C21H36N10O12	[M - H]-
8	C24H40N6O11S	[M - H]-
9	C25H41N4O12P	[M - H]-
10	C23H45N2O15P	[M - H]-
11	C16H36N12O14	[M - H]-
12	C26H36N8O10	[M - H]-
13	C23H46N2O13P2	[M - H]-
14	C22H32N14O8	[M - H]-
15	C28H44O13S	[M - H]-
16	C22H45N4O12PS	[M - H]-
17	C22H44N4O14S	[M - H]-
18	C21H44N6O11S2	[M - H]-

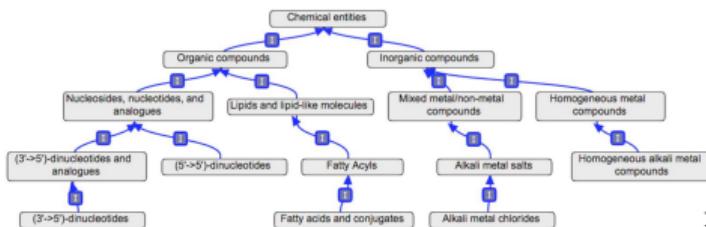
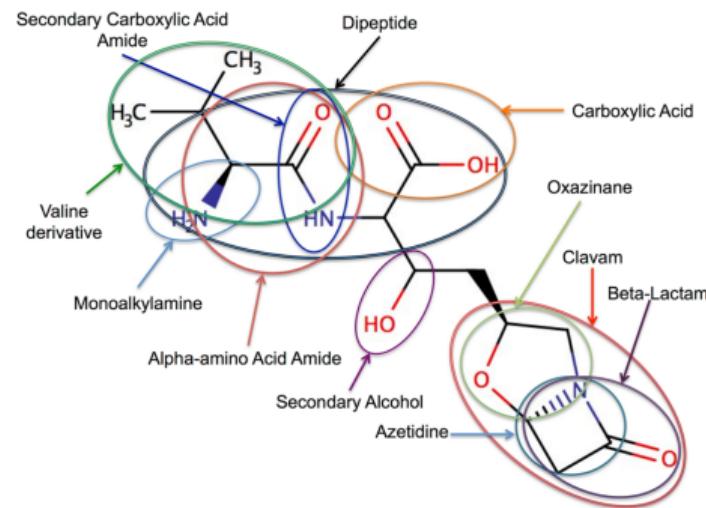
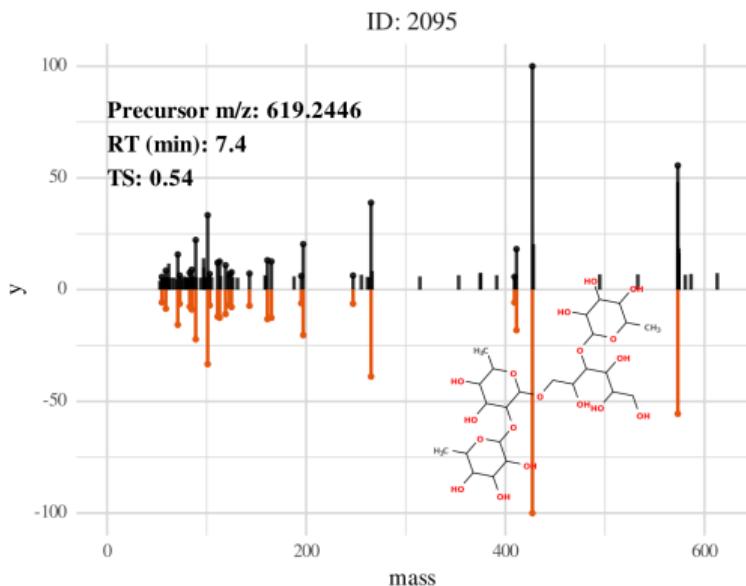
2 structure

rank	inchikkey2d
1	1_C31H40O13_[M-H]-
1	BAKYVUHOODEWGV
2	ZHRVXLJKFZCJGO
3	MWBWDSDAVQXNEL
2	2_C24H44O18_[M-H]-
1	UIZMPBNAOQKGCK
2	JILVRDEKXTWCM
3	AZUNYUJYDUSJDE
5	5_C25H40N4O14_[M-H]-
1	OKSLMPAOQCSBDG
2	CXLXWLLFLUHXJB
3	DMHICHMMWHPZHT

Classification

 in MS/MS spectra view

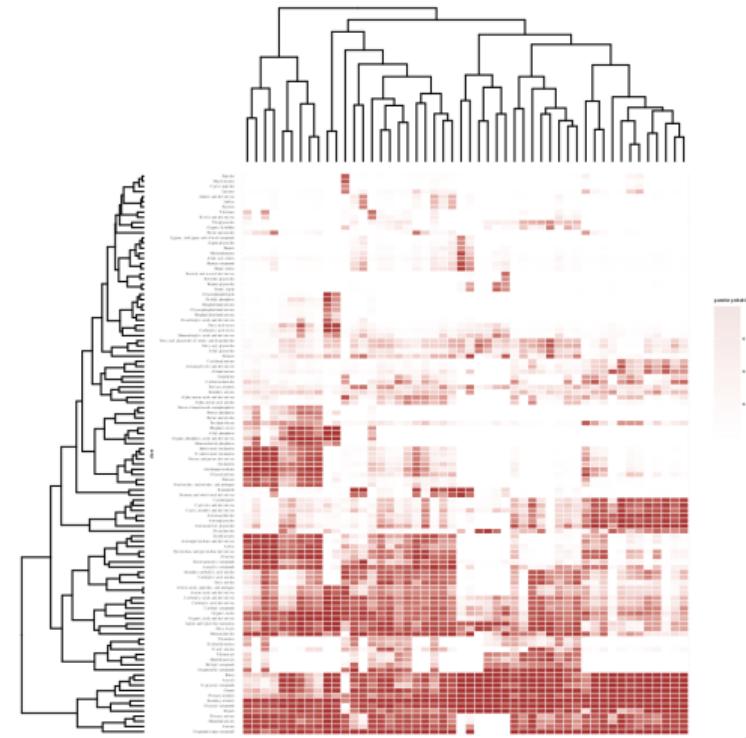
- classes for sub- or dominant structure
- Hierarchy



Classification instance: ID:2095

Prediction for all candidates

rank	molecularFormula	adduct
1	C31H40O13	[M - H]-
2	C24H44O18	[M - H]-
3	C20H40N6O16	[M - H]-
4	C23H44N2O15S	[M - H]-
5	C25H40N4O14	[M - H]-
6	C32H36N4O9	[M - H]-
7	C21H36N10O12	[M - H]-
8	C24H40N6O11S	[M - H]-
9	C25H41N4O12P	[M - H]-
10	C23H45N2O15P	[M - H]-
11	C16H36N12O14	[M - H]-
12	C26H36N8O10	[M - H]-
13	C23H46N2O13P2	[M - H]-
14	C22H32N14O8	[M - H]-
15	C28H44O13S	[M - H]-
16	C22H45N4O12PS	[M - H]-
17	C22H44N4O14S	[M - H]-
18	C21H44N6O11S2	[M - H]-



Individual feature ⇒ overall dataset

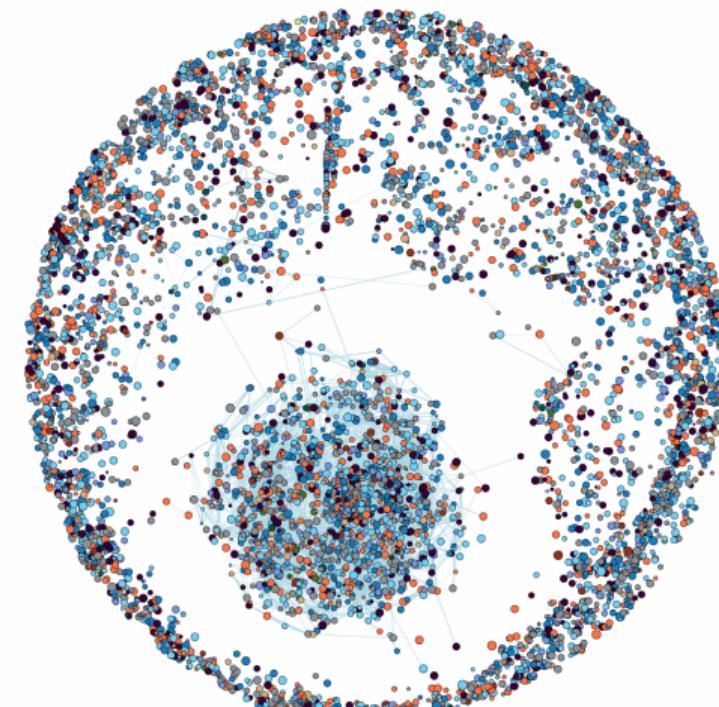
- Individual feature

- 1 formula
- 2 structure
- 3 class

↓

- Overall dataset

- 1 mutiple...

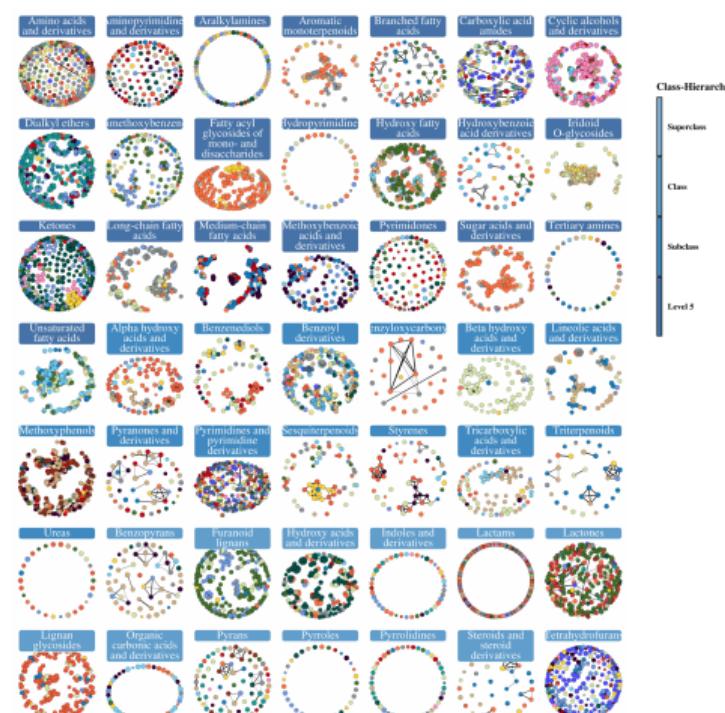


The principle of MCnebula

The purpose of MCnebula

Main steps

- 0 Candidates collection
- 1 Filter candidates
- 2 Create reference
- 3 Filter classification
 - ...
- 4 Gather for an overall
 - formula
 - structure
 - ...
- 5 Visualization in specific classes

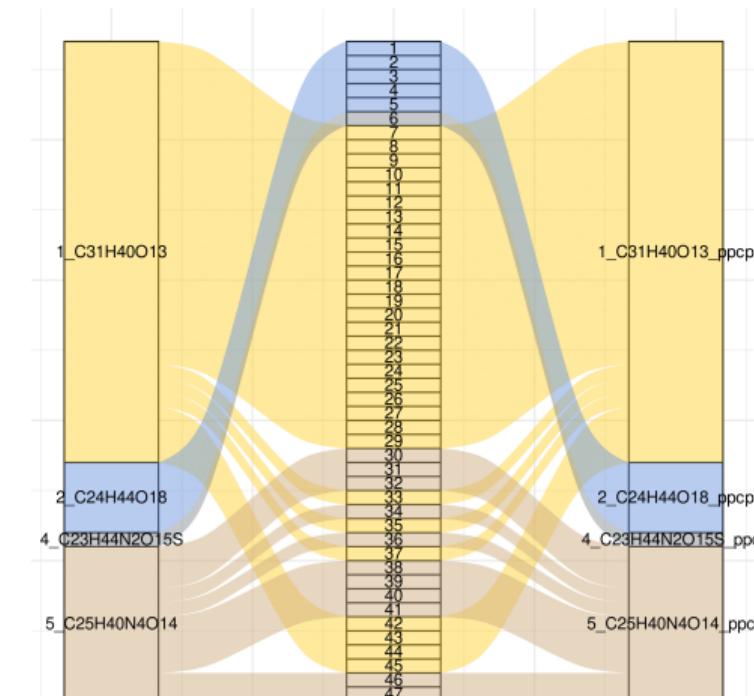


Create reference ⇒ specific candidate

Ranking methods

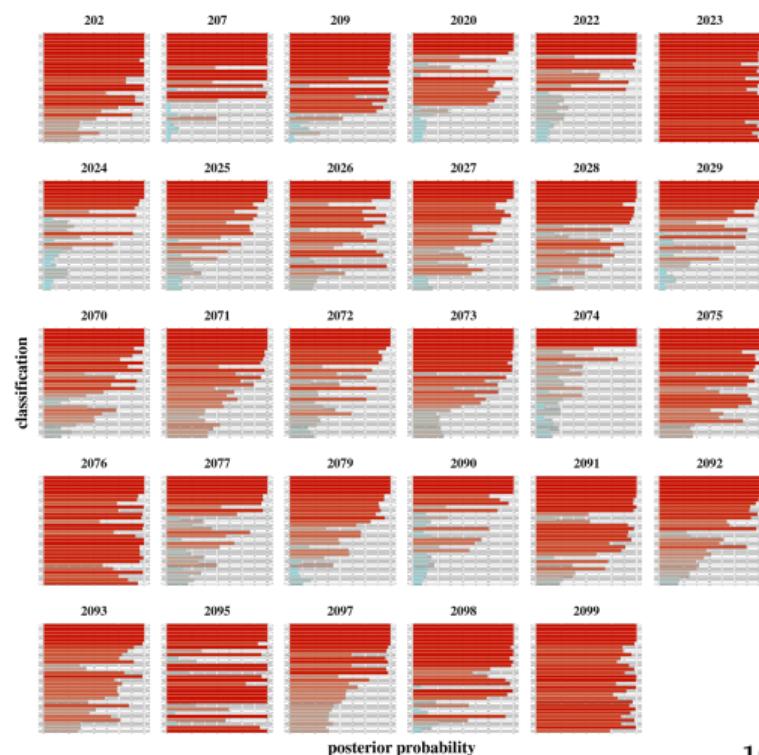
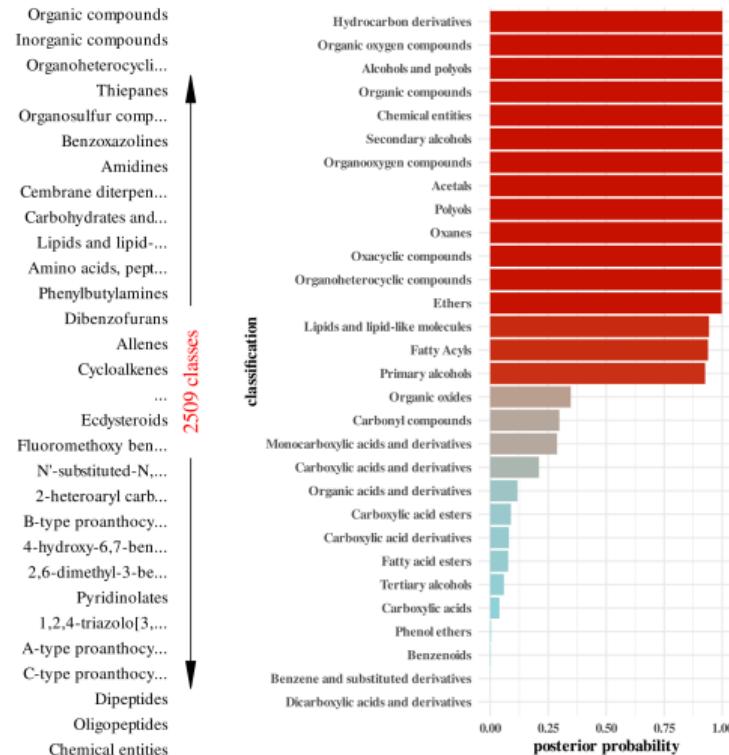
- Formulae
 - sirius.score
 - zodiac.score ◀
 - tree.score
 - iso.score
 - error.mass
 - ...
- Structure
 - csi.score ◀
 - tan.i.score
 - cosmic.score
- Classification
 - specific classes

Ranking relationship of formula, structure and classification



Filter classification

instance: ID:2095 2_C24H44O18_[M-H]- ⇒ dataset



Filter classification multiple filtering

1 Inner filter

- by posterior probability
- by class name
 - position isomerism
 - specific classes

2 Create stardust classes

3 Cross filter stardust

- Quantity
- Score
- Identical

4 Create nebula-index



Progress

Code refactoring: data processing

- S4 data structure ✓
- Extract data ✓
 - 1 project_conformation ✓
 - 2 project_metadata ✓
 - 3 project_api ✓
 - 4 project_dataset ✓
- Filter data
 - 0 collate_data ✓
 - 1 filter_formula ✓
 - 2 filter_structure ✓
 - 3 filter_ppcp ✓
 - 4 create_reference ✓

- organize data
 - 1 create_features_annotation ✓
 - 2 create_stardust_classes ✓
 - 3 cross_filter_stardust ✓
 - quantity ✓
 - score ✓
 - identical ✓
 - 4 create_nebula_index ✓
 - 5 compute_spectral_similarity ✓
 - 6 create_parent_nebula ✓
 - 7 create_child_nebulae ✓

Code refactoring: data processing

- Visualization
 - 1 Re-write in S4 object.
 - 2 Flexibility.
- Statistic analysis
 - 1 Linear regression.
 - 2 Feature selection.
 - 3 PCA etc.
- Report
 - 1 Flexibility.

Next Schedule

Overall

- Data filtering and arranging system. ✓
- Visualization system.
- Statistic system.
- Report.
- Documentation.
- Website.
- Figures for article.
- Rerite article.