

Seminar

Reporter: Lichuang Huang

2022-04-23

Supervisor: Gang Cao

1 MCnebula

2 MCnebula evaluation

3 MCnebula validation

4 E. ulmoides dataset

5 END

MCnebula

Inovation

- 1** High accuracy classes clustering for MS/MS annotation
 - over 80% accuracy clustering, even unknown compound (no structure information)
- 2** Intuitive compound classes distribution in network visualization
 - each class involves a sub-nebula to explore the compound annotation
- 3** MCnebula algorithm integrated in R
 - cover SIRIUS LC-MS workflow analysis into R pipeline
- 4** A wide range of applicability
 - not be confined to metabolome identification
 - not be confined to spectrum library, but structure library

MCnebula
oo

MCnebula evaluation
●oooooooo

MCnebula validation
oooooooooooo

E. ulmoides dataset
oooo

END
o

MCnebula evaluation

Compare with GNPS and MolNetEnhancer

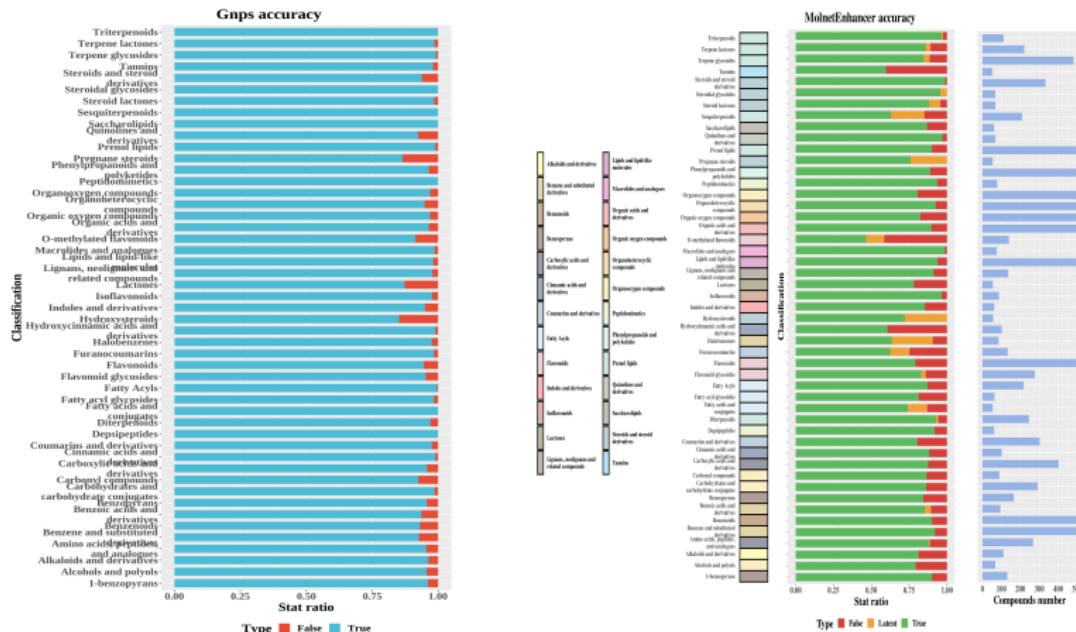


Figure 1: MolNetEnhancer

COSMIC Noise design

- We simulated a global mass shift (bias) by drawing a random number δ^* from $\mathcal{N}(0, \sigma_{mb}^2)$ and then shifting every peak mass m by $\delta^* m$. The standard deviation σ_{mb} was chosen as $\sigma_{mb} = (10/3) \times 10^{-6}$ (medium noise) or $\sigma_{mb} = (15/3) \times 10^{-6}$ (high noise), so that the $3\sigma_{mb}$ interval represents a 10-ppm shift for medium noise and a 15-ppm shift for high noise.
- We simulated individual mass deviations by drawing, for each peak with mass m individually, a random number δ from $\mathcal{N}(0, \sigma_{md}^2)$ and shifting the peak by δm . The standard deviation σ_{md} was chosen so that the $3\sigma_{md}$ interval represents a 10-ppm shift for medium noise and a 20-ppm shift for high noise.
- We simulated intensity variations in the spectrum: each peak intensity was multiplied by an individual random number ϵ drawn from $\mathcal{N}(1, \sigma_{id}^2)$. Variance was chosen as $\sigma_{id}^2 = 1$ for medium noise and $\sigma_{id}^2 = 2$ for high noise. Furthermore, 0.03 times the maximum peak intensity of the spectrum was subtracted from each peak intensity. If a peak intensity fell below the threshold of one thousands of the maximum intensity in the spectrum, the peak was discarded.
- Finally, we added ‘noise peaks’ to the spectrum. As uniformly choosing the mass of a noise peak would result in peaks that are too easy to spot and sort out by our subsequent analysis¹², we, instead, used peaks that appeared in other measured spectra. In pre-processing, a pool of ‘noise peaks’ was gathered from the fragmentation spectra, using all peaks that did not have a molecular subformula decomposition of the known molecular formula of the precursor. For each spectrum, αn of these ‘noise peaks’ were added to the spectrum, where n is the number of peaks in the spectrum and $\alpha = 0.2$ for medium noise and $\alpha = 0.4$ for high noise. Intensities of ‘noise peaks’ were adjusted for maximum peak intensities in the contributing and receiving spectrum.

Add noise: main method

- Mass global shift
- Mass individual shift
- Intensity shift
- Add noise peak from other peak

Noise model

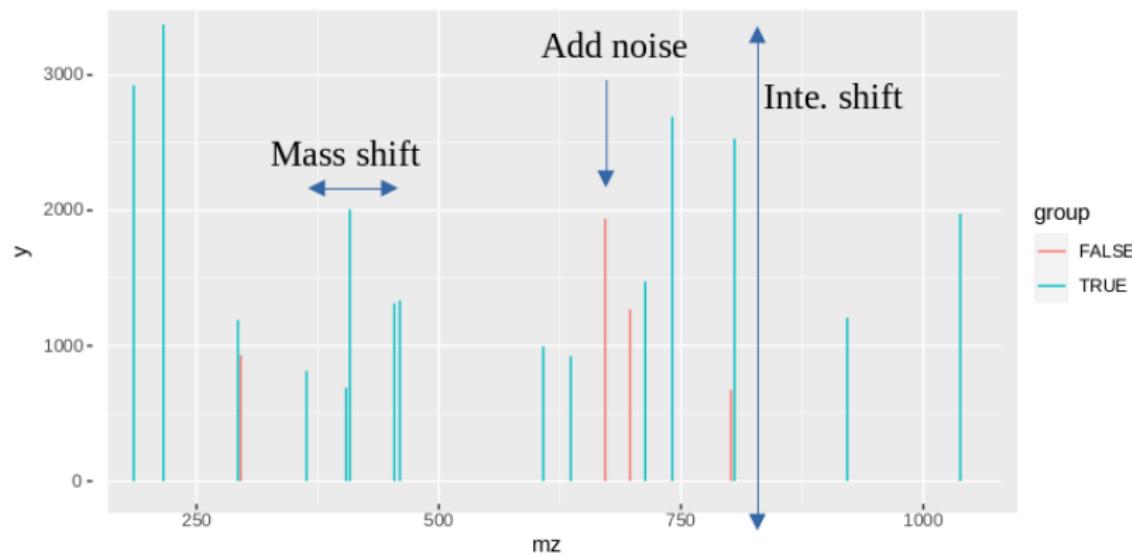
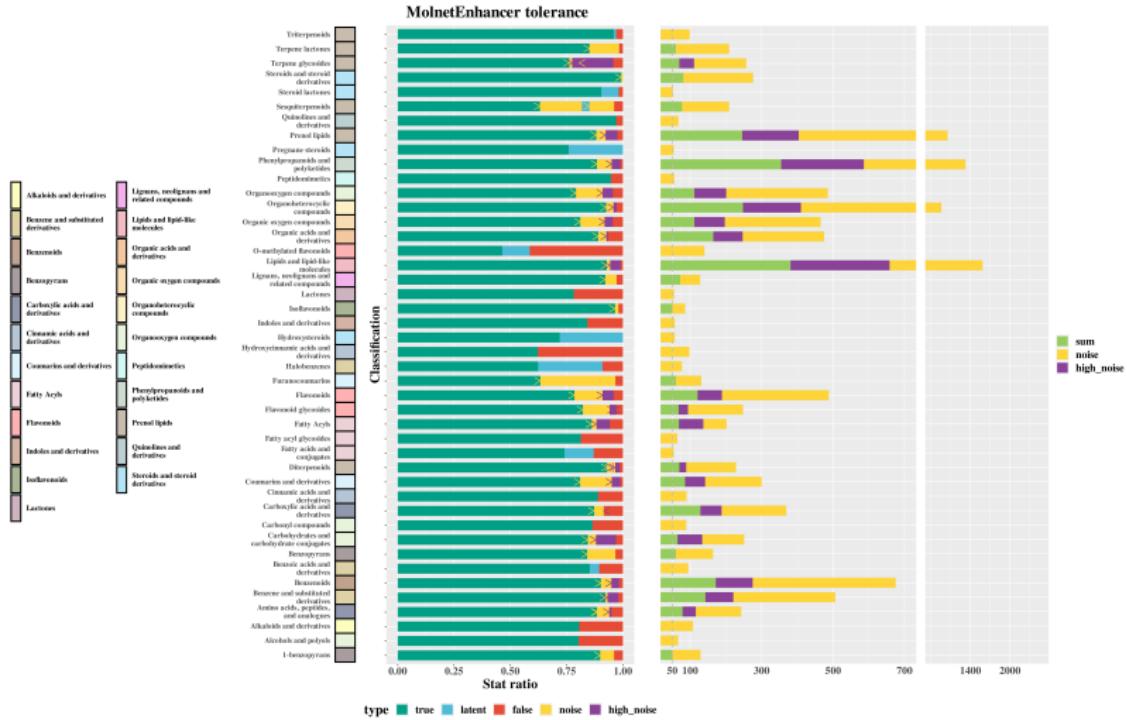
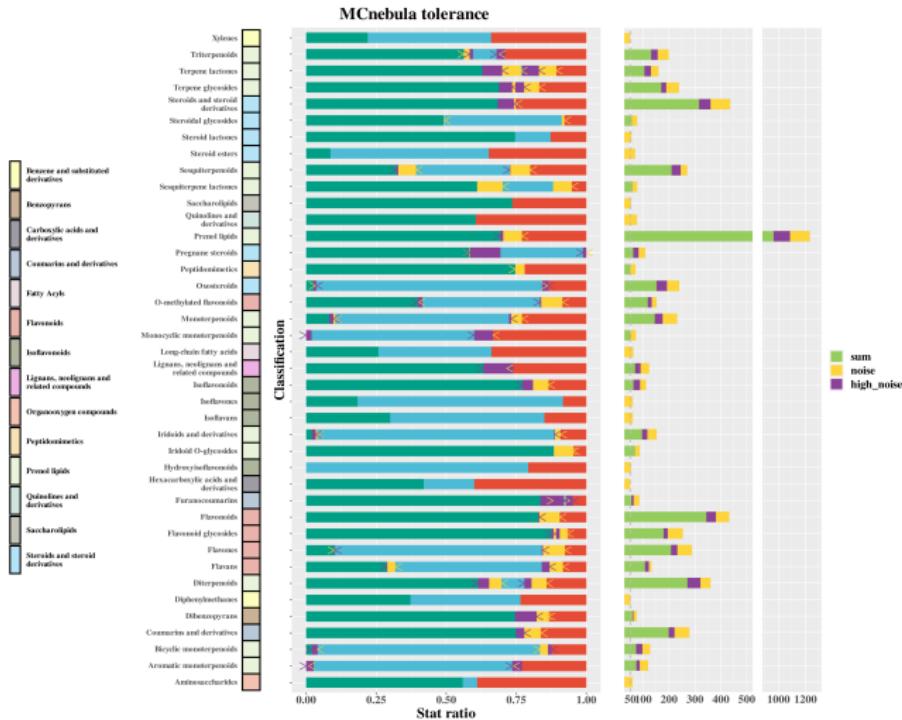


Figure 3: noise model

Results: MolNetEnhancer



Results: MCnebula



MCnebula
oo

MCnebula evaluation
oooooooo

MCnebula validation
●oooooooo

E. ulmoides dataset
oooo

END
○

MCnebula validation

Re-analyse a research

Resource

Mortality Risk Profiling of *Staphylococcus aureus* Bacteremia by Multi-omic Serum Analysis Reveals Early Predictive and Pathogenic Signatures

Jacob M. Wozniak,^{1,2,3,4} Robert H. Mills,^{1,2,3,4,5,6,7} Joshua Olson,^{2,3,5} J.R. Caldera,^{3,5,9} Gregory D. Sepich-Poore,⁶ Marvic Carrillo-Terrazas,^{1,2,3,4} Chih-Ming Tsai,^{3,5} Fernando Vargas,^{2,8} Rob Knight,^{3,4,5,6,7} Pieter C. Dorrestein,^{2,4} George Y. Liu,^{3,6} Victor Nizet,^{2,3,5} George Sakoulas,^{3,5} Warren Rose,^{10,11} and David J. Gonzalez^{1,2,3,4,12,*}

¹Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093, USA

²Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA

³Collaborative to Halt Antibiotic-Resistant Microbes, University of California, San Diego, La Jolla, CA 92093, USA

⁴Center for Microbiome Innovation, University of California, San Diego, La Jolla, CA 92093, USA

⁵Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA

⁶Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA

⁷Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA 92093, USA

⁸Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA

⁹Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

¹⁰School of Pharmacy, School of Medicine and Public Health University of Wisconsin-Madison, Madison, WI 53705, USA

¹¹Department of Medicine, School of Medicine and Public Health University of Wisconsin-Madison, Madison, WI 53705, USA

¹²Lead Contact

*Correspondence: djgonzalez@ucsd.edu

<https://doi.org/10.1016/j.cell.2020.07.040>

Figure 6: an article of cell

Analysis workflow (part 1)

- 1 Overview of Multi-omic data
- 2 Find High-Confidence Biomarkers
 - EFS algorithm for Proteomics, Metabolomics
- 3 Post-translational modifications (PTMs)-Tolerant analysis...
 - GNPS molecular networking, function analysis...
- 4 K-means clustering
 - for Proteomics
 - function analysis
 - for Metabolomics (GNPS MN, library match, CSI:fingerID, classyfire)
 - acyl-carnitines, steroid derivatives, indoles

Analysis workflow (part 2)

5 Data integration

- LASSO logistic classification algorithm
 - SERPIND1
 - thyroxine (T4)
- mixOmics R: protein-metabolite relationships

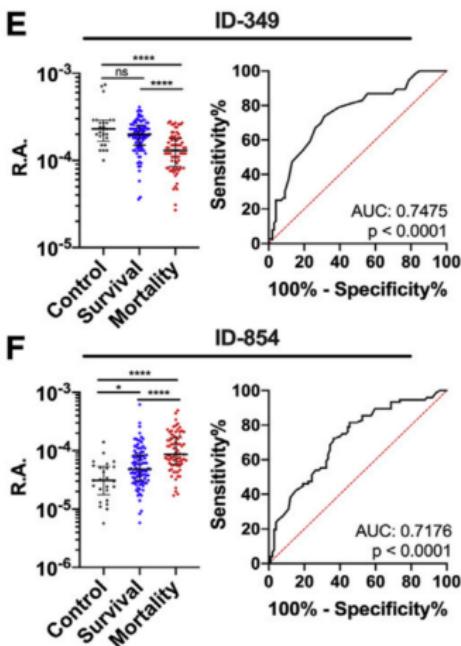
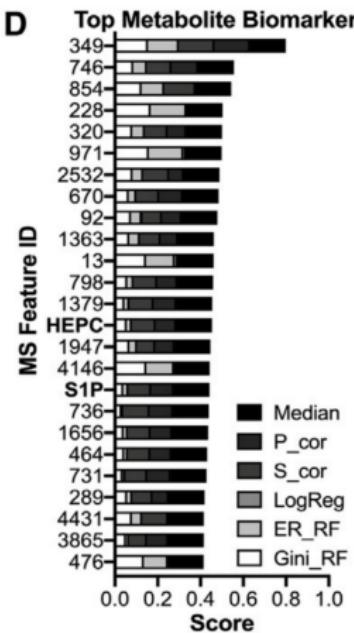
6 Global evaluation

- IGF signalling

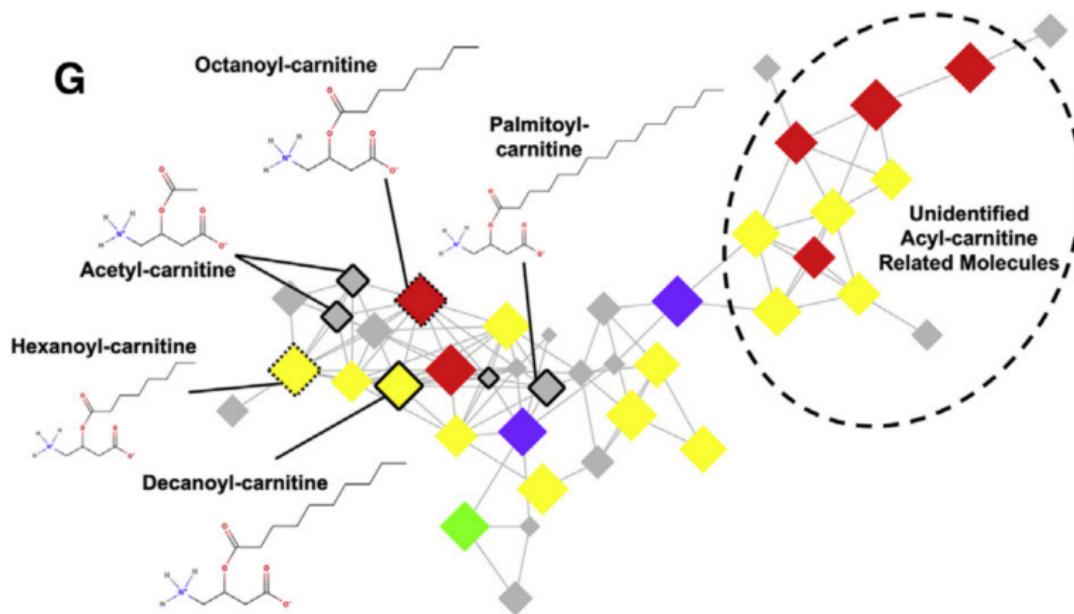
7 Knowledge-Based Analysis (Underlying Cytokine)

8 In vivo validation

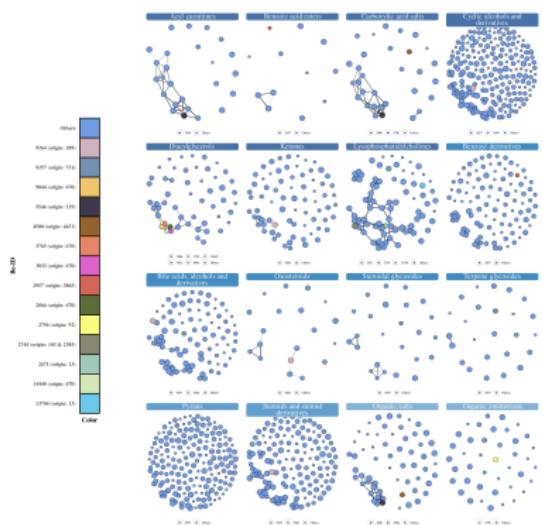
key partion: identification of top feature



key partition: identification in classes

G

Re-analyse via MCnebula



Lysophosphatidylcholines

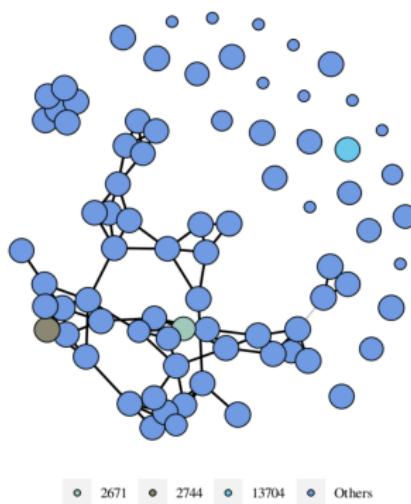
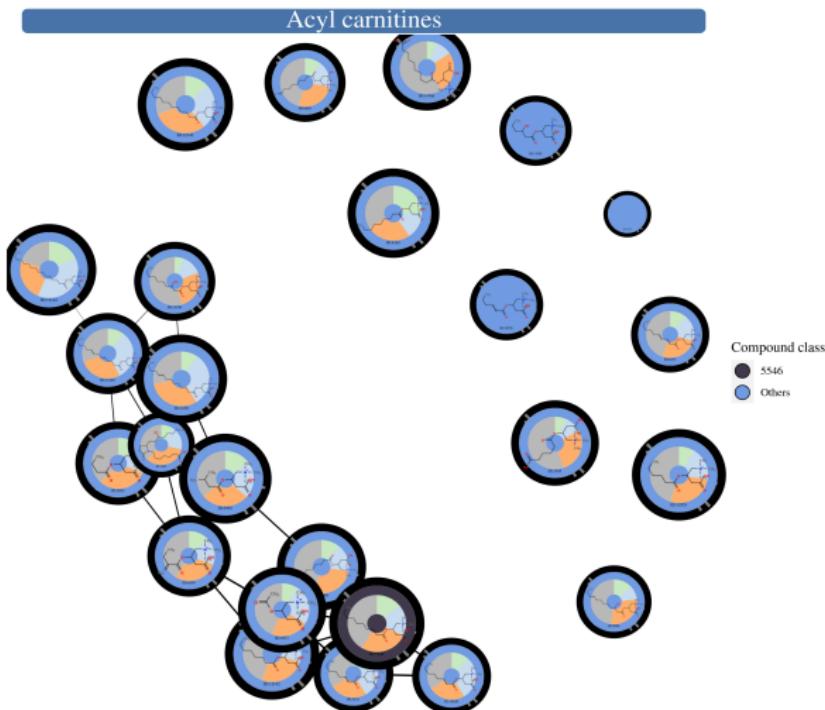


Figure 9: bio-tracing

ACs identification



Identification of T4

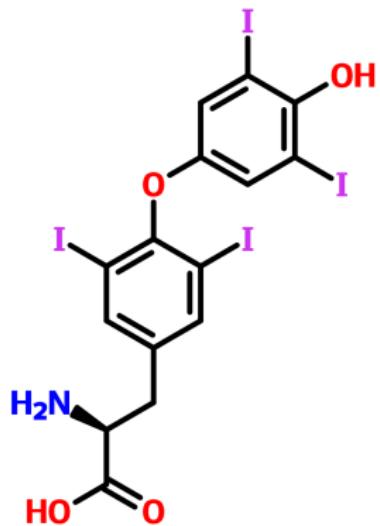


Figure 11: T4

MCnebula
oo

MCnebula evaluation
oooooooo

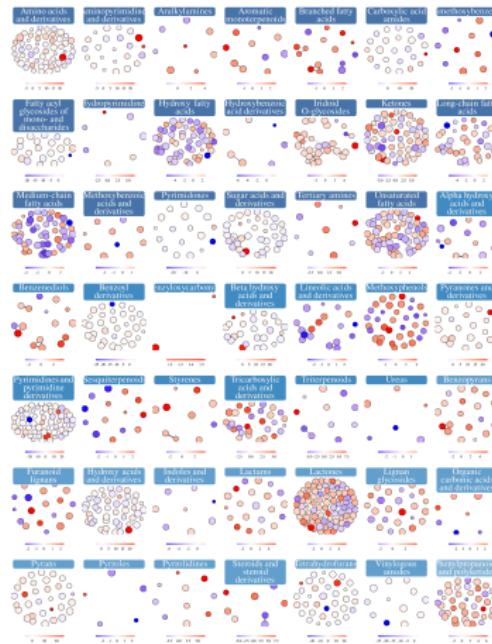
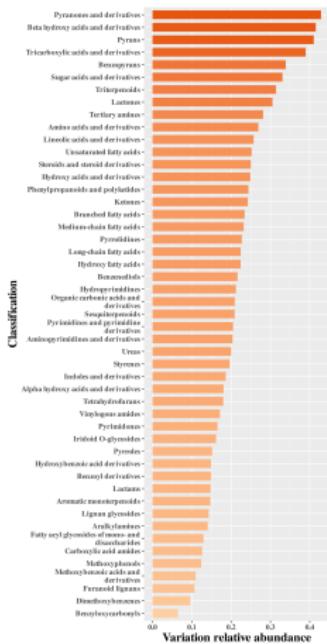
MCnebula validation
oooooooooooo

E. ulmoides dataset
●ooo

END
○

E. ulmoides dataset

Compounds with $|\log_2(\text{FC})| > 1$



Compounds identified summary

E. ulmoides compounds summary¹

LC-MS in negative ion mode

No	Name ²	Id ³	Precursor m/z	Mass error (ppm) ⁴	RT (min)	Formula ⁵	Tanimoto similarity ⁶	Variation ⁷	InChIKey planar ⁸
Alcohols and polyols									
1	(1R,3R,4S,5R)-3-[(E)-3-(3,4-dihydroxyphenyl)prop-2-enoyl]oxy-4-[(E)-3-(3,4-dihydroxyphenyl)prop-2-enyl]peroxy-1,5-dihydroxycyclohexane-1-carboxylic acid	3005	517.1344	-1.2627399	18.7	C25H26O12	0.5648323	1111	CVBGAIVSRYDJKM
2	(1R,3R,4S,5R)-4-[(E)-3-(3,4-dihydroxyphenyl)prop-2-enoyl]oxy-1,3-dihydroxy-5-[(E)-3-(4-hydroxyphenyl)prop-2-enoyl]oxycyclohexane-1-carboxylic acid	2701	499.1240	-1.1267325	18.6	C25H24O11	0.5999618	-	WMSGJFOCCDGOK
3	(1S,3R,4R,5R)-3-[(E)-3-[2-[2-[(E)-3-[(1R,2R,3R,5S)-5-carboxy-2,3,5-trihydroxycyclohexyl]oxy-3-oxoprop-1-enyl]-4,5-dihydroxyphenyl]-4,5-dihydroxyphenyl]prop-2-enoyl]oxy-1,4,5-trihydroxycyclohexane-1-carboxylic acid	4801	705.1667	-0.8827578	9.7	C32H34O18	0.6064440	-	UCIQBCLGNROEGN
4	(2-hydroxy-2-oxo-1,3,2-lambda5-dioxaphospholan-4-yl)methyl hexadec-9-ynoate	4268	387.1912	-7.6559595	20.6	C19H33O6P	0.5414530	-	YGKAQLJCSMOCQJ
5	(2S,3S,6S)-3-hydroxy-6-propan-2-yloxyxane-2-carboxylic acid	2955	203.0917	-4.1582093	2.3	C9H16O5	0.5163548	-	IWYCTGOIGRELFV
6	(3Z)-3-(2-hydroxycyclobutylidene)-2-methylpropanoic acid	1440	155.0703	-7.2741248	10.0	C8H12O3	0.5088873	-	OQEYAWOJ1QEKFQN
7	1,10-dihydroxy-1,2,6b,9,9,12a-hexamethyl-3,4,5,6,6a,6a,7,8,8a,10,11,12,13,14b-tetradecaahydro-2H-picene-4a-carboxylic acid	3138	457.3316	-1.1984657	25.2	C29H46O4	0.6570784	1	FBVNLSSRTTDAPMX
8	14-hydroxy-5,14-dimethyltetraacyclo[11.2.1.0,10.0,4,9]hexadecane-5,9-dicarboxylic acid	2654	349.1998	-5.9620093	21.2	C20H30O5	0.5341241	-	CDKNUBLSEZJSAU
9	2-(1-cyclobutylethyl)propanedioic acid	282	185.0809	-5.4547976	6.3	C9H14O4	0.5324641	-	IRIBCBZPYUOOSN

Schedule

- MCnebula
 - 1 collate data and figures
 - 2 revise article

MCnebula
oo

MCnebula evaluation
oooooooo

MCnebula validation
oooooooooooo

E. ulmoides dataset
oooo

END
●

END