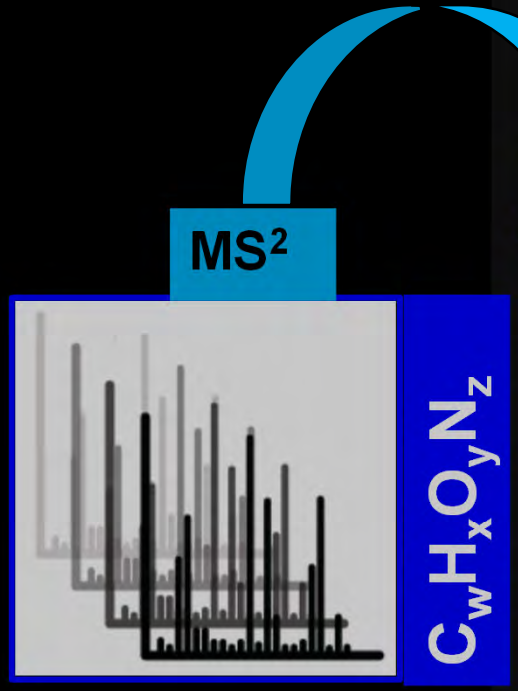
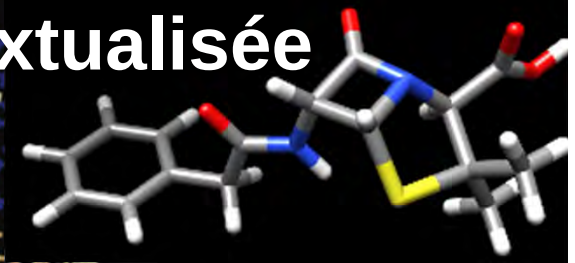
Les Diablerets, 16 Sept 2021

Journées scientifiques
Labo 2030, si loin, si près

CCCTA

Comment identifier un métabolite sans accès à une base de donnée dédiée en MS ? Des pistes pour les produits naturels par une approche métabolomique contextualisée





Jean-Luc Wolfender,
Pierre Marie Allard, Adriano Rutz, Arnaud Gaudry,
Luis Guerrero, Louis Felix Nothias
School of Pharmaceutical Sciences,
University of Geneva, Switzerland

ISPSO
INSTITUT DES SCIENCES PHARMACEUTIQUES DE SUISSE OCCIDENTALE

UNIVERSITÉ DE GENÈVE

Looking at a natural extract with LC-MS based metabolite profiling

Existing ~unknown

In silico prediction

Known NPs ~300'000

MS/MS data ~25'000

with

orking

OMICS

Or the full monty
2D NMR
CD
X ray…

NPs

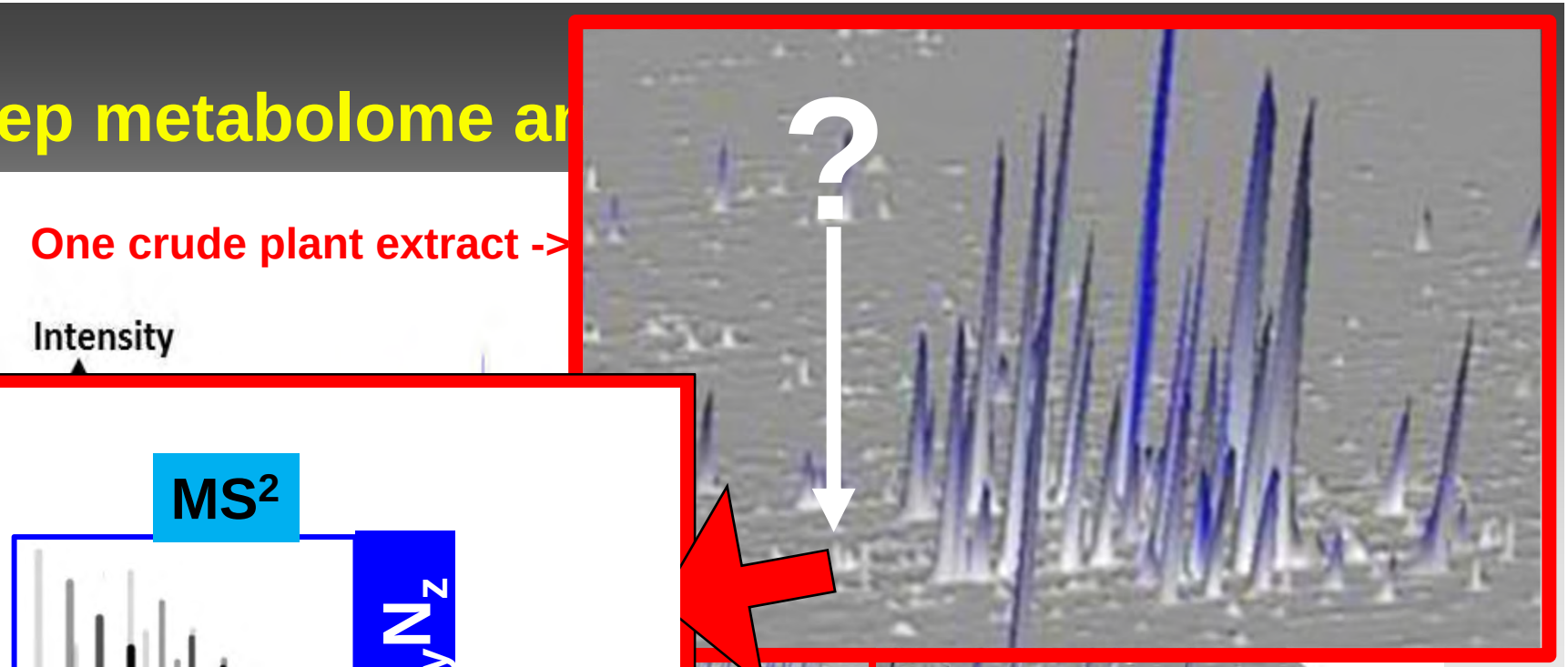
Ridder, L., et al. *Anal. Chem.* **2014**, *86*, 4767-4774.

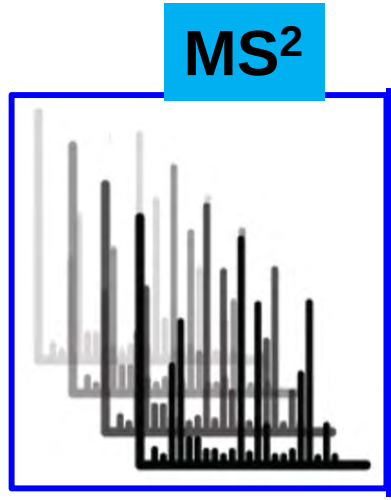
Deep metabolome an

One crude plant extract ->

Intensity

?

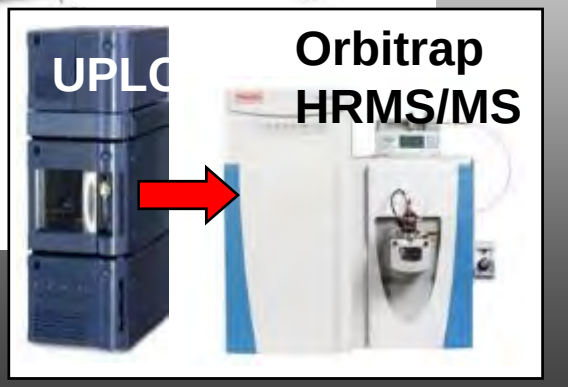




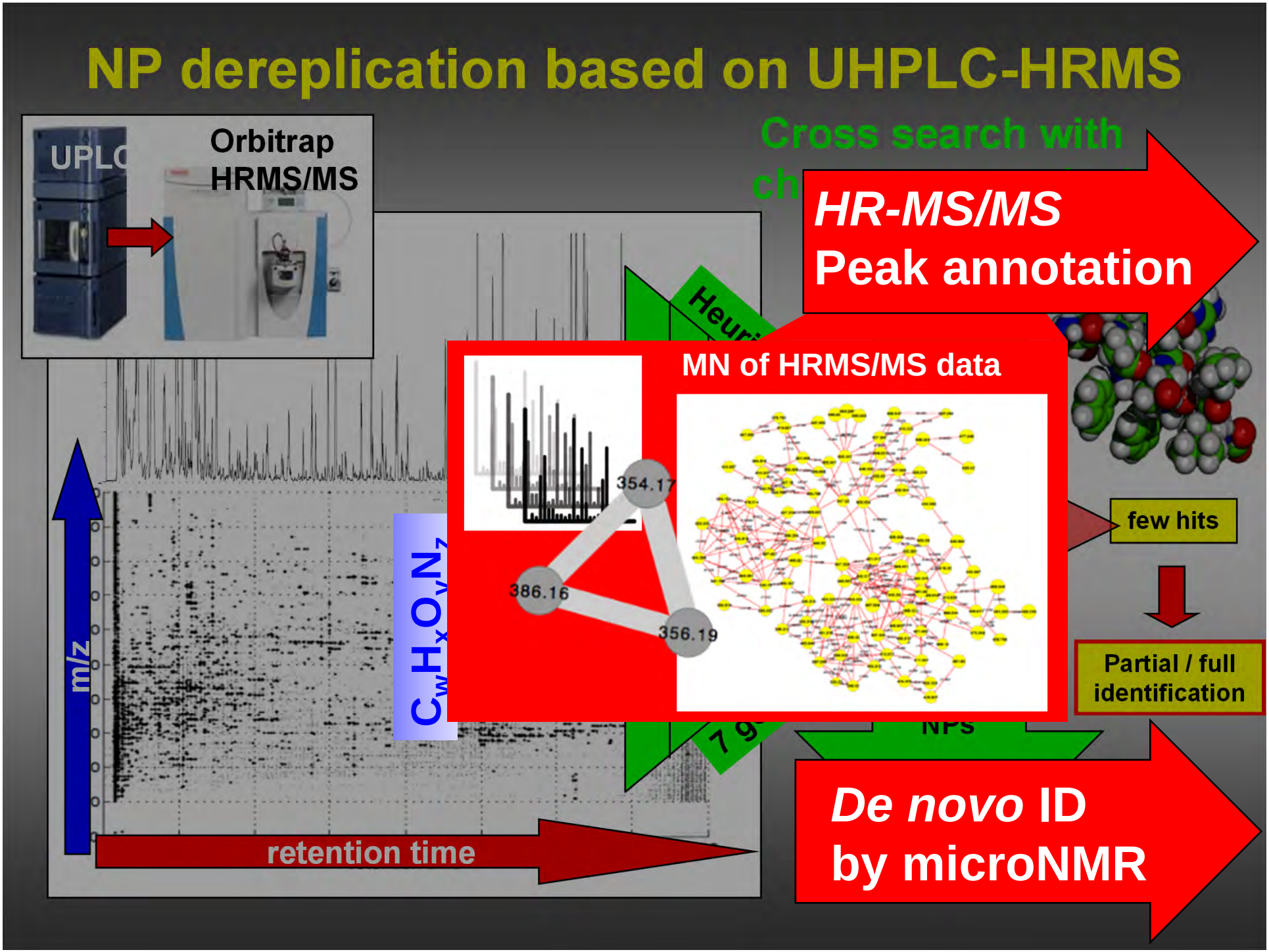
metabolite spectral fingerprints

15
Retention Time

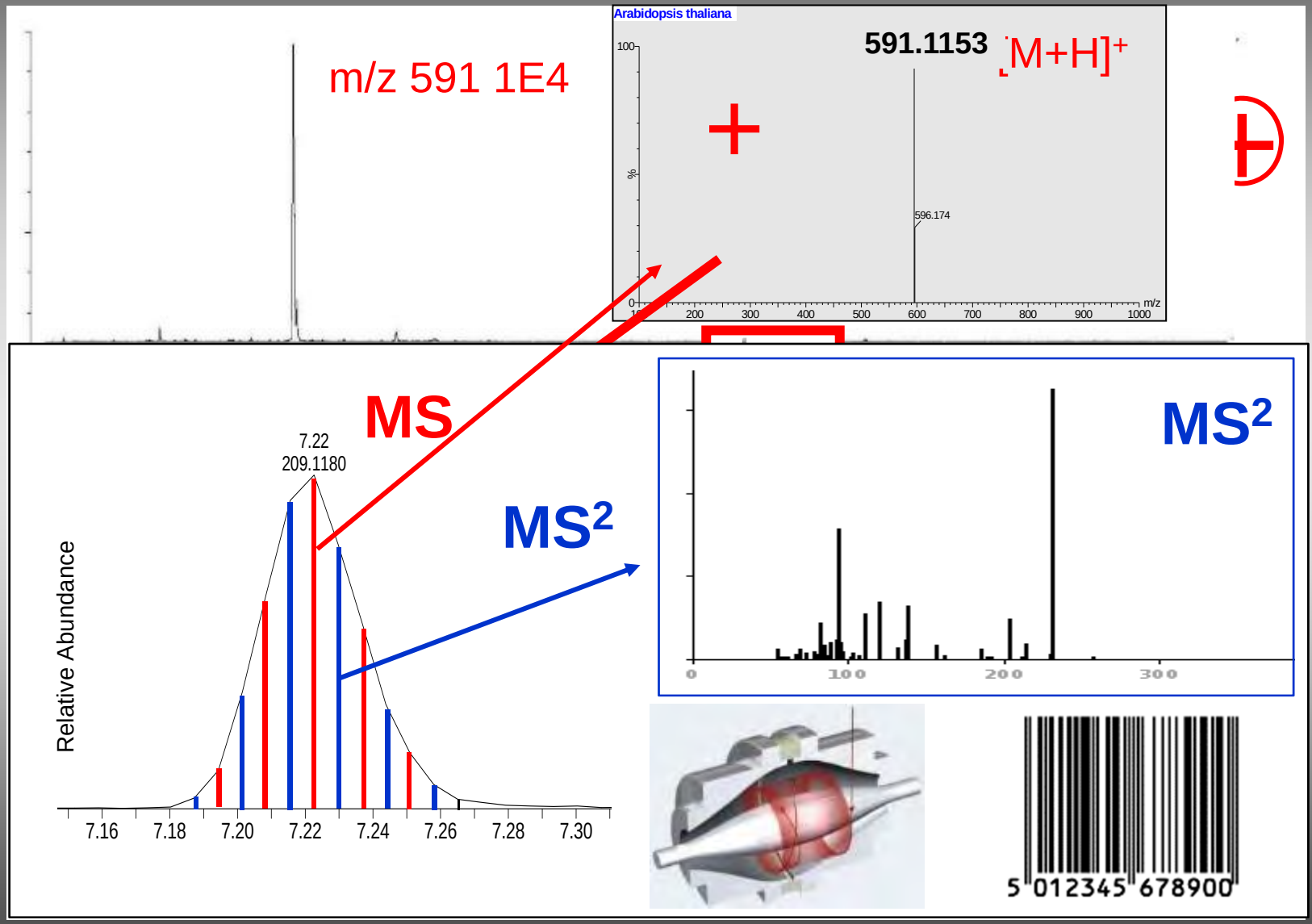
25



NP dereplication based on UHPLC-HRMS



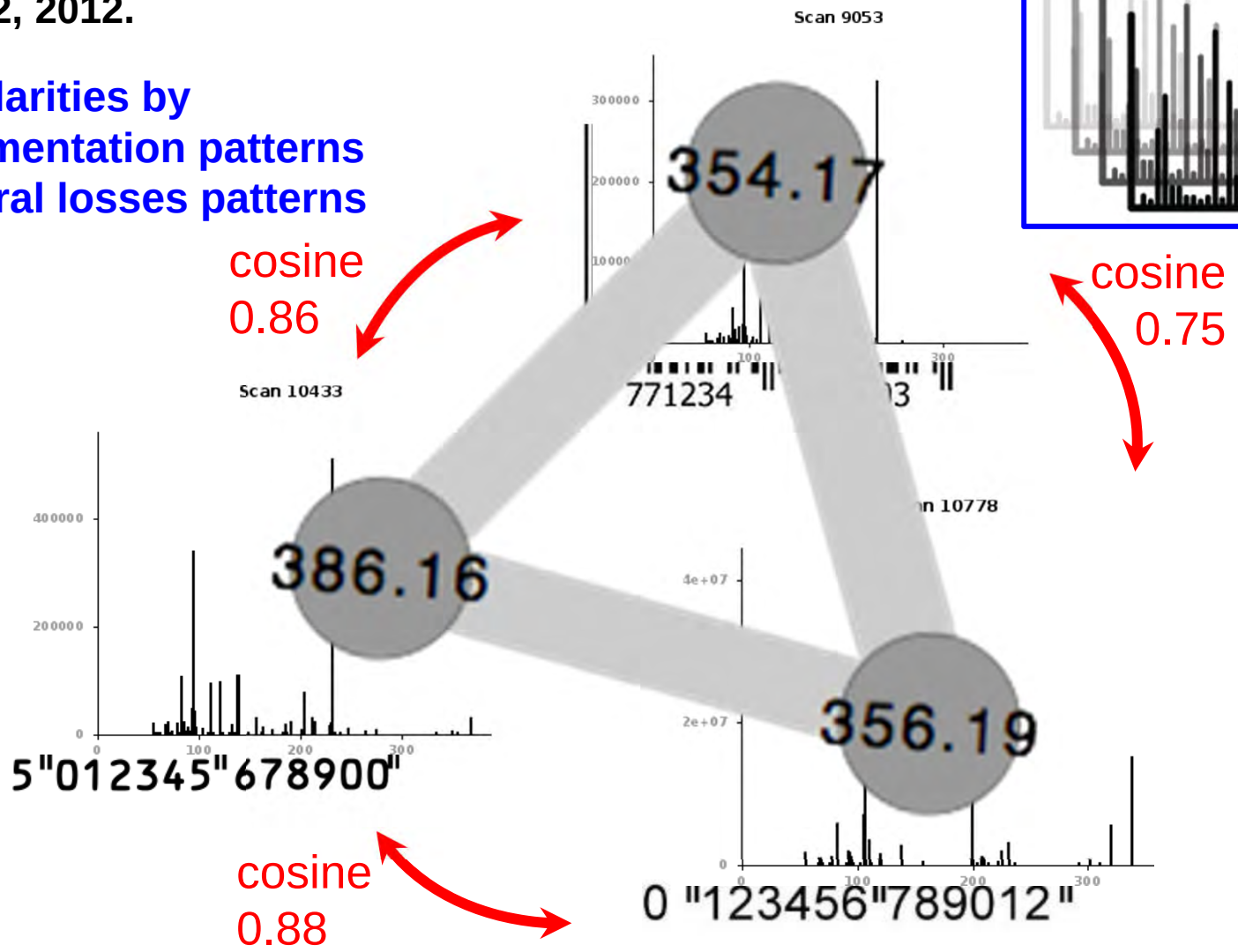
Information from UHPLC-HRMS^n profiling

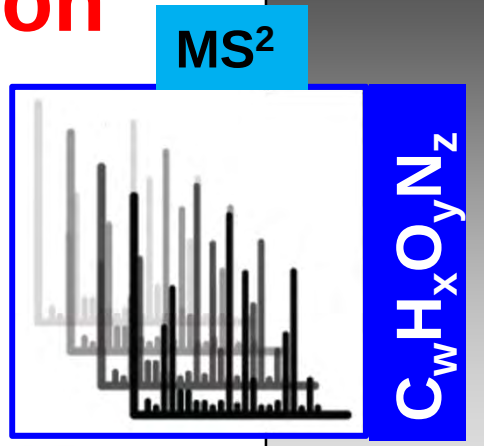


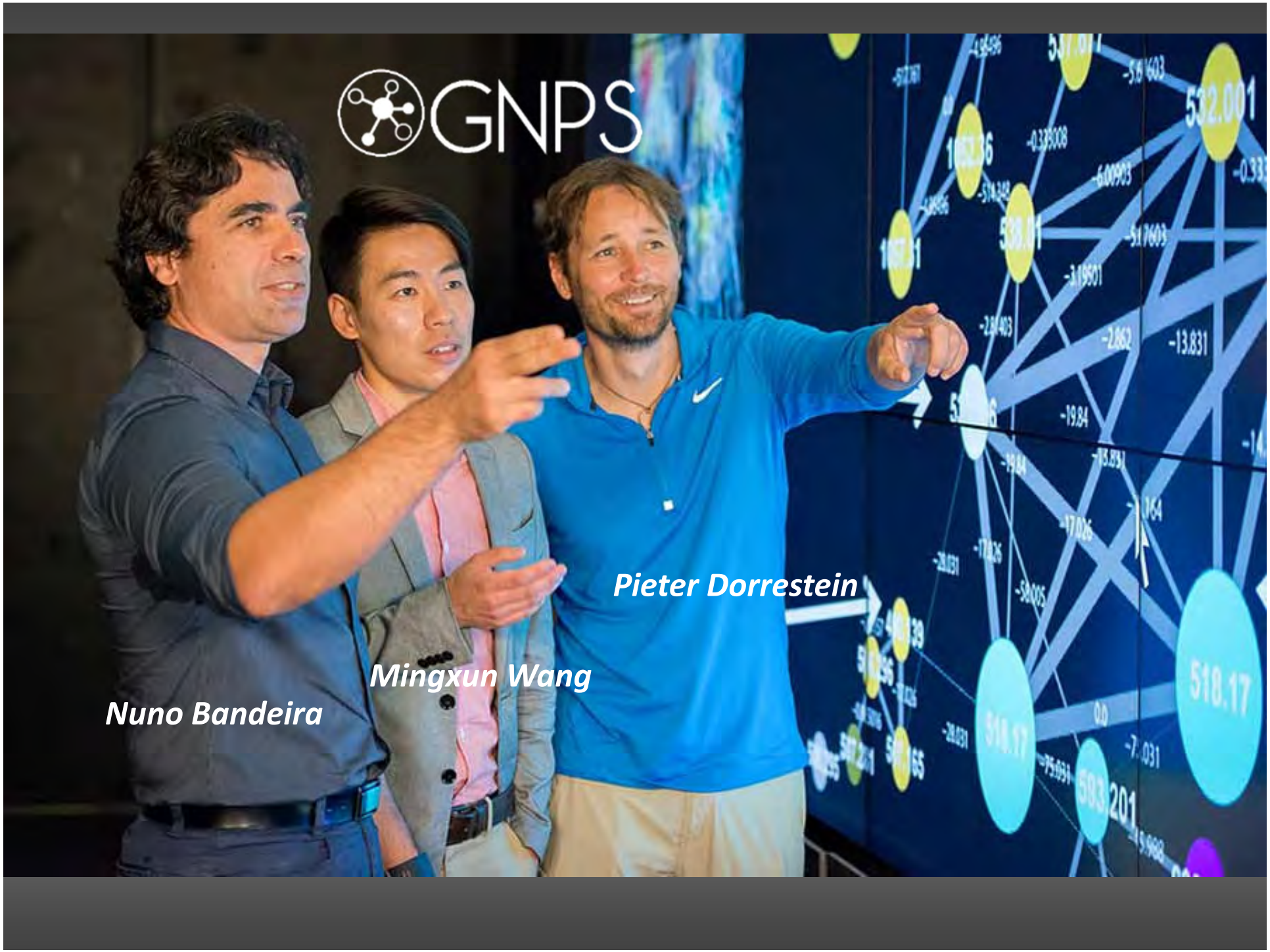
MS/MS network for interpretation

J. Watrous et al. “Mass spectral molecular networking of living microbial colonies.,” *PNAS*, vol. 109, no. 26, pp. E1743–52, 2012.

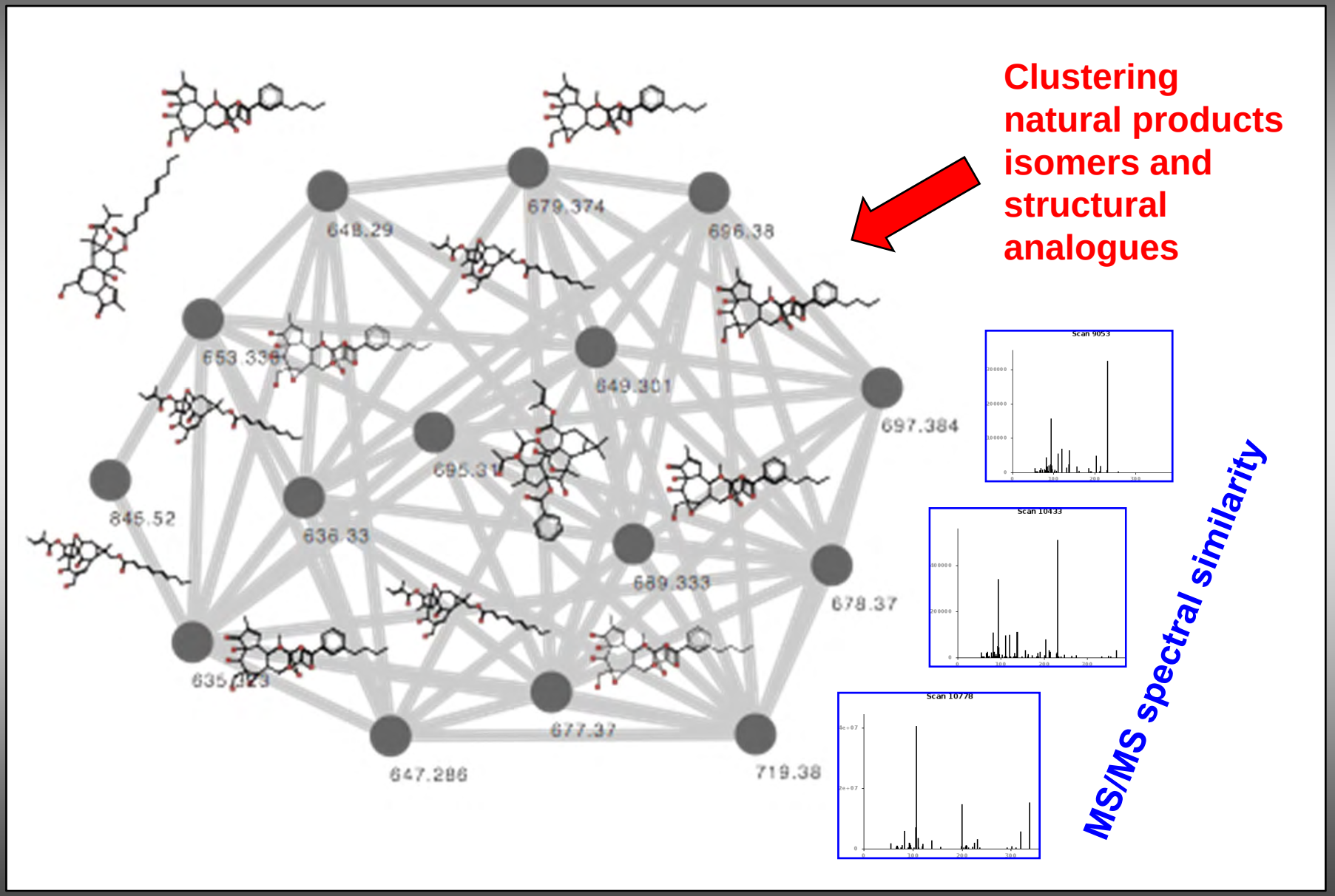
Similarities by
fragmentation patterns
neutral losses patterns





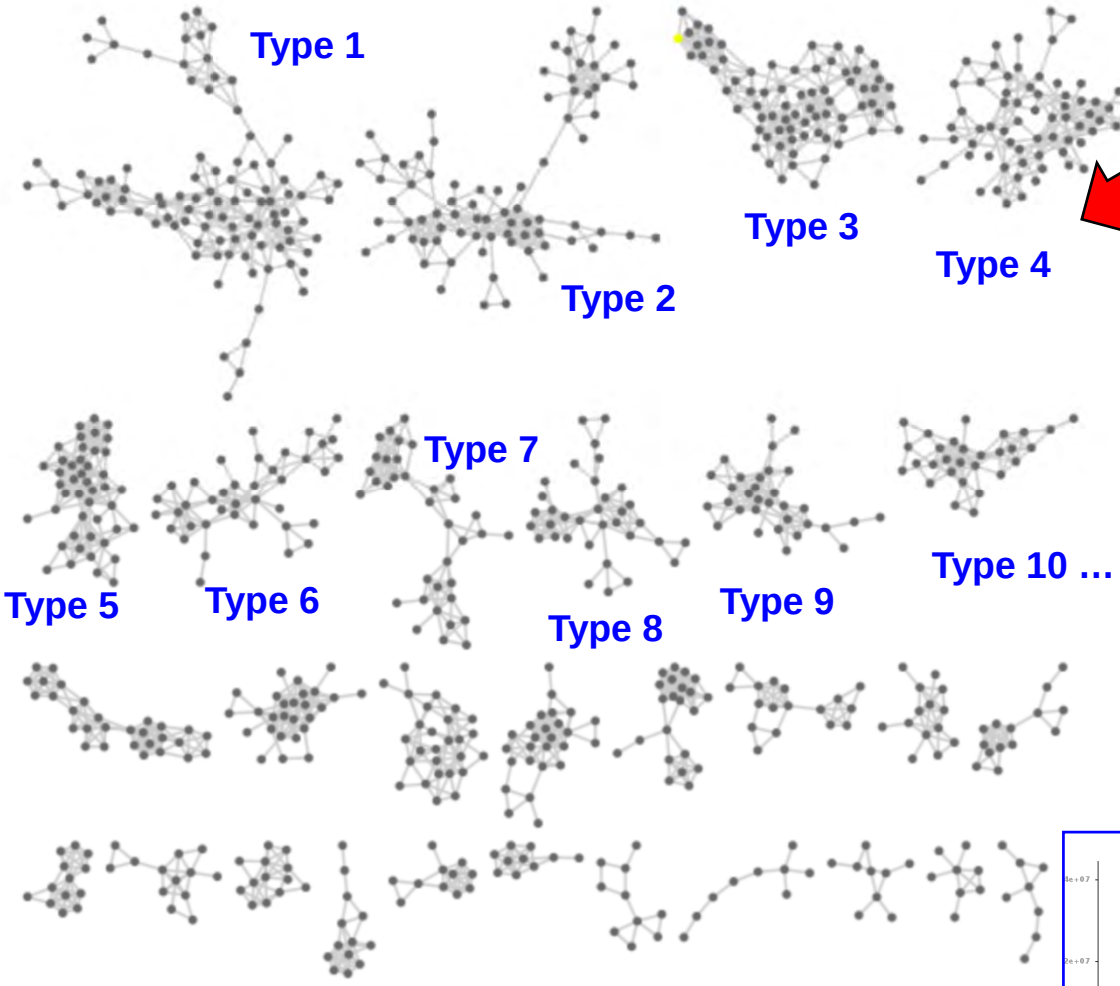

GNPS
Pieter Dorrestein
Mingxun Wang
Nuno Bandeira

MS/MS molecular networking

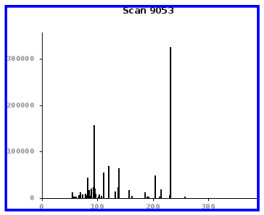


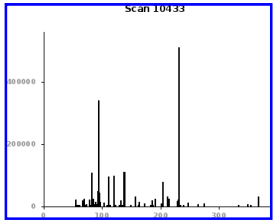
MS/MS molecular networking

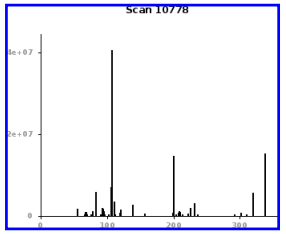
Different type of NPs



Clustering natural products isomers and structural analogues

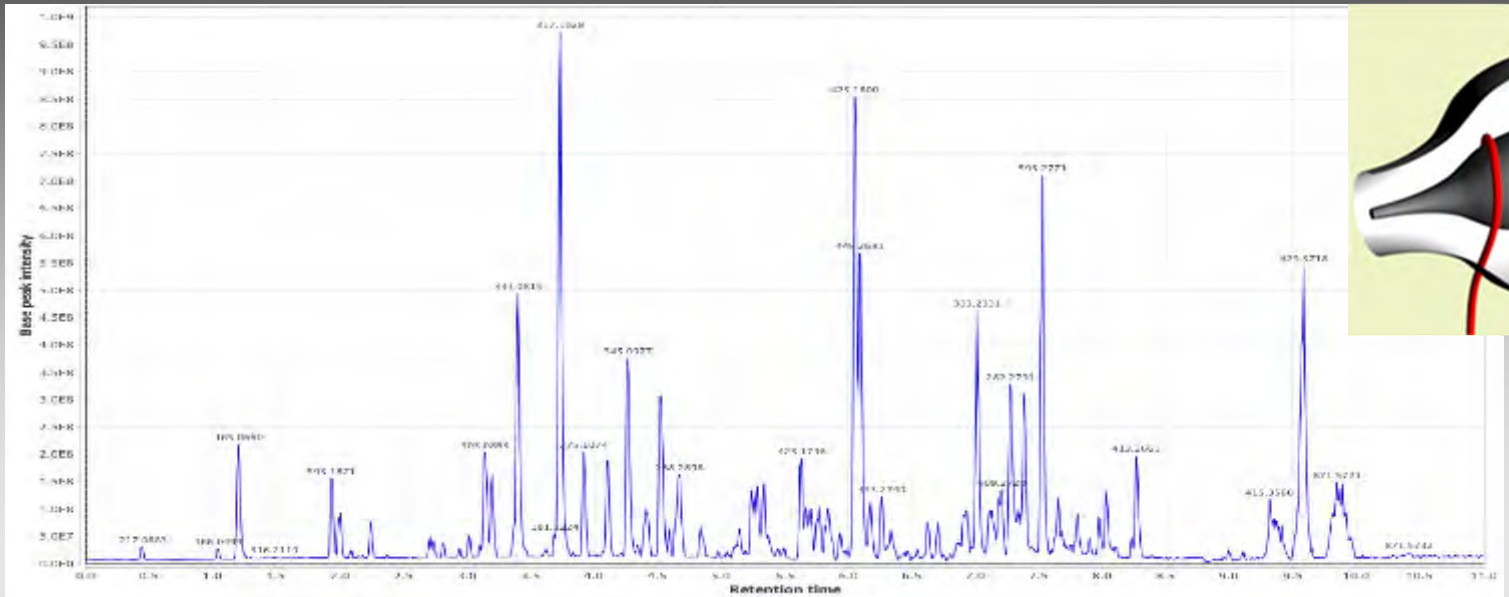


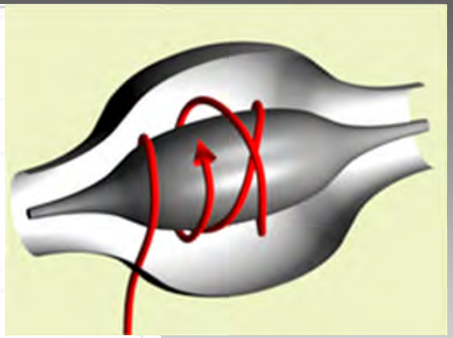


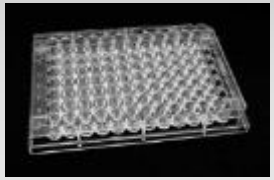


MS/MS spectral similarity

Screening of Euphorbiaceae for chemodiversity

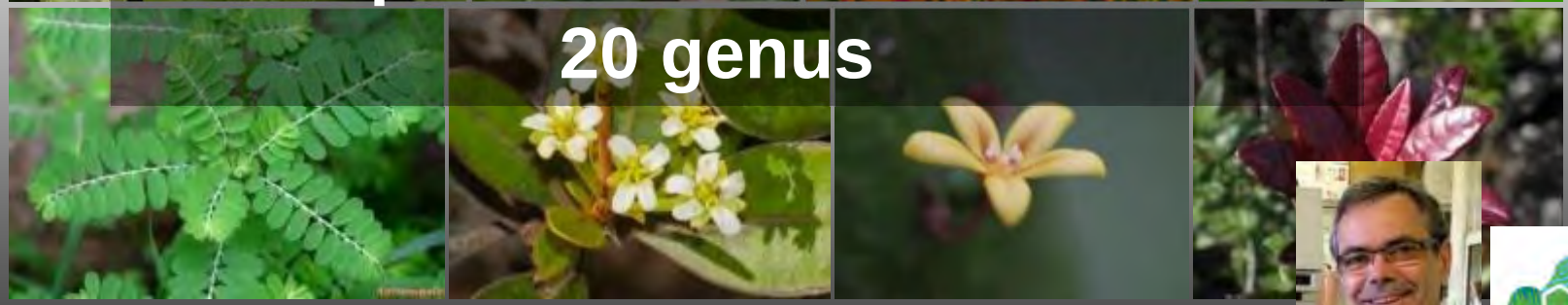






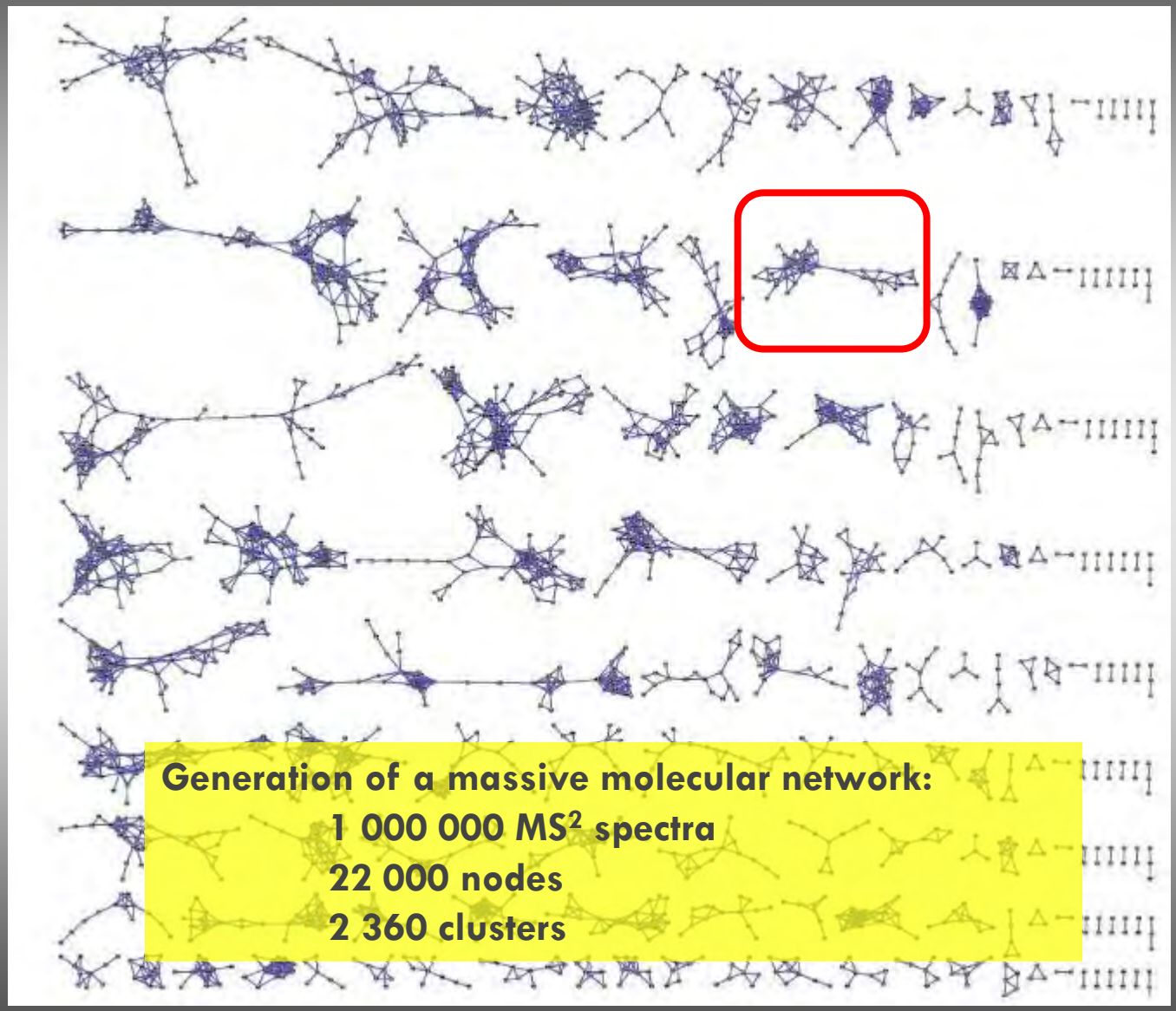
290 Euphorbiaceae EtOAc extracts

20 genus

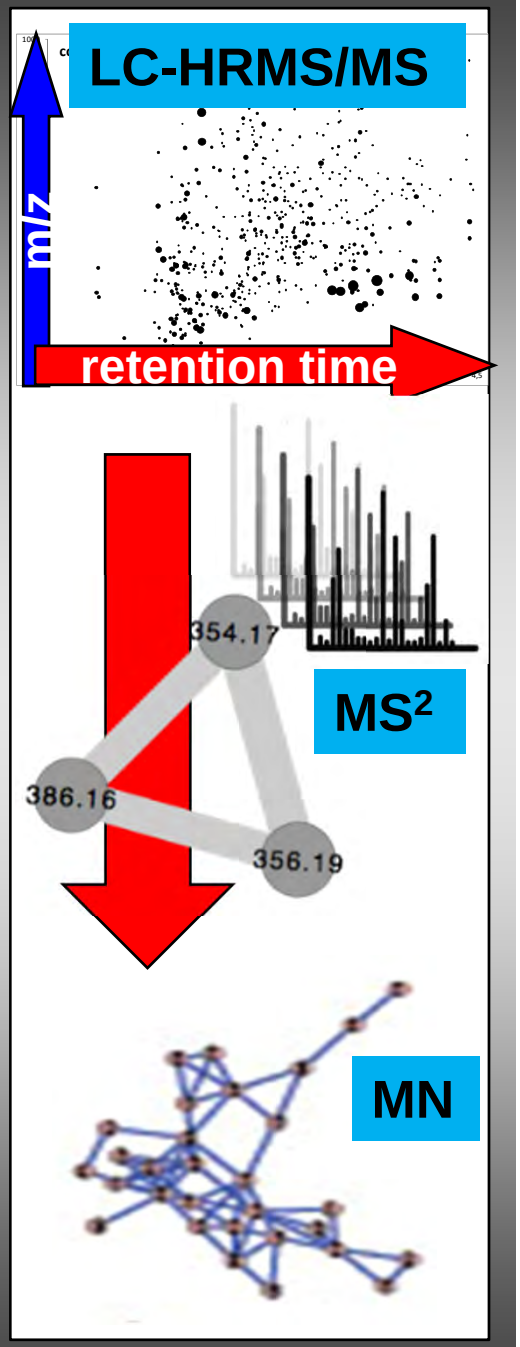


Collaborations with Marc LITAUDON (ICSN – CNRS)

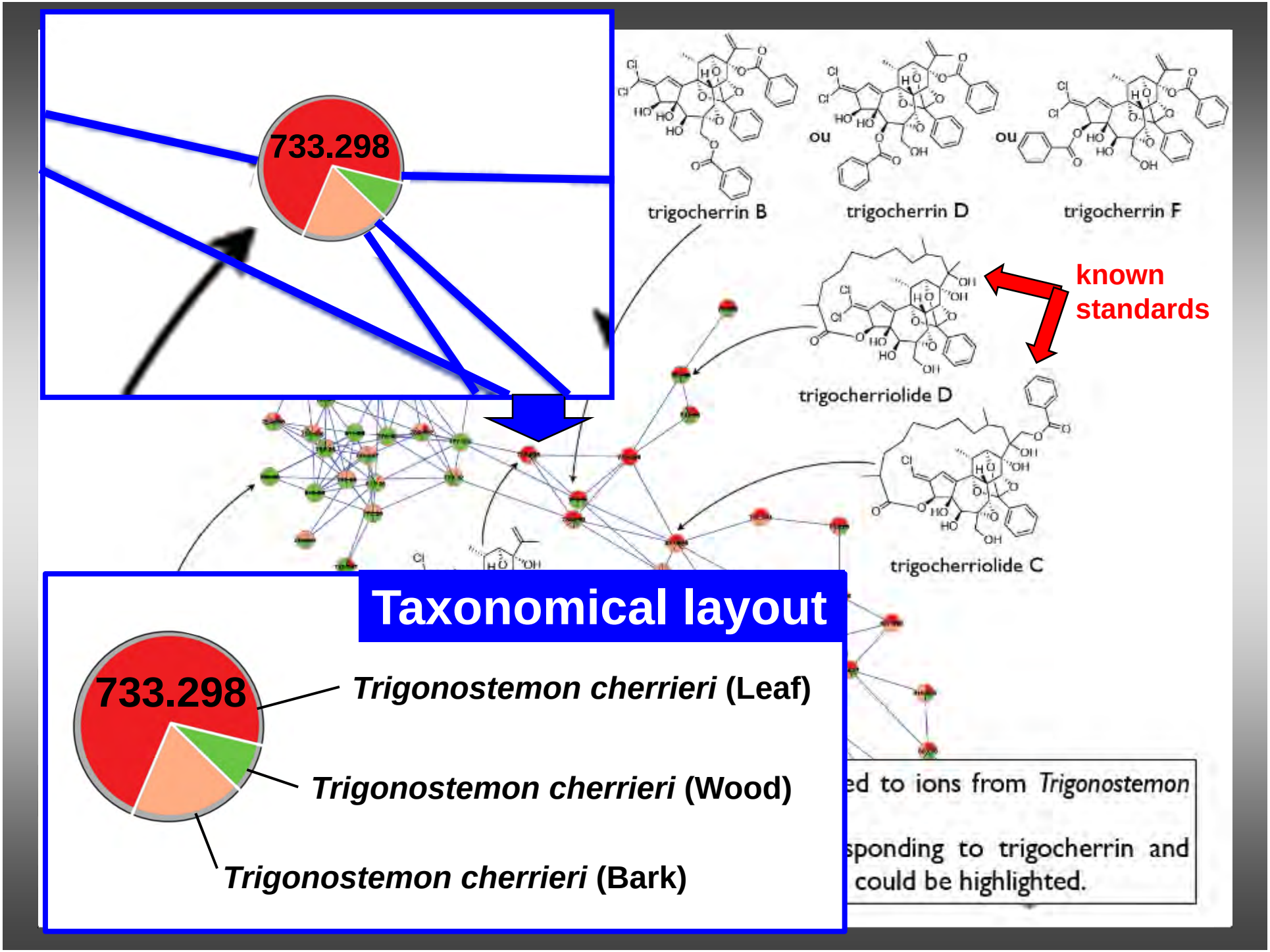
MS/MS molecular network



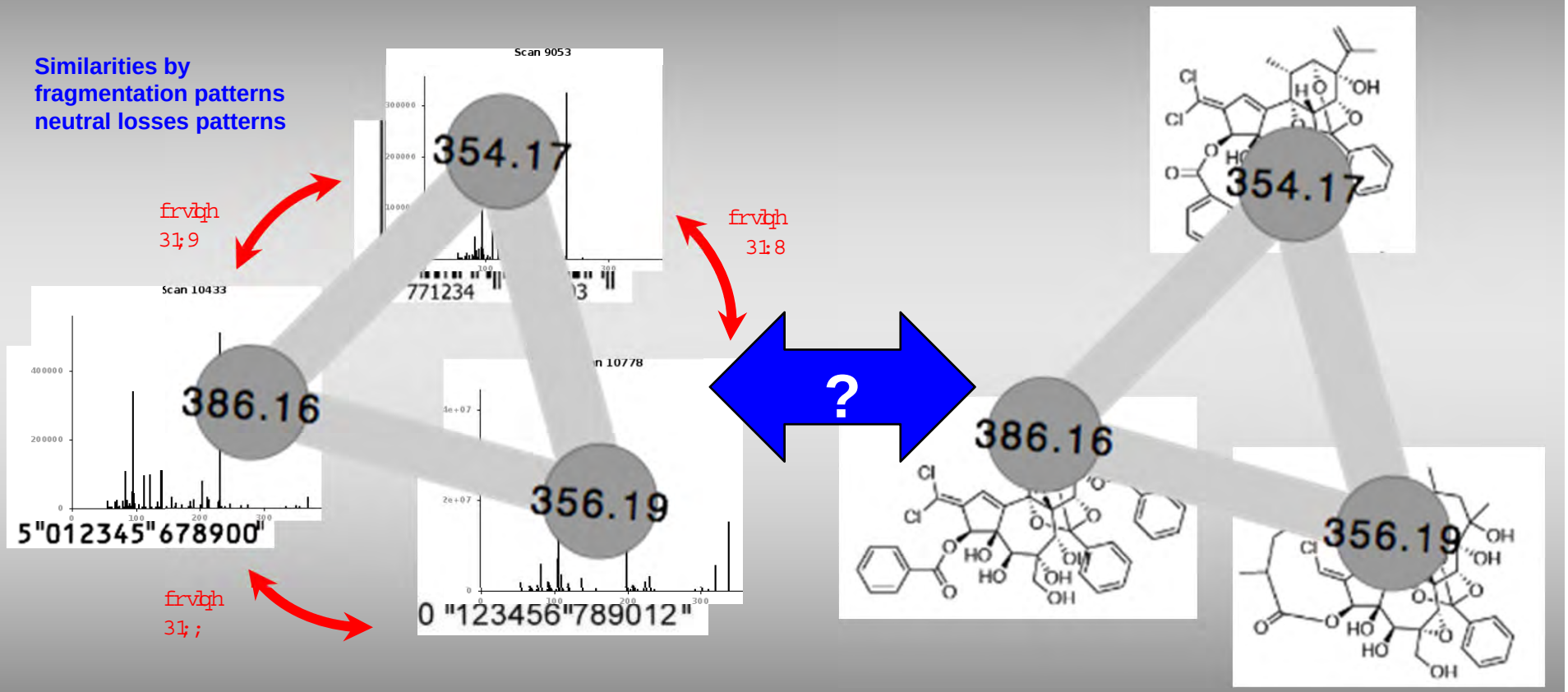
Generation of a massive molecular network:
1 000 000 MS^2 spectra
22 000 nodes
2 360 clusters



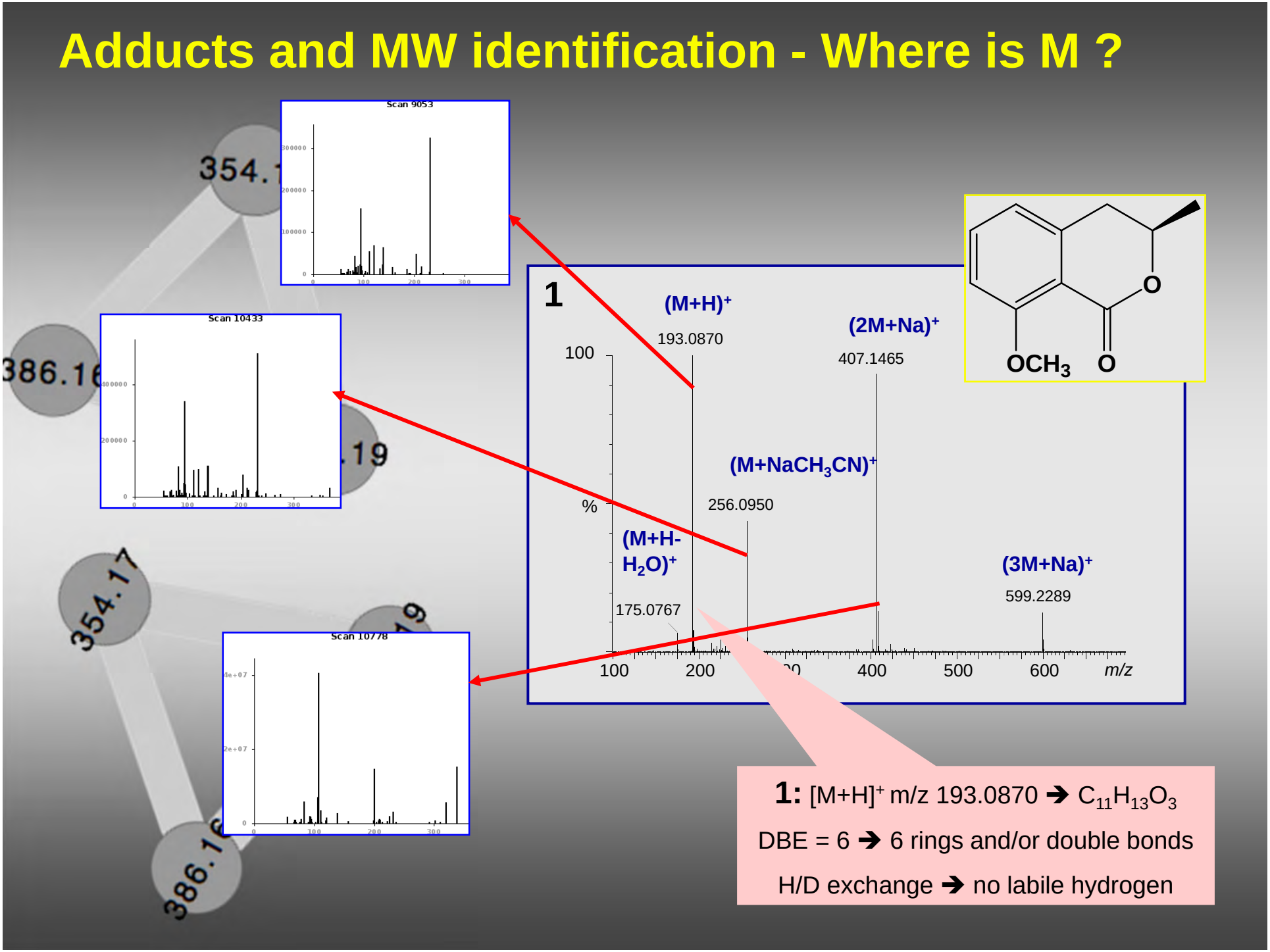
Creation of MN of 290 Euphorbiaceae extracts


733.298
trigocherrin B
ou
trigocherrin D
ou
trigocherrin F
known standards
trigocherriolide D
trigocherriolide C
Taxonomical layout
733.298
Trigonostemon cherrieri (Leaf)
Trigonostemon cherrieri (Wood)
Trigonostemon cherrieri (Bark)
ed to ions from Trigonostemon
sponding to trigocherrin and
could be highlighted.

Spectral networking = Molecular networking ?

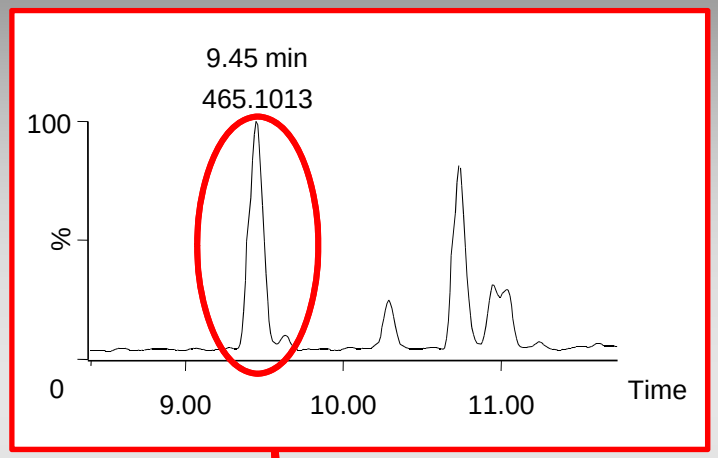


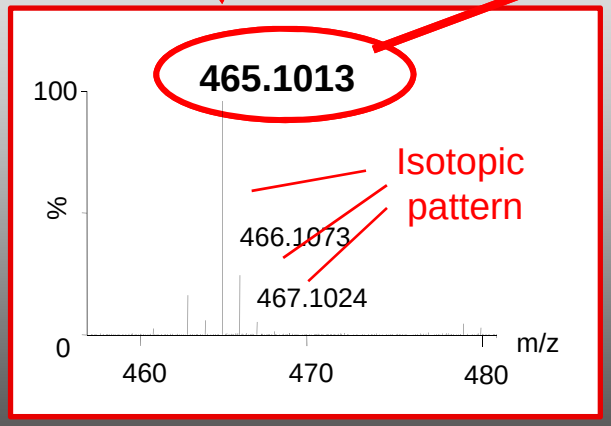
Adducts and MW identification - Where is M ?



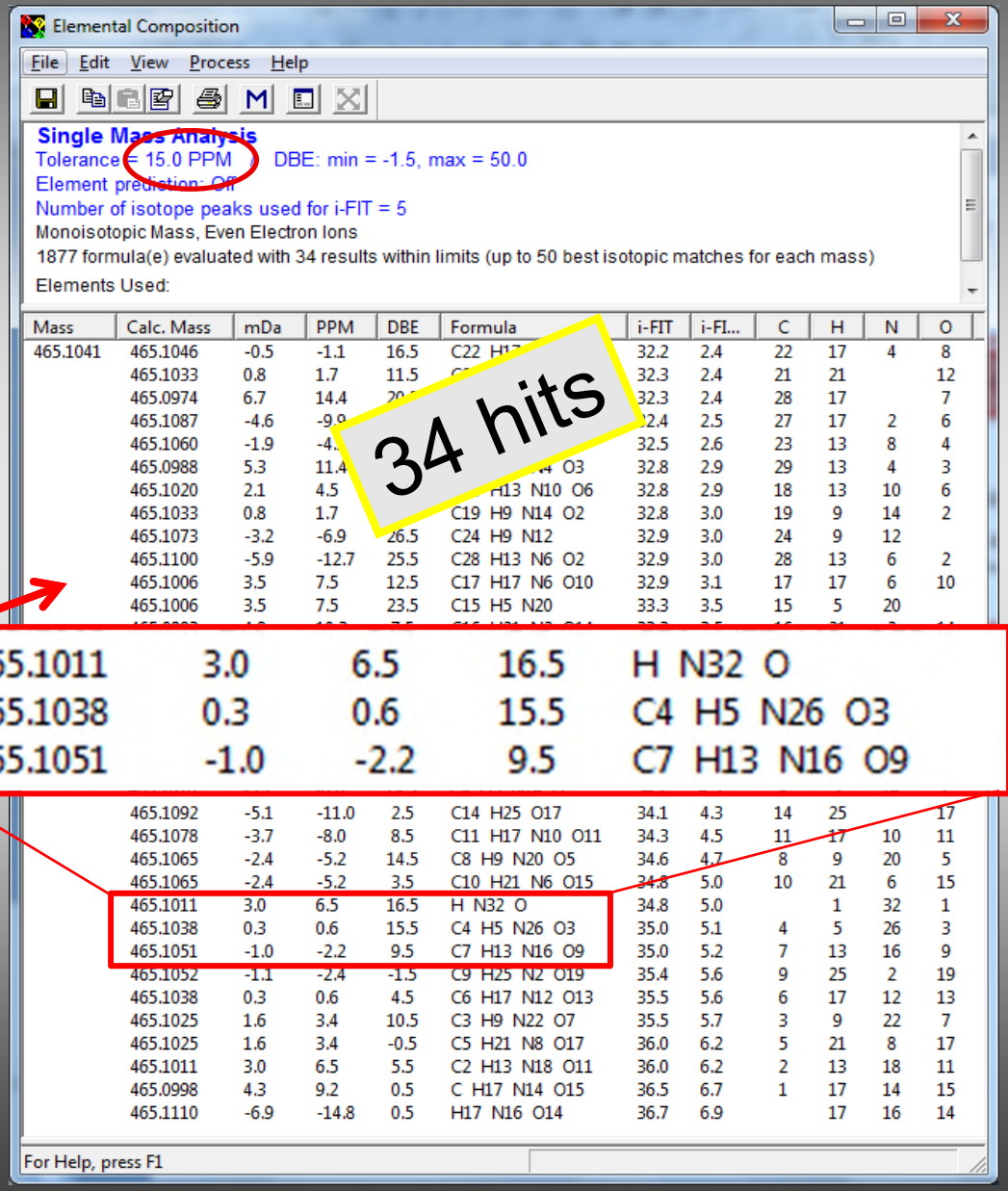
What is the correct molecular formula

1 peak



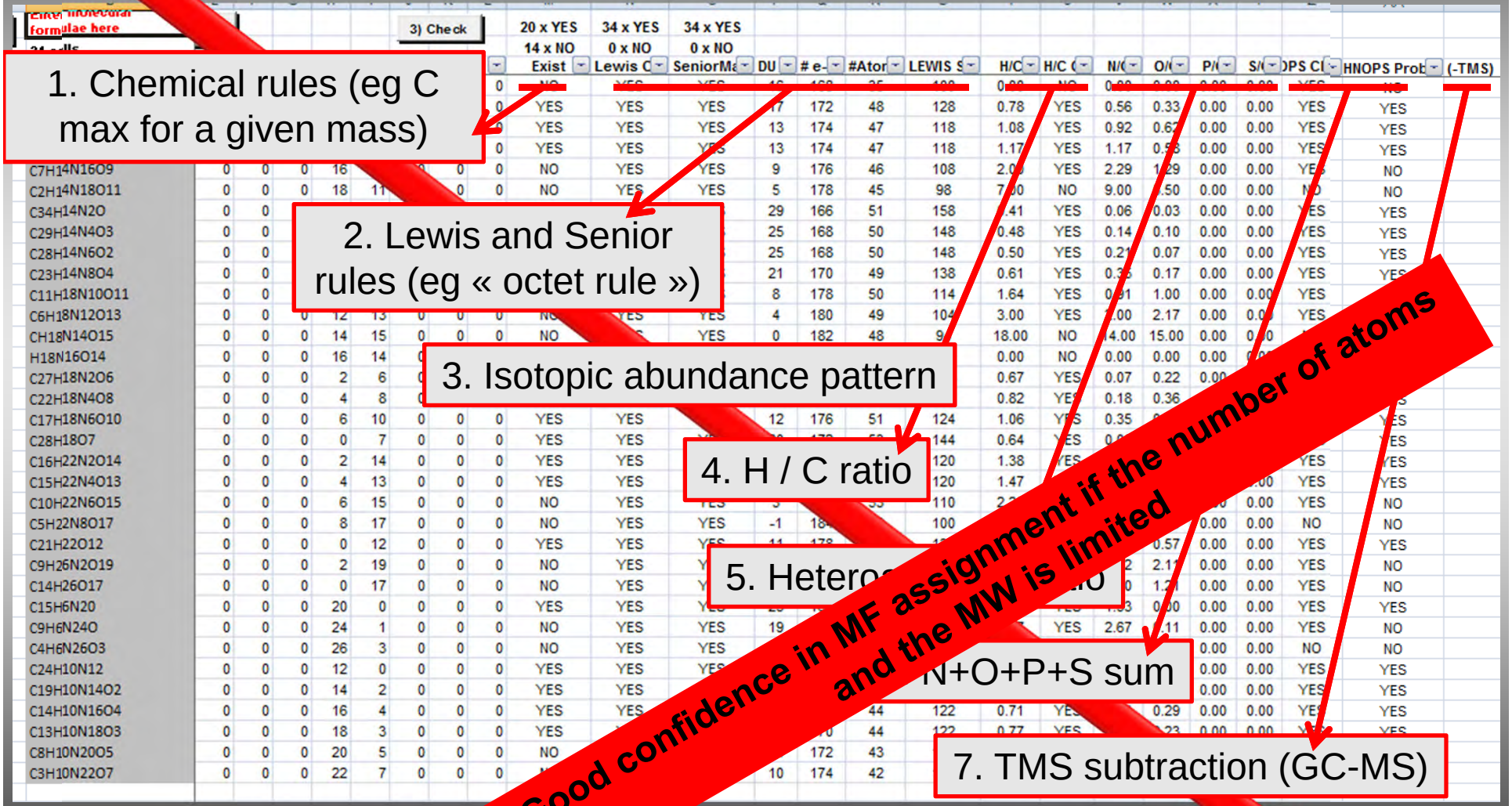


TOF-MS, negative mode



The 7 golden rules

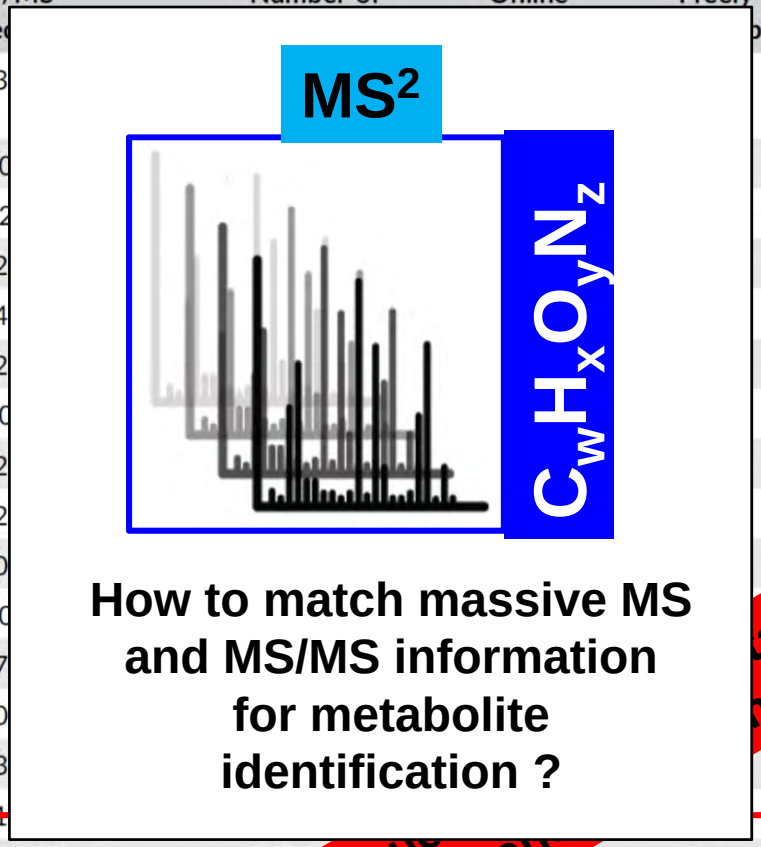
34 hits



6 hits

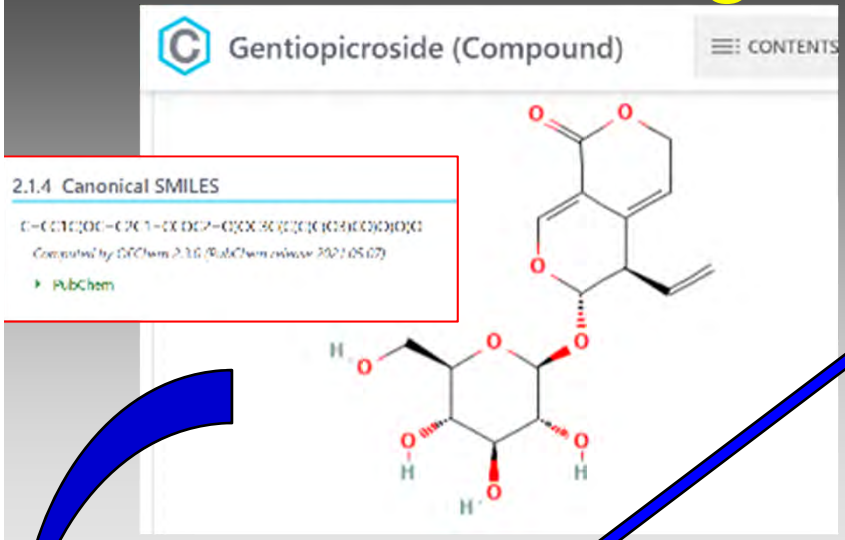
Tandem mass spectral databases for small molecule identification

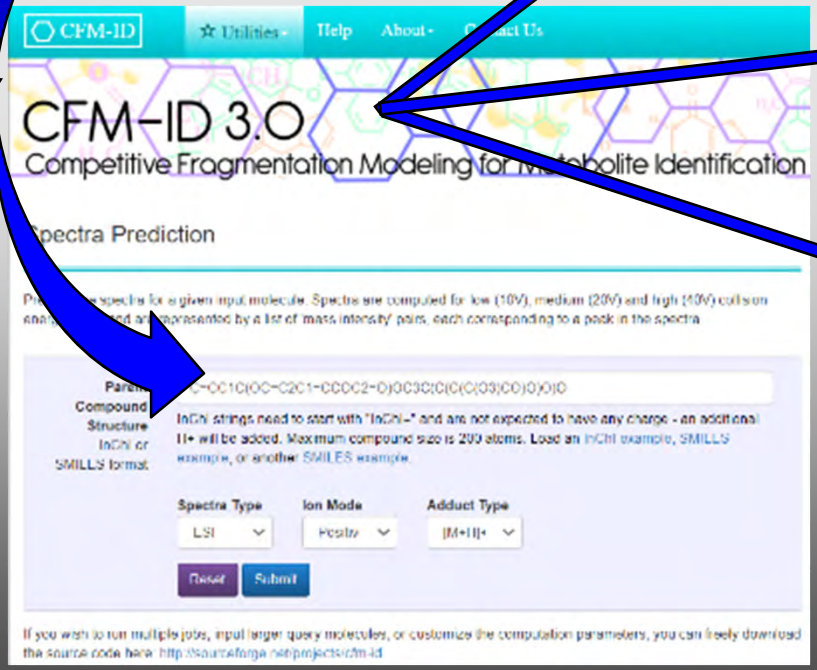
#	Name	Number of MS/MS spec	Number of	Online	Freely ble	Instrument diversity	In silico
1	NIST14 MS/MS	193				+++	–
2	MassBank	22 0				+++	–
3	METLIN	72 2				++	+/–
4	LipidBlast	212				++	+
5	MoNA	194				+++	
6	mzCloud	182				+	
7	MetaboBASE	26 0				+	
8	GNPS	212				++	
9	Spektraris	262					–
10	ReSpect	900					–
11	MSforID	20 0					–
12	HMDB	577				+	+/–
13	MetaMS	150				+	–
14	Sumner Library	173				+	–
15	ChemicalSoft	161				+	–
16	UNPD-ISDB	170 602	170		+	+	+
17	Designer drugs	10 000		–	$	+	–
18	Drugs/poisons	10 000		–	$	+	–

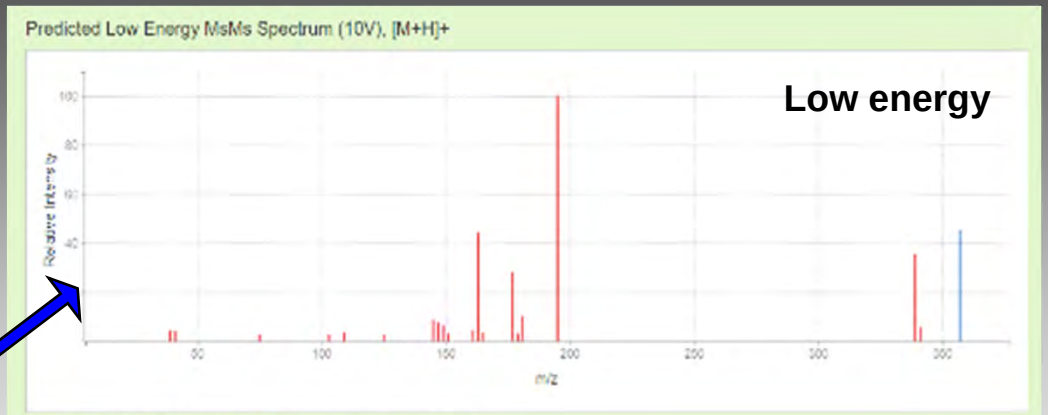


Kind T, *et al.* (2017) Identification of s olecules using accurate mass MS/MS search. *Mass Spectrom. Rev.*: doi 10.1002/mas.21535

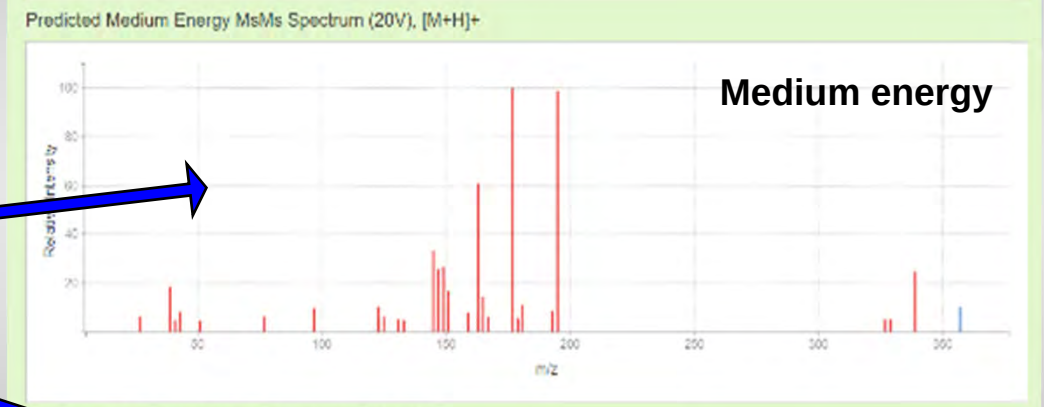
In silico fragmentation with CFM-ID







Low energy

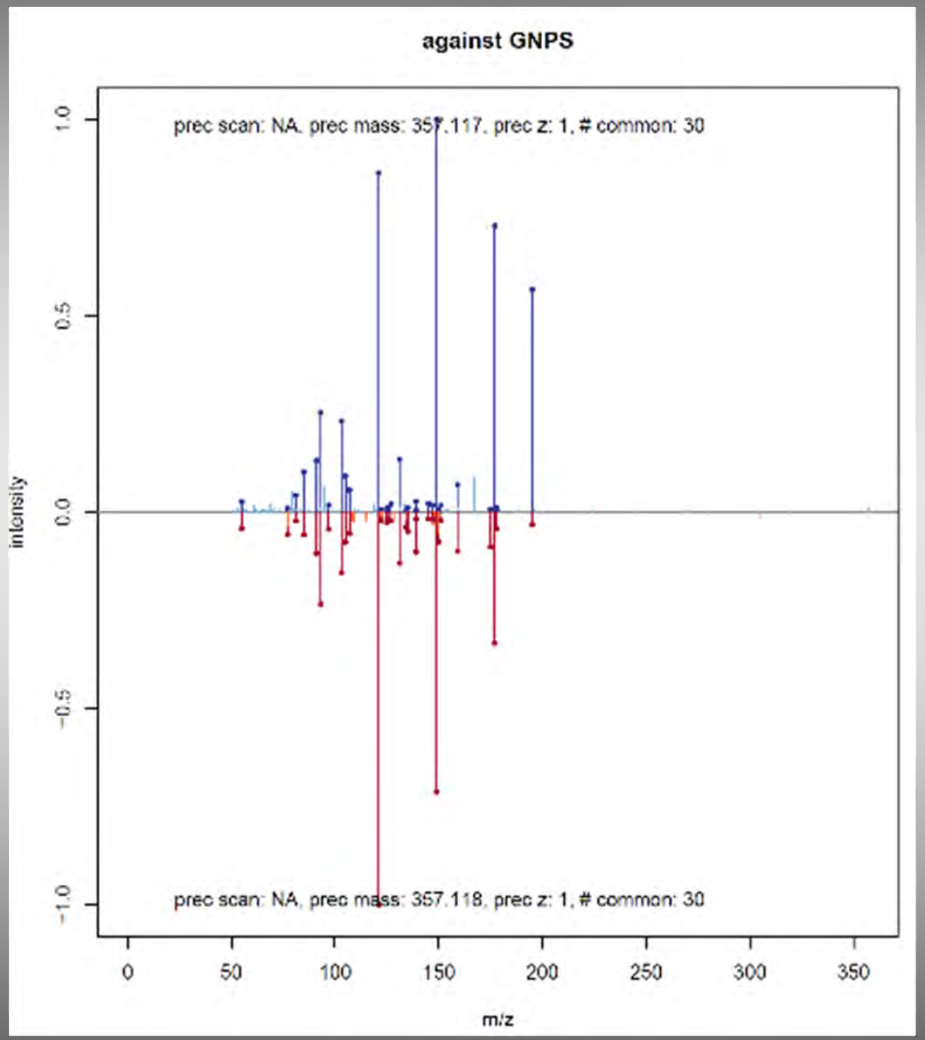


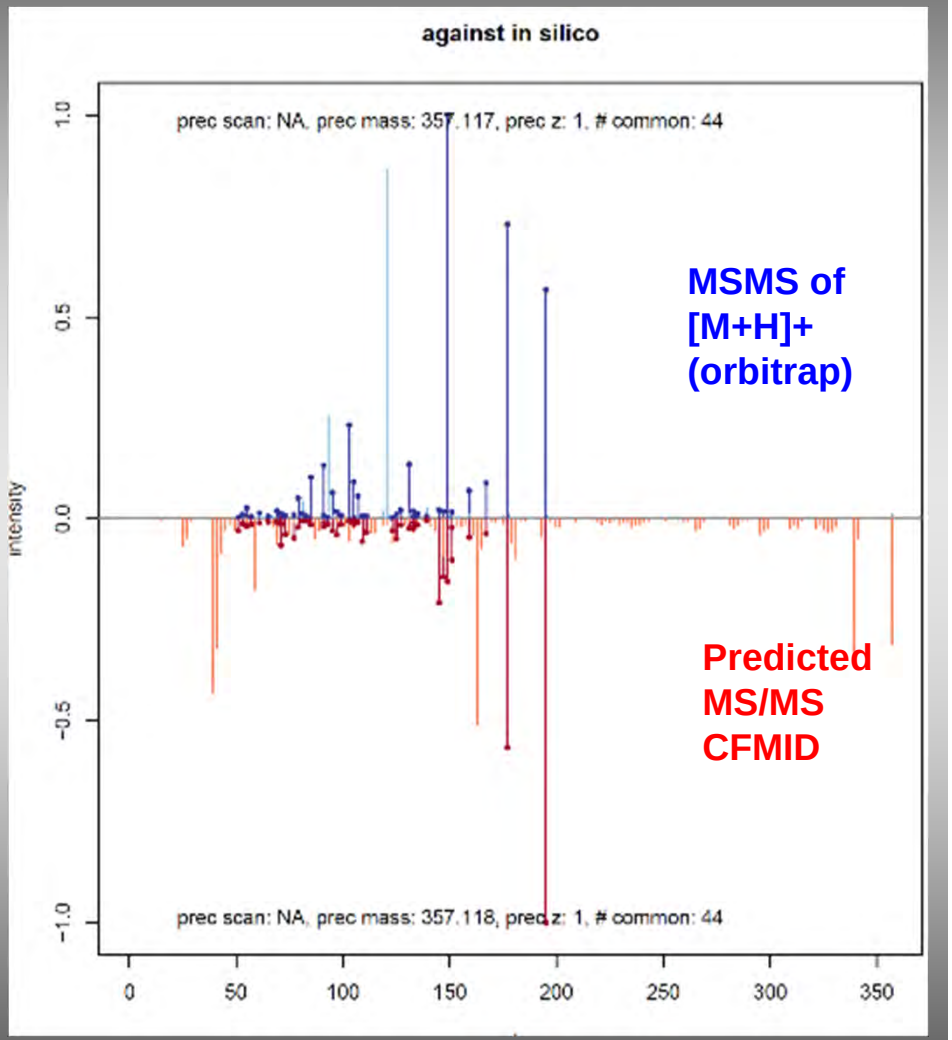
Medium energy

High energy

Djoumbou-Feunang, Y., et al. *Metabolites* 2019, *9*.

In silico fragmentation with CFM-ID





MSMS of [M+H]+ (orbitrap)

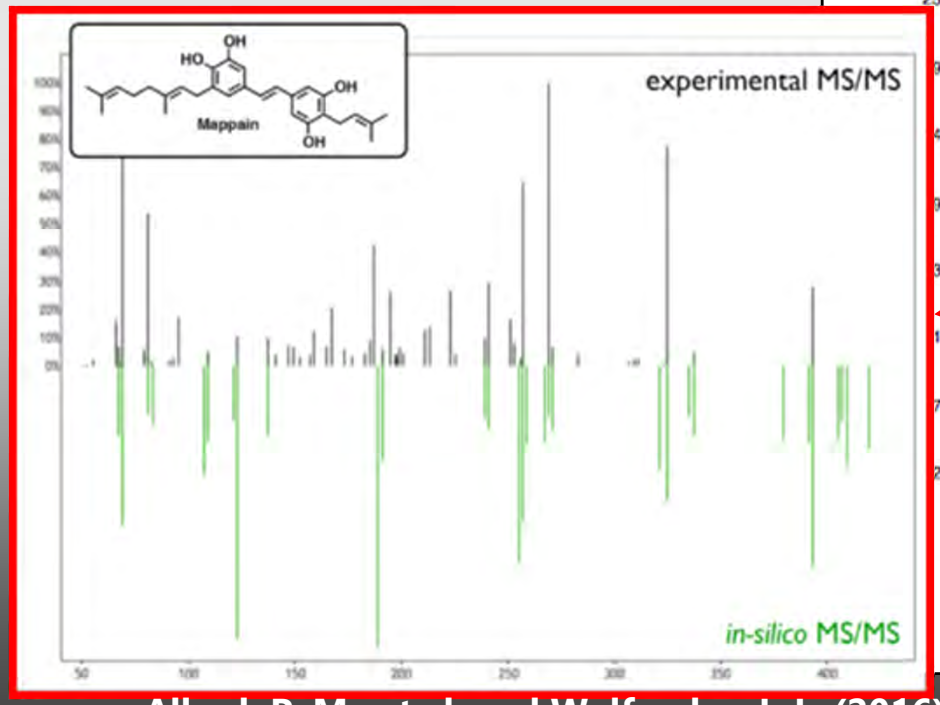
Predicted MS/MS CFMID

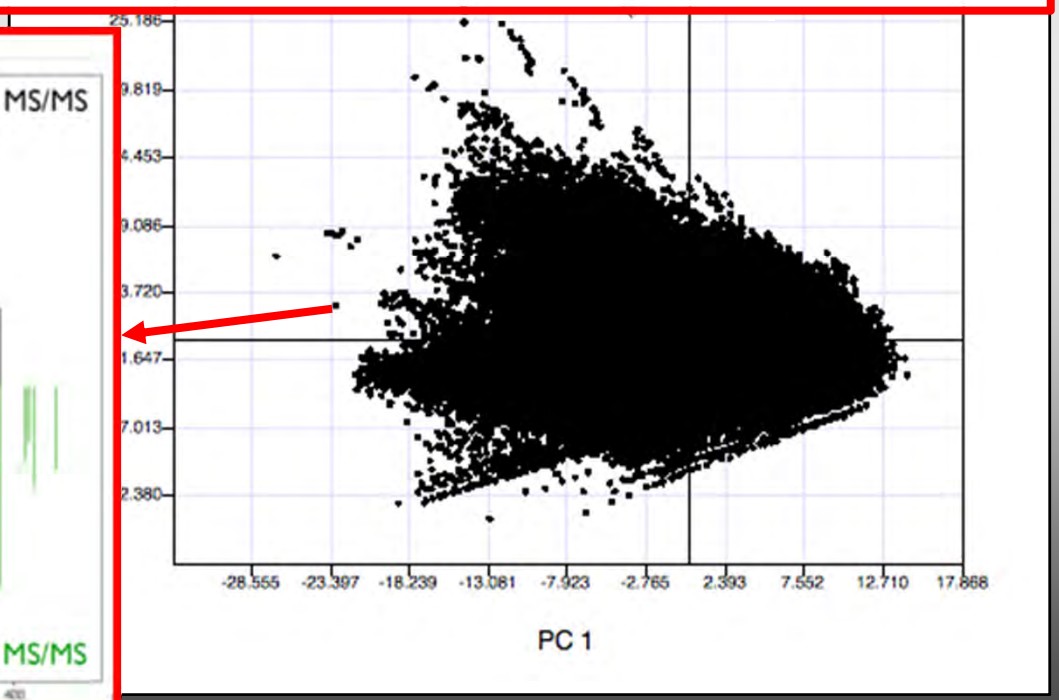
GNPS

CFM-ID 3.0

Thesis A. Rutz, UNIGE

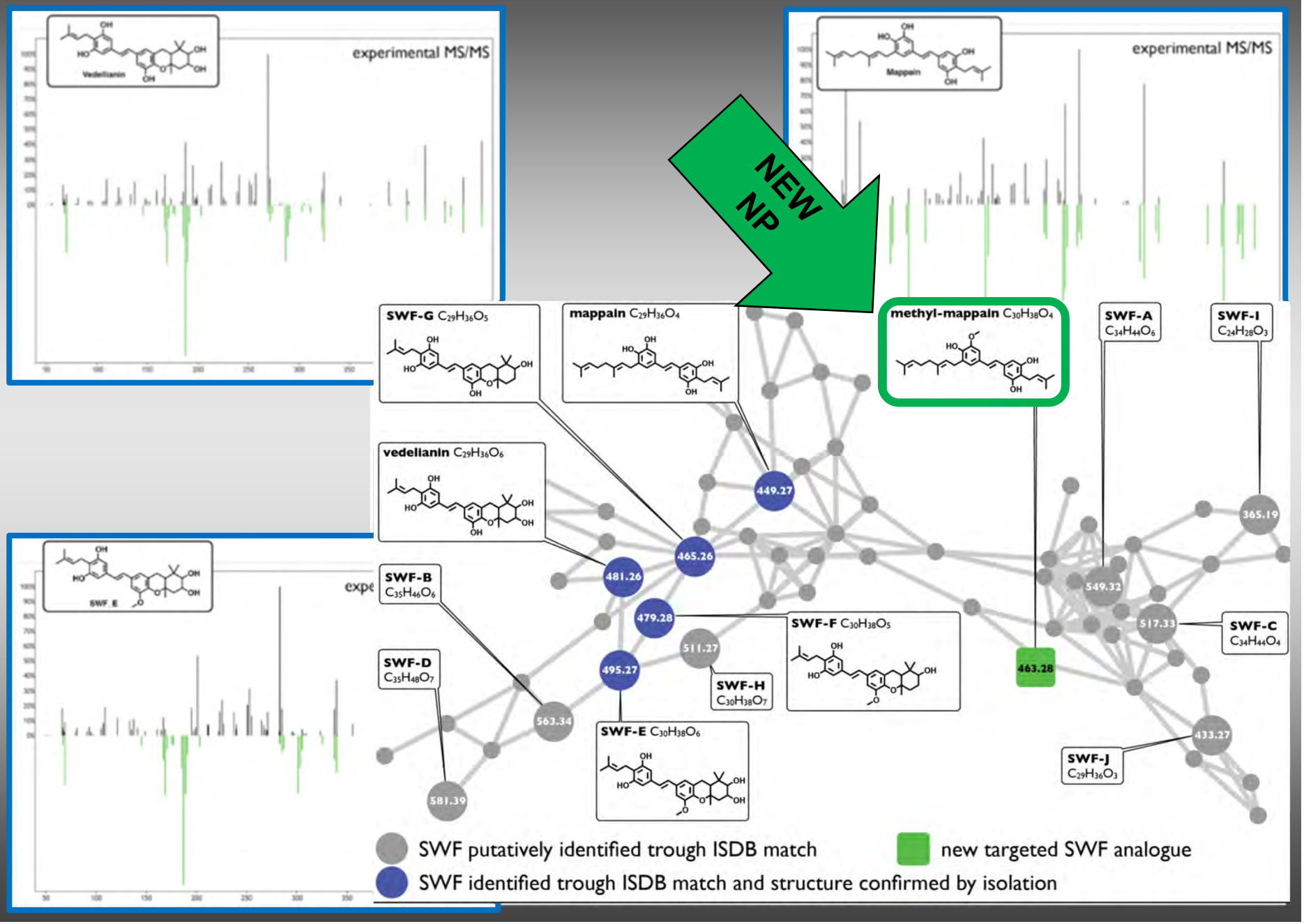
NP In silico MS/MS database ISDB





Allard, P.-M., et al. and Wolfender, J.-L. (2016). *Anal. Chem.*10.1021/acs.analchem.5b04804

Prenylated stilbenes in various Macaranga spp.

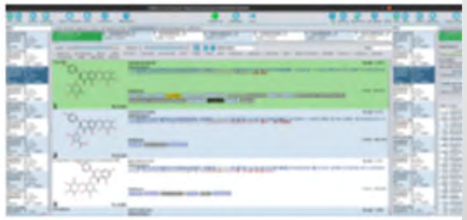


Additionnal computational tools

SIRIUS+CSI:FingerID

CSI:FingerID (Compound Structure Identification: FingerID) is a method for searching a MS/MS of a small molecule (metabolite) in a database of molecular structures.

SIRIUS



ZODIAC

ZODIAC performs *de novo* molecular formula annotation on complete biological datasets (high-resolution, high mass accuracy LC-MS/MS runs).

COSMIC

COSMIC is a workflow that allows you to assign confidence to structure annotations.

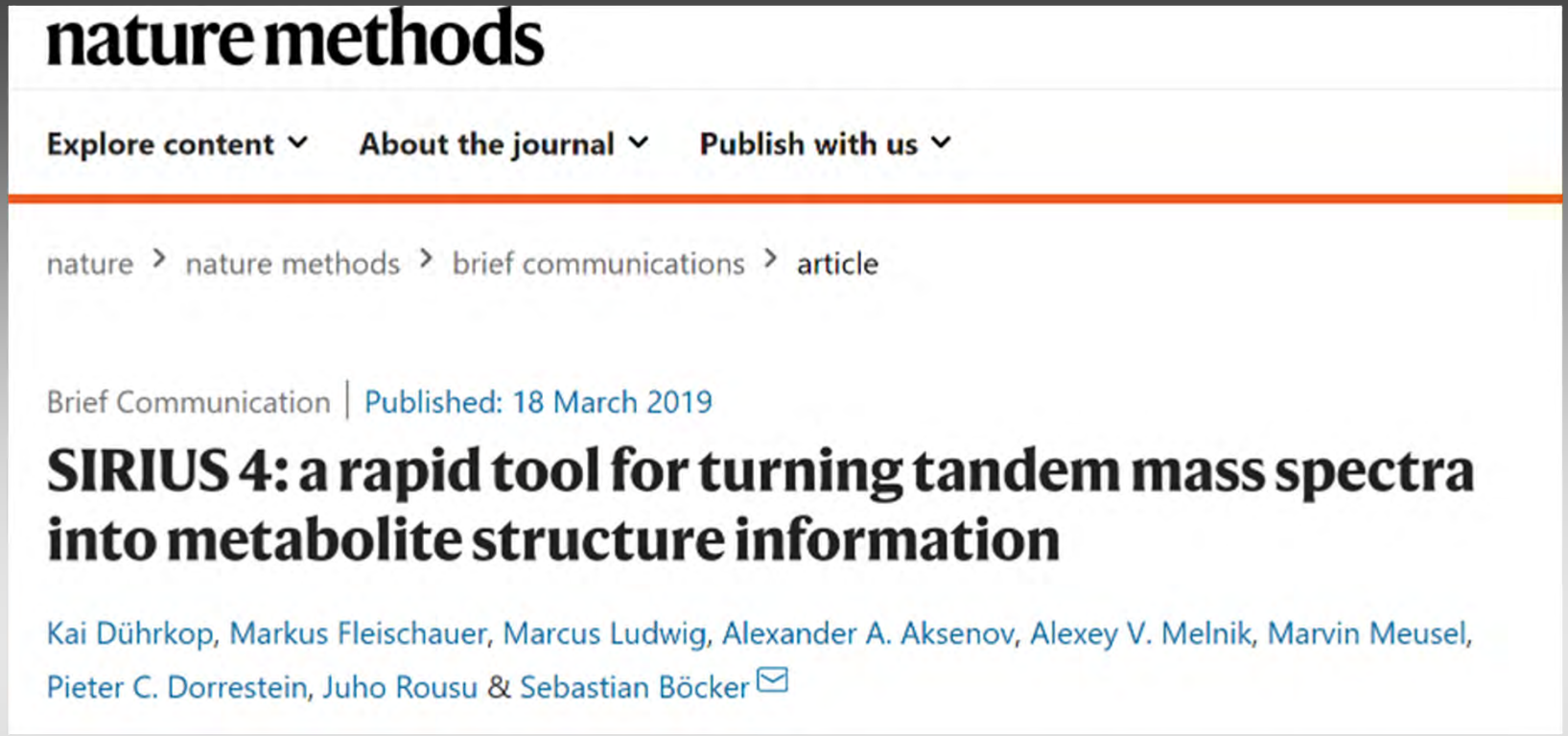
CANOPUS

CANOPUS is a tool for predicting compound classes directly from MS/MS with no DB

Lehrstuhl Bioinformatik Jena

Friedrich-Schiller-Universität Jena, Fakultät für Mathematik und Informatik

Duhrkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., Dorrestein, P. C., Rousu, J. and Bocker, S. (2019). SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Meth.* **16**, 299-+.



nature methods

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature methods › brief communications › article

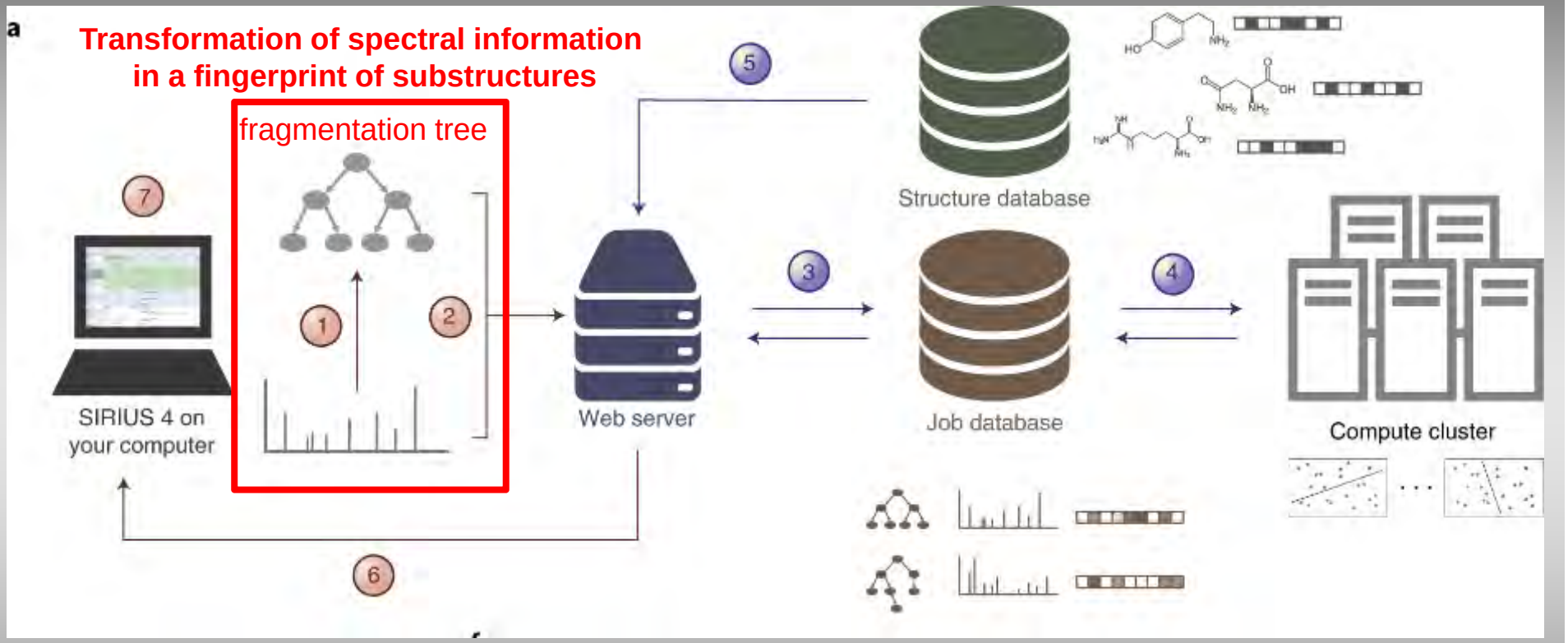
Brief Communication | Published: 18 March 2019

SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information

Kai Dührkop, Markus Fleischauer, Marcus Ludwig, Alexander A. Aksenov, Alexey V. Melnik, Marvin Meusel, Pieter C. Dorrestein, Juho Rousu & Sebastian Böcker

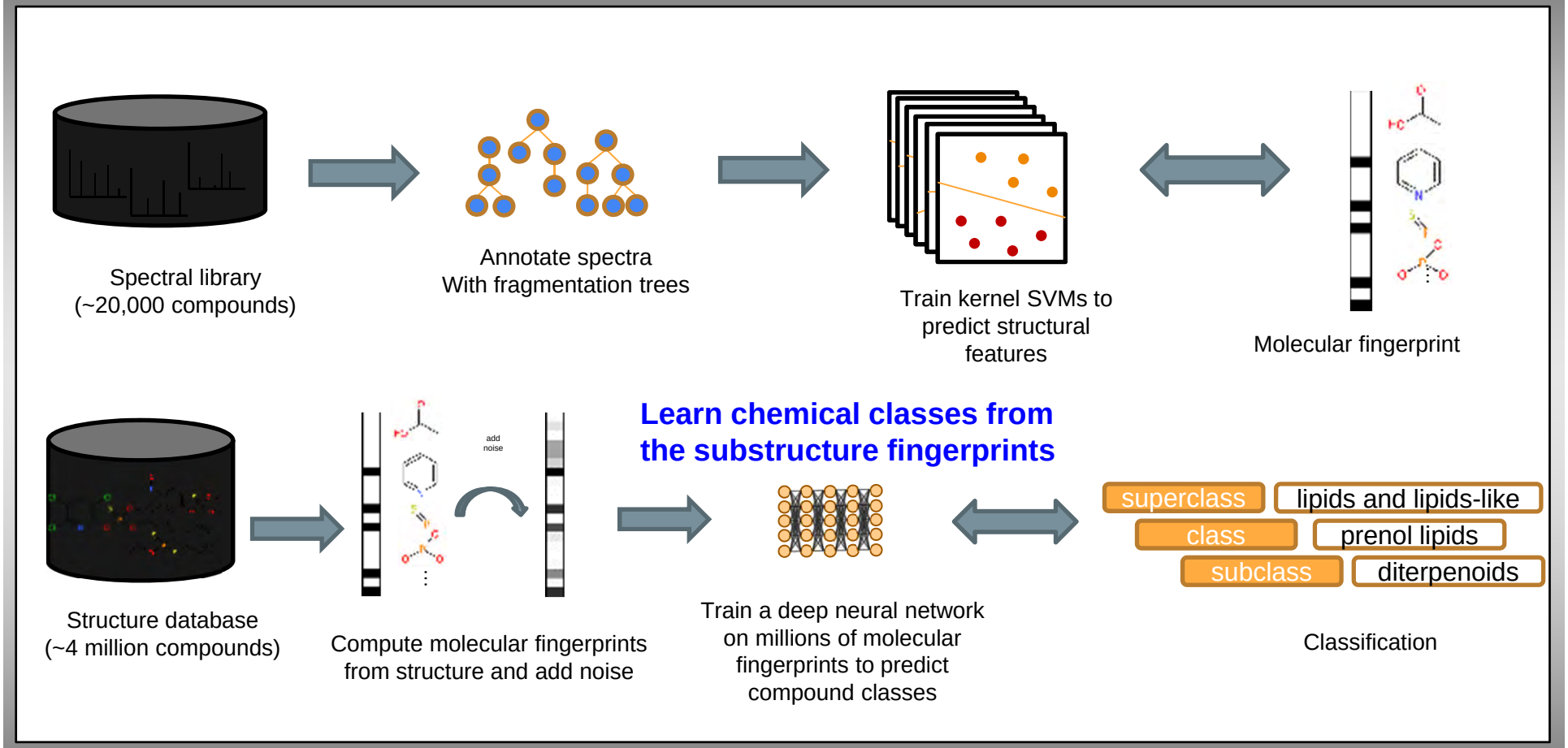
- What is the molecular formula of the query compound among all molecular formulas, both previously observed and unobserved?
- Given a database of molecular structures, what is the structure that best explains the experimental data?

SIRIUS



SIRIUS 4 aims to identify the molecular formula of the query compound and annotate the MS/MS spectrum with a fragmentation tree.

CANOPUS: compound class annotation

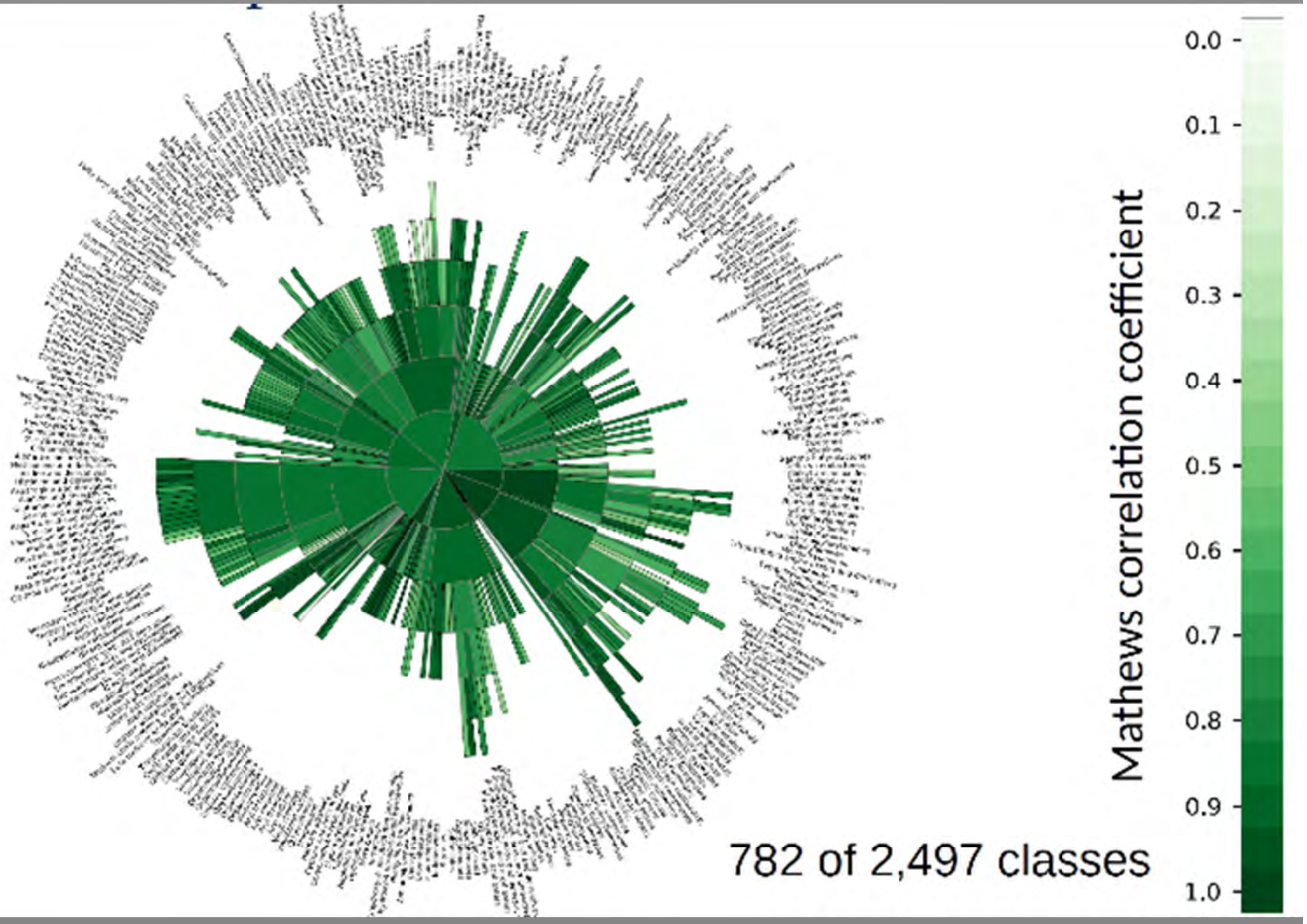


Dührkop, K., Nothias, L. F., Fleischauer, M., Reher, R., Ludwig, M., Hoffmann, M. A., Petras, D., Gerwick, W. H., Rousu, J., Dorrestein, P. C. and Böcker, S. (2021). Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **39**, 462-471.

Natural product class prediction by canopus

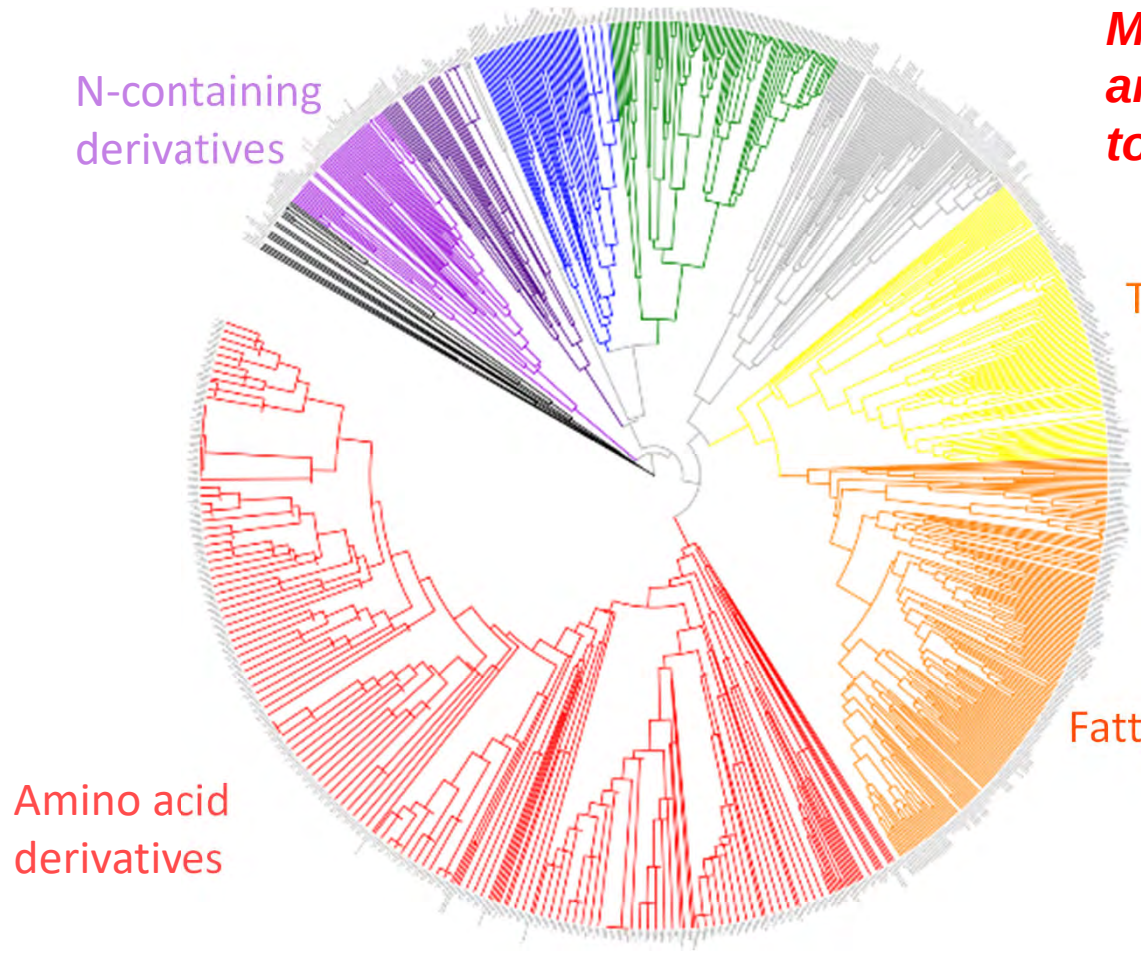
Compound class	accuracy
Flavonoid glycosides	99.96%
Flavonols	99.96%
Glycosyl compounds	99.37%
Phosphocholines	99.95%
Beta lactams	99.85%
Chromones	99.63%
Pyrroloquinolines	99.83%
Steroidal glycosides	99.95%
Aminosaccharides	99.78%
Pentoses	99.73%
Oxosteroids	98.96%
Cholines	99.91%
Oligosaccharides	99.91%
Olefins	99.96%

… and much more



Dührkop, K., Nothias, L. F., Fleischauer, M., Reher, R., Ludwig, M., Hoffmann, M. A., Petras, D., Gerwick, W. H., Rousu, J., Dorrestein, P. C. and Böcker, S. (2021). Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **39**, 462-471.

CANOPUS: compound class annotation



Proof of principle Metabolites annotated in tomato

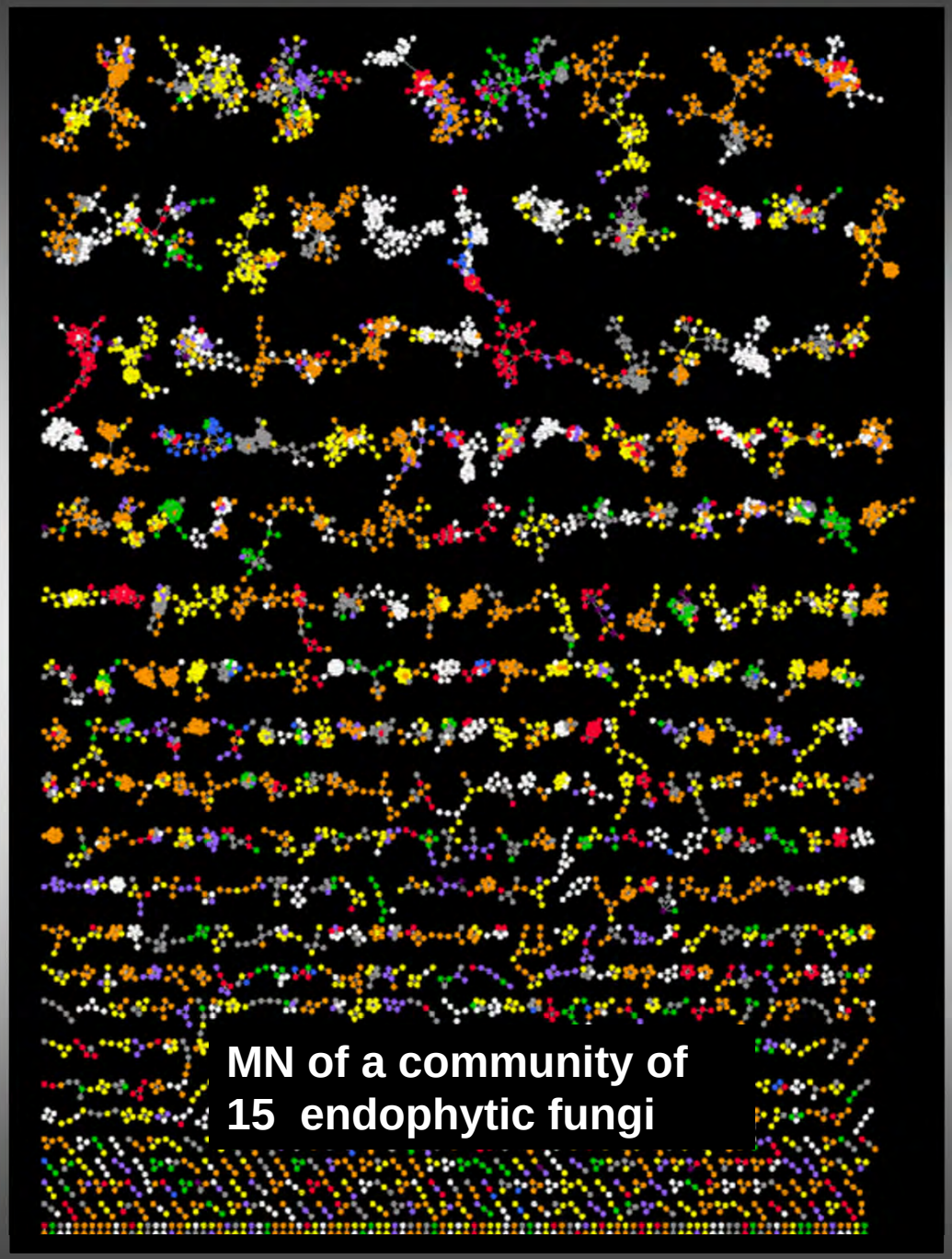
Crude methanolic extract of seedlings

https://itol.embl.de/tree/7680178332386153928643 1#

Tripathi, A., et al. *Nature Chemical Biology* 2021, *17*, 146-151.

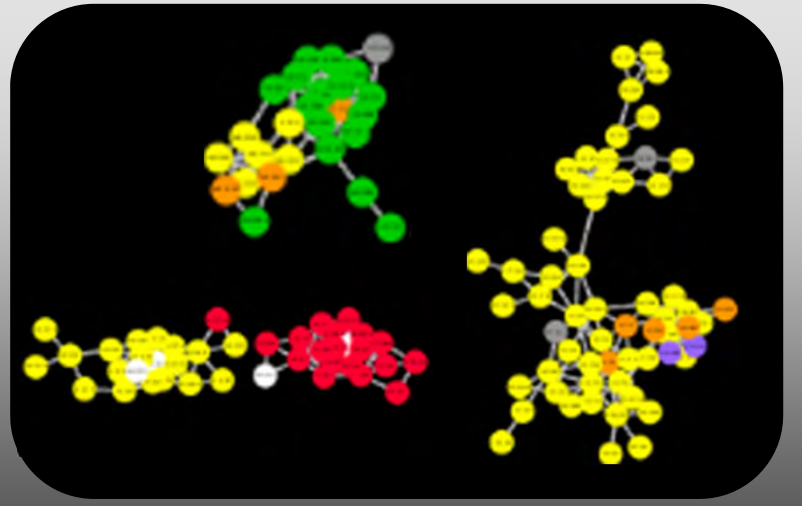
L. F. Nothias

CANOPUS class annotation in MN



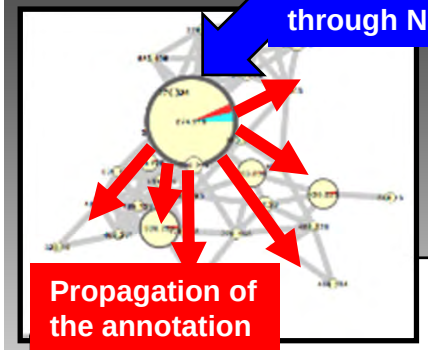
MN of a community of 15 endophytic fungi

- Alkaloids
- Amino acid derivatives
- Carbohydrates
- Fatty acid derivatives
- Polyketides
- Phenylpropanoids
- Terpenoids



Thesis L. Pélissier UNIGE

Combined NMR and MN strategy for deep metabolome annotation



Standards

Prenol lipids: 208 metabolites

Fatty acyls: 152 metabolites

Flavonoids:129 metabolites

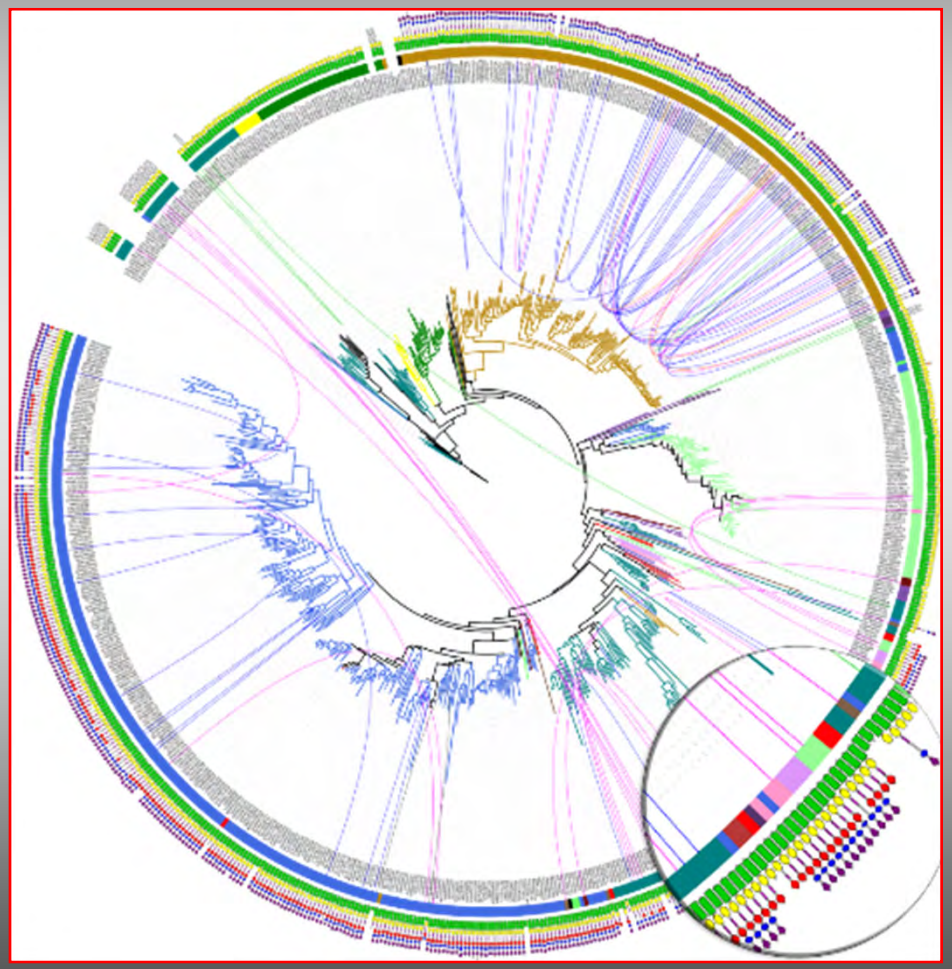
Organooxygen compounds: 95 metabolites

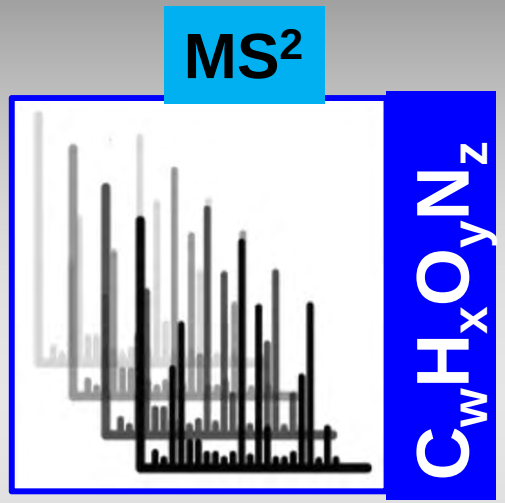
F. deltoidea Jack (Moraceae)

"Mas cotek"- golden tiny spots

24 Standards isolated ➡ **550 annotations**

Improve annotation by contextualization of the data through the taxonomy of the producing organisms



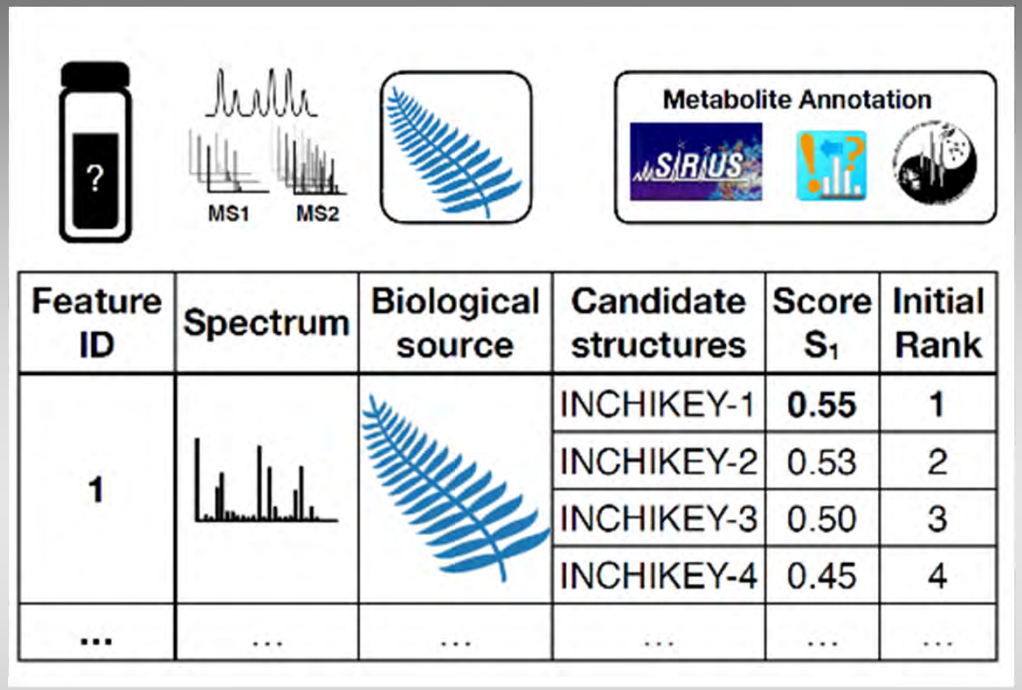


A given MS feature in a specific:

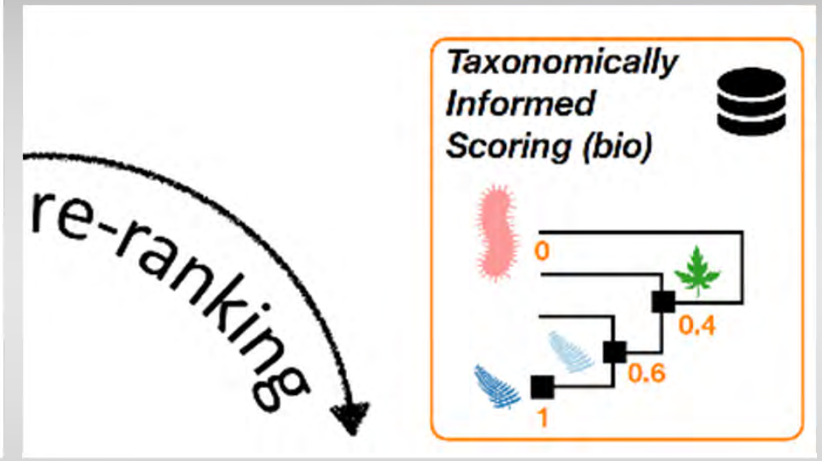
1) Species: *G. lutea*
2) Genus: *Gentiana*
3) Family: *Gentianaceae*

Rutz, A., et al. *Front Plant Sci* 2019, *10*, 1329.

Re-reranking of annotation candidates



Feature ID	Spectrum	Biological source	Candidate structures	Score S_1	Initial Rank
1			INCHIKEY-1	**0.55**	1
			INCHIKEY-2	0.53	2
			INCHIKEY-3	0.50	3
			INCHIKEY-4	0.45	4
...	...	...	...	...	...



Feature ID	Spectrum	Biological source	Candidate structures	Score S_1	Initial Rank	Candidate biological source	Score S_2	Combined Score $(S_1 + S_2)$ /2	Final Rank
1			INCHIKEY-4	**0.45**	4		1	**0.78**	1
			INCHIKEY-1	0.55	1		0.6	0.58	2
			INCHIKEY-2	0.53	2		0.4	0.47	3
			INCHIKEY-3	0.50	3		0	0.25	4
...	...	...	...	...	...	...	...	...	...

Rutz, A., et al. ***Front Plant Sci*** **2019**, *10*, 1329.

Re-reranking of annotation candidates

Flavonol glycoside (GNPS lib.)
Flavone glycoside
Taxonomical rank 2 (In silico)

m/z 287.0
Kaempferol?
Luteolin?

Flavone cluster CI 11 (PI)

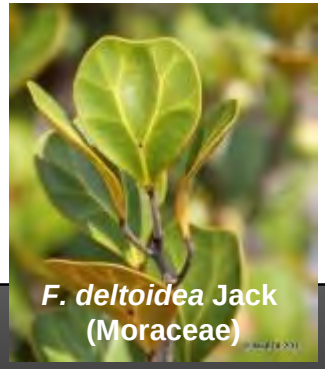
Flavone glycoside
Taxonomical rank 1 (In silico)

Flavone glycoside

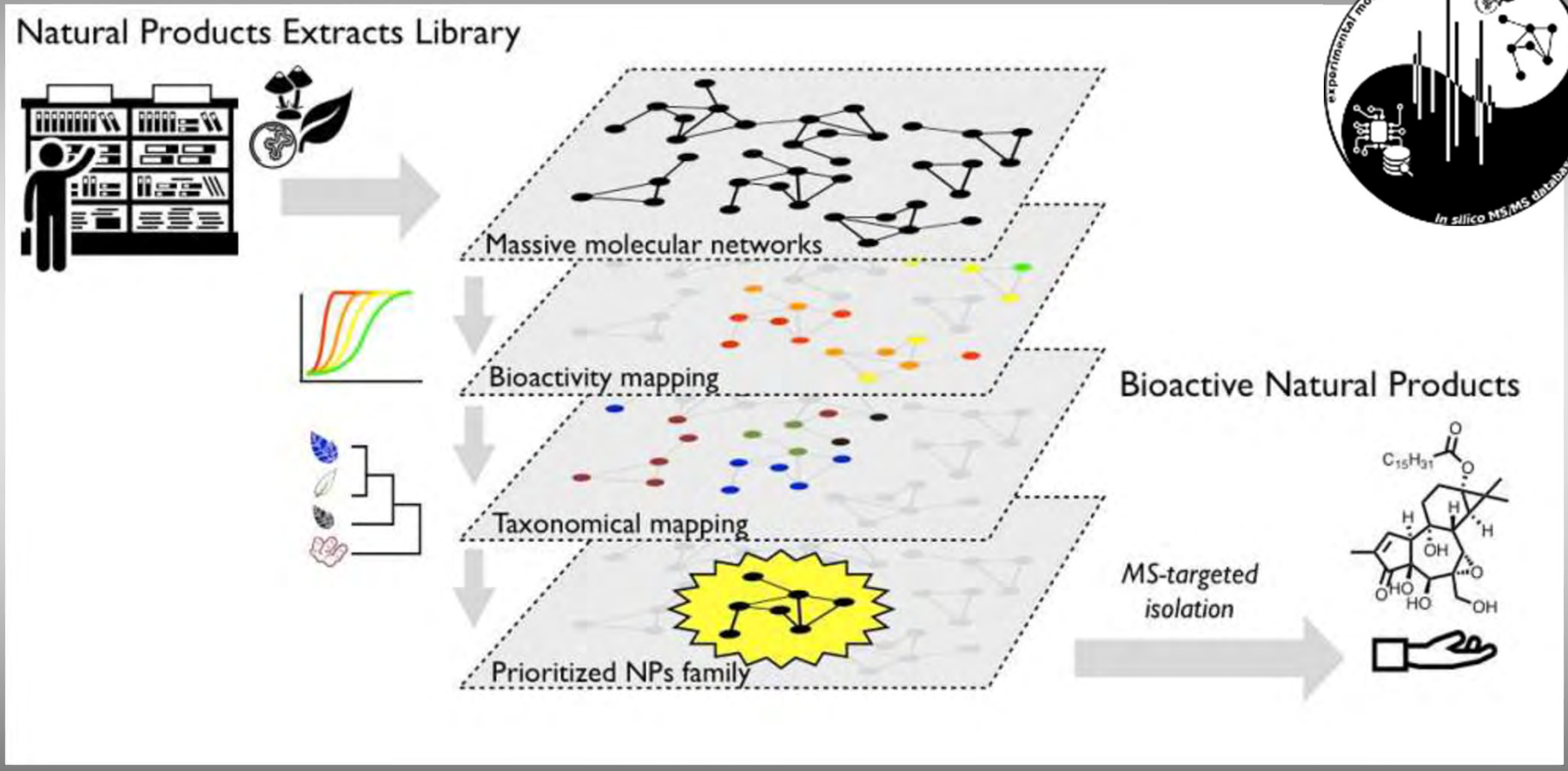
433.11;14.7
595.16;13.3
449.10;14.3
565.15;14.5
S24
579.17;15.4
595.16;14.2
449.10;12.4
1157.3;15.4
287.05;18.2
271.05;19.5
283.05;23.3

Flavone glycoside
Rhoifolin
$C_{27}H_{30}O_{14}$

Flavone aglycone

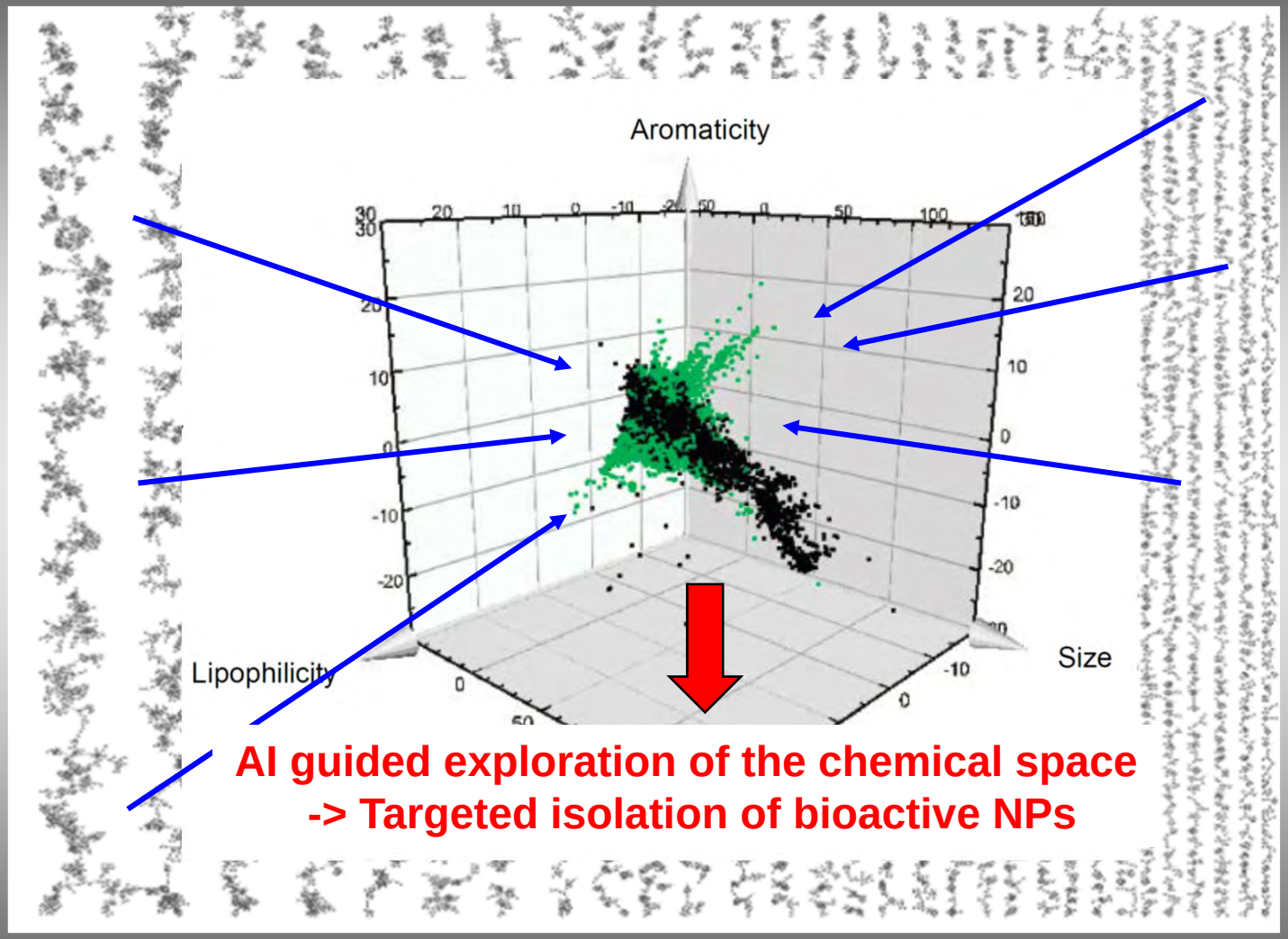


Bioactive Natural Products Prioritization Using Massive Multiinformational Molecular Networks



Olivon F, Allard PM, Koval A, Righi D, Genta-Jouve G, Neyts J, Apel C, Pannecouque C, Nothias LF, Cachet X, Marcourt L, Roussi F, Katanaev VL, Touboul D, Wolfender JL, Litaudon M. Bioactive Natural Products Prioritization Using Massive Multi-informational Molecular Networks. *ACS Chem Biol* 2017.

MN of the biodiverse Pierre Fabre collection



AI guided exploration of the chemical space
-> Targeted isolation of bioactive NPs

The massive PF Molecular Network

- **1600 Samples (159 Families, 533 Genus, 767 species)**
- **64'001 nodes**
- **>15'000 spectral clusters(from 100 to 2 nodes)**

Pierre Fabre

ISPSO
INSTITUT DES SCIENCES PHARMACEUTIQUES DE SUISSE OCCIDENTALE

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

MN of the biodiverse Pierre Fabre collection

The massive PF Molecular Network

- **1600 Samples (159 Families, 533 Genus, 767 species)**
- **64’001 nodes**
- **>15’000 spectral clusters(from 100 to 2 nodes)**

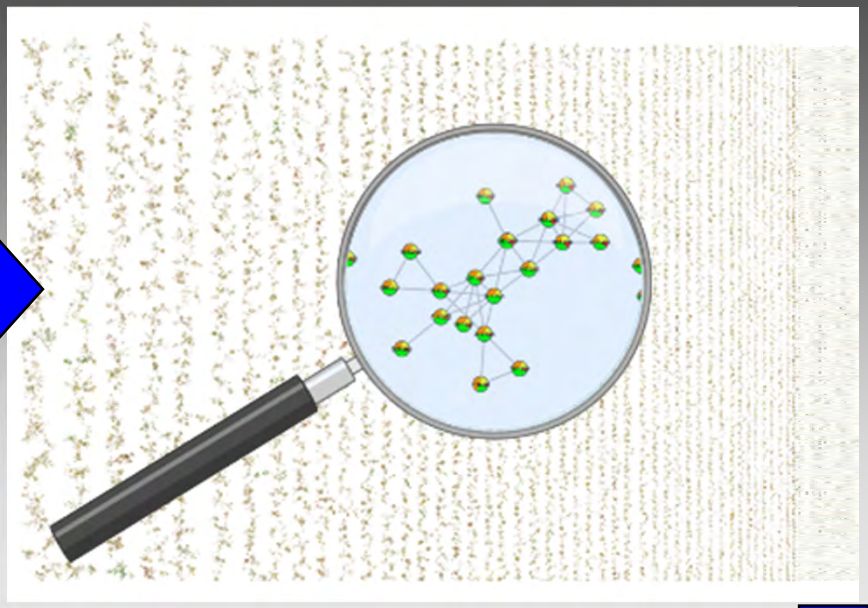
Pierre Fabre

ISPSO
INSTITUT DES SCIENCES PHARMACEUTIQUES DE SUISSE OCCIDENTALE

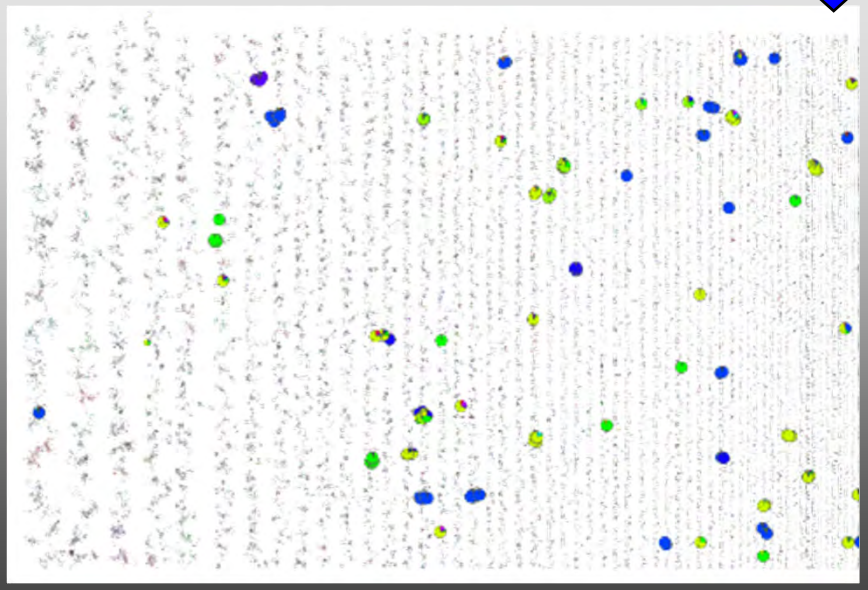
FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Taxonomical visualization

Pie chart mapping with bioactivity



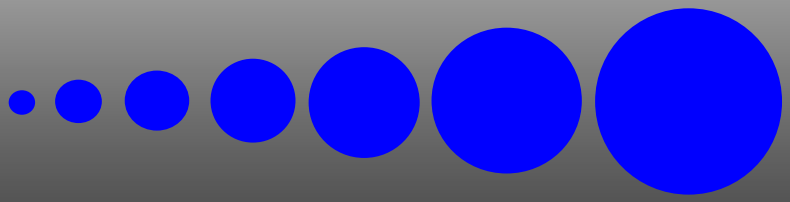
Bioactivity score representation



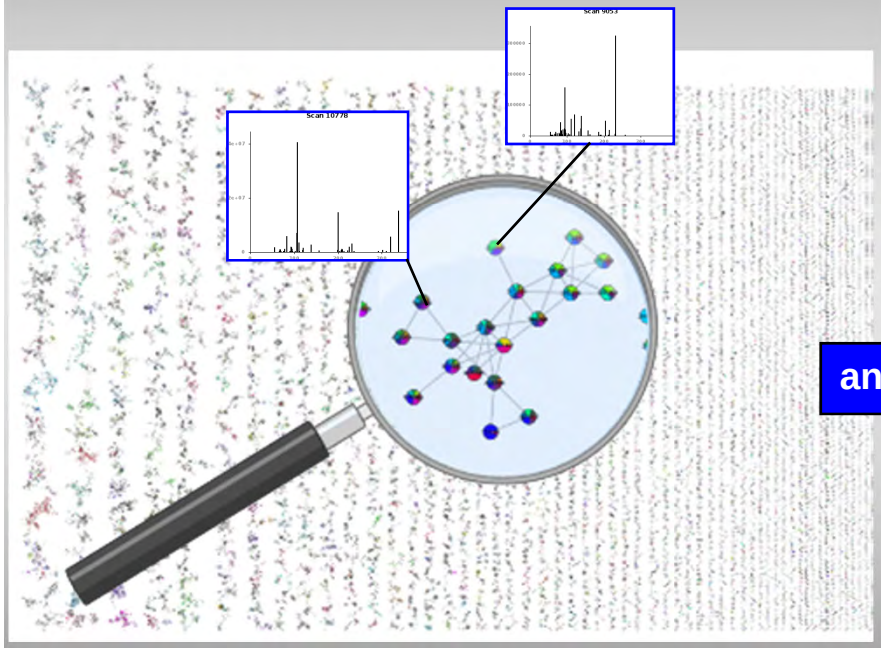
Bioactivity mapping: bioactivity score

-10,000:
Feature present in extracts with 100% cytotoxicity and 0% activity

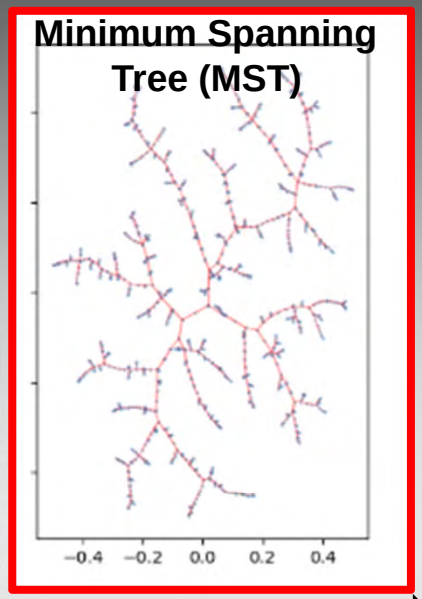
+10,000:
Feature present in extracts with 0% cytotoxicity and 100% activity



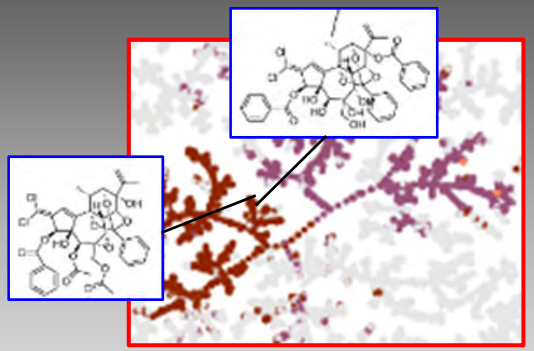
From spectral network to chemical space

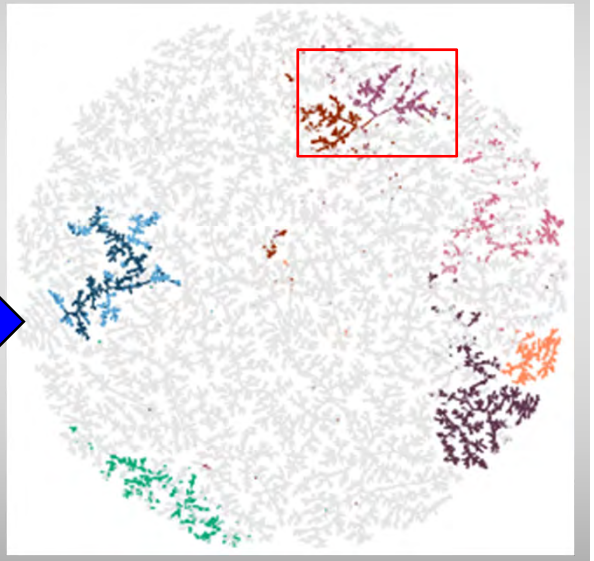


spectral networks



annotated structures





Explorable chemical space
For searching Bioactives NPs

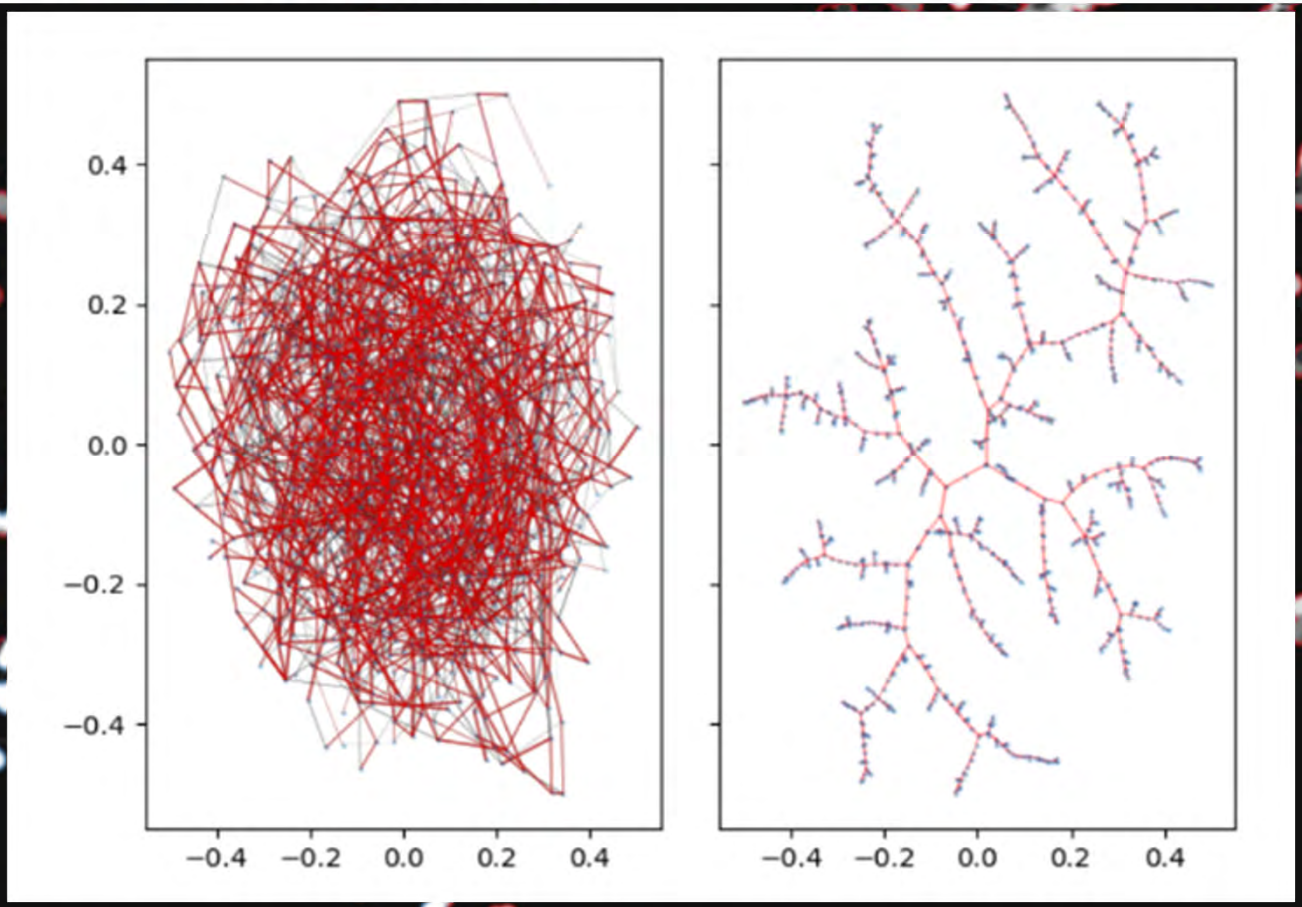
GNPS

SIRIUS

RESEARCH ARTICLE Open Access

Visualization of very large high-dimensional data sets as minimum spanning trees

Daniel Probst* and Jean-Louis Reymond*



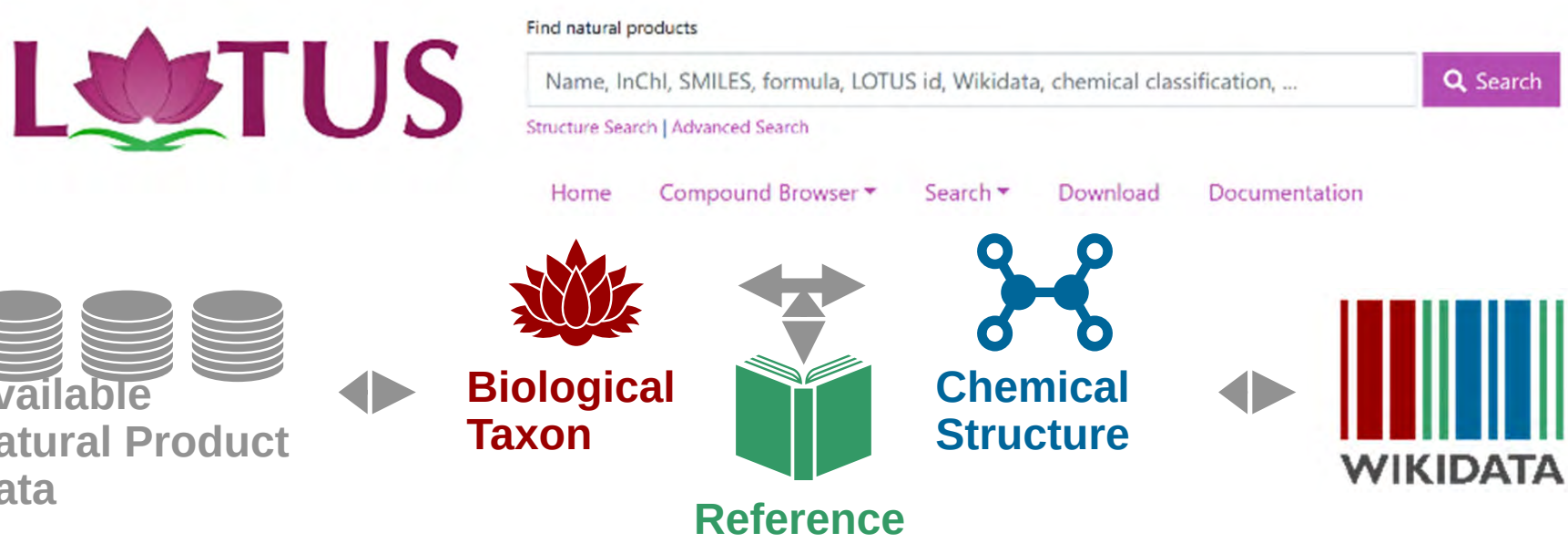
k-nearest neighbor graph(k-NN)

Minimum Spanning Tree (MST)

Unique link between closely chemically related structure

Probst, Daniel, and Jean-Louis Reymond. 2020. "Visualization of Very Large High-Dimensional Data Sets as Minimum Spanning Trees." *Journal of Cheminformatics* 12 (1). https://doi.org/10.1186/s13321-020-0416-x.https://tmap.gdb.tools/

The LOTUS Initiative for Knowledge Sharing in Natural Products Research



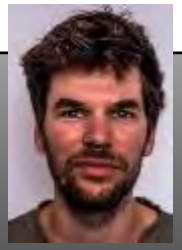
Unstructured, redundant and noisy data

Cleaned and documented data

Collaboratively curated and structured data

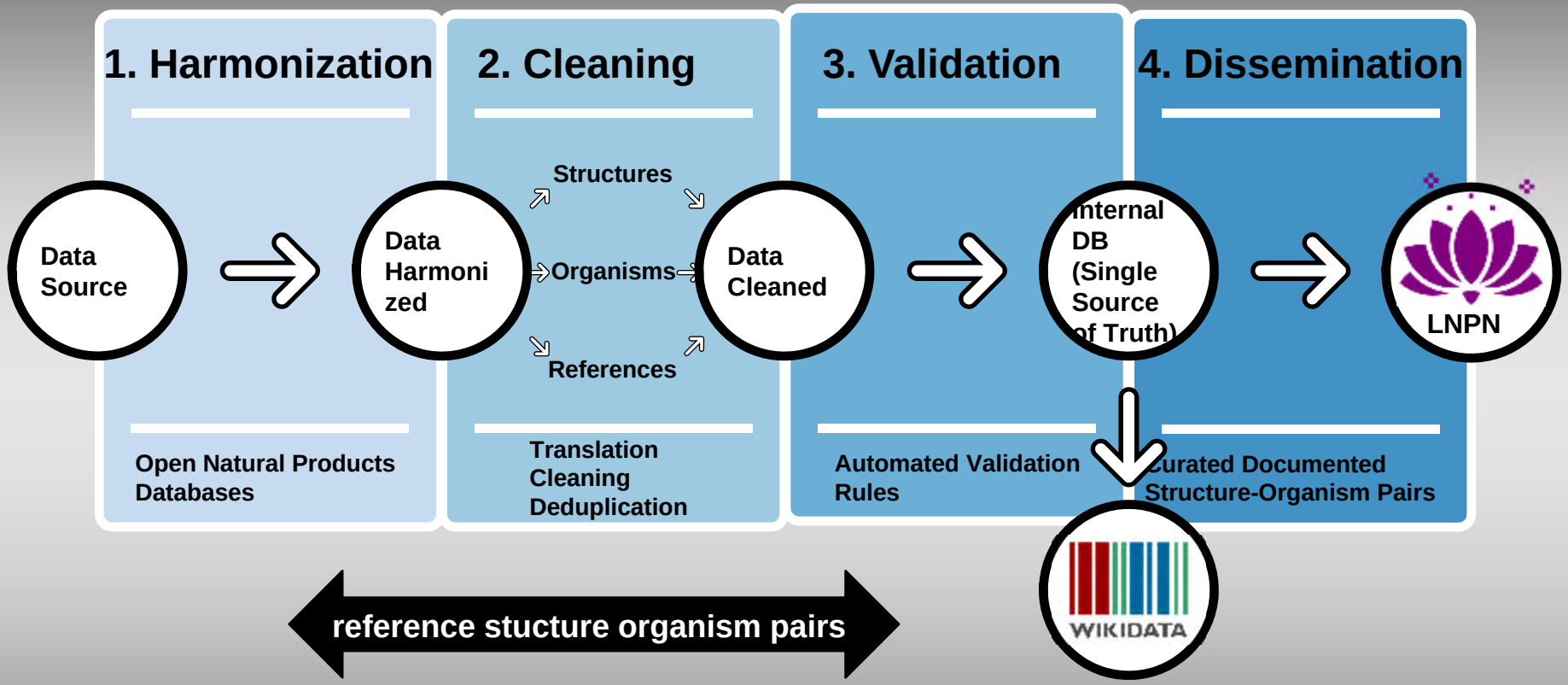
establish relationships between distinct molecular structures and the living organisms from which they were identified.

700,000+ referenced structure organism pairs



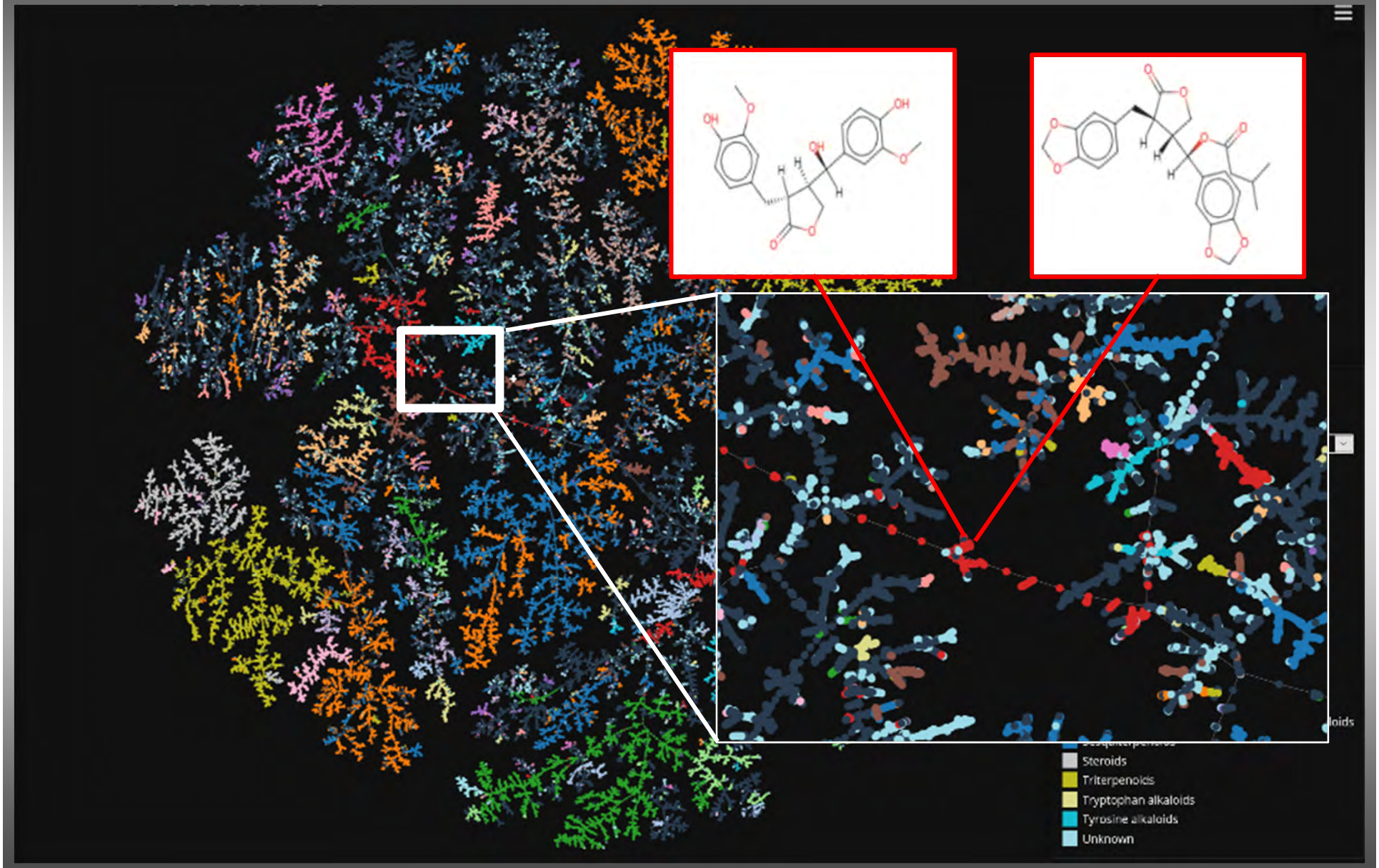
Rutz, A., Sorokina, M., Galgonek, J., Mietchen, D., Willighagen, E., Gaudry, A., Graham, J. G., Stephan, R., Page, R., Vondrášek, J., Steinbeck, C., Pauli, G. F., Wolfender, J.-L., Bisson, J. and Allard, P.-M. (2021). The LOTUS Initiative for Open Natural Products Research: Knowledge Management through Wikidata. *bioRxiv* 2021.2002.2028.433265.

The LOTUS Initiative for Knowledge Sharing in Natural Products Research



	Structure	Organism	Reference
Before curation	Cyathocaline	Stem bark of Cyathocalyx zeylanica CHAMP. ex HOOK. f. & THOMS. (Annonaceae)	Wijeratne E. M. K., de Silva L. B., Kikuchi T., Tezuka Y., Gunatilaka A. A. L., Kingston D. G. I., J. Nat. Prod., 58, 459-462 (1995).
After curation	**VFIIVOHWCNHINZ-UHFFFAOYSA-N**	***Cyathocalyx zeylanicus***	**10.1021/NP50117A020**

LOTUS chemical space organized as a TMAP

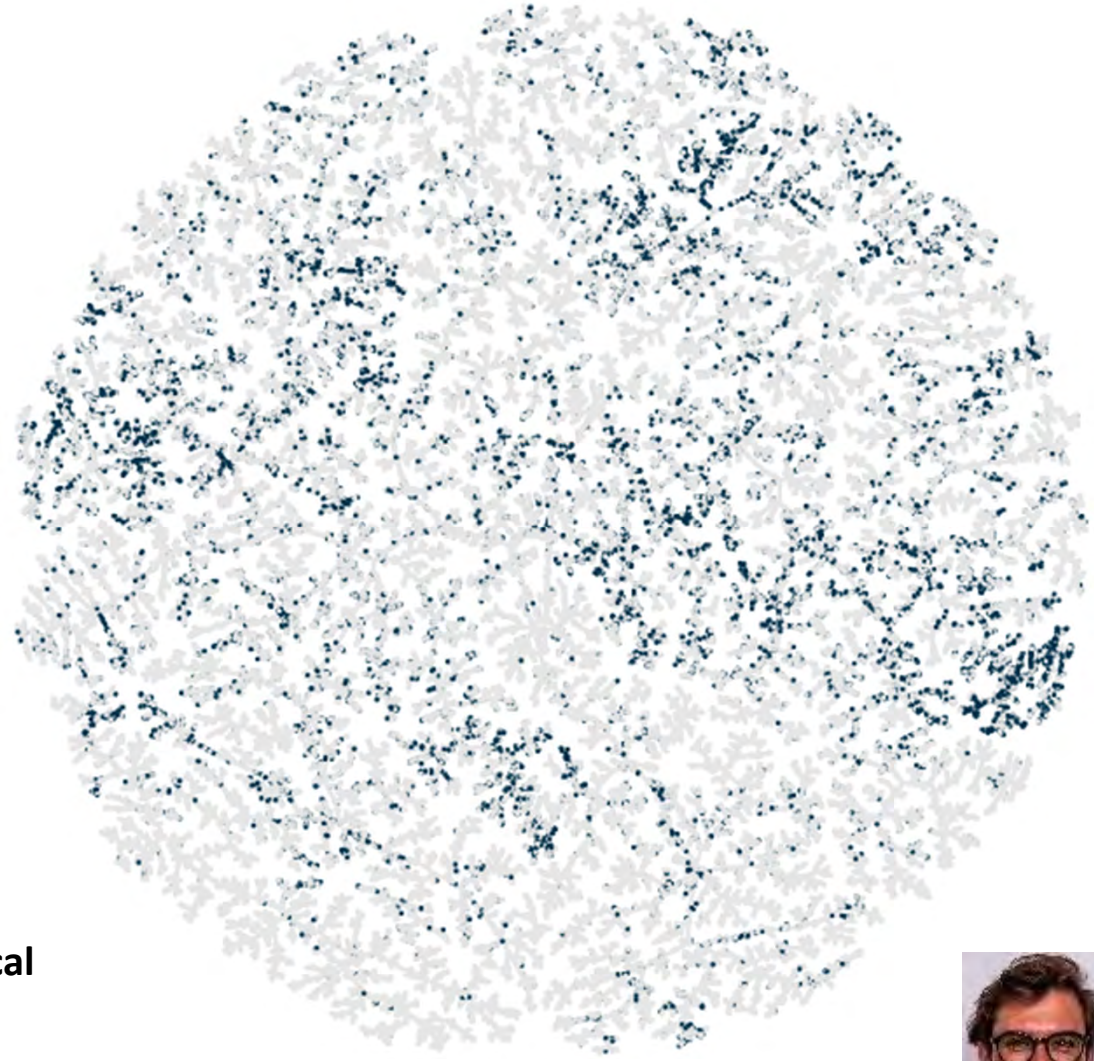


Annotations of the PF library on LOTUS TMAP

Annotation status

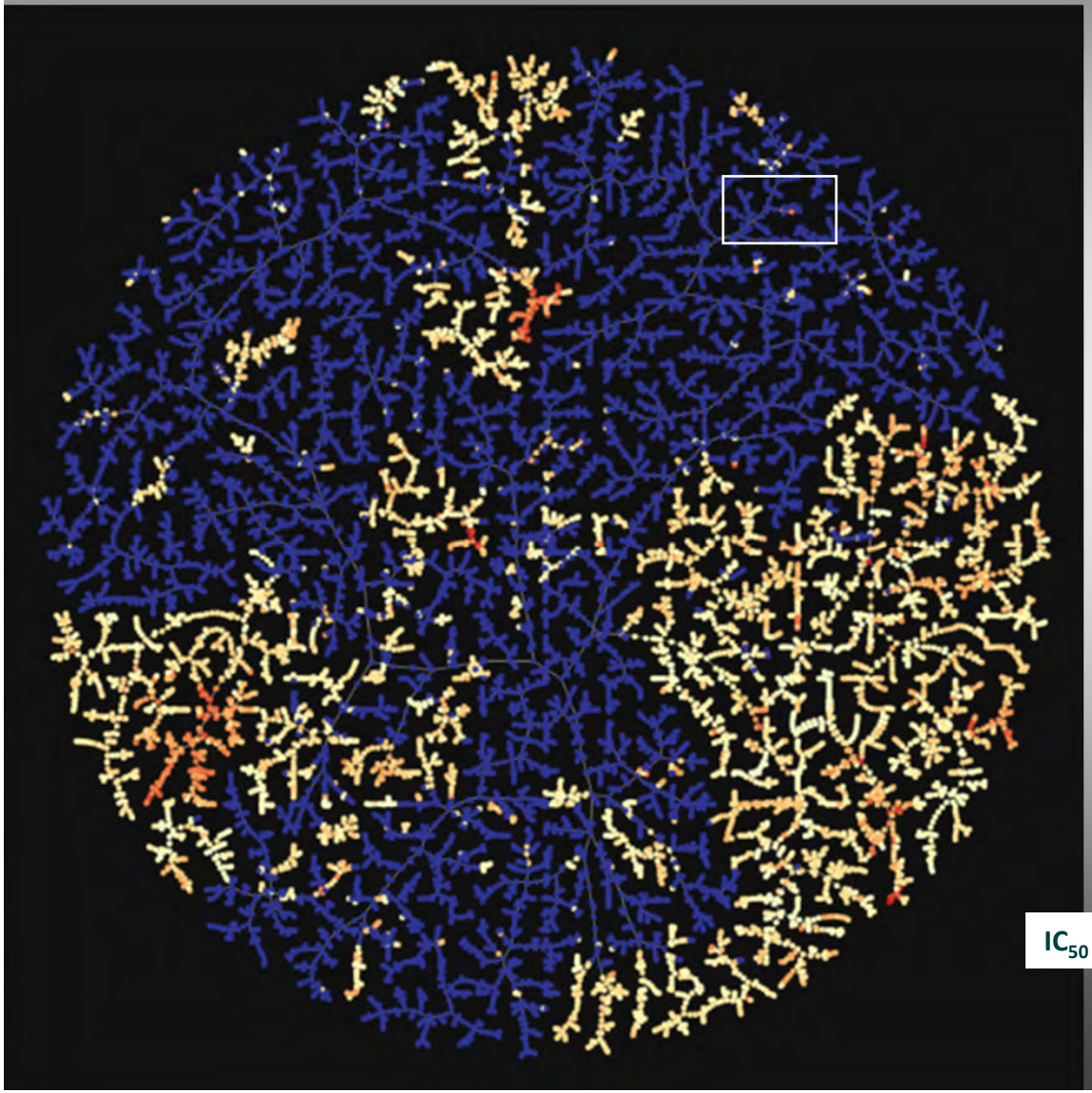
Not annotated

Annotated



Visualization of annotated chemical classes vs unannotated chemical classes

Cherry picking of intersecting NPs in the chemical space of annotations



- Blue
- Yellow

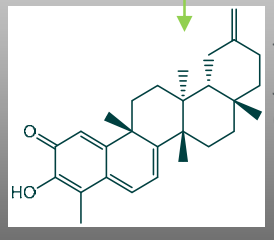
BL hits

= C ished activity against *L. donovani*

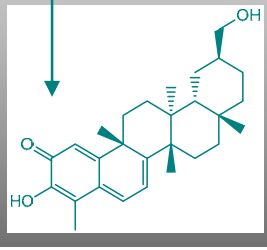
LCMS features for targeted isolation

★

m/z @RT



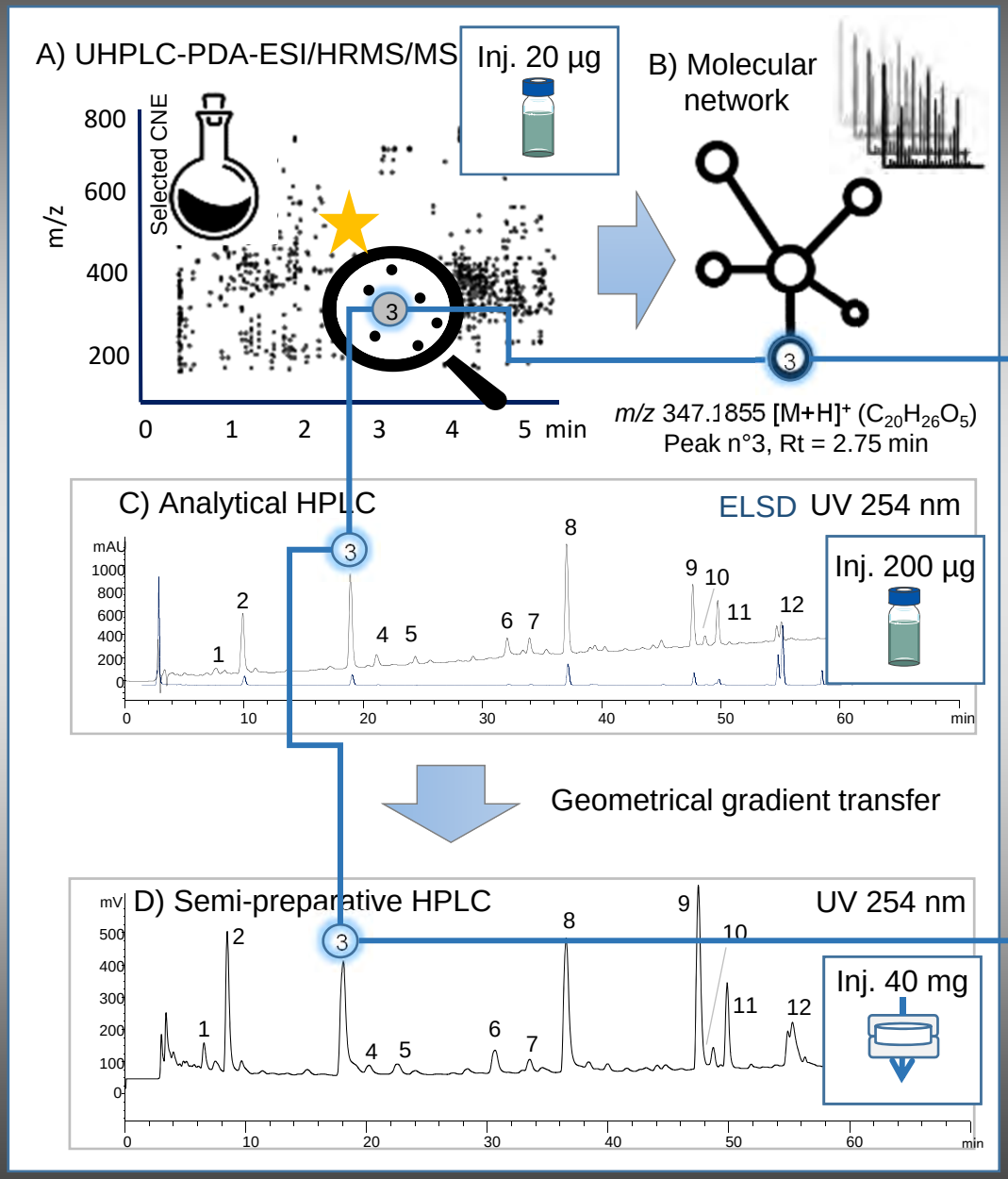
Thiem *et al*, JNP (2005) 68:251-254

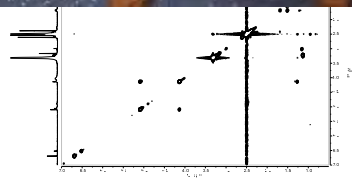


IC_{50} 0.036 µg/mL, SI = 50

IC_{50} 0.027 µg/mL, SI = 31

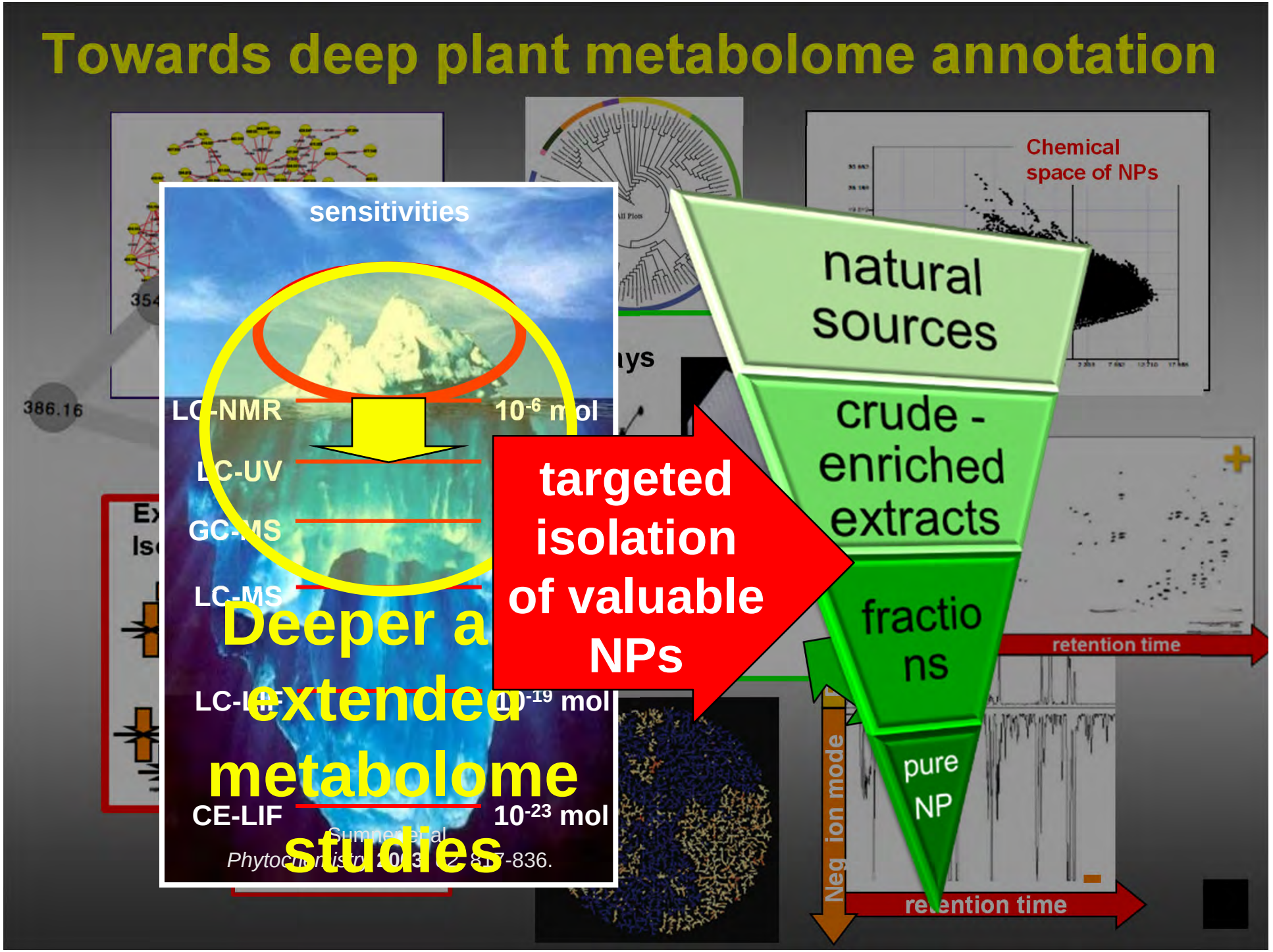
High resolution targeted isolation





COSY NMR spectrum of peak 3 in DMSO-d_6

Towards deep plant metabolome annotation



Acknowledgments

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES

ISPSO
INSTITUT DES SCIENCES PHARMACEUTIQUES DE SUISSE OCCIDENTALE

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

PLANT SURVIVAL
National Centre of Competence in Research

AGENCE NATIONALE DE LA RECHERCHE
ANR