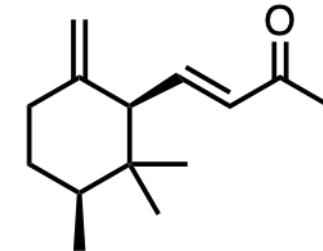


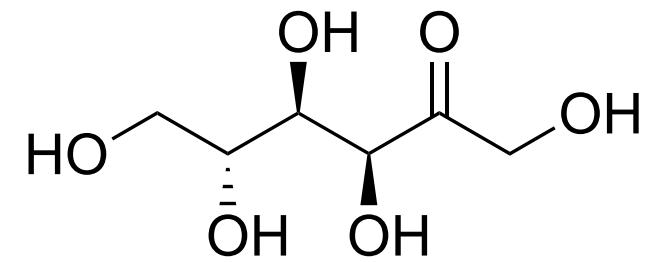
# Prior knowledge as a compass guiding untargeted metabolomics

Jena Centre for Bioinformatics (JCB)  
Ernst-Abbe-Platz 2

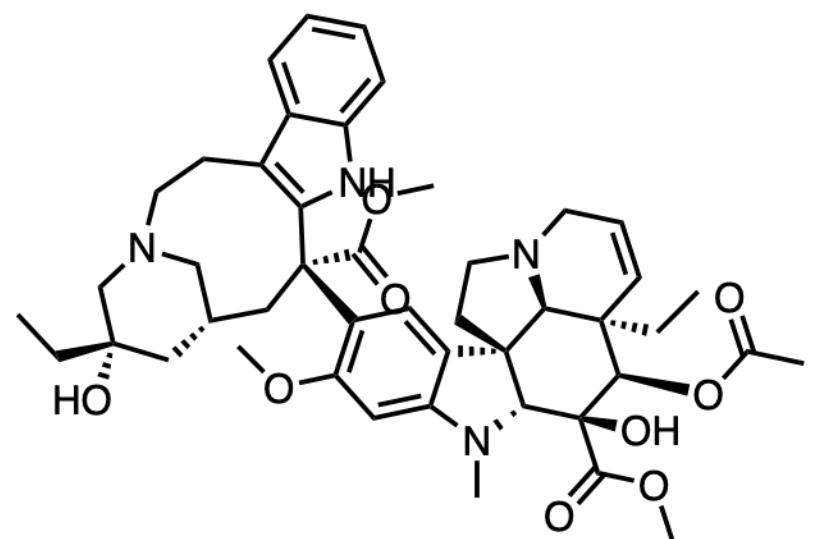
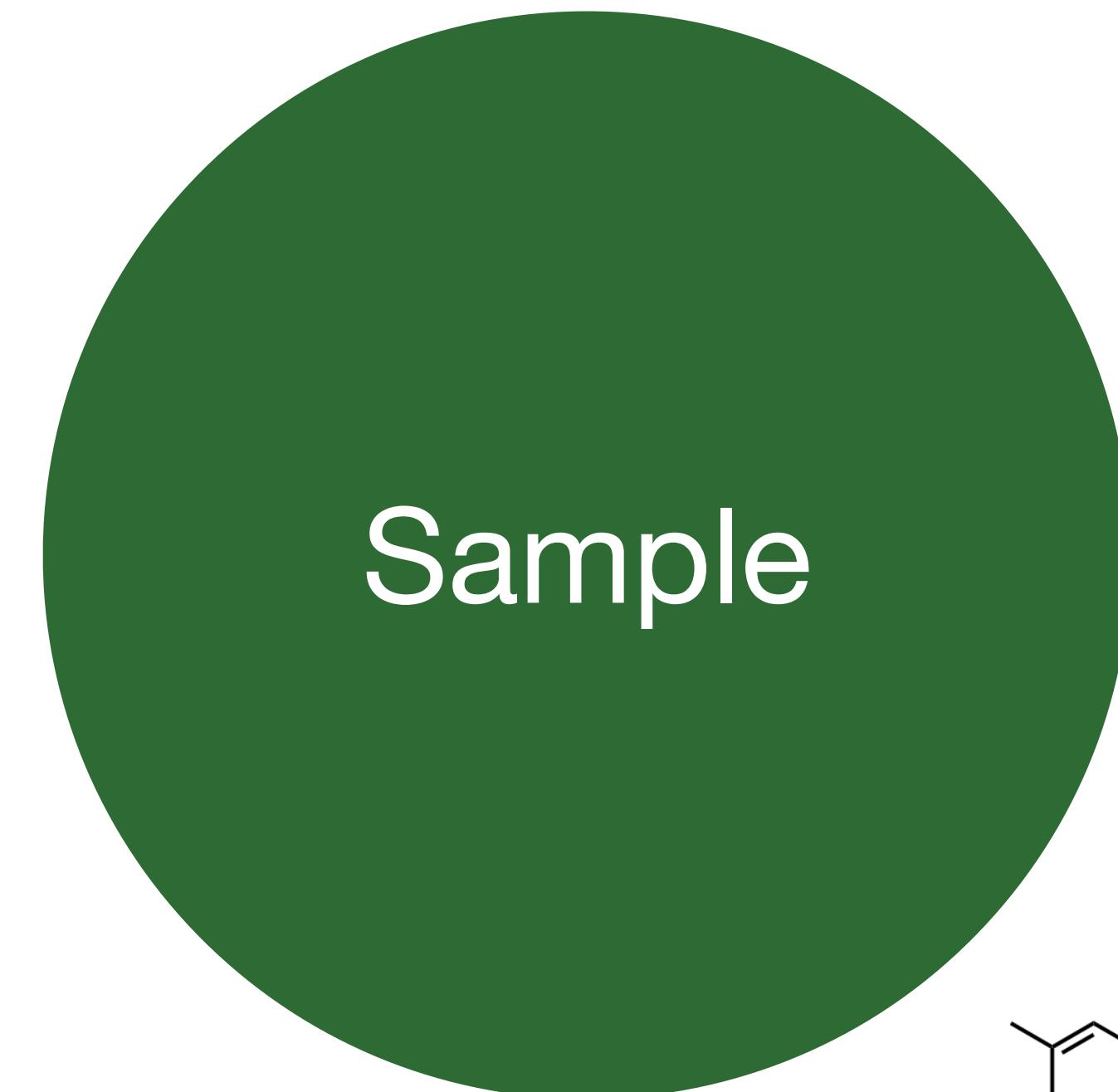
# Chemical analysis



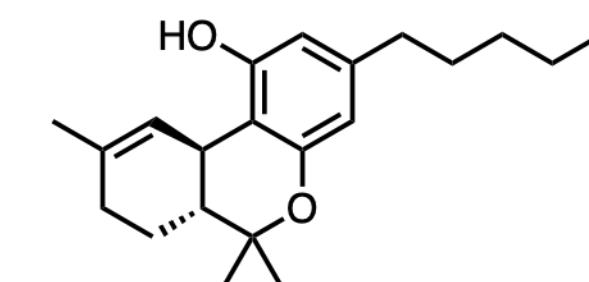
Volatiles



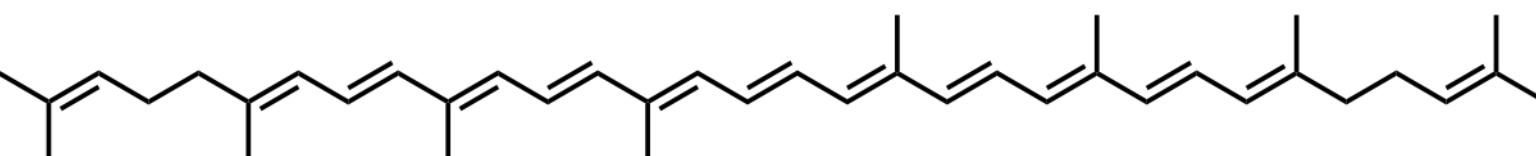
Sugars



Natural products



Exposome

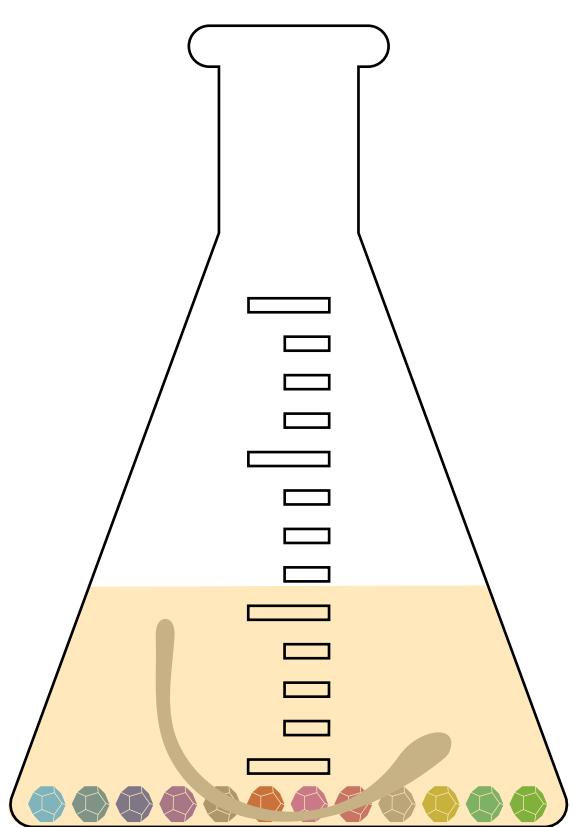
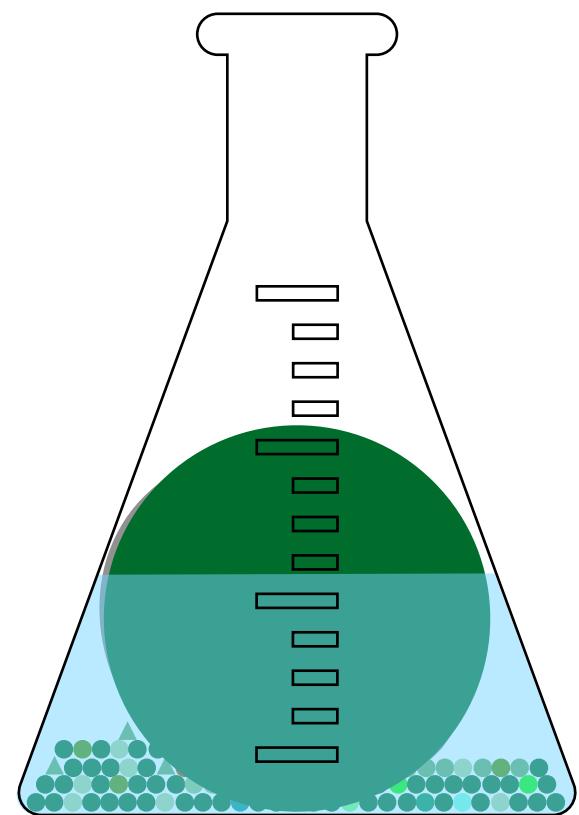


Lipids

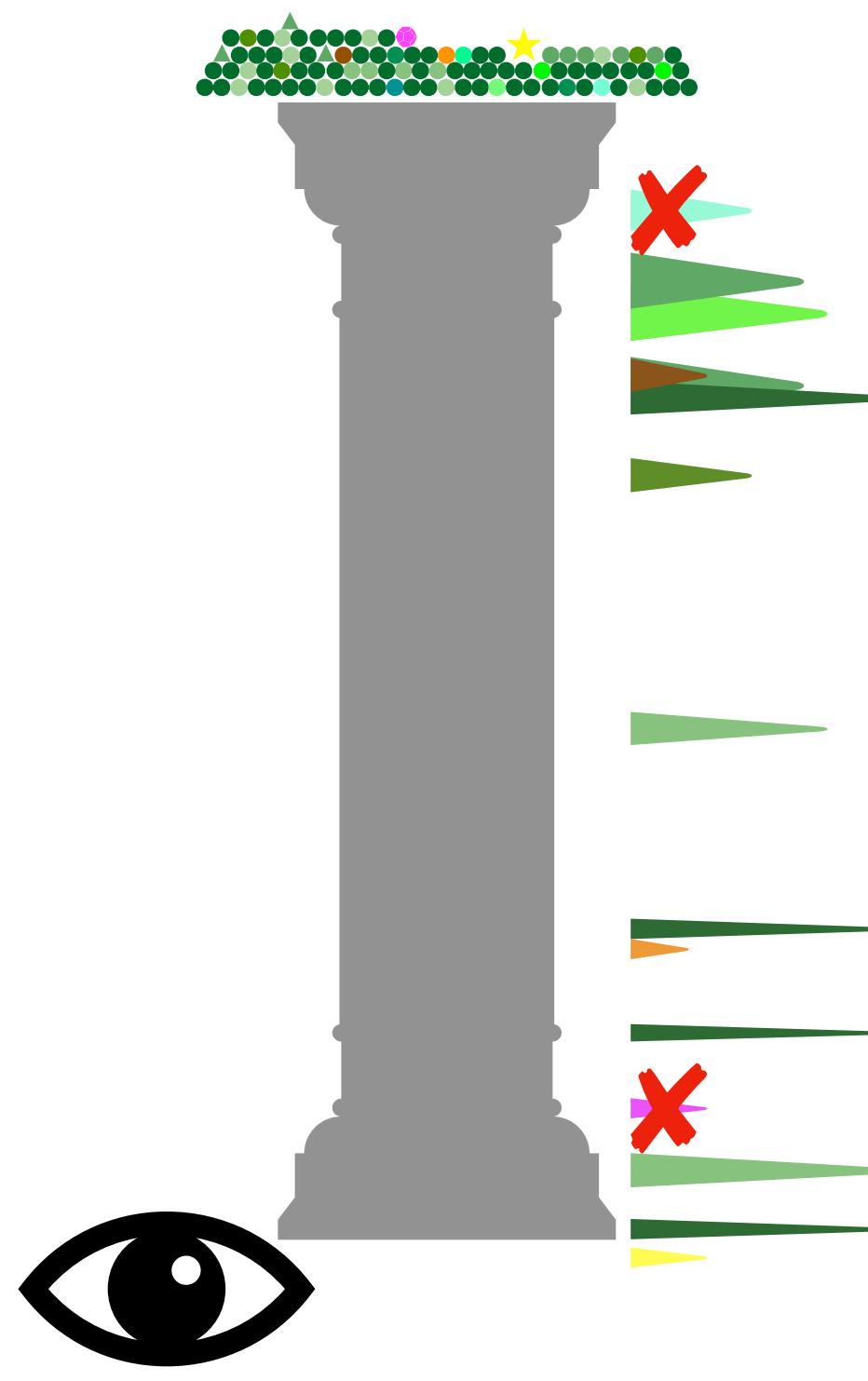
# Chemical analysis



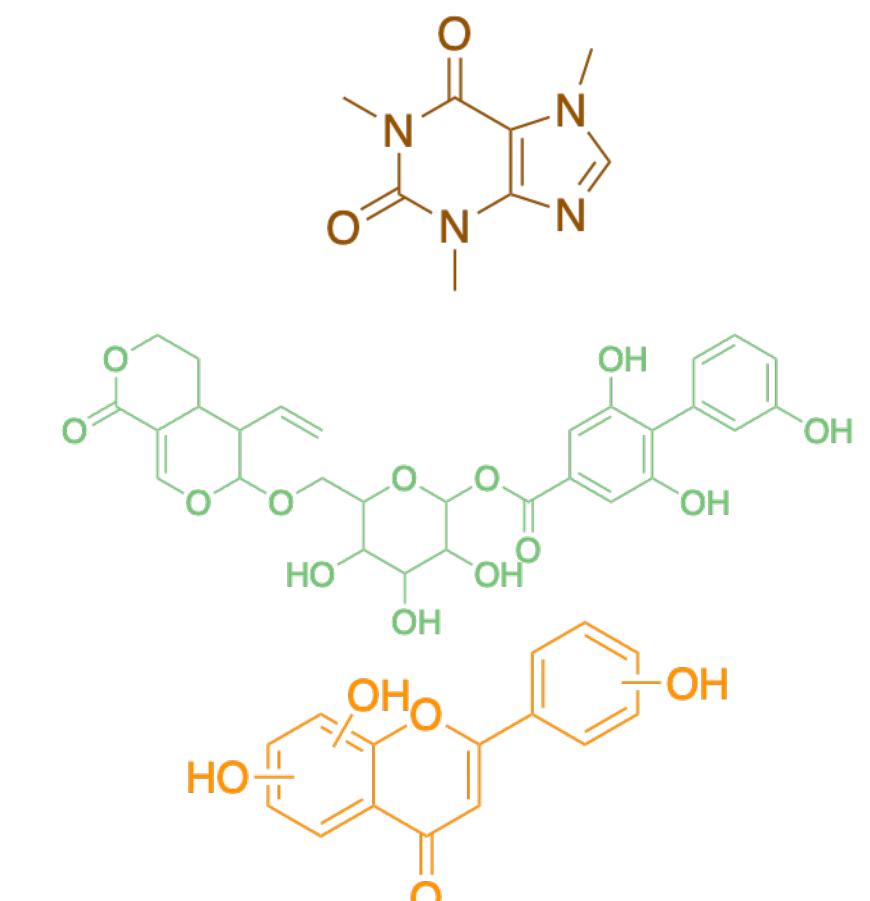
Extract(s)



Chromatography

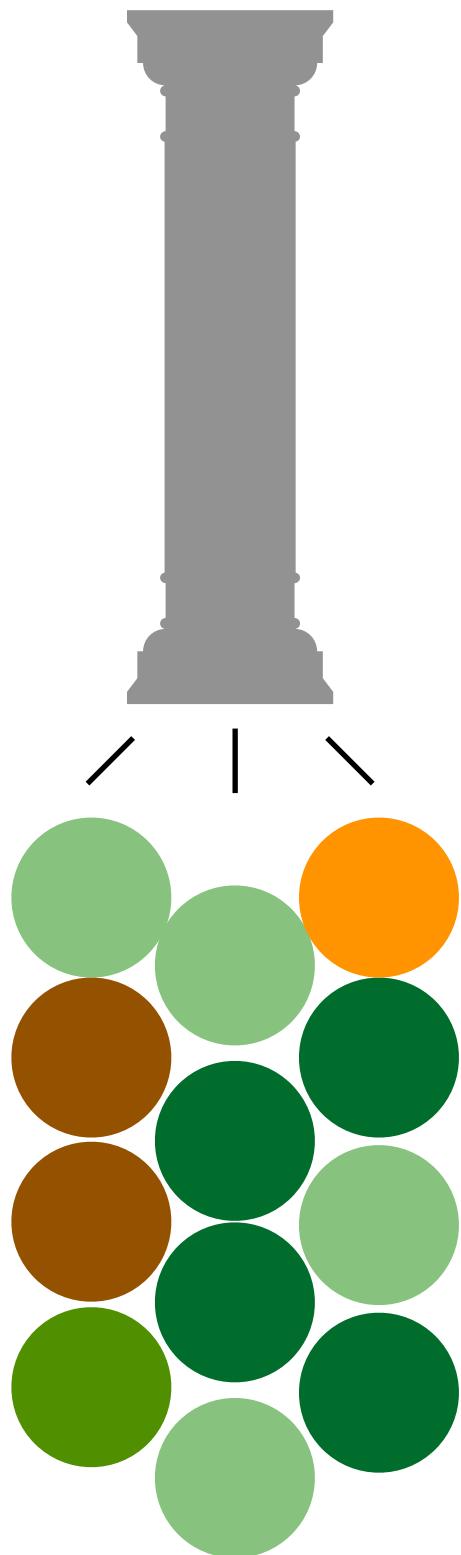


Annotation

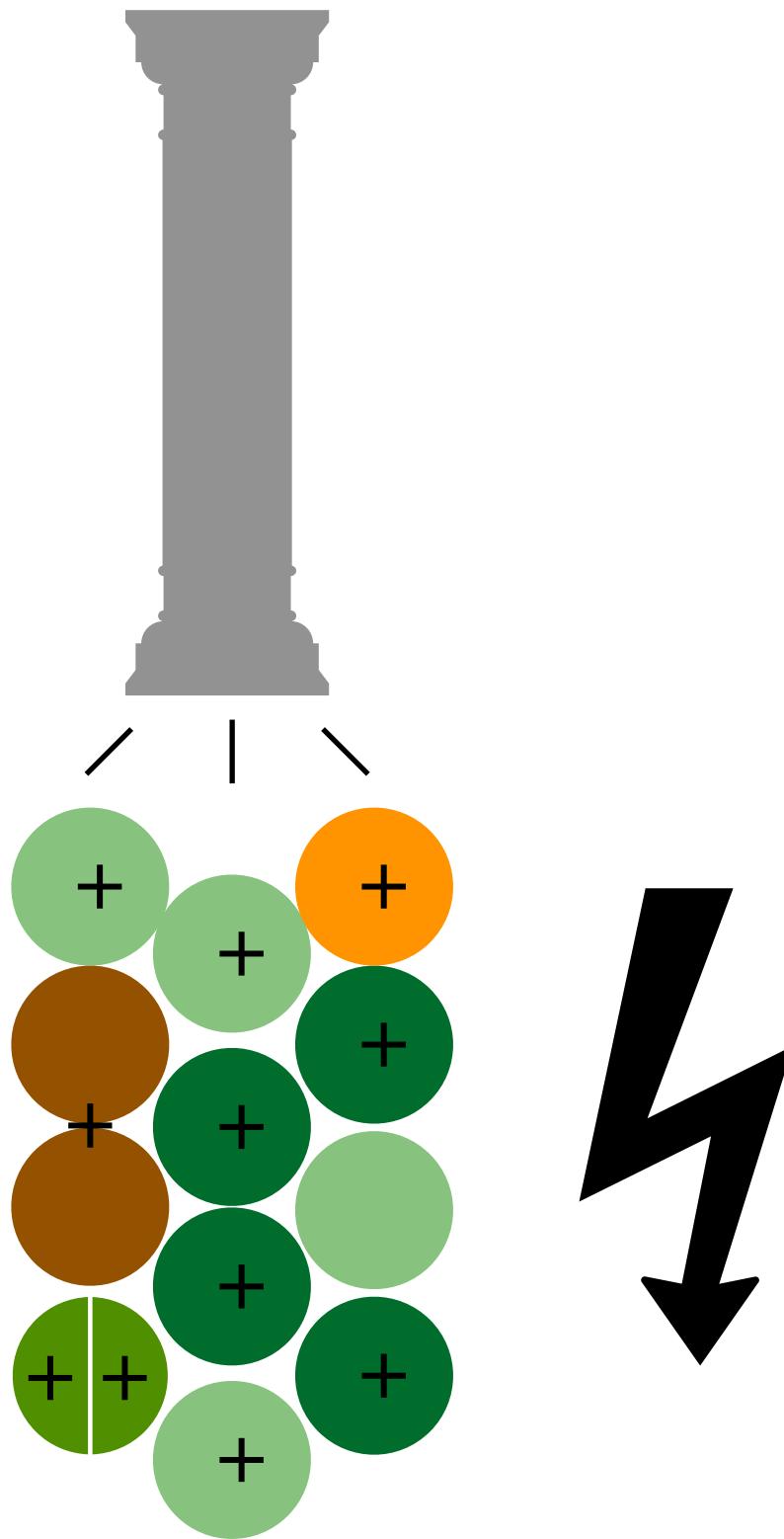


*Fragmenting like a terpenoid*  
?

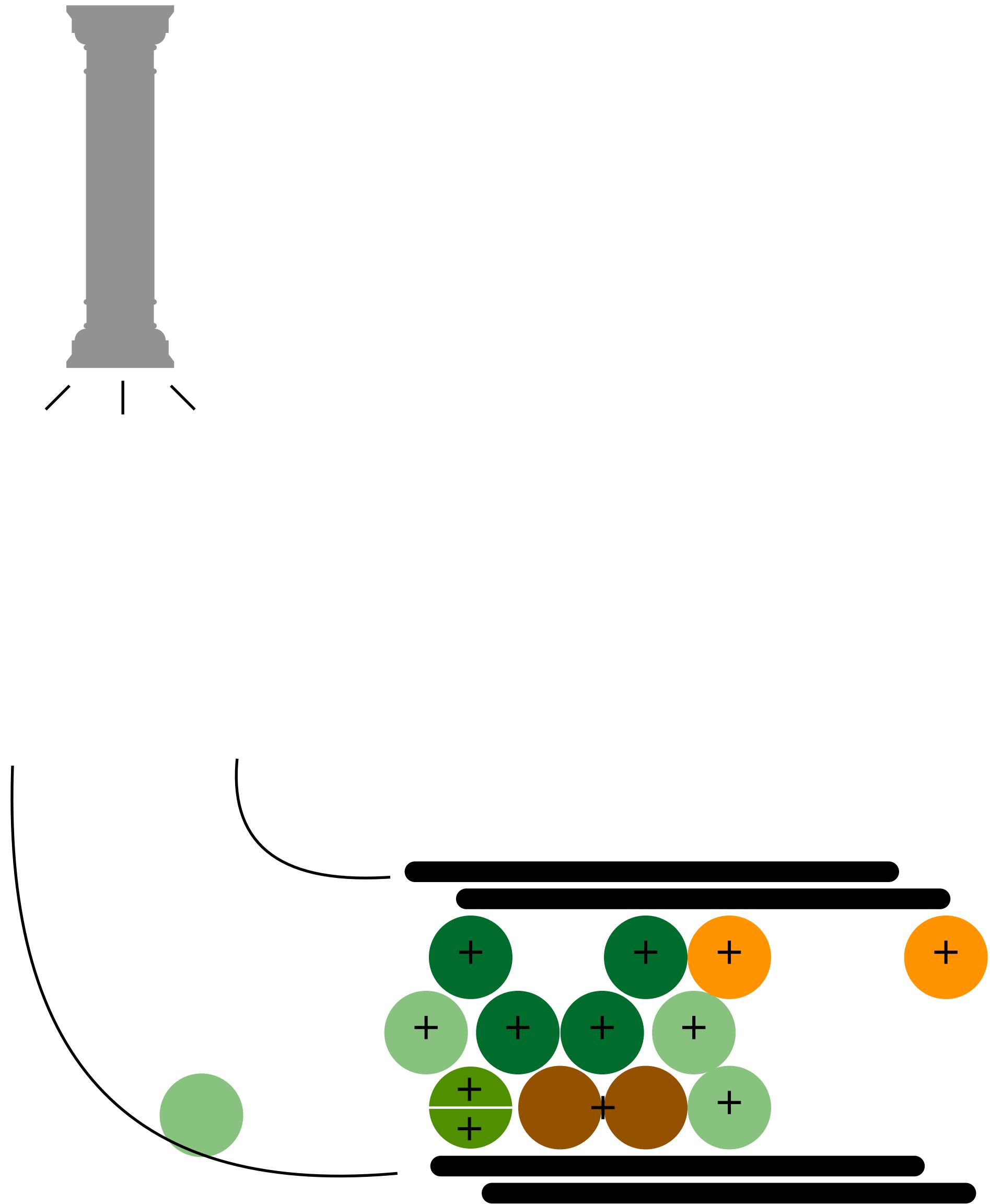
# Electrospray ionization mass spectrometry



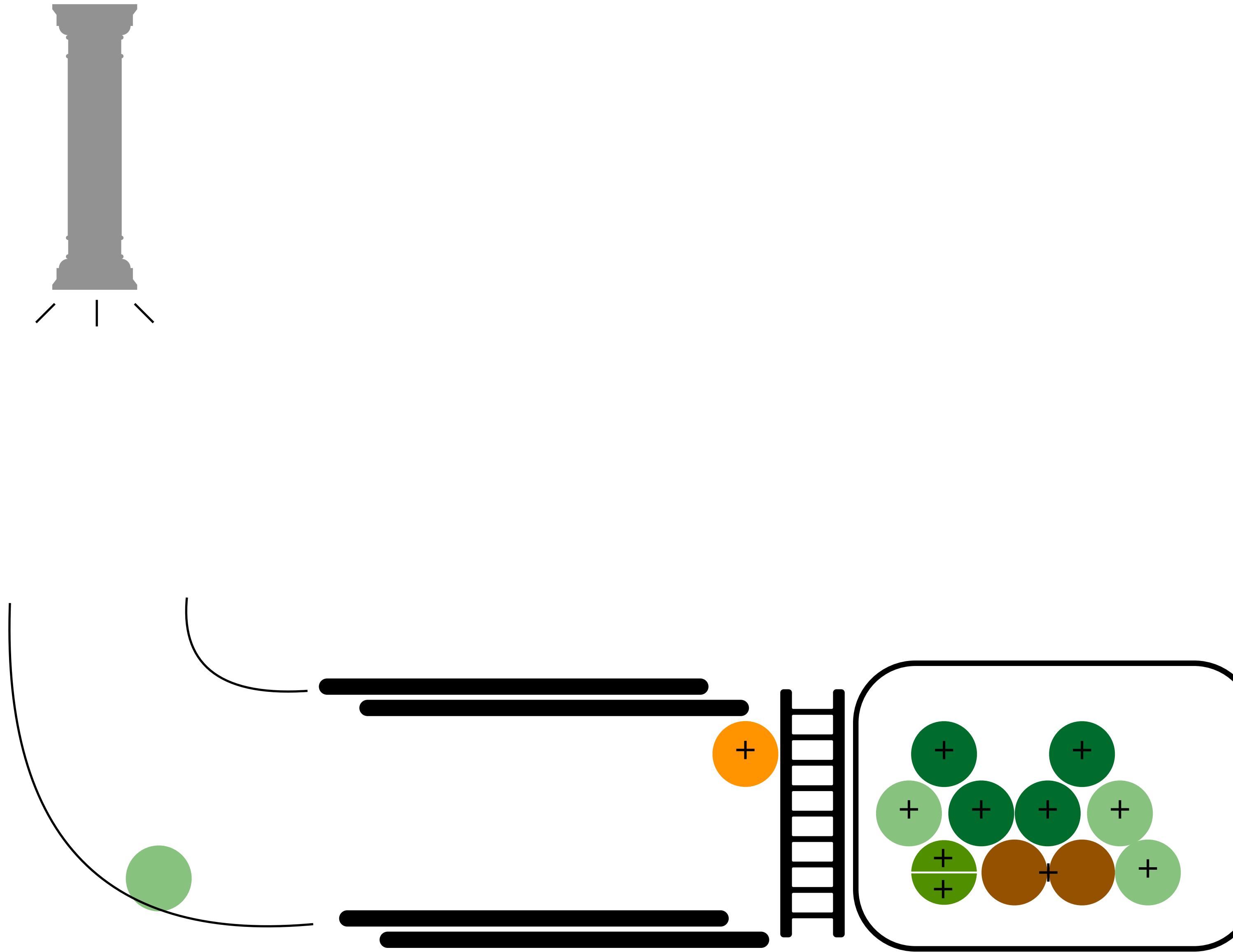
# Electrospray ionization mass spectrometry



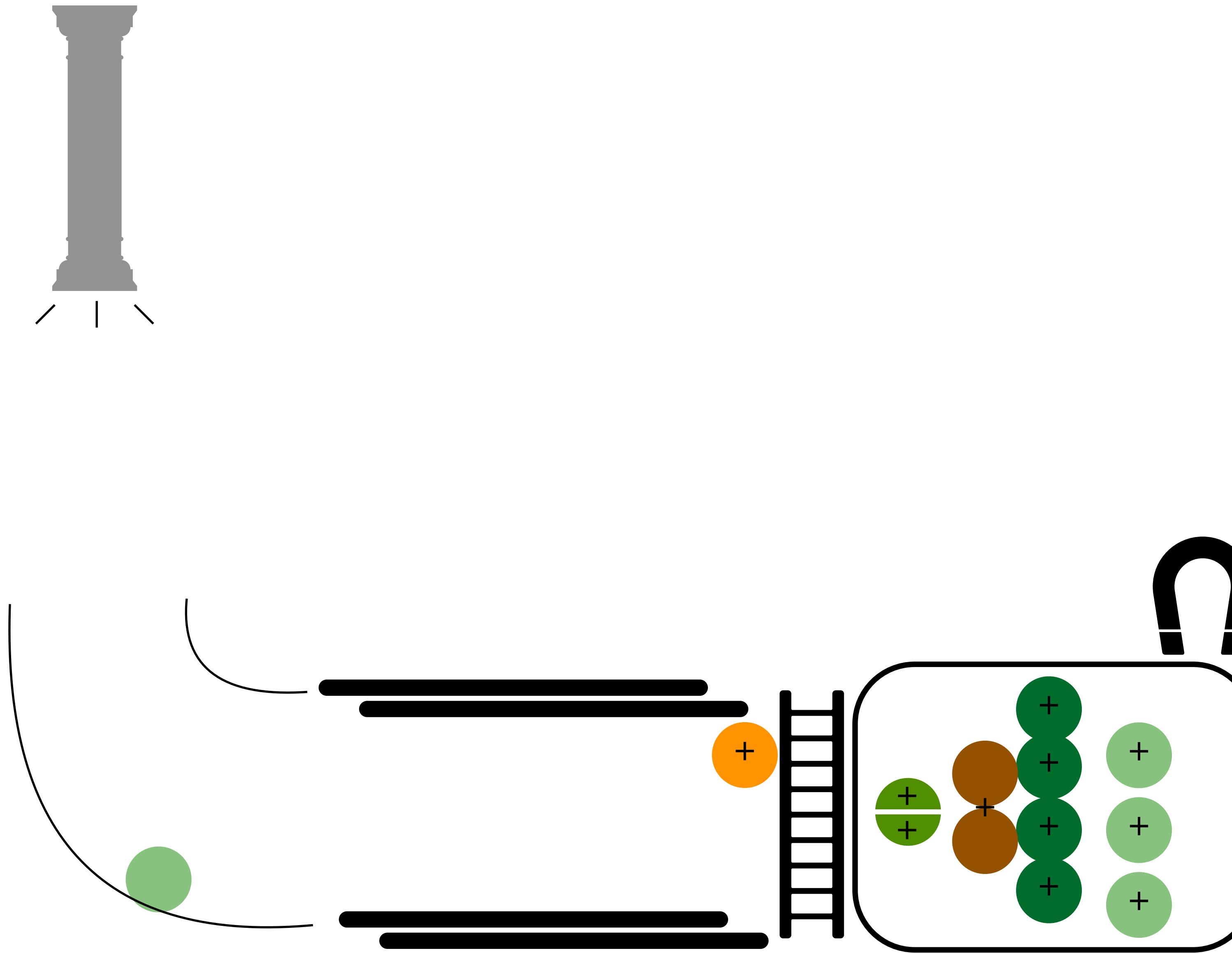
# Electrospray ionization mass spectrometry



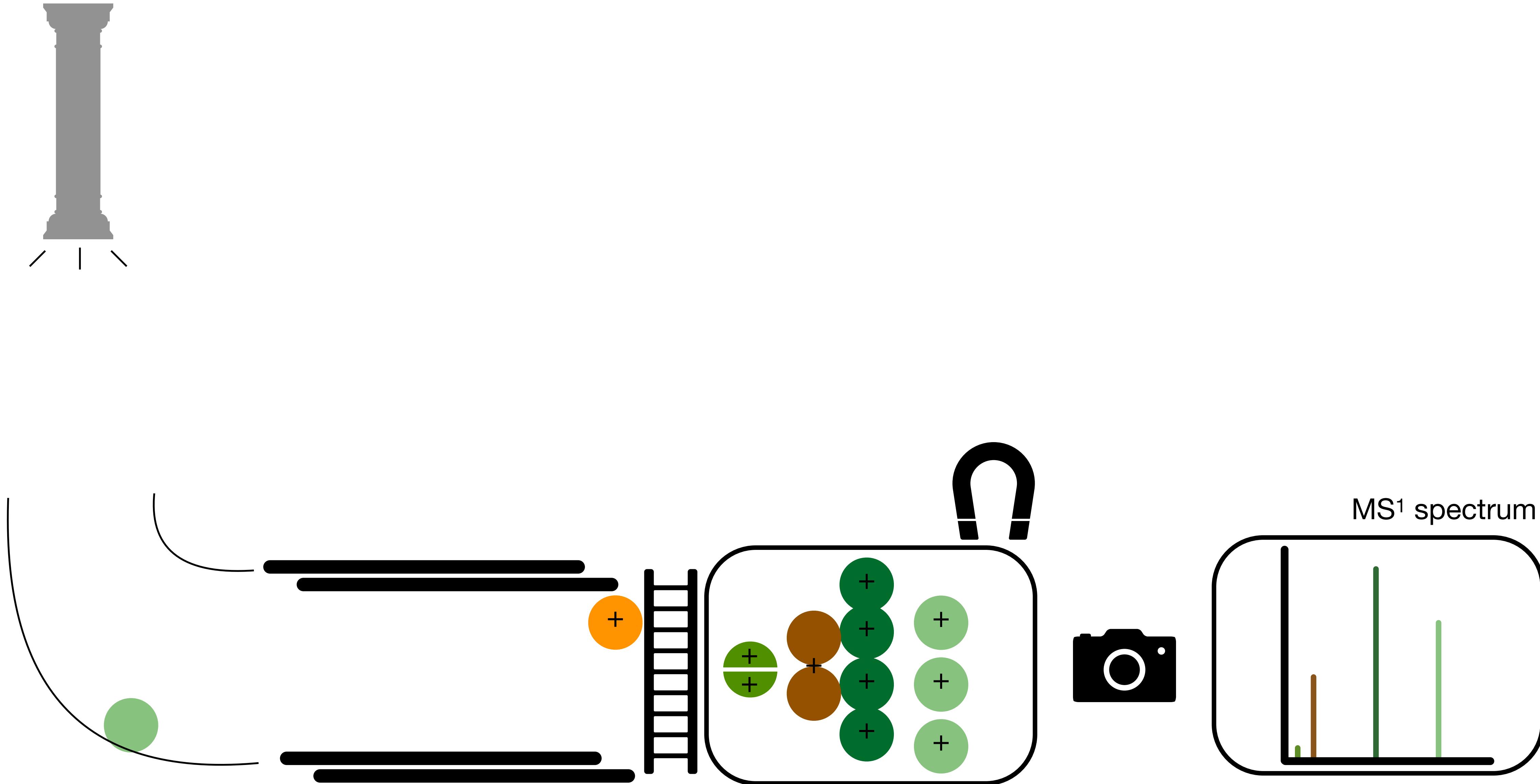
# Electrospray ionization mass spectrometry



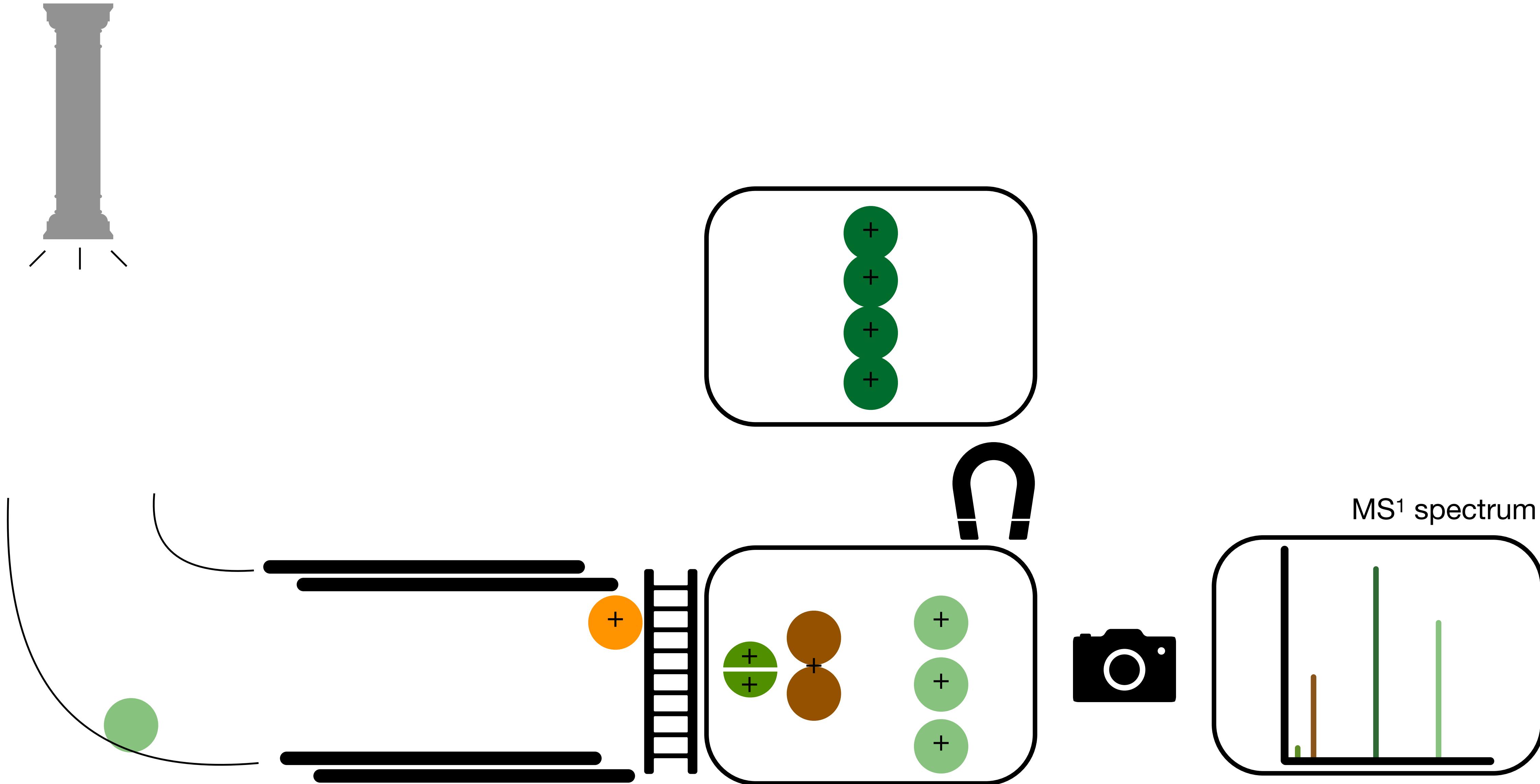
# Electrospray ionization mass spectrometry



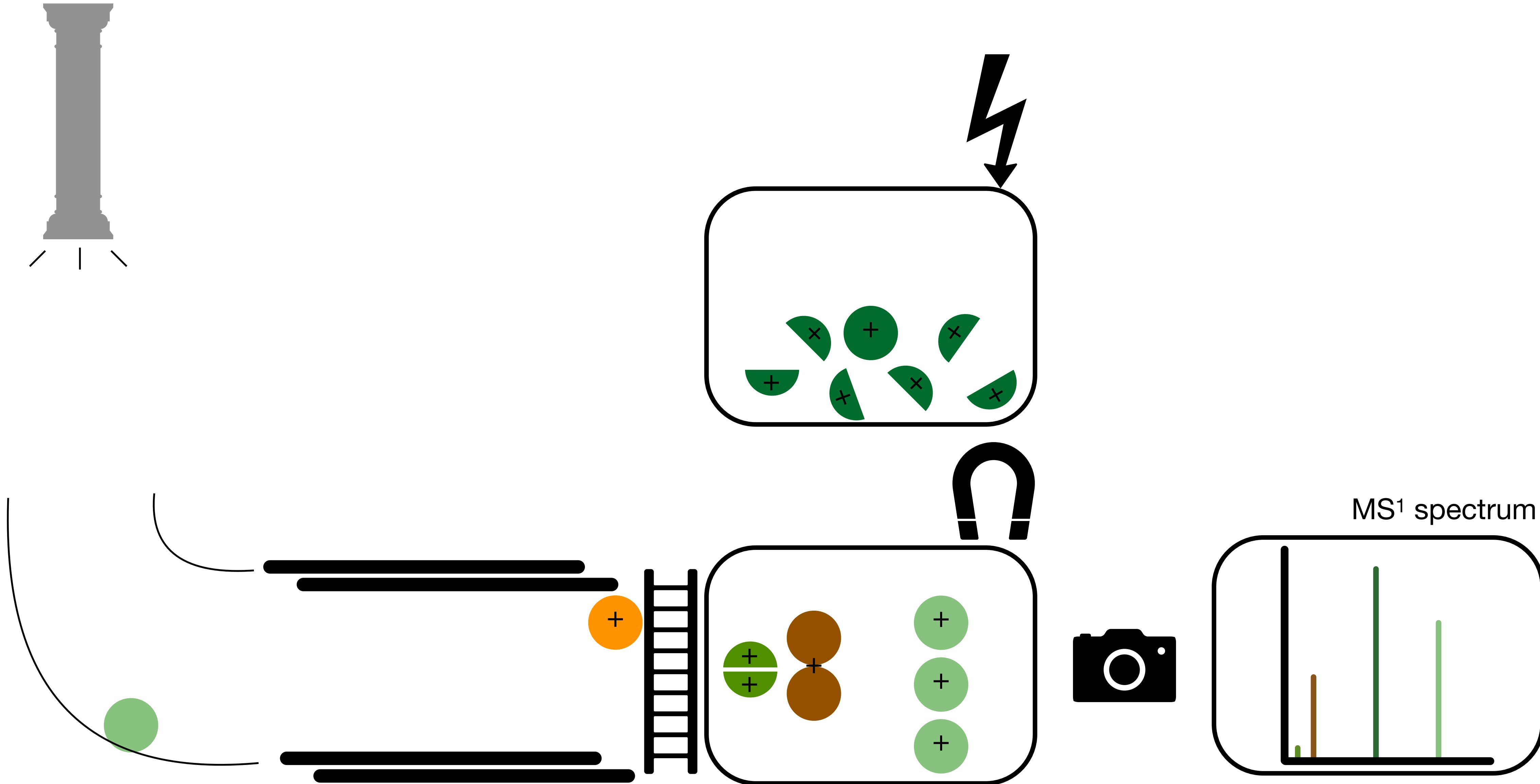
# Electrospray ionization mass spectrometry



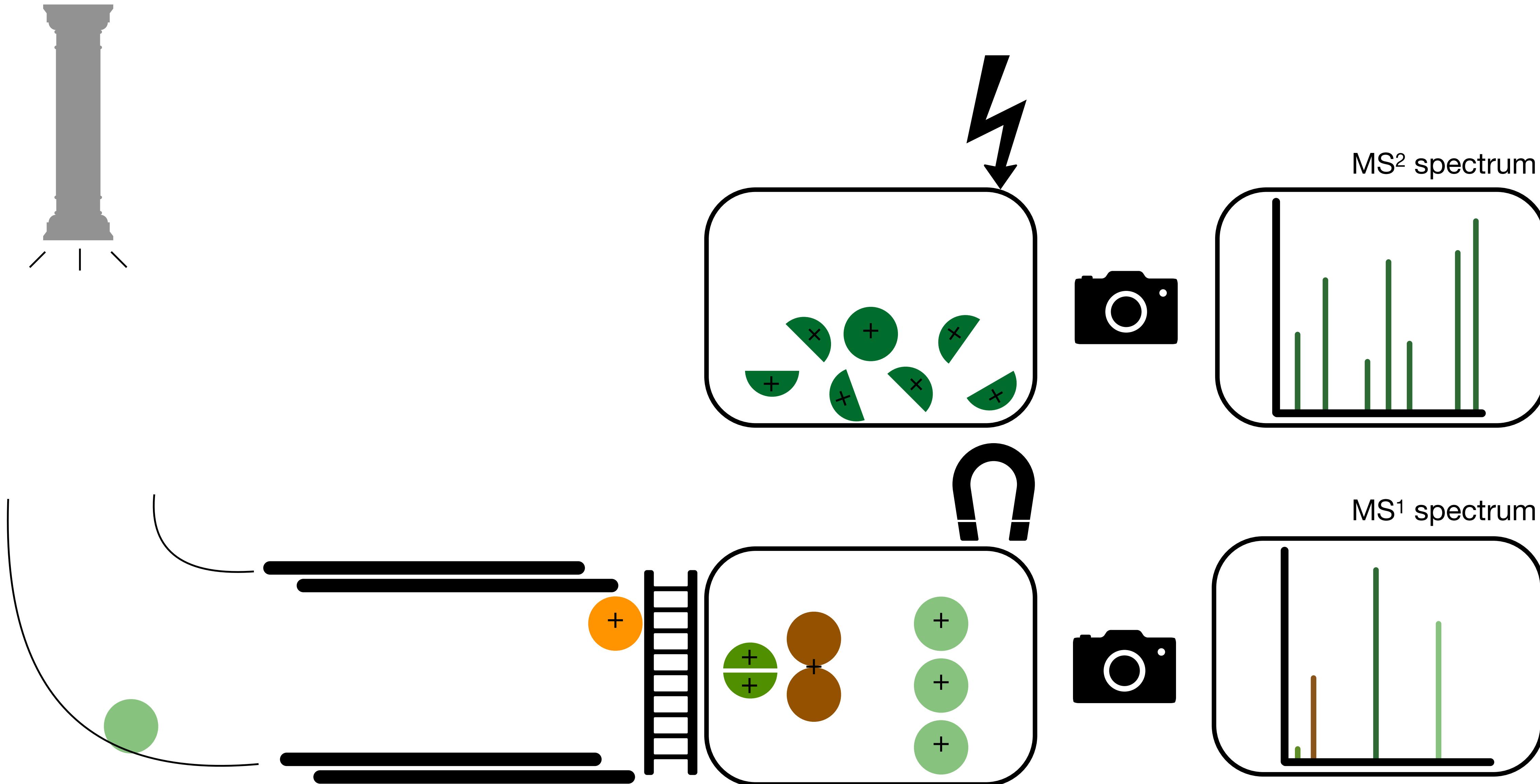
# Electrospray ionization mass spectrometry



# Electrospray ionization mass spectrometry

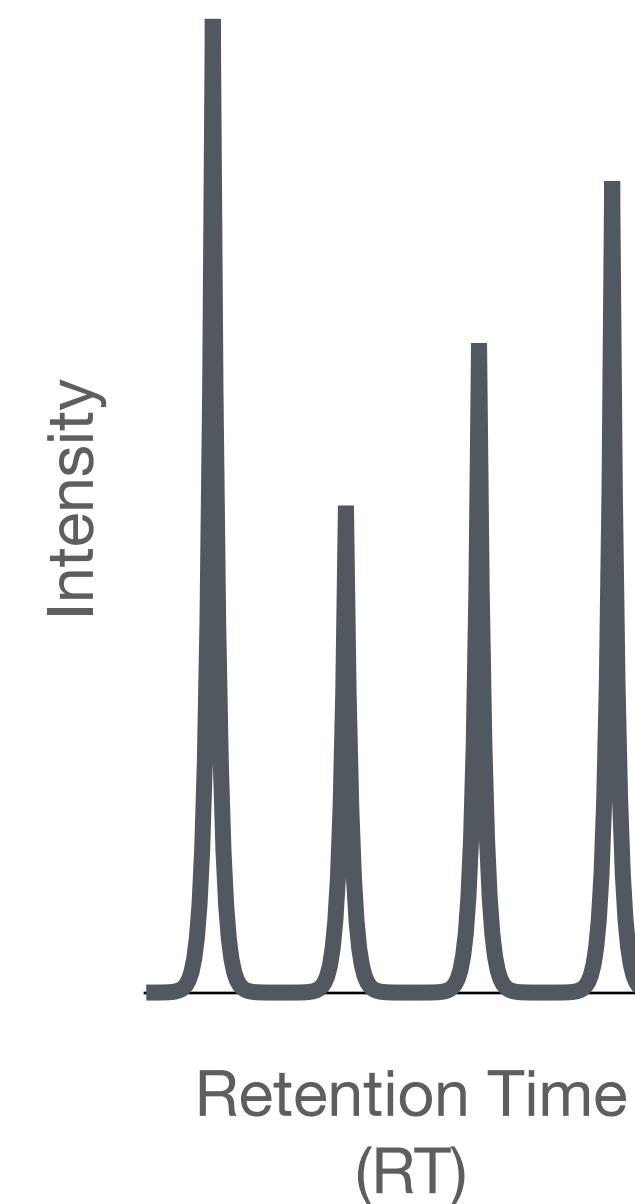


# Electrospray ionization mass spectrometry

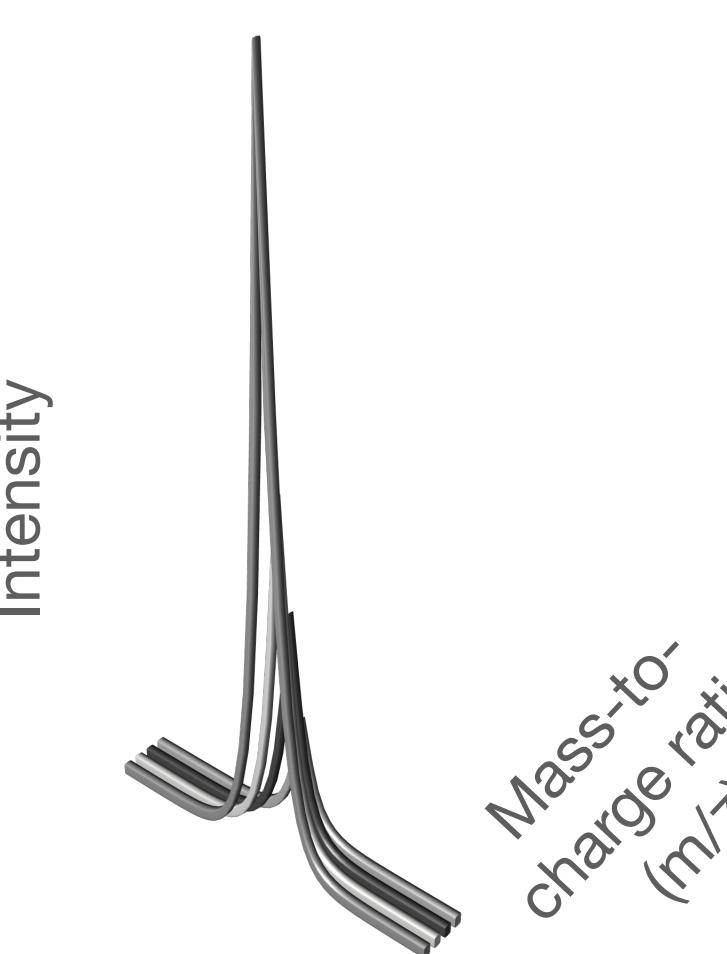


# Glossary

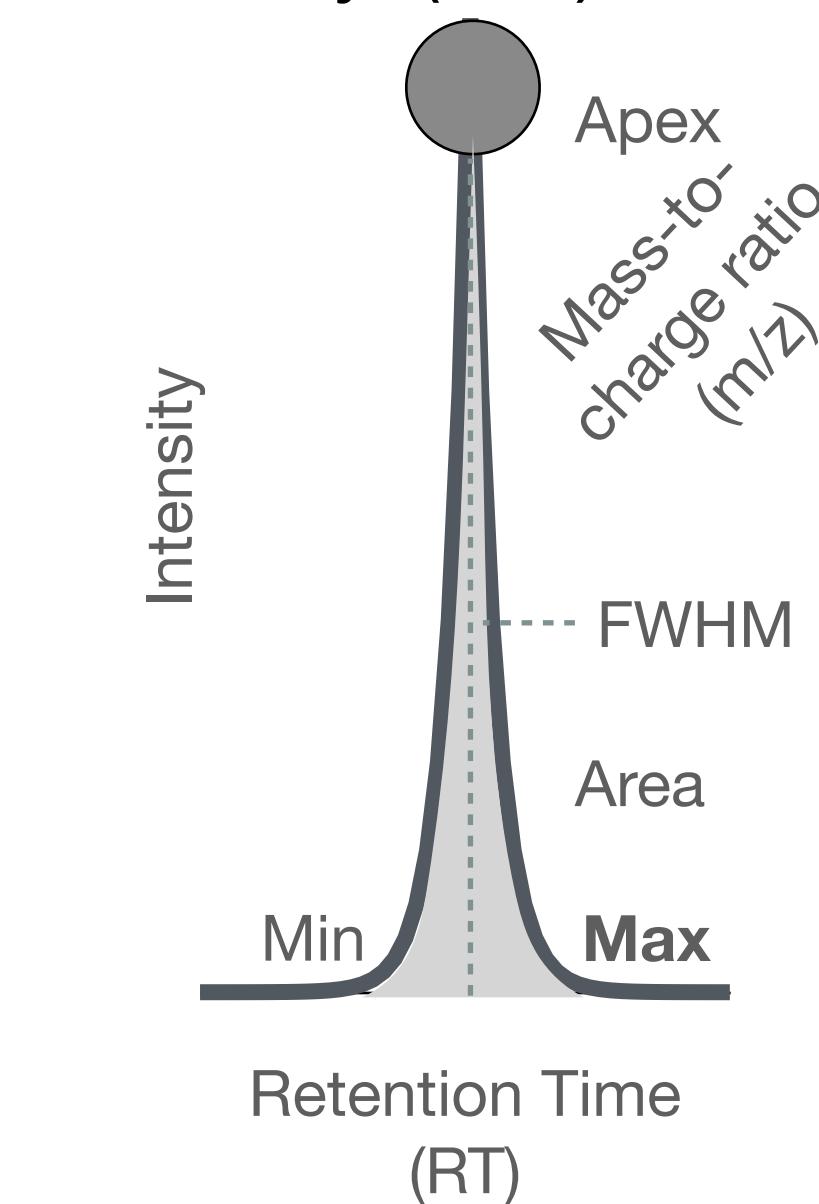
Chromatogram:  
Intensity =  $f(RT)$



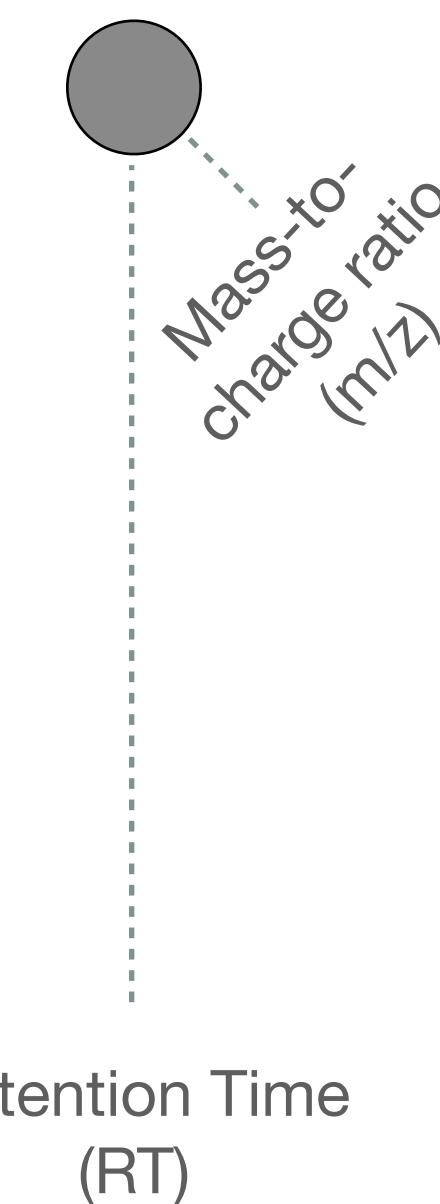
Spectrum:  
Intensity =  $f(m/z)$



Peak:  
Parameters with  
intensity, ( $m/z$ ) and RT



Feature:  
RT @  $m/z$  pointer

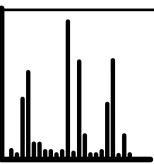


Feature list (or table):

ID	RT	$m/z$	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	

# Metabolite annotation

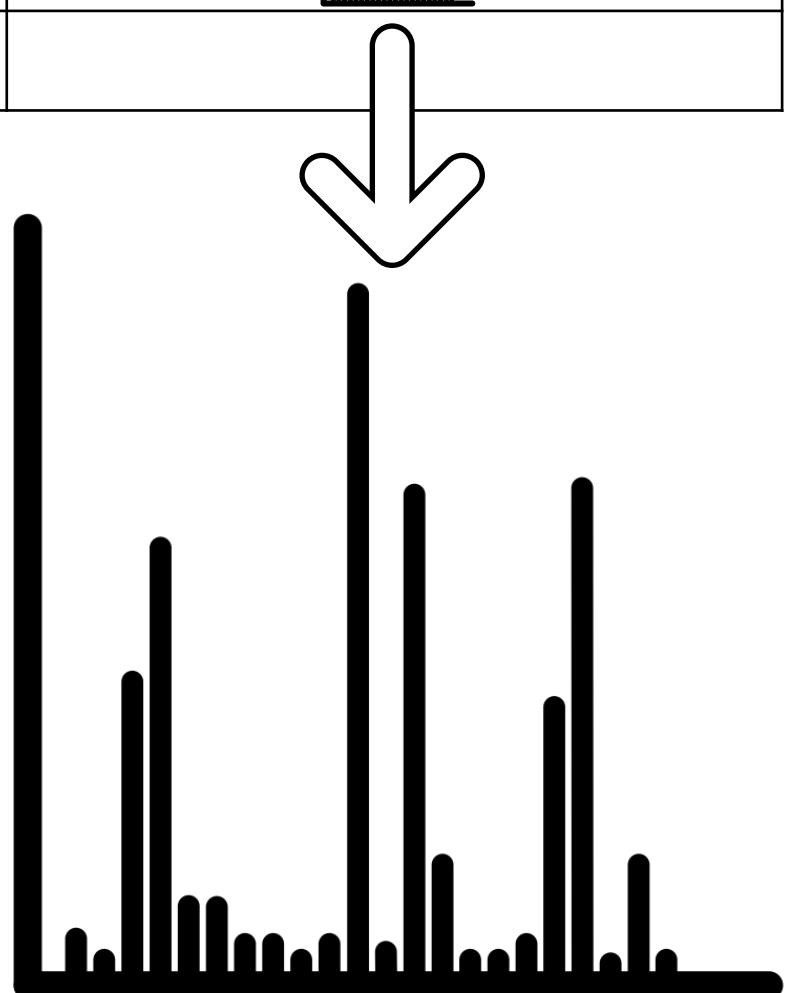
Feature list (or table):

ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	....	

# Metabolite annotation

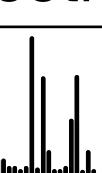
Feature list (or table):

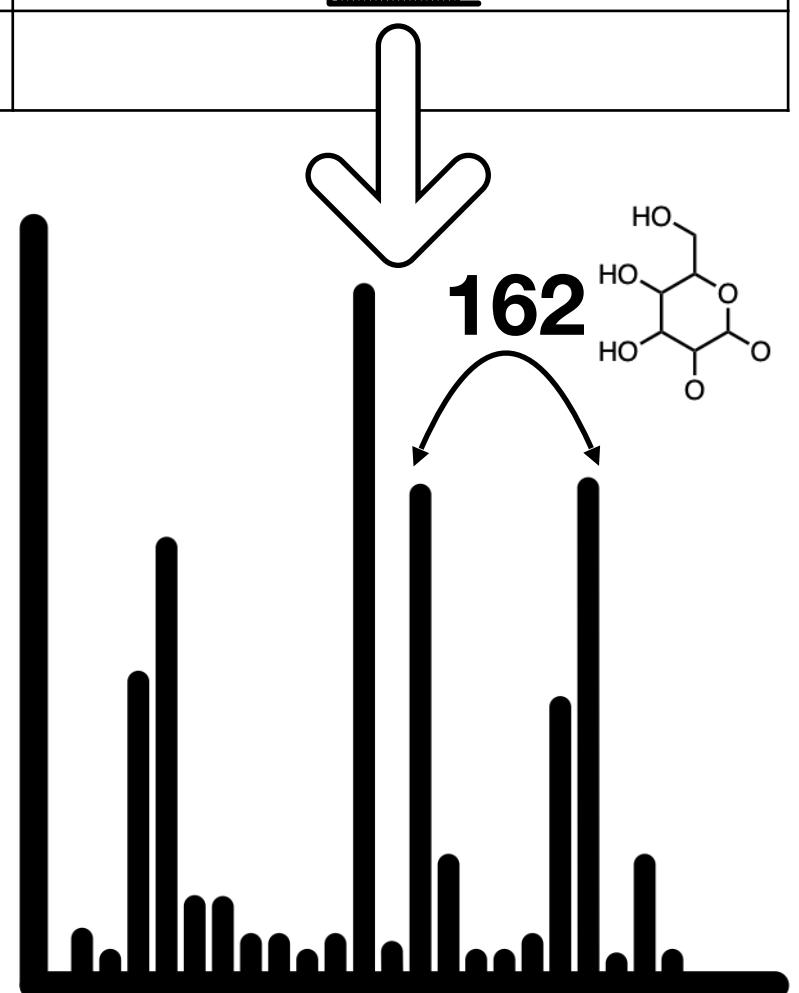
ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



# Metabolite annotation

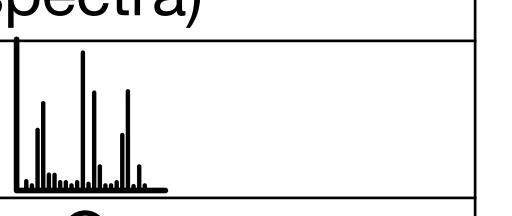
Feature list (or table):

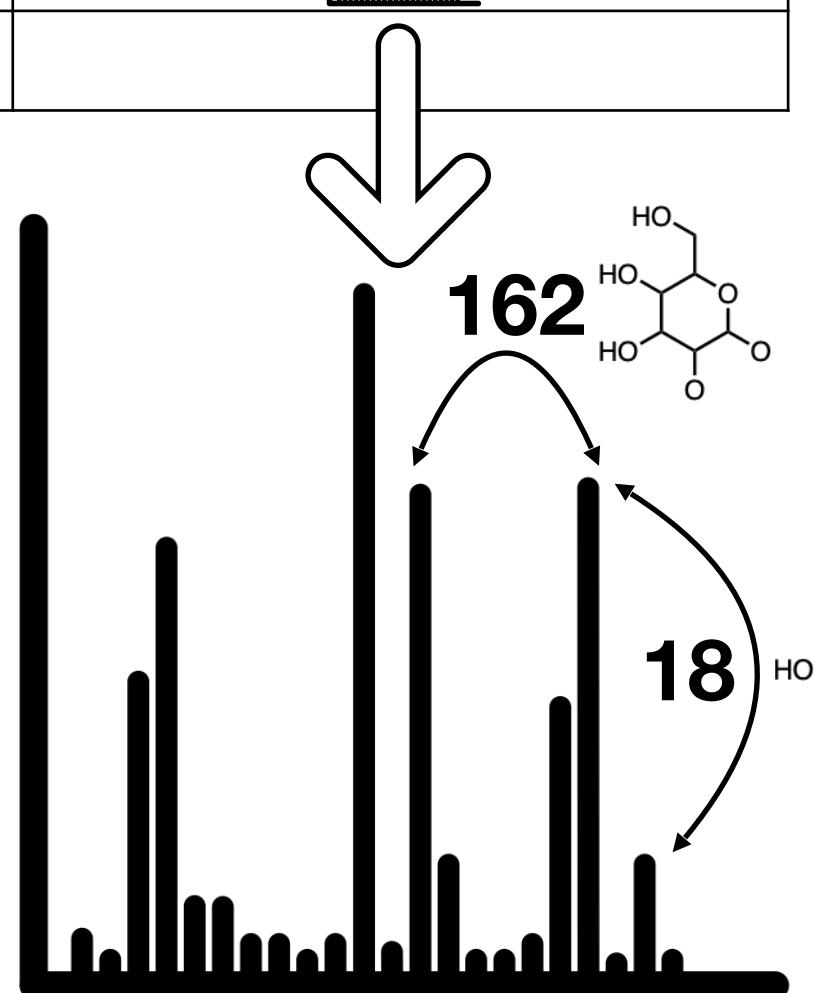
ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



# Metabolite annotation

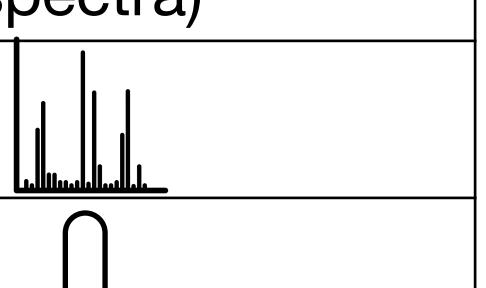
Feature list (or table):

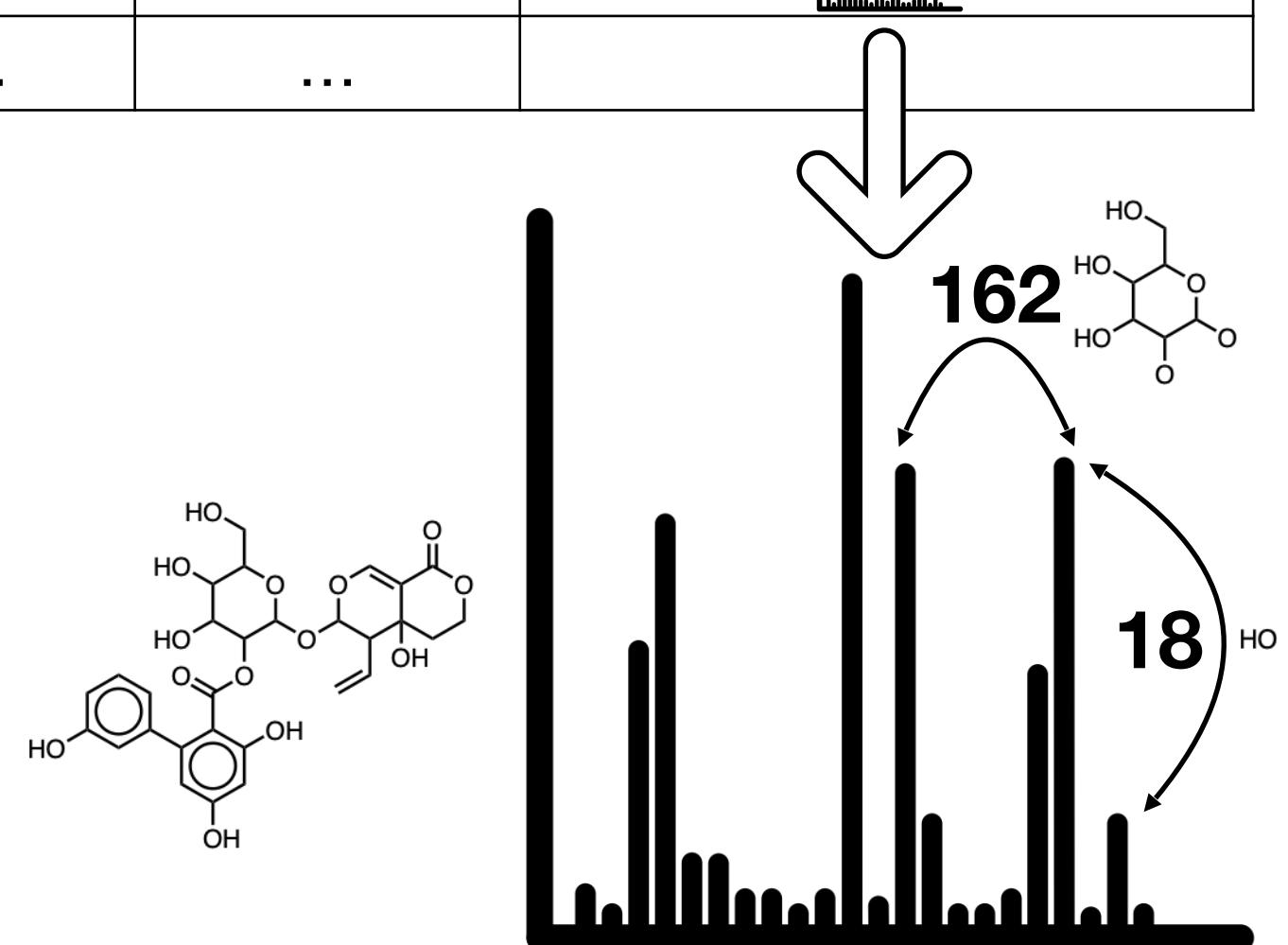
ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



# Metabolite annotation

Feature list (or table):

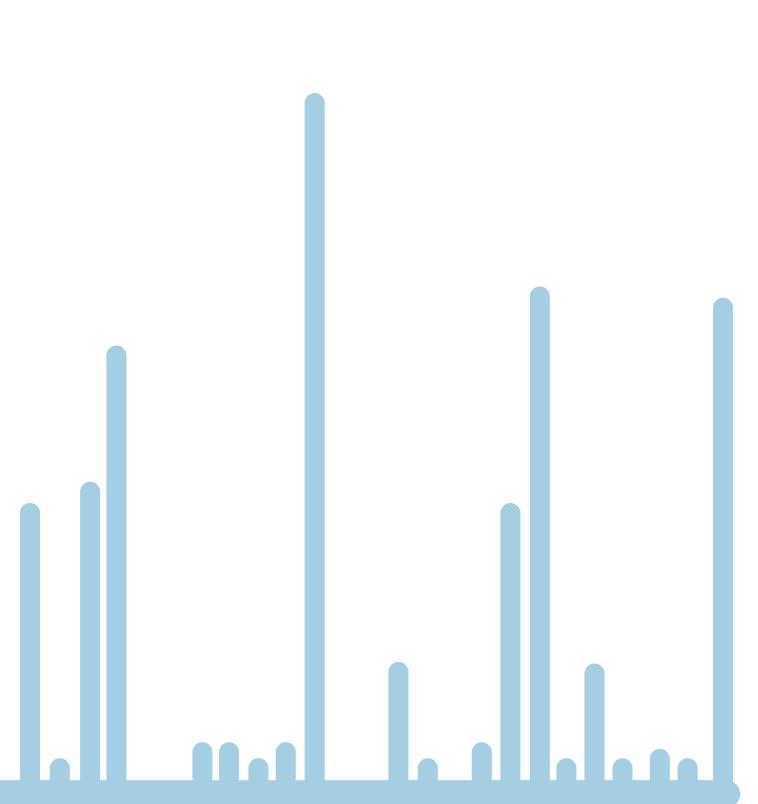
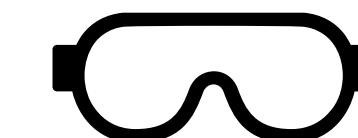
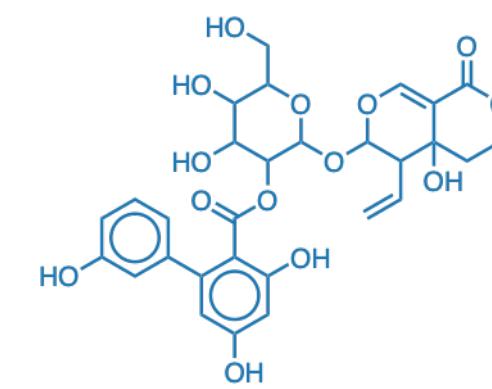
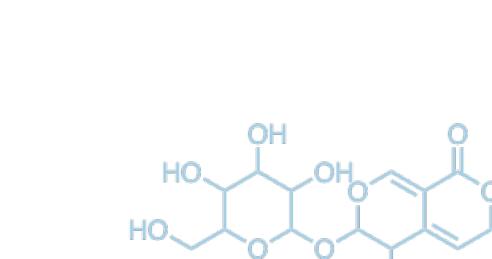
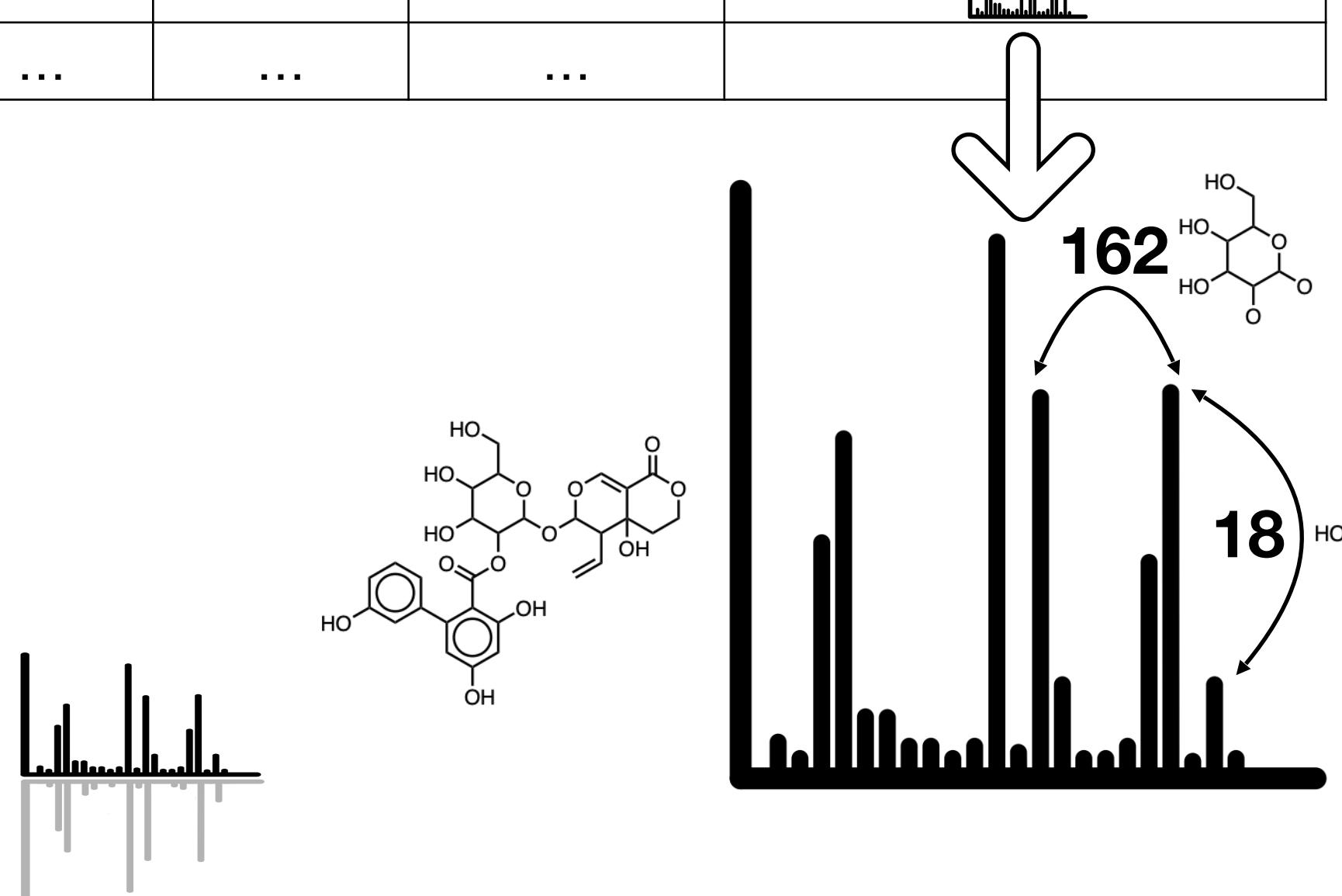
ID	RT	<i>m/z</i>	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



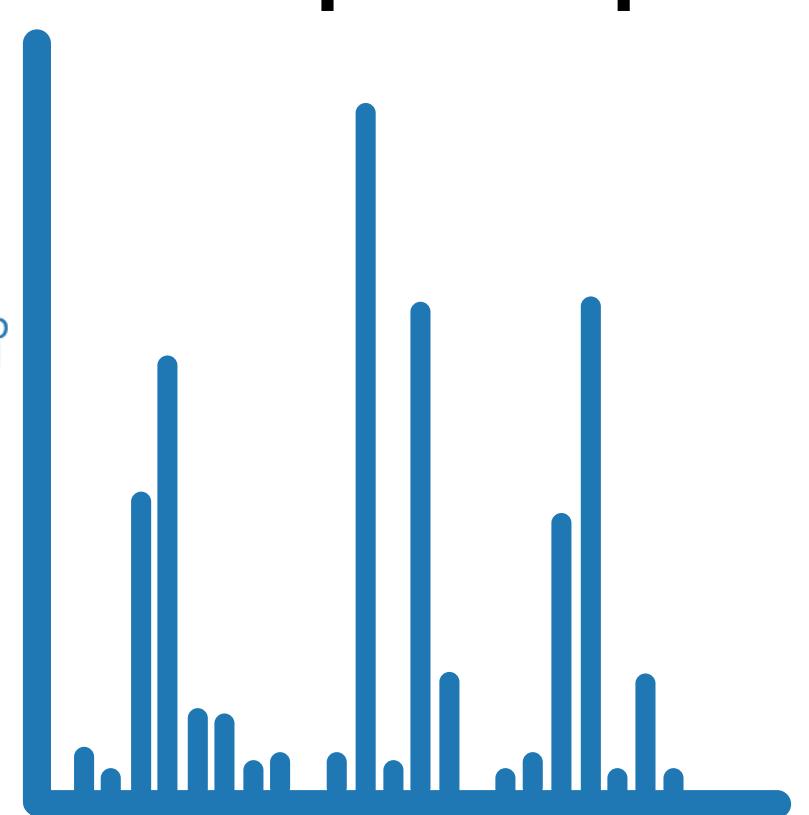
# Metabolite annotation

Feature list (or table):

ID	RT	$m/z$	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



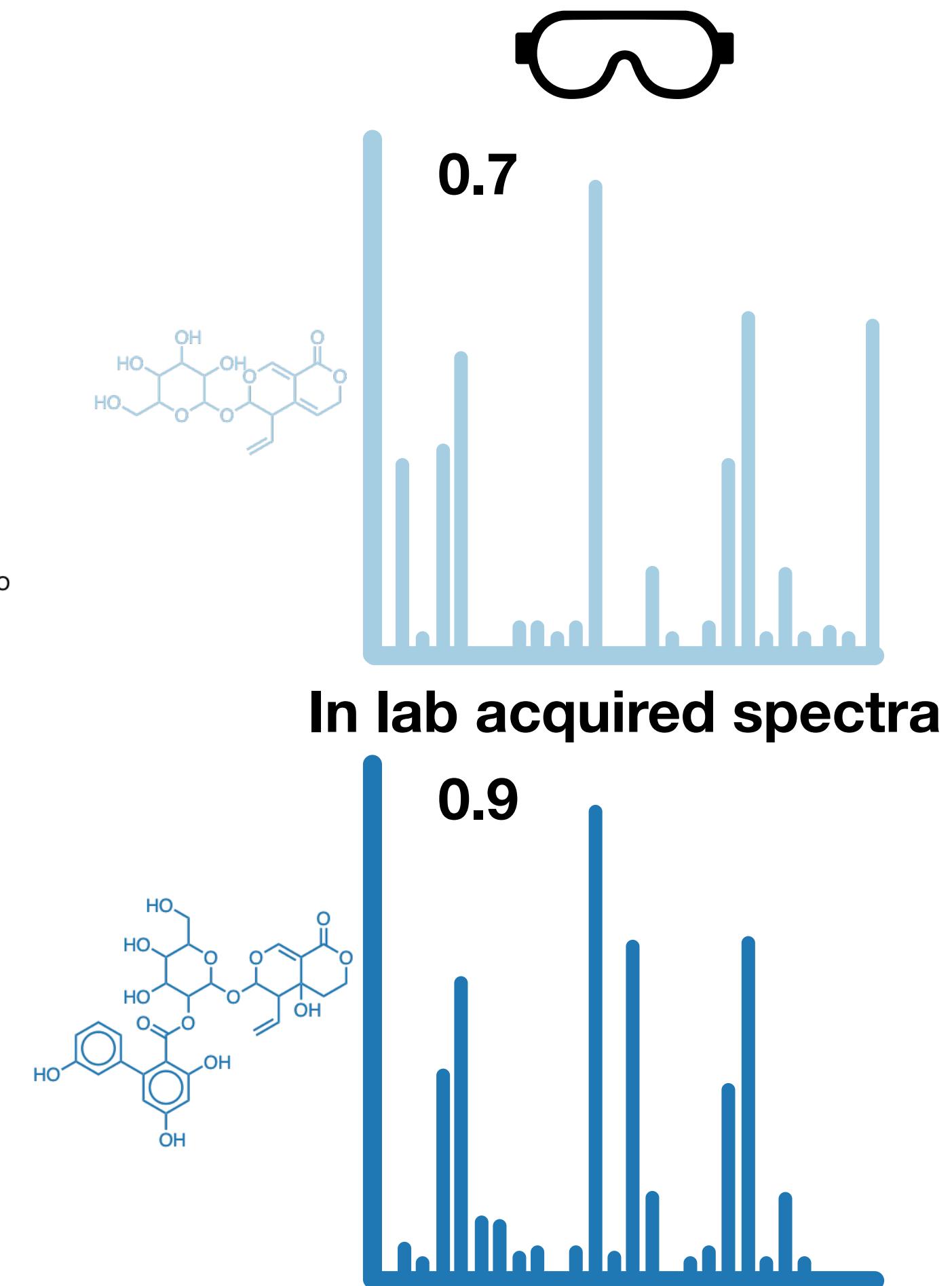
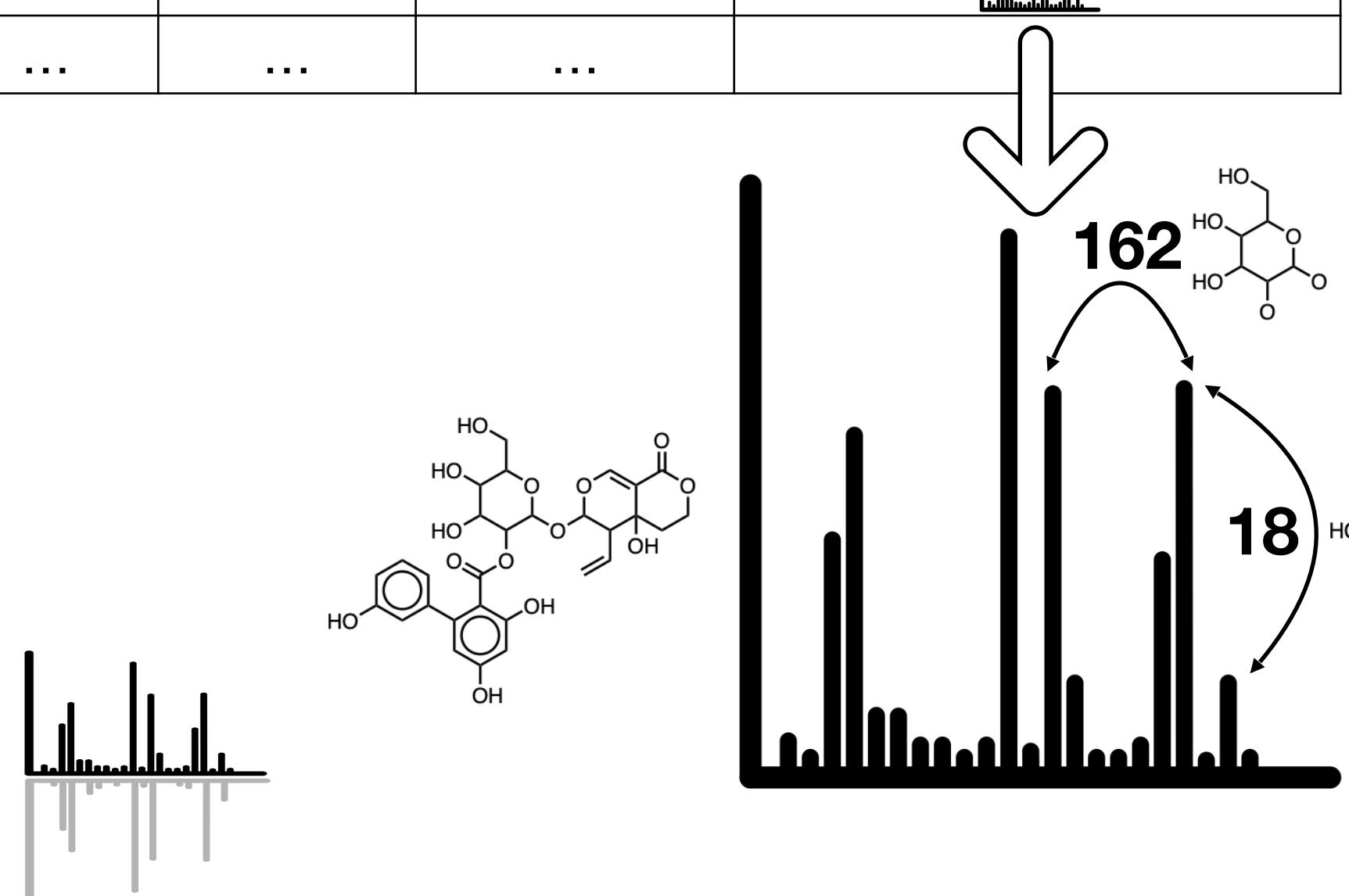
In lab acquired spectra



# Metabolite annotation

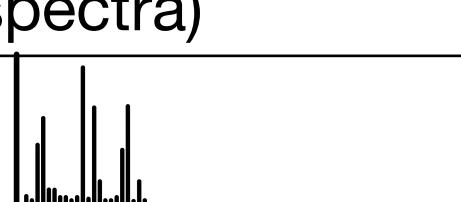
Feature list (or table):

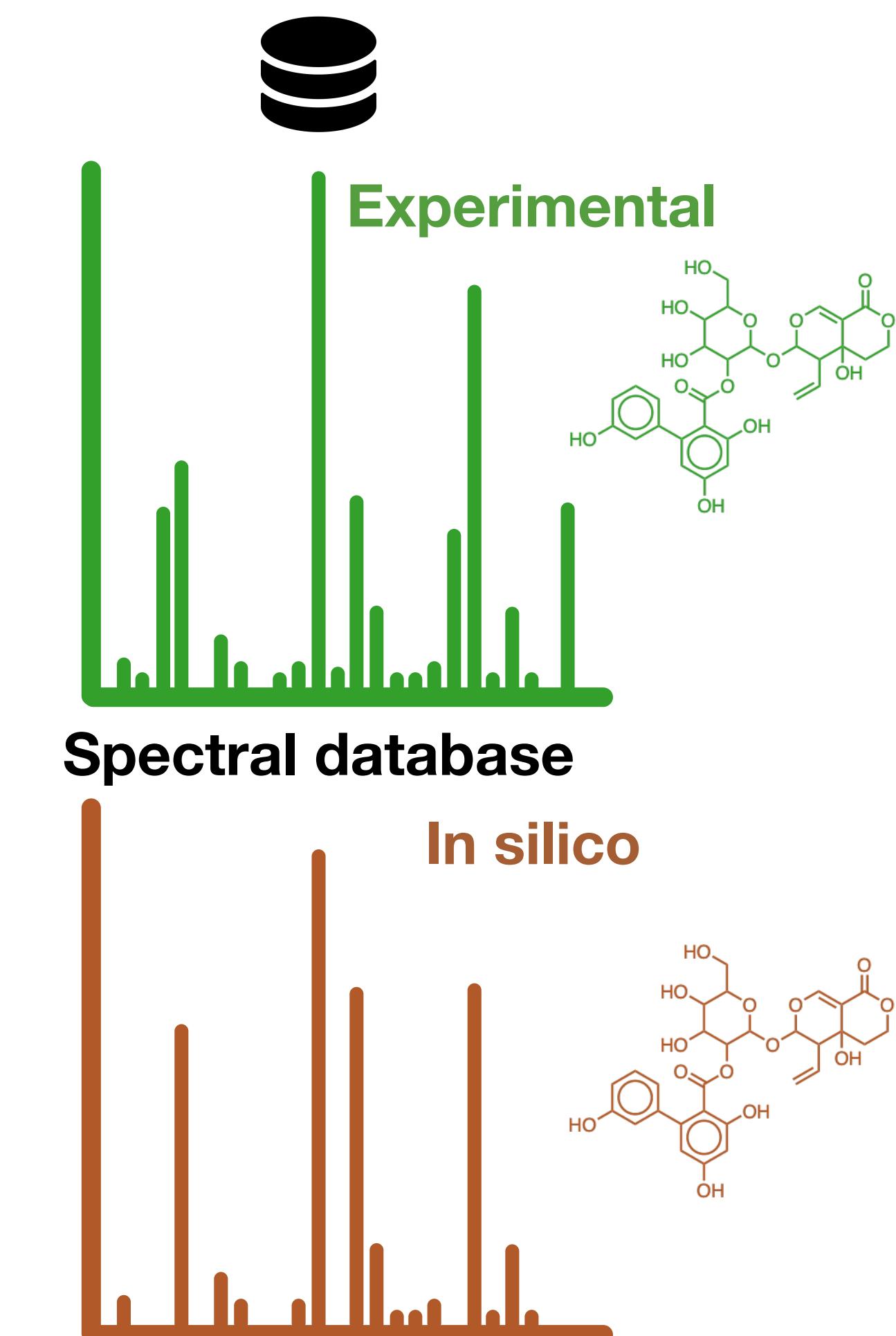
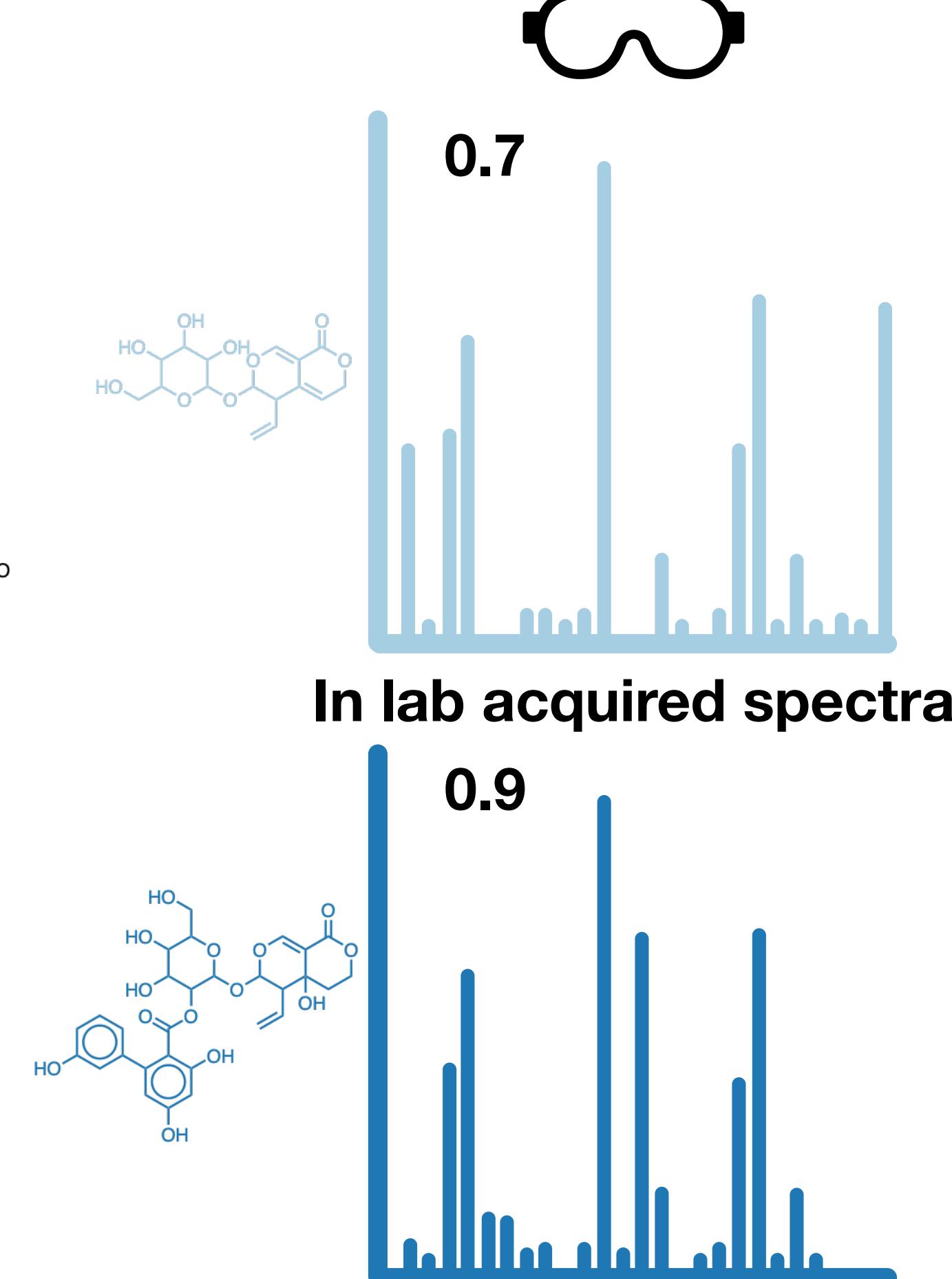
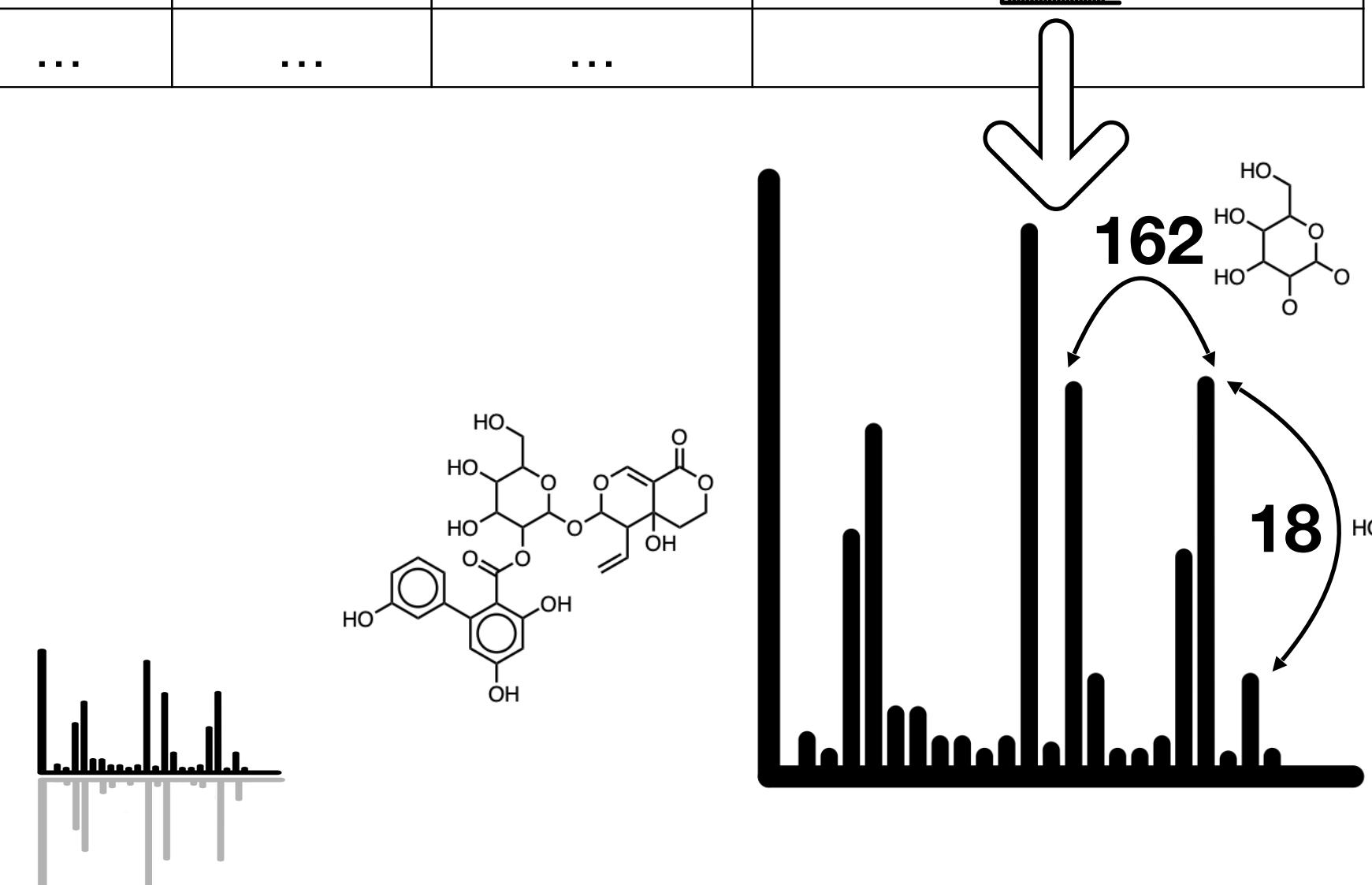
ID	RT	$m/z$	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



# Metabolite annotation

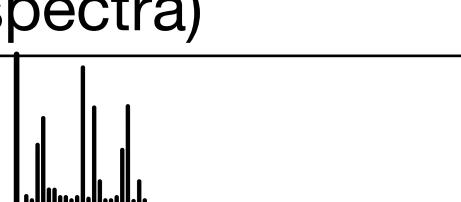
Feature list (or table):

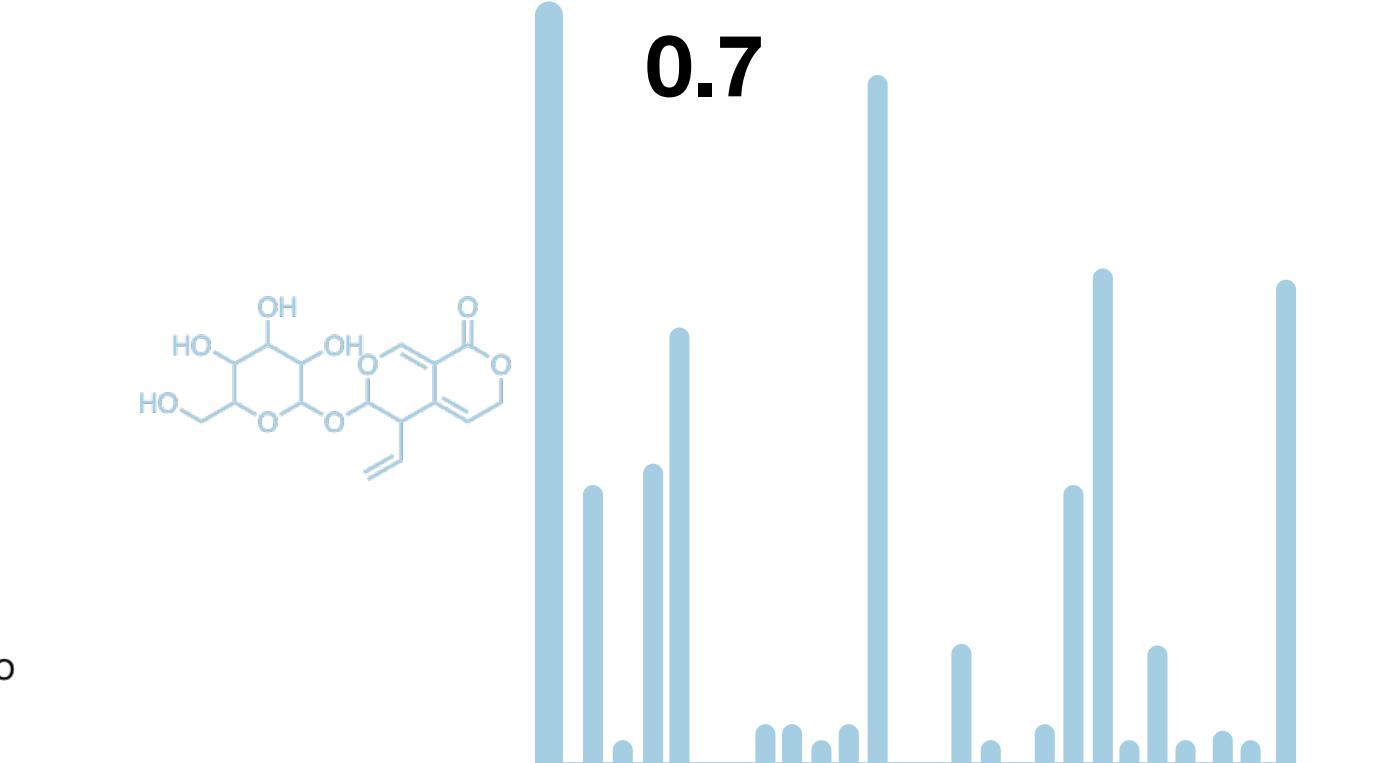
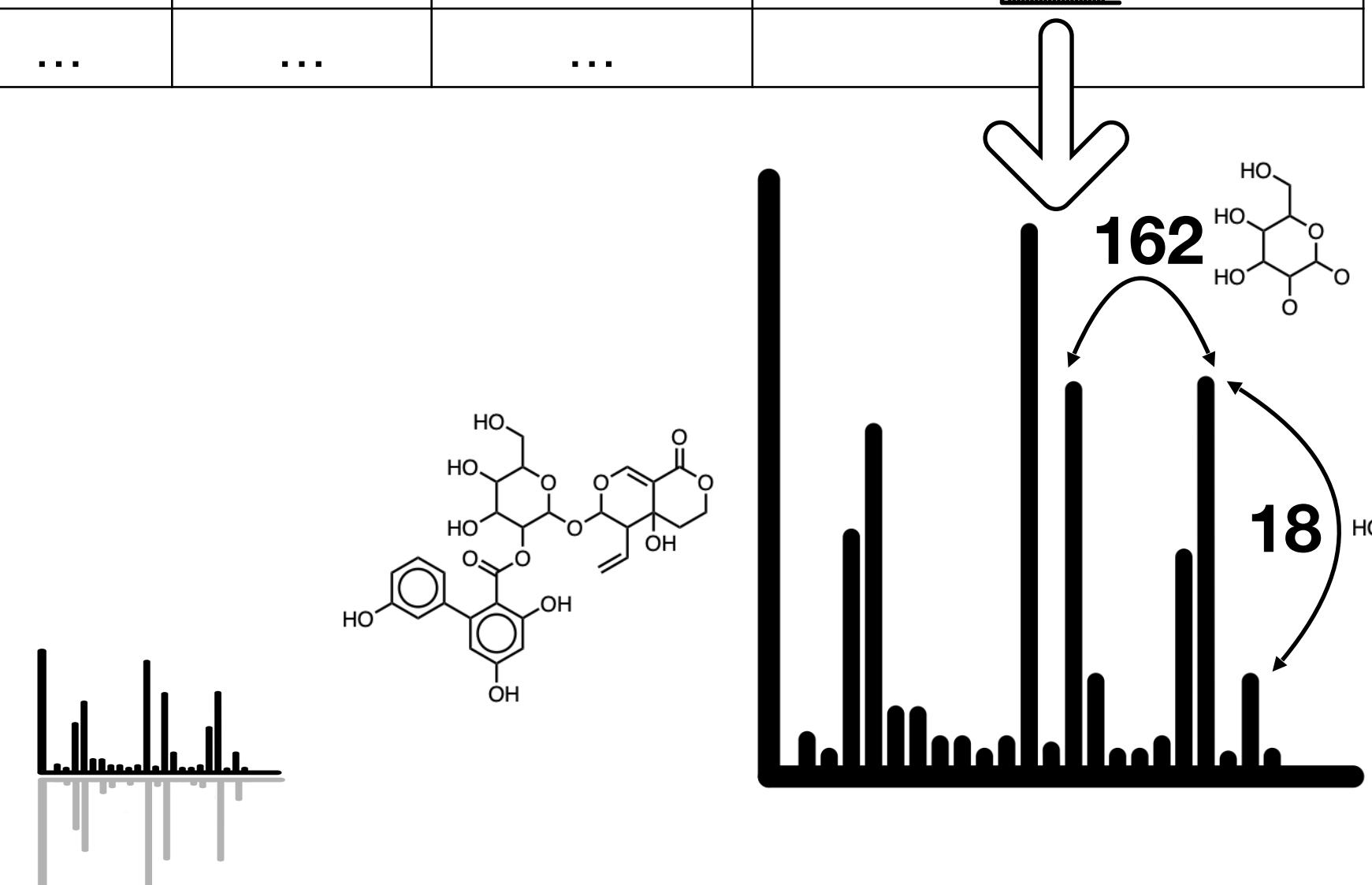
ID	RT	$m/z$	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



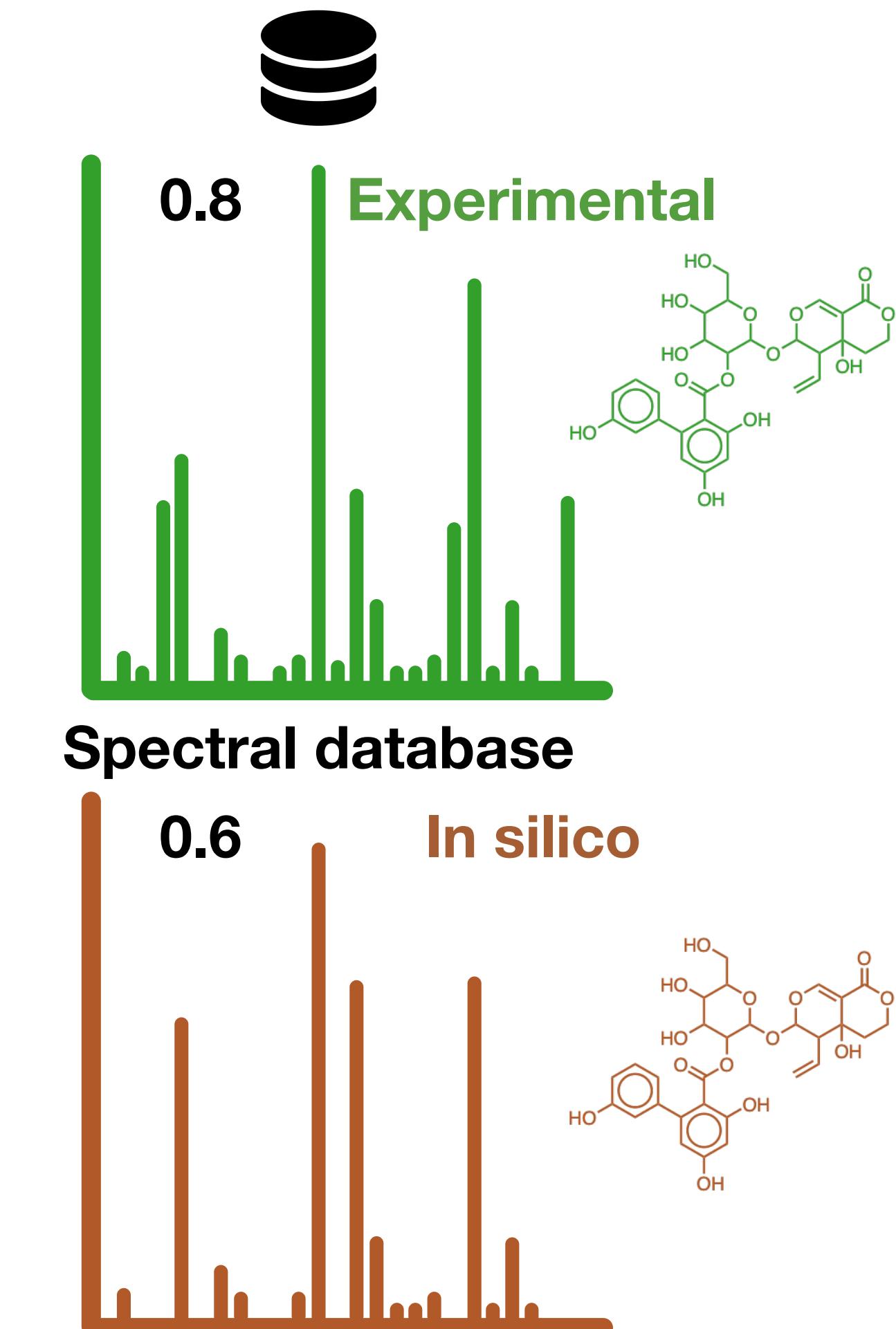
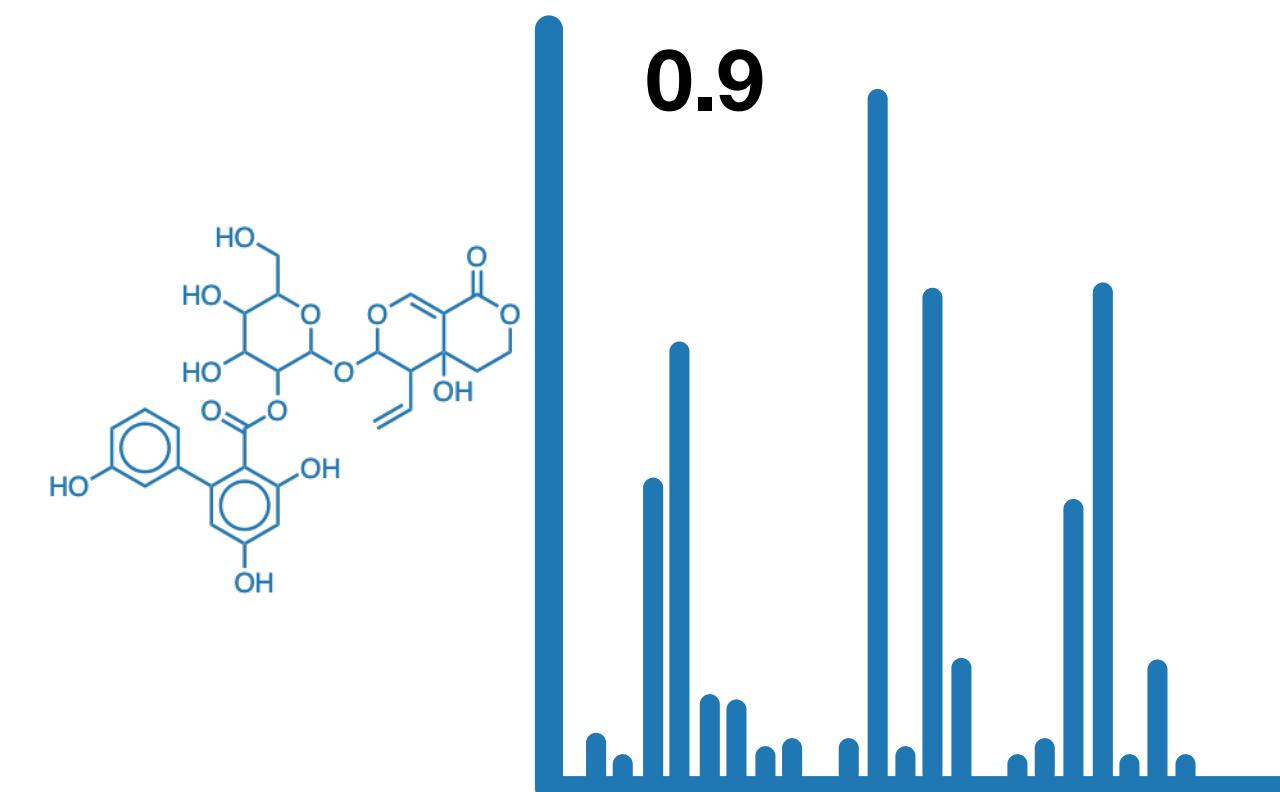
# Metabolite annotation

Feature list (or table):

ID	RT	$m/z$	Peak area	Associated spectrum (or spectra)
1	123.45	123.4567	9876543.21	
...	...	...	...	



In lab acquired spectra

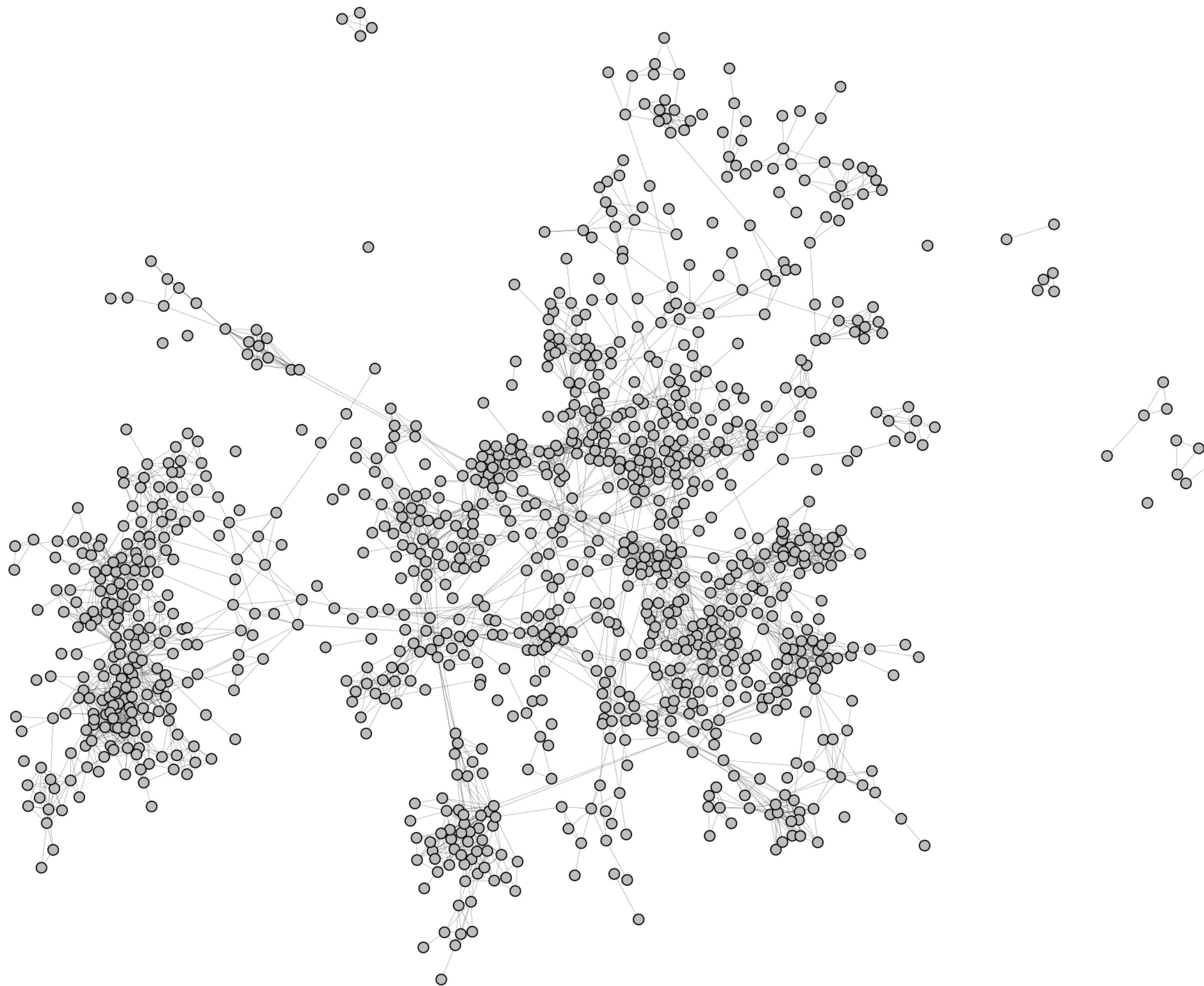
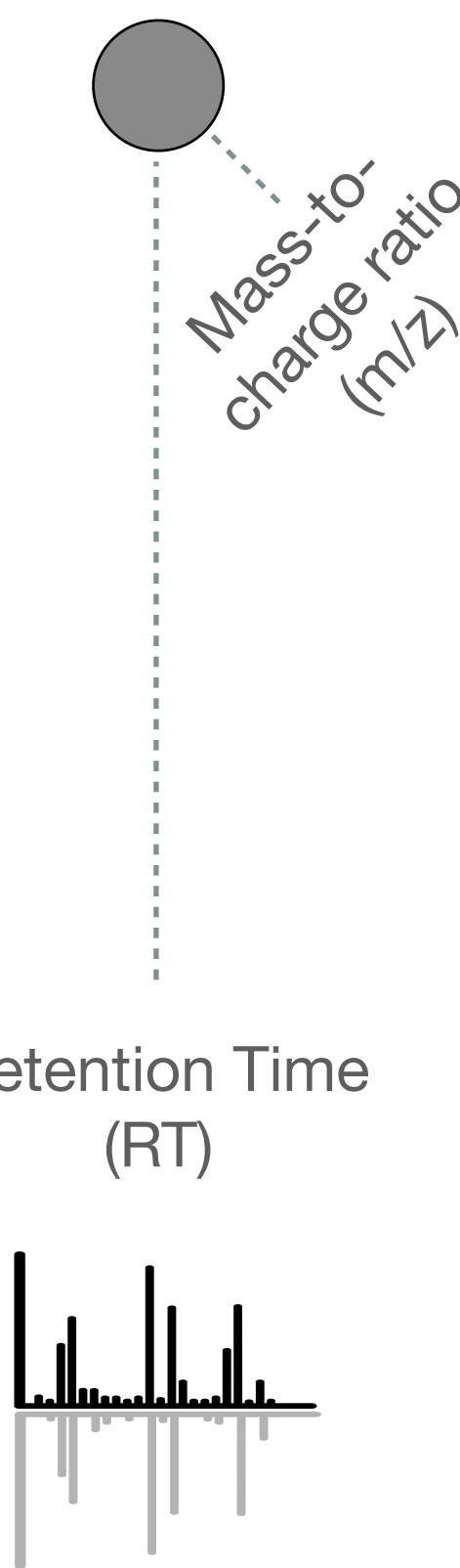


# Metabolite annotation

Feature ID	Spectrum	Candidate structure	Score $S_1$	Initial rank
1			0.9	1
			0.8	2
			0.7	3
			0.6	4
...	...	...	...	...

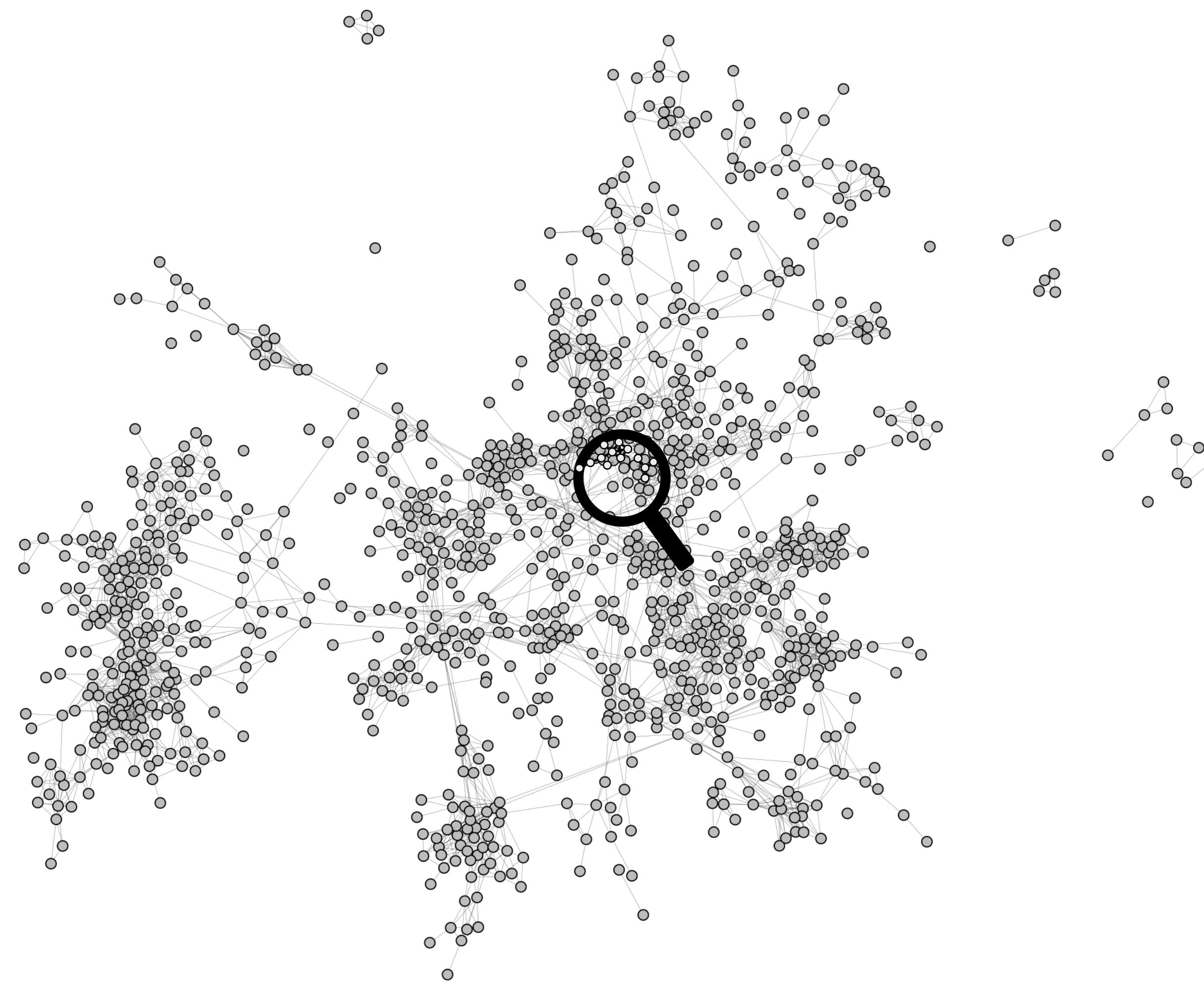
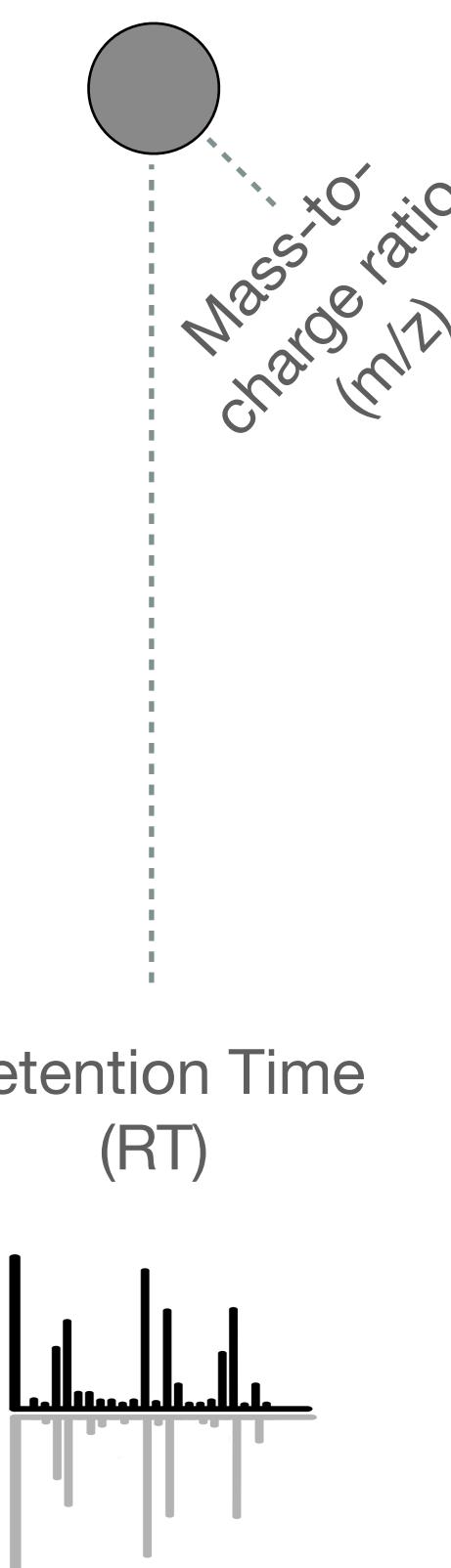
# Metabolite annotation

Feature:  
RT @  $m/z$  pointer

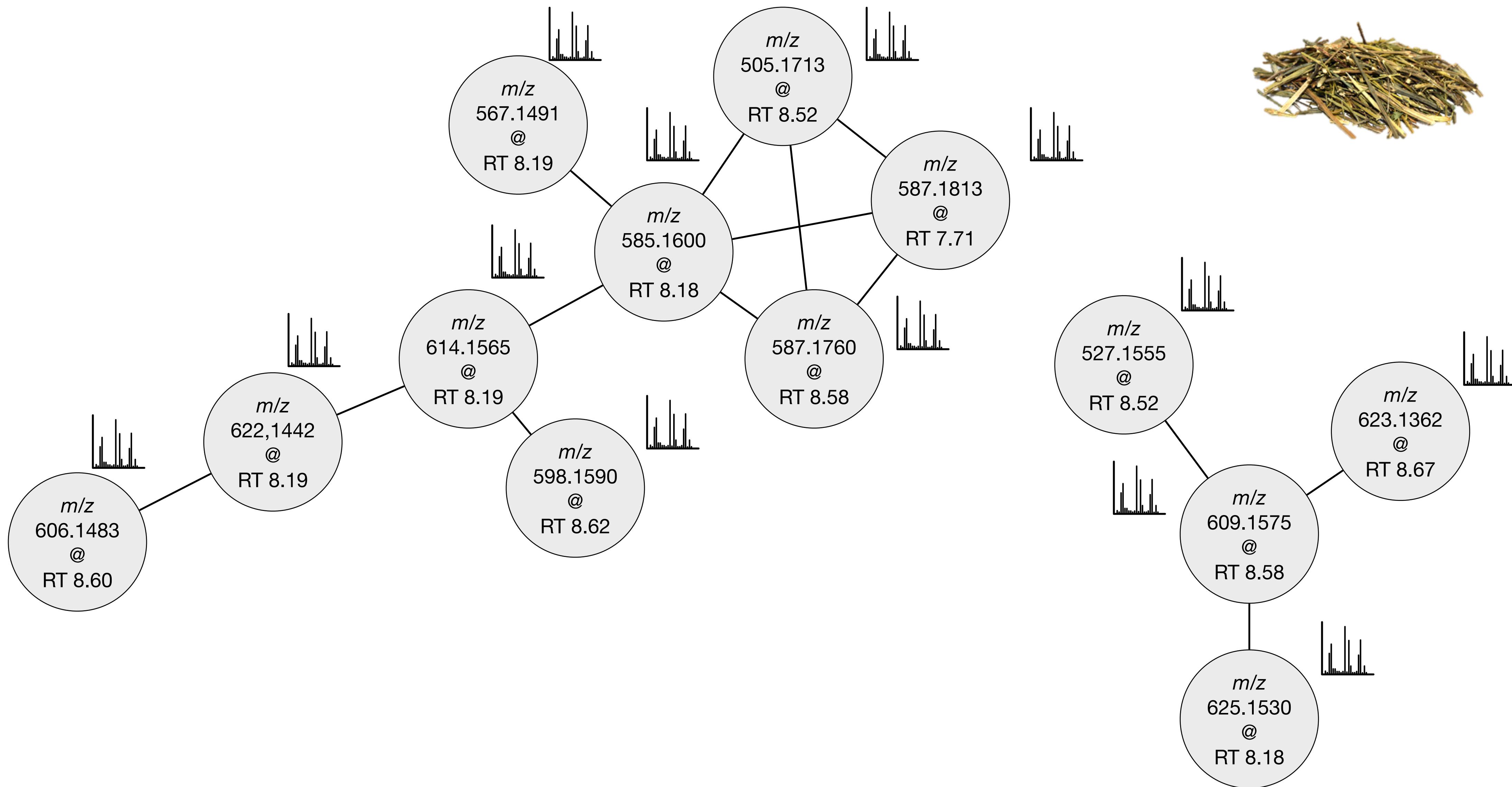


# Metabolite annotation

Feature:  
RT @  $m/z$  pointer

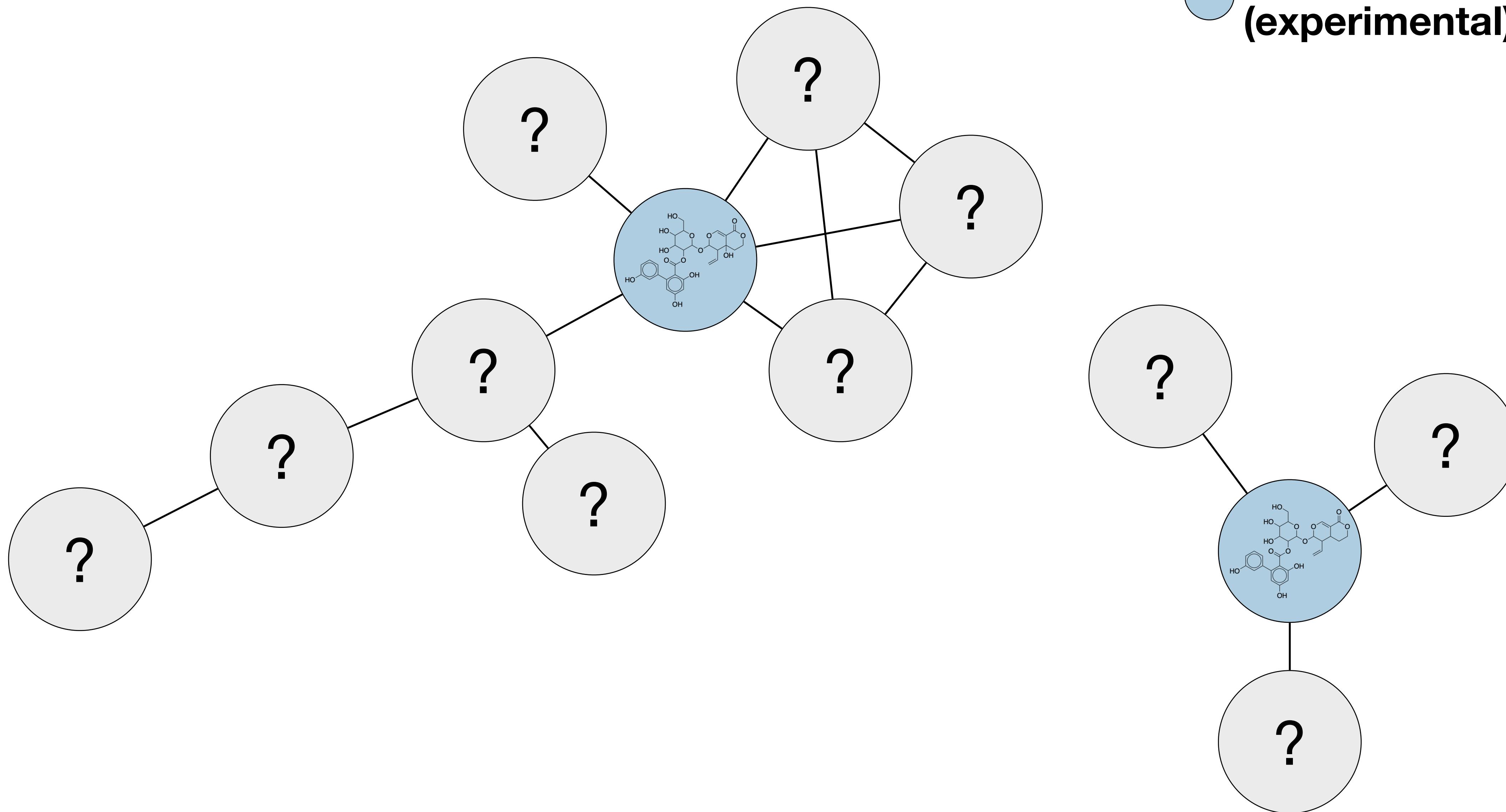


# Metabolite annotation

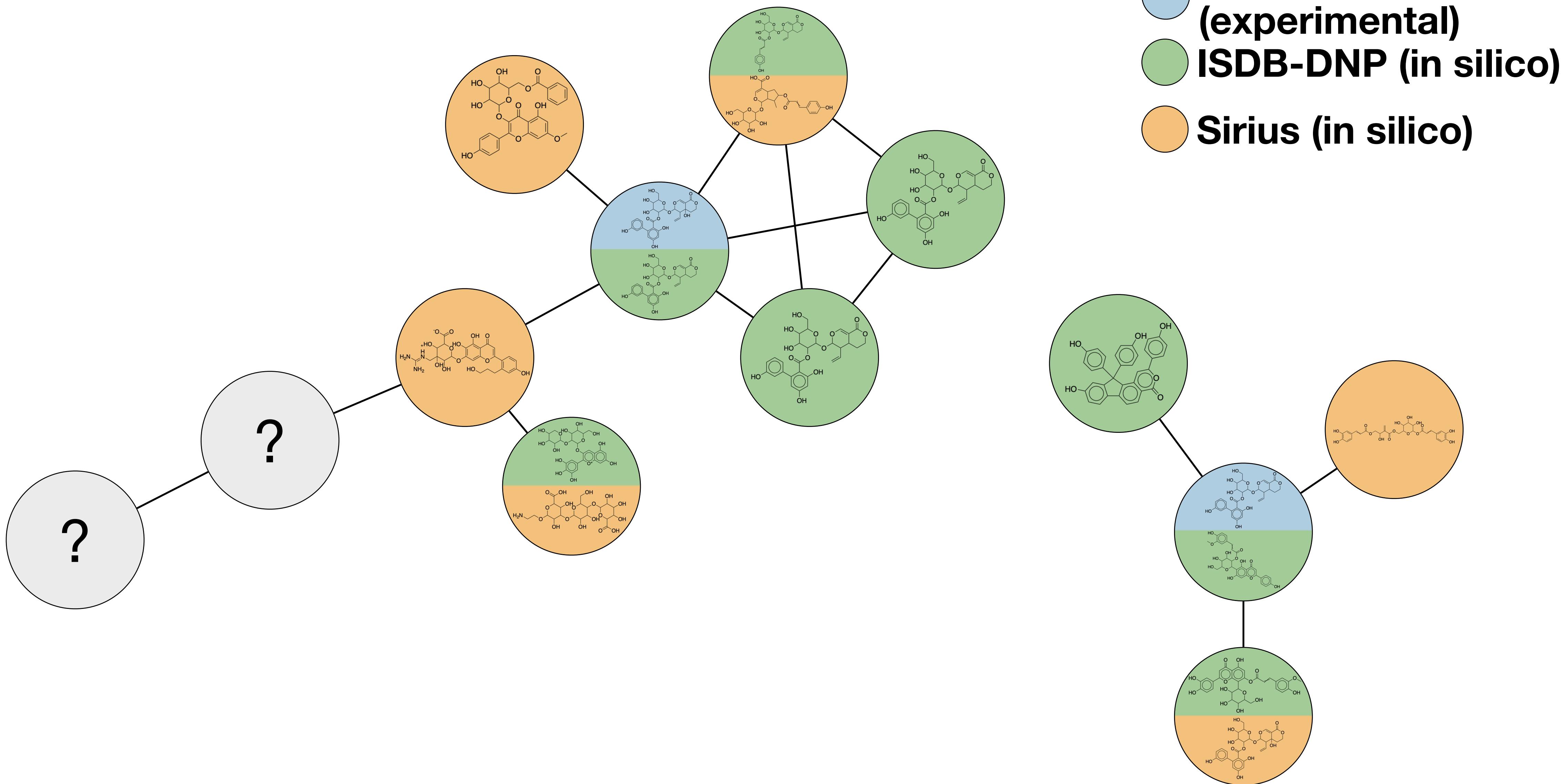


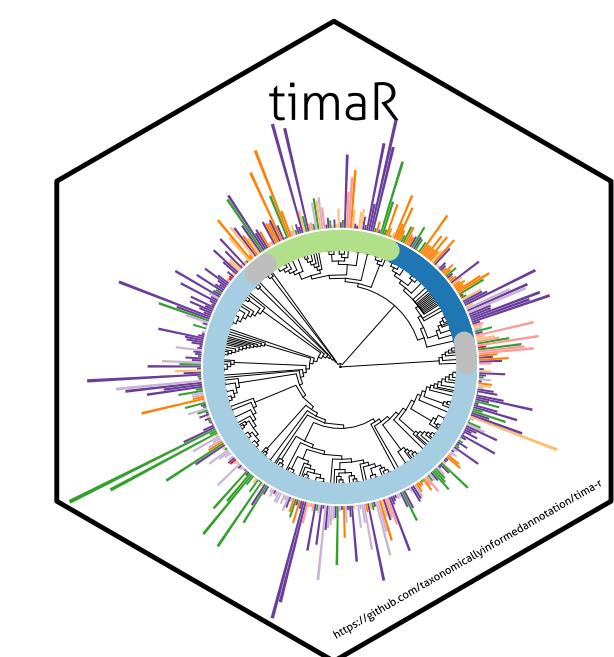
# Metabolite annotation

GNPS  
(experimental)



# Metabolite annotation

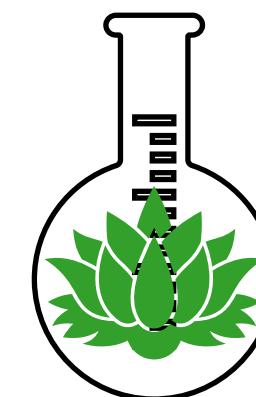




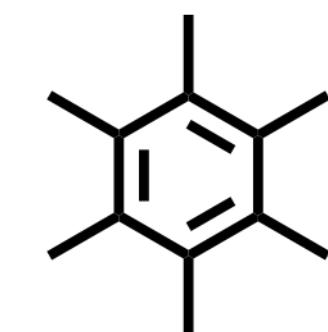
# Taxonomy as prior

- Strongly informative
- Orthogonal
- Highly interoperable
- Long history, well defined, used to handle both identifiers and ambiguity

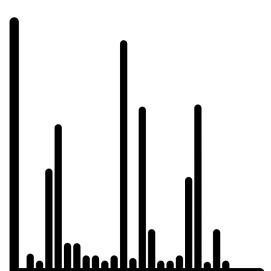
## Metabolomics



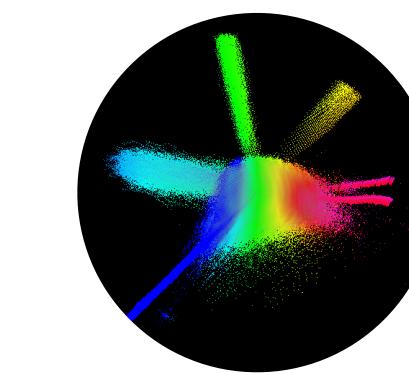
Natural Extract



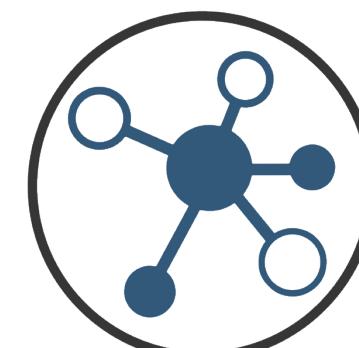
Structure



MS<sup>2</sup>  
spectrum



Chemical  
Space



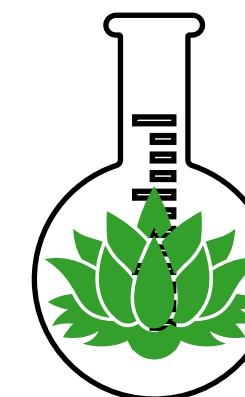
Molecular  
Network

# Taxonomically Informed Metabolite Annotation

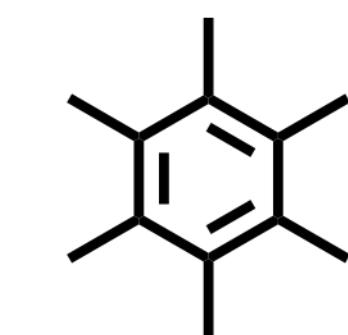


Biological organism

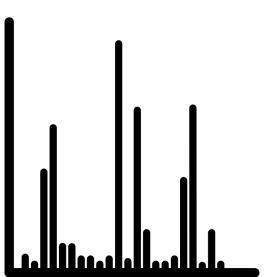
Metabolomics



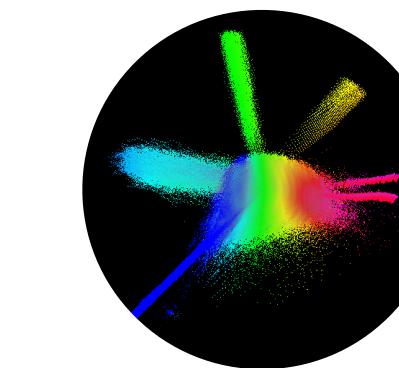
Natural Extract



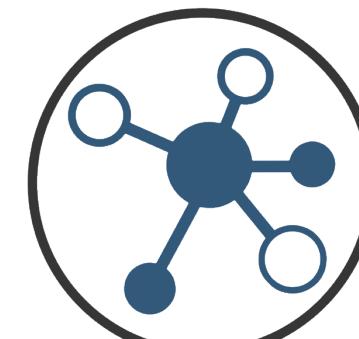
Structure



$MS^2$   
spectrum

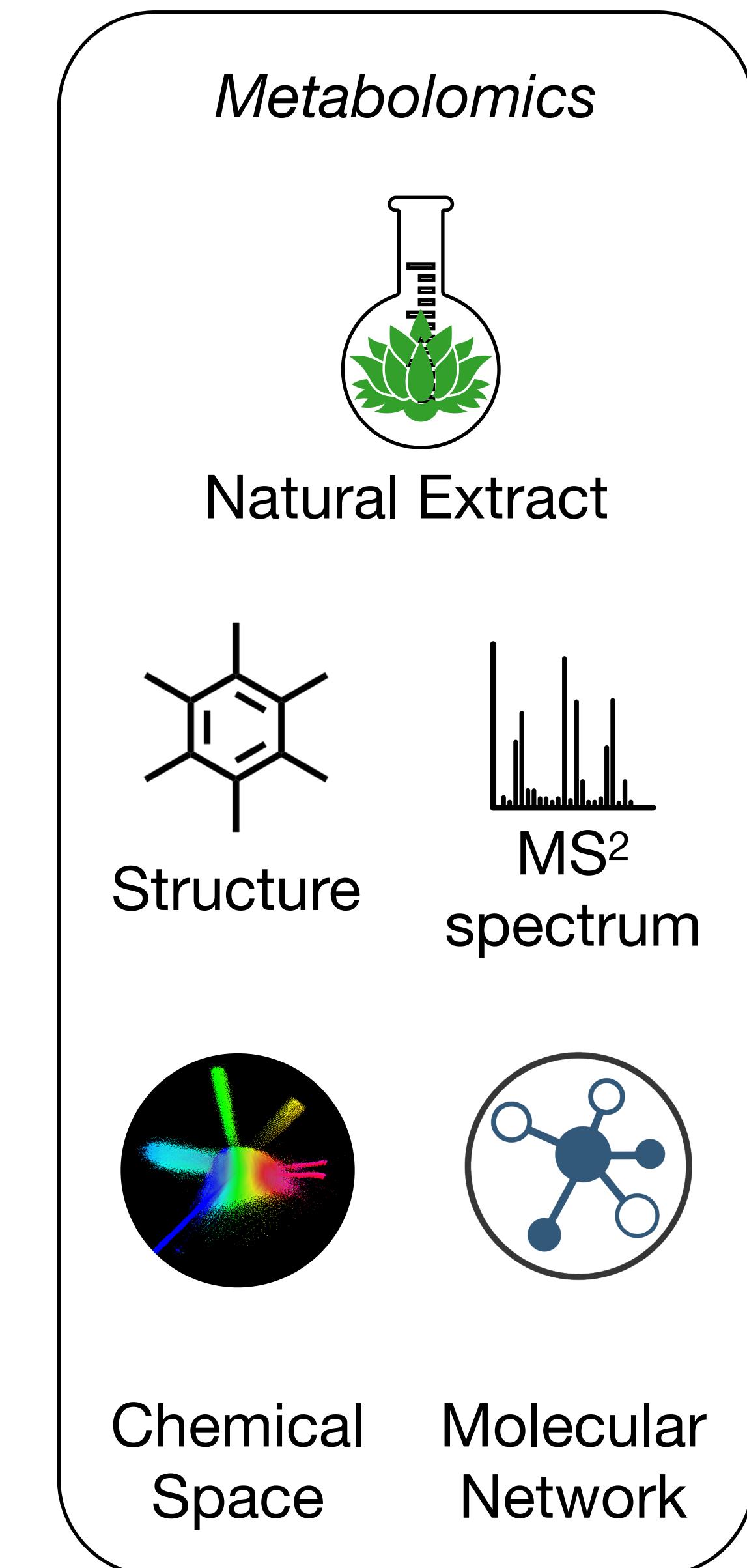
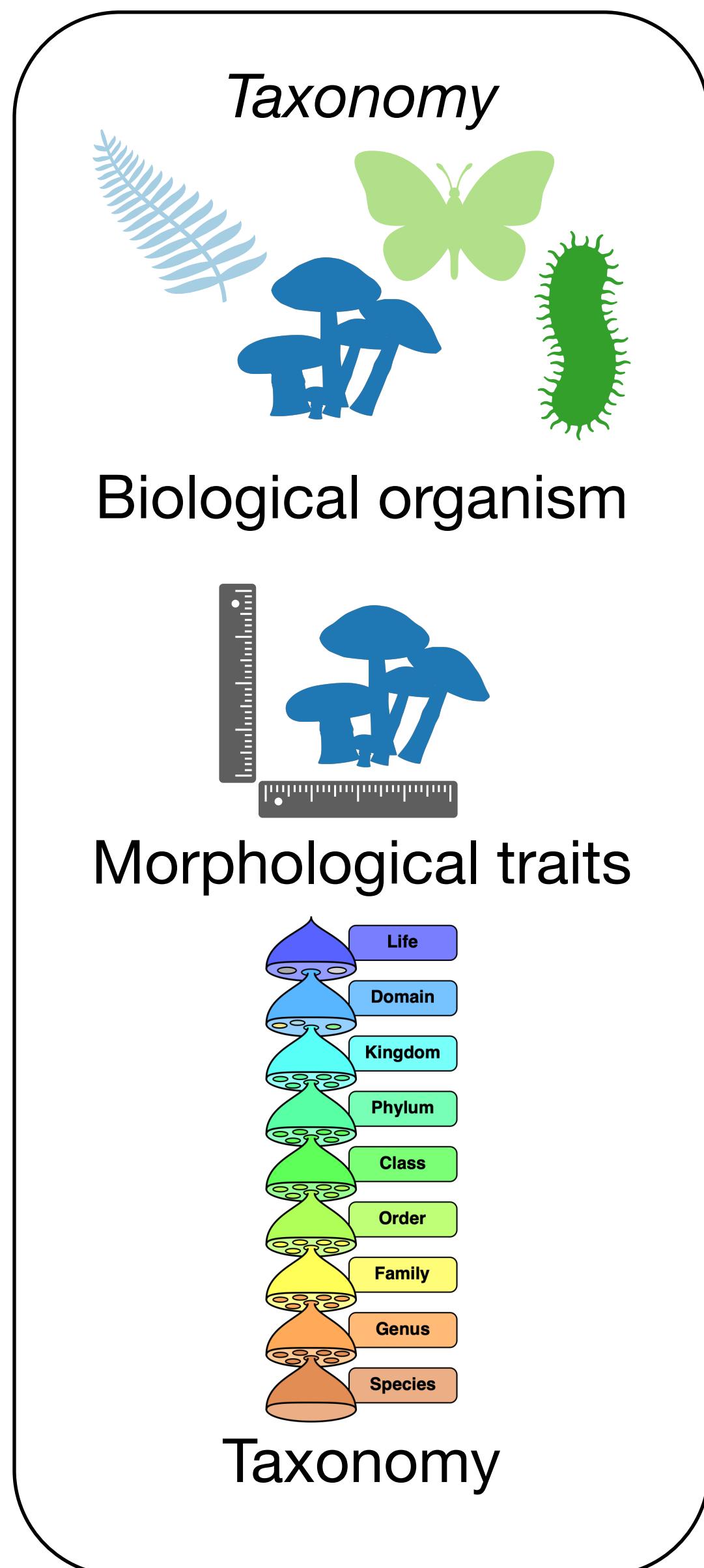


Chemical  
Space

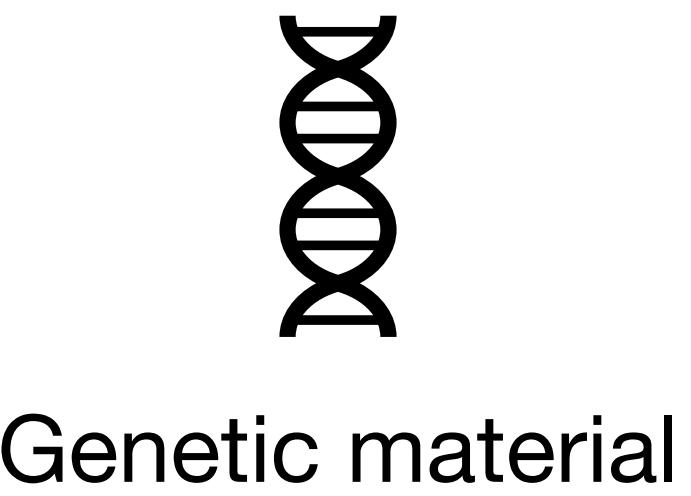
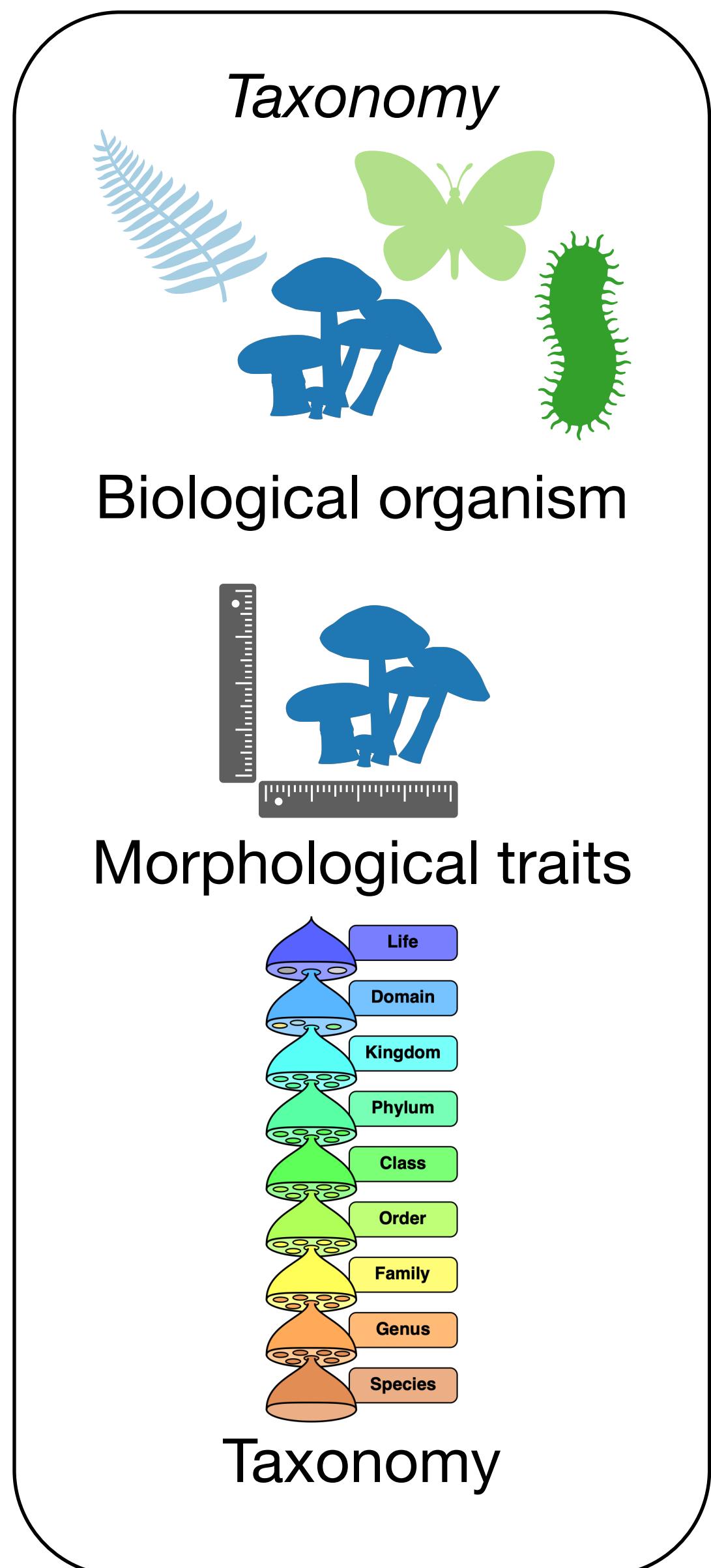


Molecular  
Network

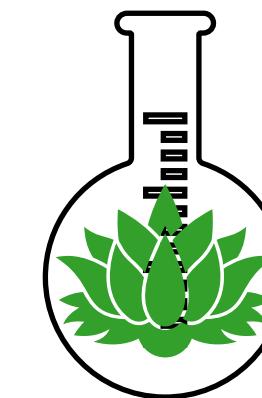
# Taxonomically Informed Metabolite Annotation



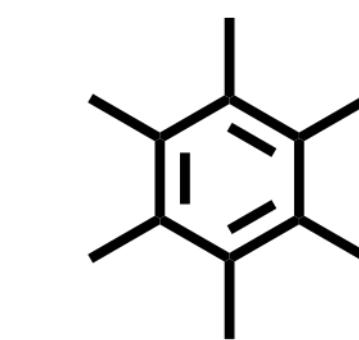
# Taxonomically Informed Metabolite Annotation



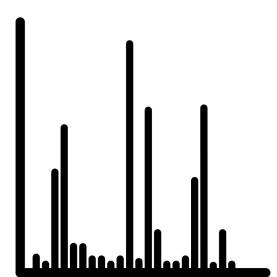
## Metabolomics



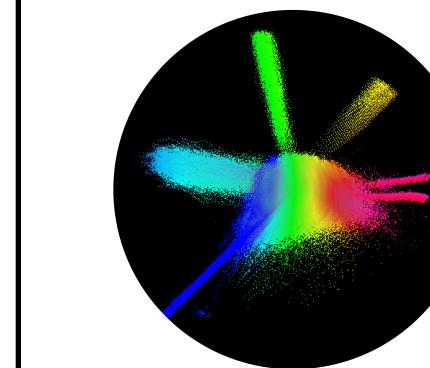
Natural Extract



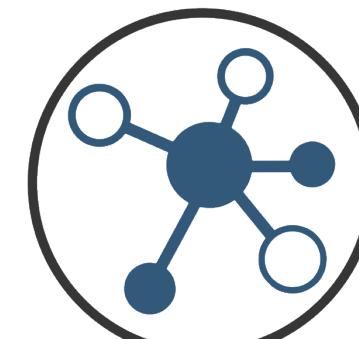
Structure



MS<sup>2</sup>  
spectrum

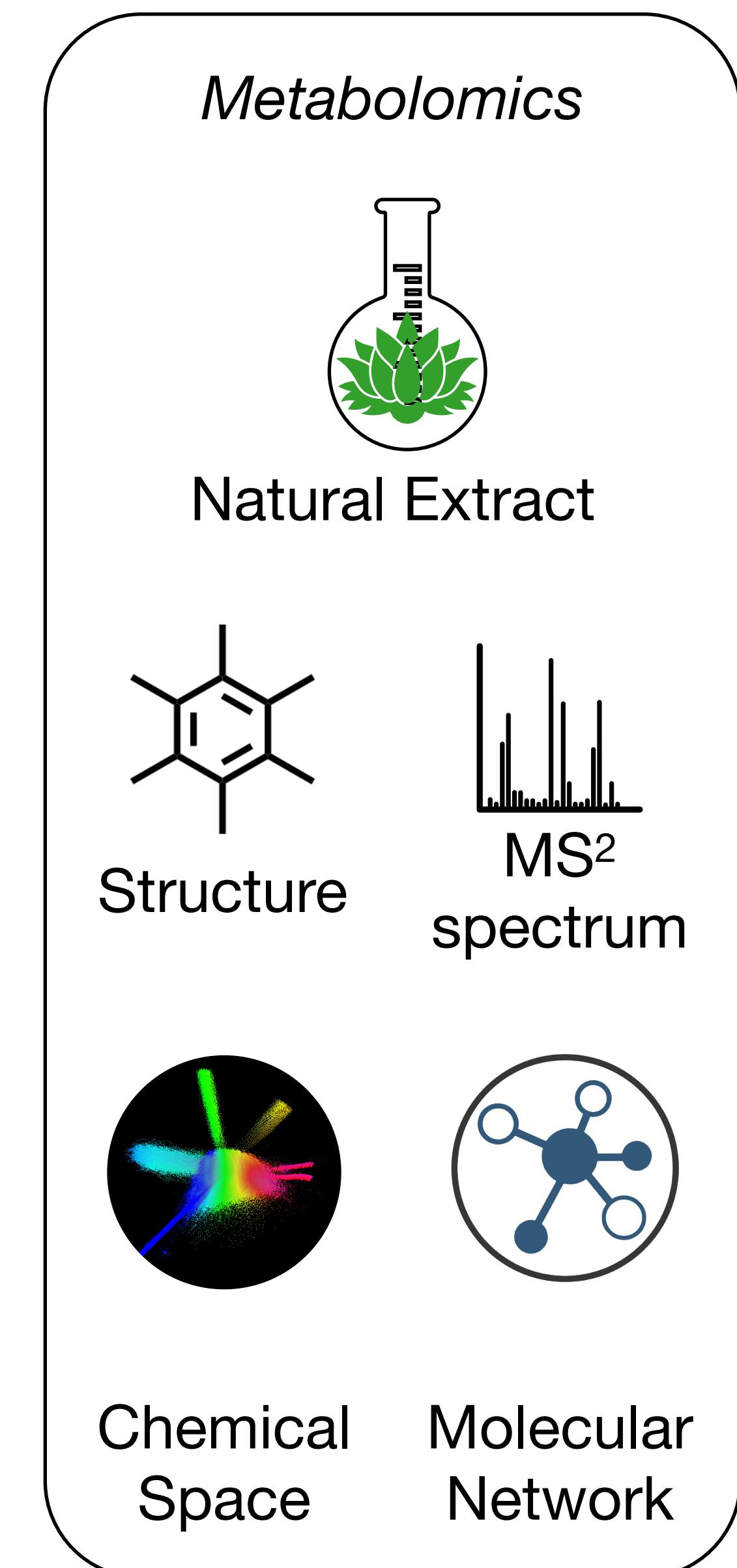
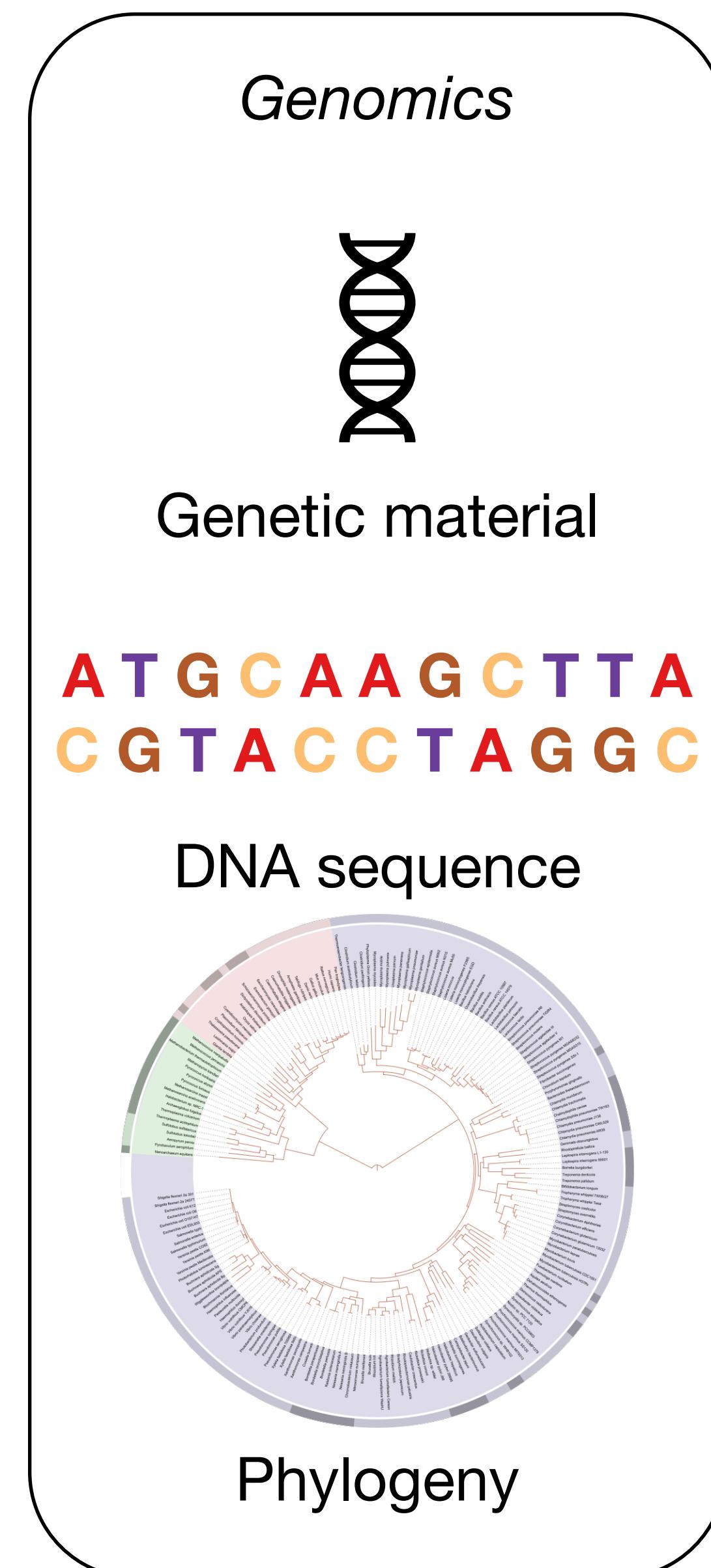
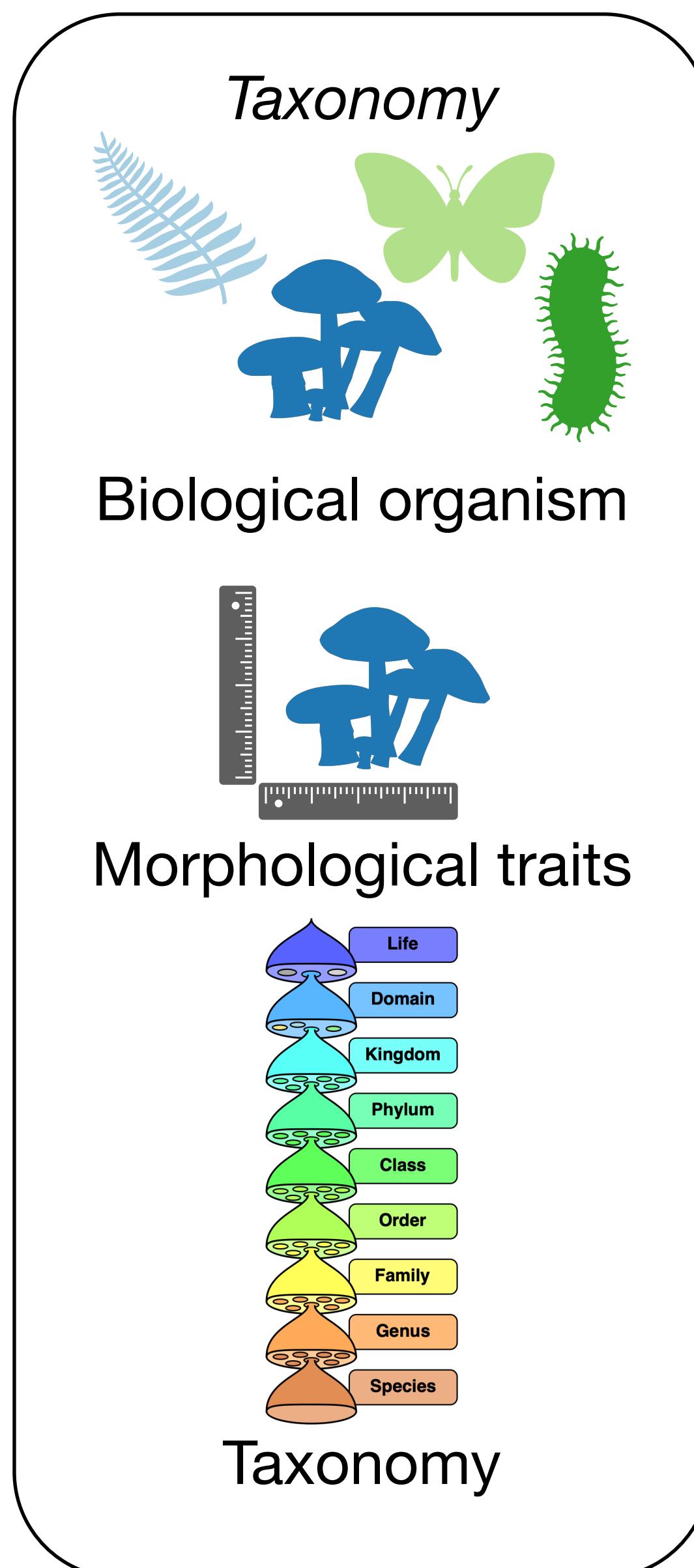


Chemical  
Space

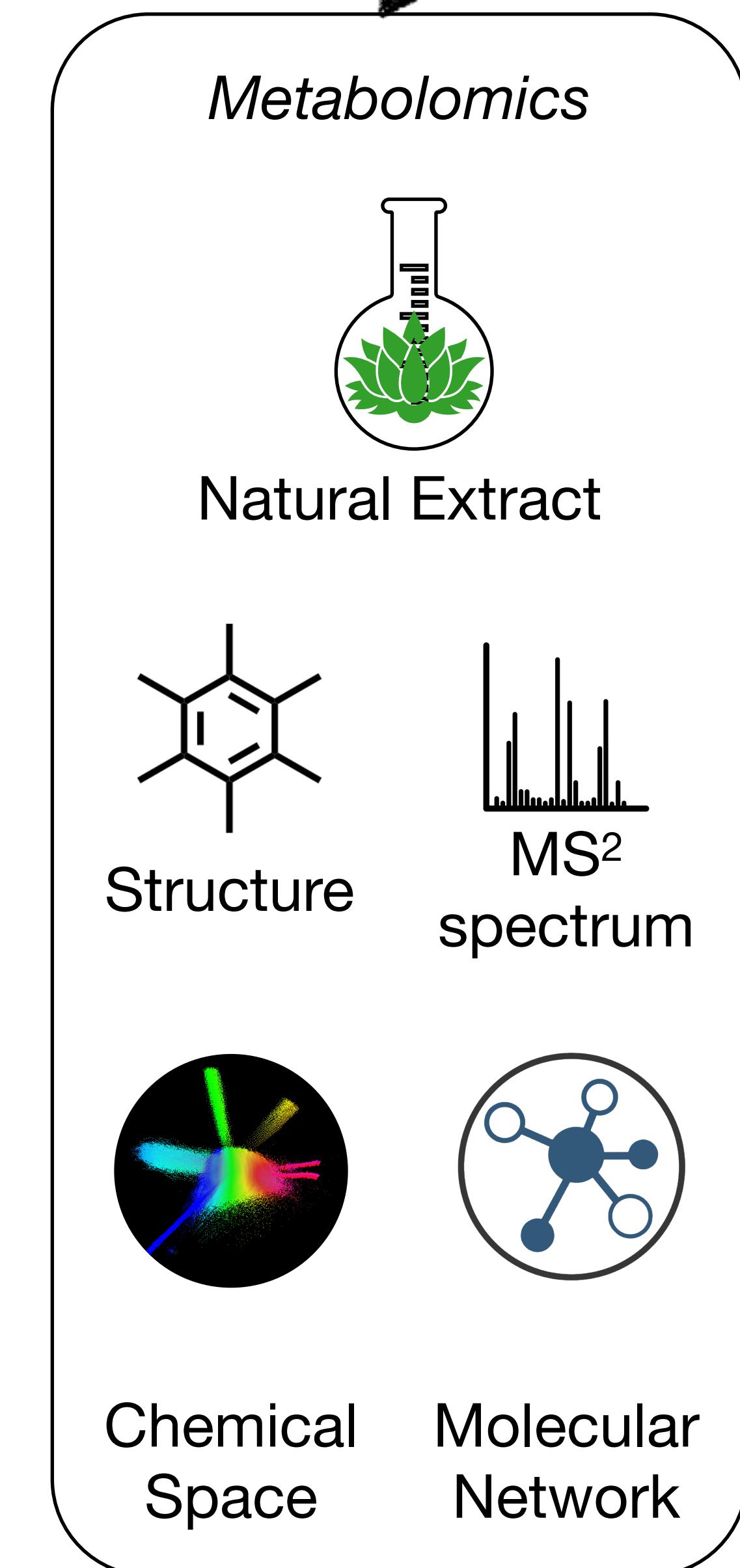
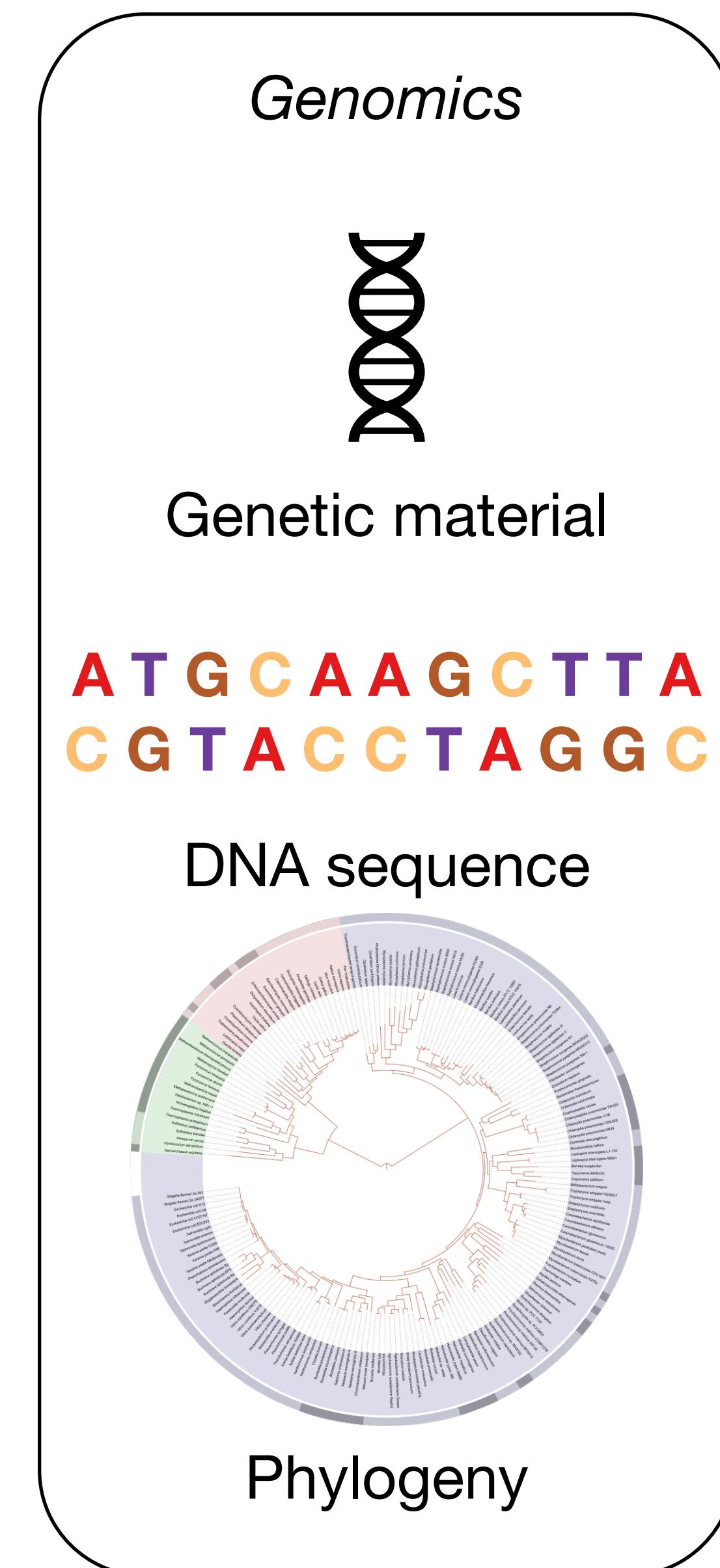
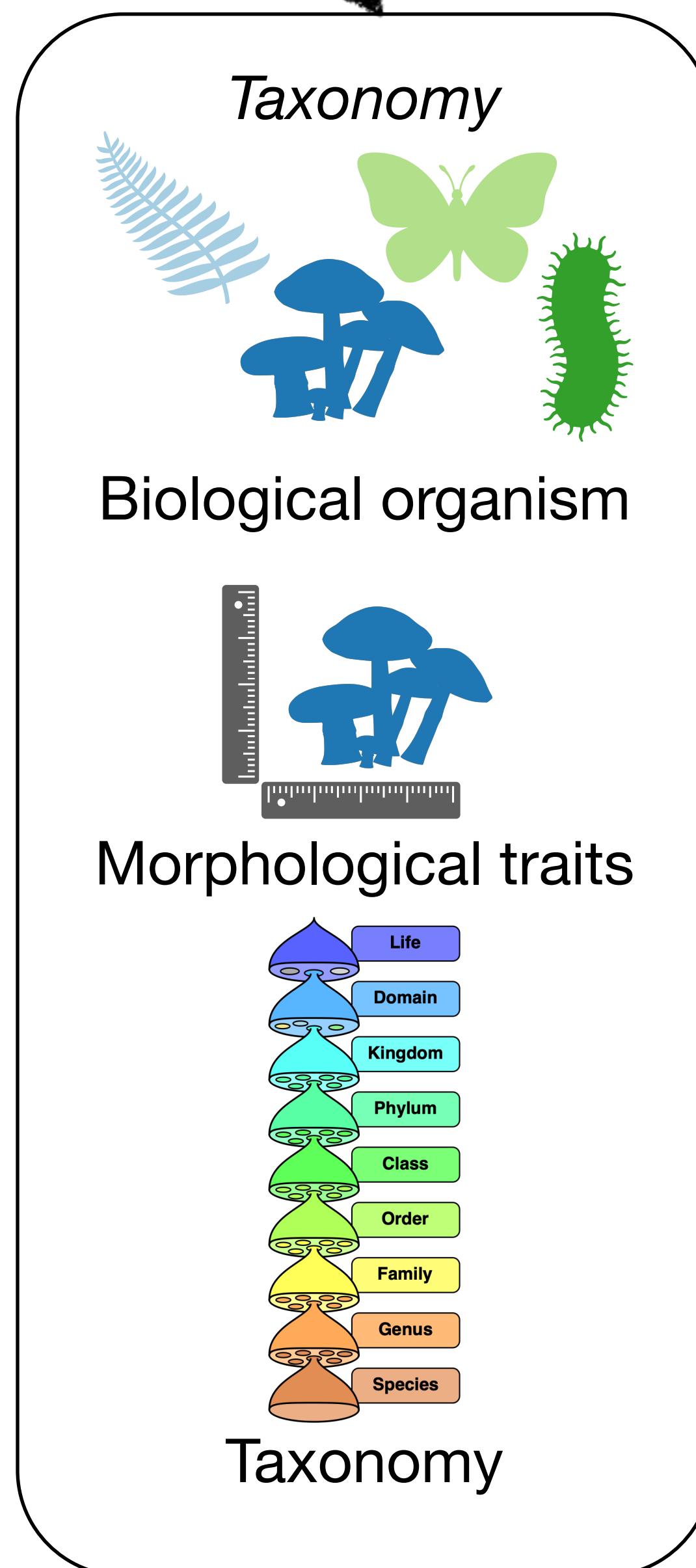


Molecular  
Network

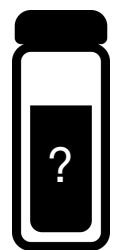
# Taxonomically Informed Metabolite Annotation



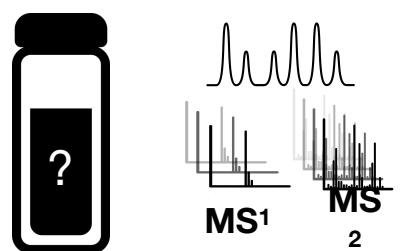
# Taxonomically Informed Metabolite Annotation



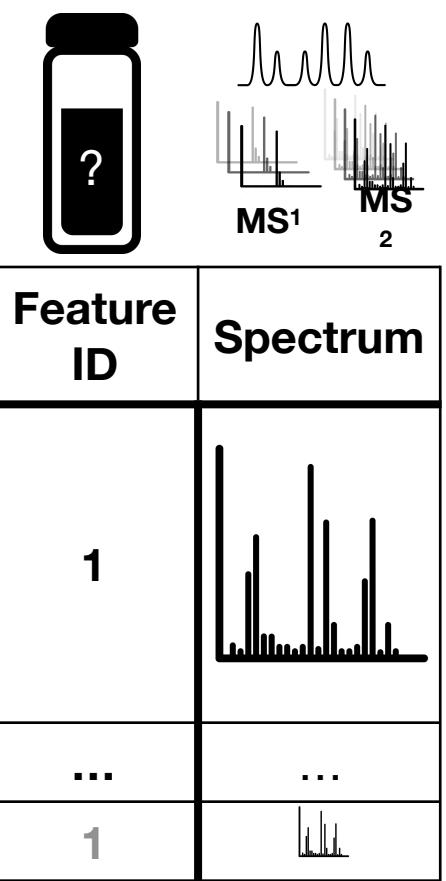
# Candidates re-ranking



# Candidates re-ranking



# Candidates re-ranking



Candidate structure	Score $S_1$	Initial rank
structure-1	0.55	1
structure-2	0.53	2
structure-3	0.50	3
structure-4	0.45	4
...	...	...
structure-9	0.00	9

# Candidates re-ranking

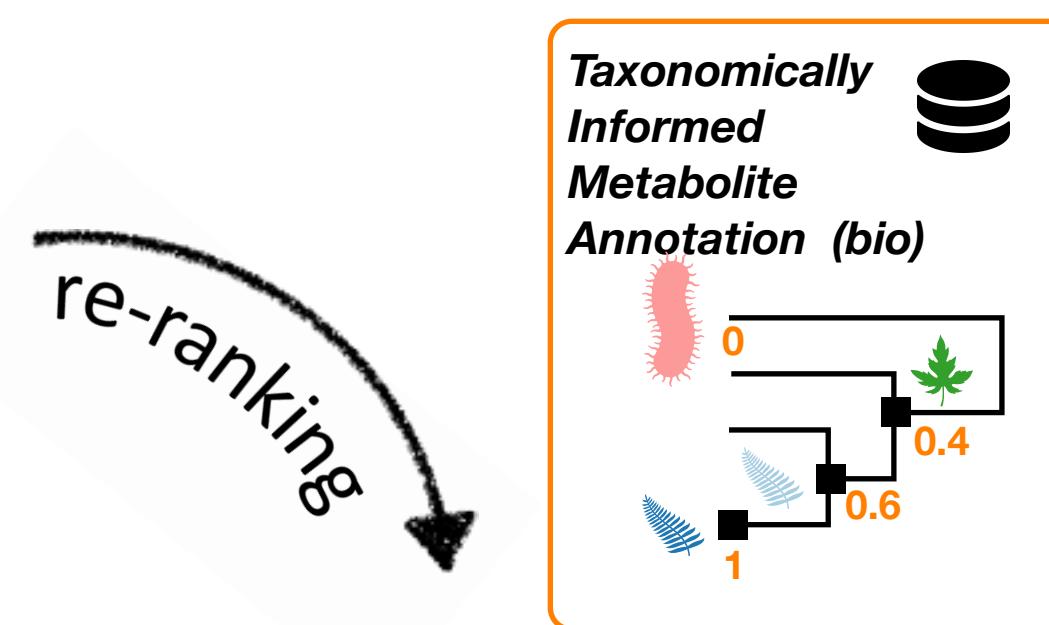


Feature ID	Spectrum	Biological source	Candidate structure	Score S <sub>1</sub>	Initial rank
1			structure-1	0.55	1
			structure-2	0.53	2
			structure-3	0.50	3
			structure-4	0.45	4
...	...	...	...	...	...
1			structure-9	0.00	9

# Candidates re-ranking

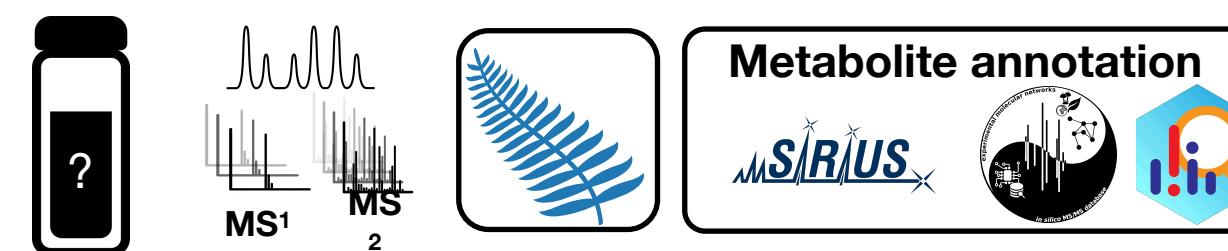


Feature ID	Spectrum	Biological source	Candidate structure	Score $S_1$	Initial rank
1			structure-1	0.55	1
			structure-2	0.53	2
			structure-3	0.50	3
			structure-4	0.45	4
...	...	...	...	...	...
1			structure-9	0.00	9

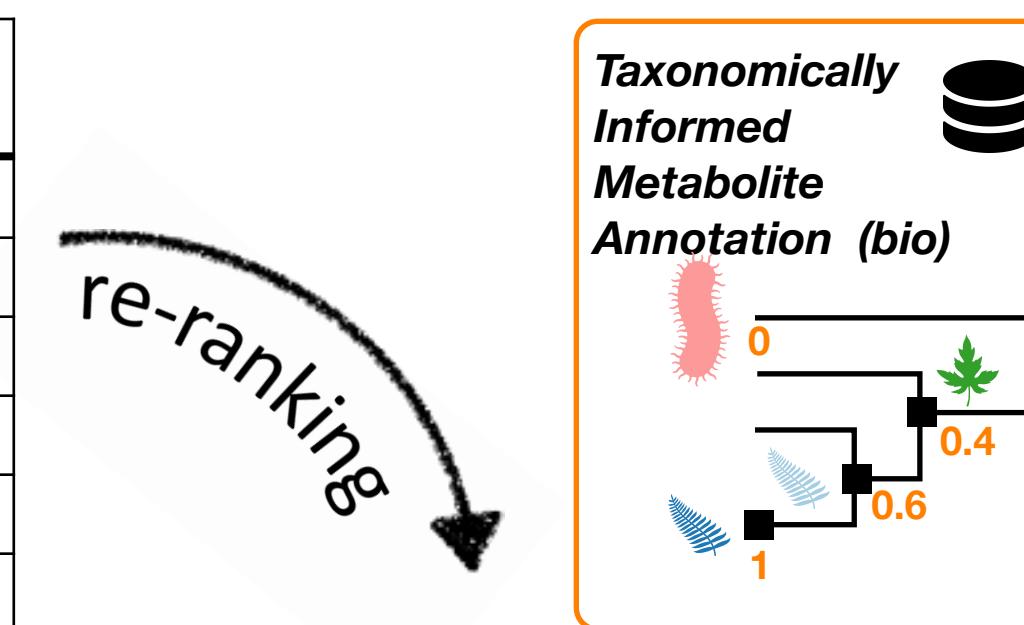


Feature ID	Spectrum	Biological source	Candidate structure	Score $S_1$	Initial rank	Candidate biological source	Score $S_2$	Combined score	Final rank
1			structure-4	0.45	4		1.00	0.78	1
			structure-1	0.55	1		0.60	0.58	2
			structure-2	0.53	2		0.40	0.47	3
			structure-9	0.00	9		0.60	0.30	4
...	...	...	structure-3	0.50	3		0.00	0.25	5

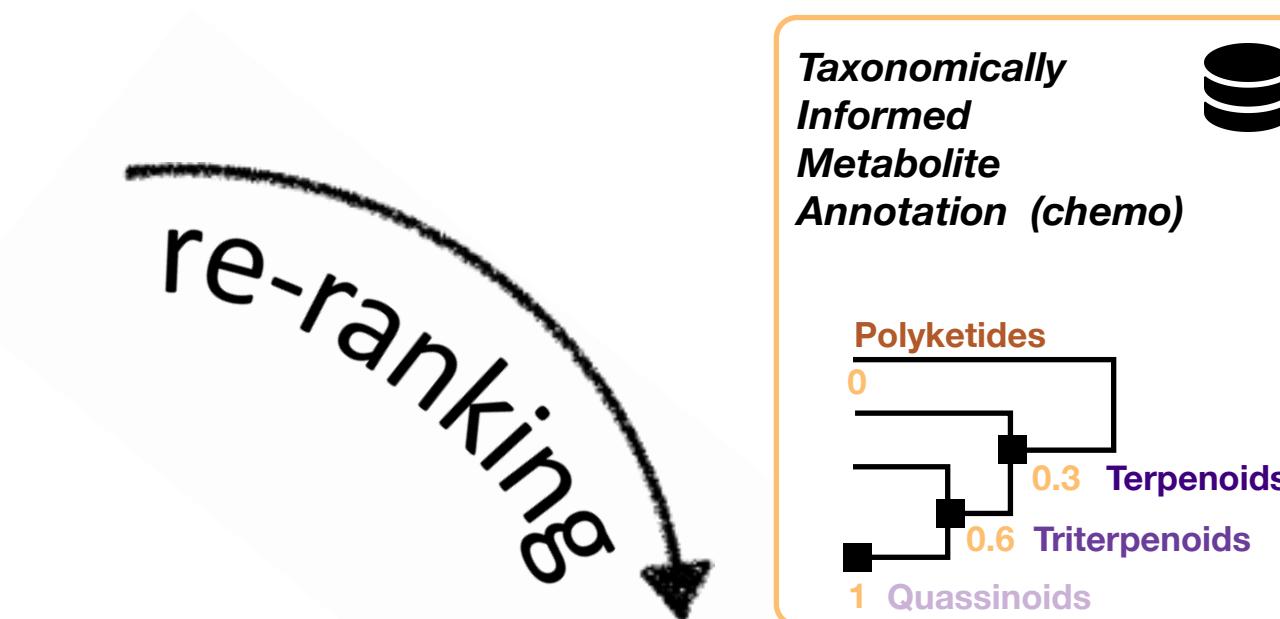
# Candidates re-ranking



Feature ID	Spectrum	Biological source	Candidate structure	Score $S_1$	Initial rank
1			structure-1	0.55	1
			structure-2	0.53	2
			structure-3	0.50	3
			structure-4	0.45	4
...	...	...	...	...	...
1			structure-9	0.00	9



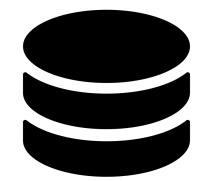
Feature ID	Spectrum	Biological source	Candidate structure	Score $S_1$	Initial rank	Candidate biological source	Score $S_2$	Combined score	Final rank
1			structure-4	0.45	4		1.00	0.78	1
			structure-1	0.55	1		0.60	0.58	2
			structure-2	0.53	2		0.40	0.47	3
			structure-9	0.00	9		0.60	0.30	4
...	...	...	structure-3	0.50	3		0.00	0.25	5



Feature ID	Spectrum	Biological source	Candidate structure	Score $S_1$	Initial rank	Candidate biological source	Score $S_2$	Combined score	2nd rank	Attributed chemical class	Candidate chemical class	Score $S_3$	Combined score	Final rank
1			structure-4	0.45	4		1.00	0.78	1	Quassinoïd	Quassinoïd	1.00	0.82	1
			structure-1	0.55	1		0.60	0.58	2		Triterpenoids	0.60	0.58	2
			structure-2	0.53	2		0.40	0.47	3		Terpenoids	0.30	0.41	3
			structure-9	0.00	9		0.60	0.30	4		Terpenoids	0.30	0.30	4
...	...	...	structure-3	0.50	3		0.00	0.25	5	...	Polyketides	0.00	0.17	5

# Benchmark

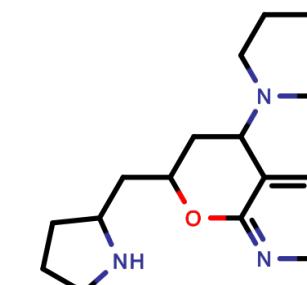
## EXPERIMENTAL SPECTRA



GNPS <https://gnps.ucsd.edu>  
public and third parties libraries

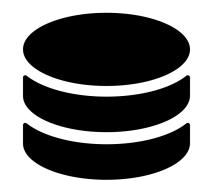
66,646 experimental spectra

## STRUCTURES



# Benchmark

## EXPERIMENTAL SPECTRA

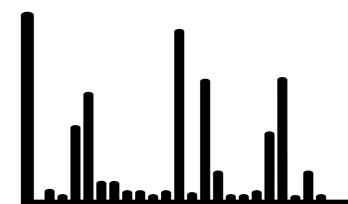


GNPS <https://gnps.ucsd.edu>  
public and third parties libraries

66,646 experimental spectra



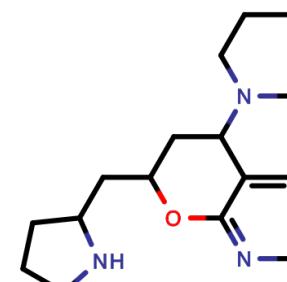
at least 6 fragments  
max 500 fragments  
 $100 \text{ Da} < x < 1500 \text{ Da}$   
[M+H]<sup>+</sup> adduct filtering



40,138 cleaned spectra



## STRUCTURES

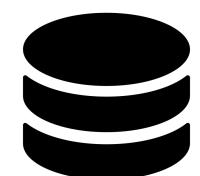


chemical translation  
and sanitization

Cleaned  
2D structures

# Benchmark

## EXPERIMENTAL SPECTRA

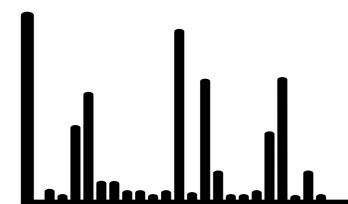


GNPS <https://gnps.ucsd.edu>  
public and third parties libraries

66,646 experimental spectra



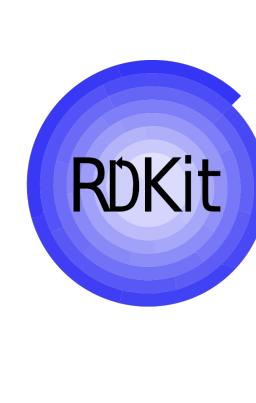
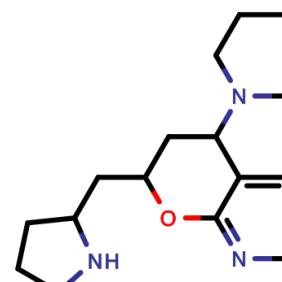
at least 6 fragments  
max 500 fragments  
 $100 \text{ Da} < x < 1500 \text{ Da}$   
[M+H]<sup>+</sup> adduct filtering



40,138 cleaned spectra



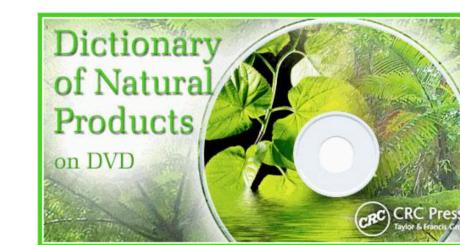
## STRUCTURES



chemical translation  
and sanitization

Cleaned  
2D structures

## BIOLOGICAL SOURCES



"Alkaloid from *Brunfelsia hopeana*"

# Benchmark

## EXPERIMENTAL SPECTRA



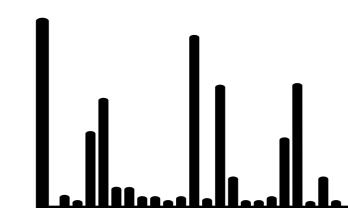
GNPS <https://gnps.ucsd.edu>  
public and third parties libraries

66,646 experimental spectra



jupyter

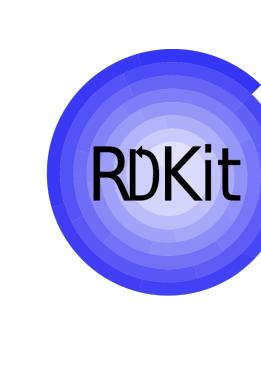
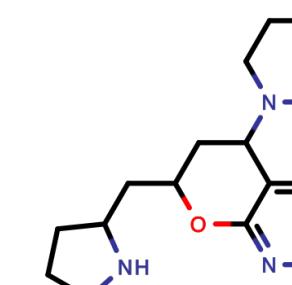
at least 6 fragments  
max 500 fragments  
 $100 \text{ Da} < x < 1500 \text{ Da}$   
[M+H]<sup>+</sup> adduct filtering



40,138 cleaned spectra



## STRUCTURES

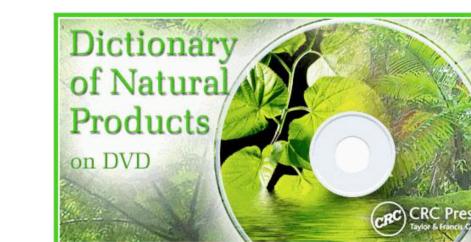


chemical translation  
and sanitization

Cleaned  
2D structures



## BIOLOGICAL SOURCES



"Alkaloid from *Brunfelsia hopeana*"



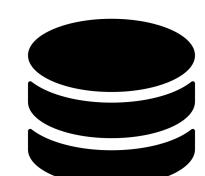
[http://  
globalnames.org](http://globalnames.org)

text recognition  
matching and  
resolving against the  
Catalogue of Life

Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

# Benchmark

## EXPERIMENTAL SPECTRA



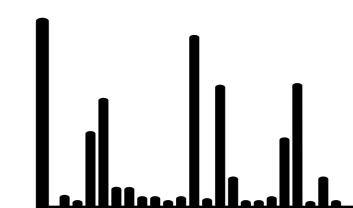
GNPS <https://gnps.ucsd.edu>  
public and third parties libraries

66,646 experimental spectra



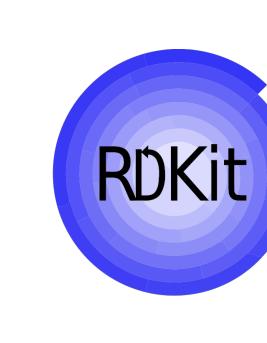
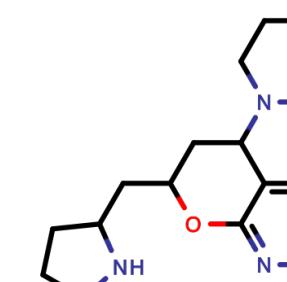
jupyter

- at least 6 fragments
- max 500 fragments
- $100 \text{ Da} < x < 1500 \text{ Da}$
- $[\text{M}+\text{H}]^+$  adduct filtering



40,138 cleaned spectra

## STRUCTURES



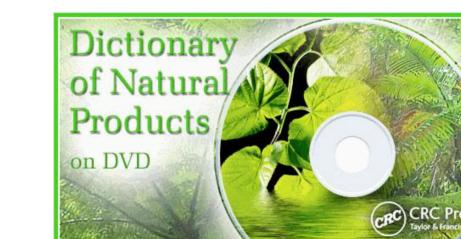
chemical translation  
and sanitization

Cleaned  
2D structures



2107 unique entries

## BIOLOGICAL SOURCES



"Alkaloid from *Brunfelsia hopeana*"



Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

text recognition  
matching and  
resolving against the  
Catalogue of Life

# Evaluation

 ISDB-DNP

 Sirius

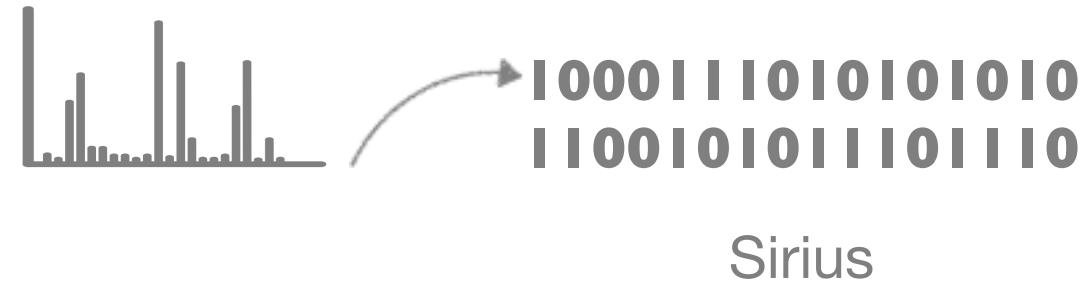
 MS-Finder

# Evaluation

## Spectral similarity



## Fingerprint similarity



**ISDB-DNP**

**Sirius**

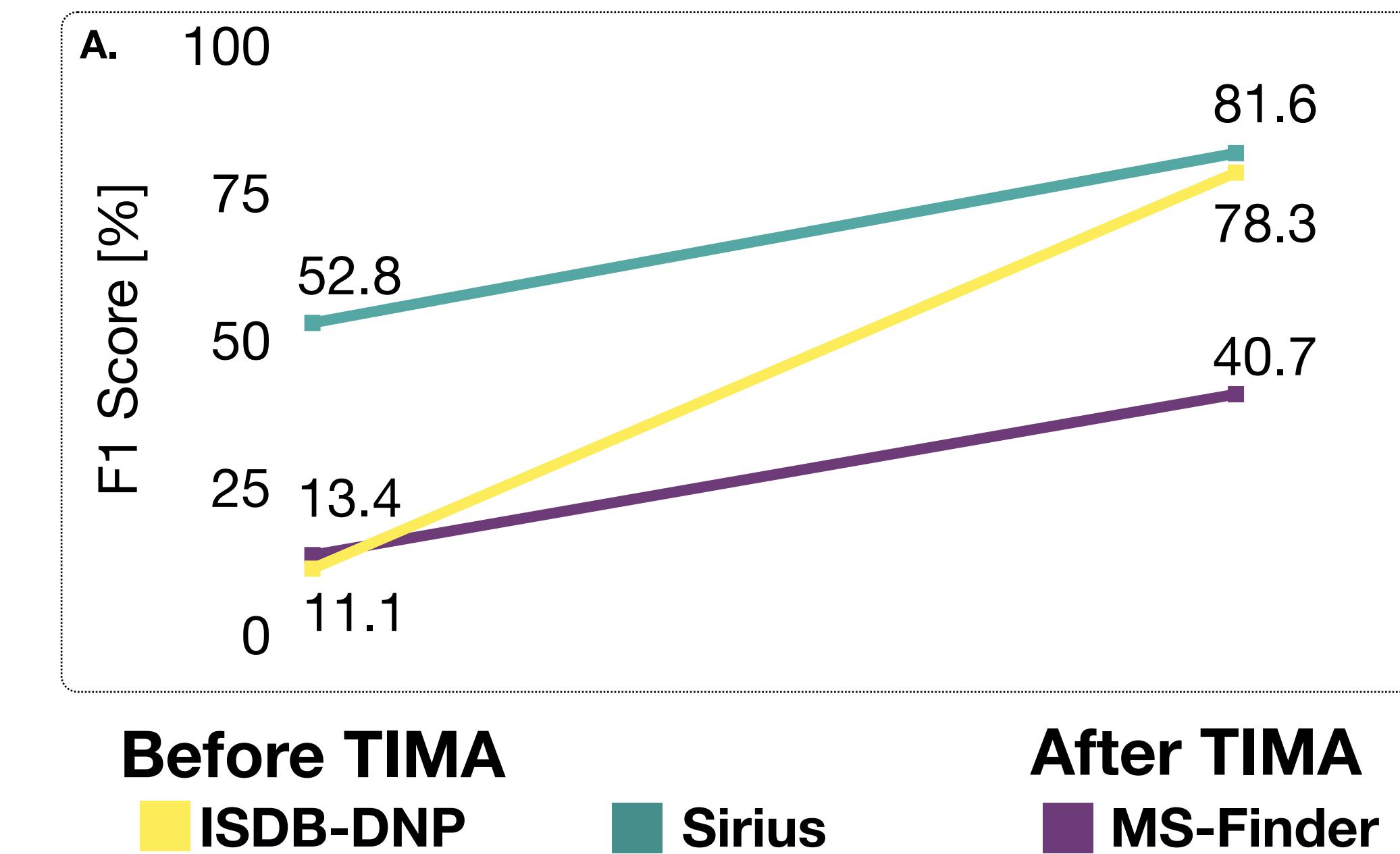
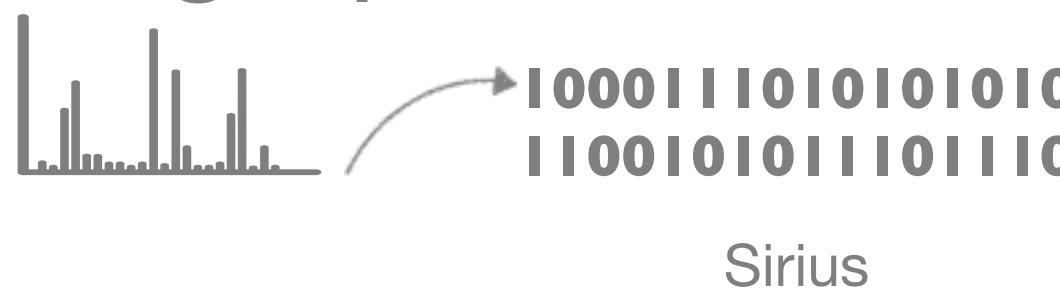
**MS-Finder**

# Evaluation

Spectral similarity



Fingerprint similarity

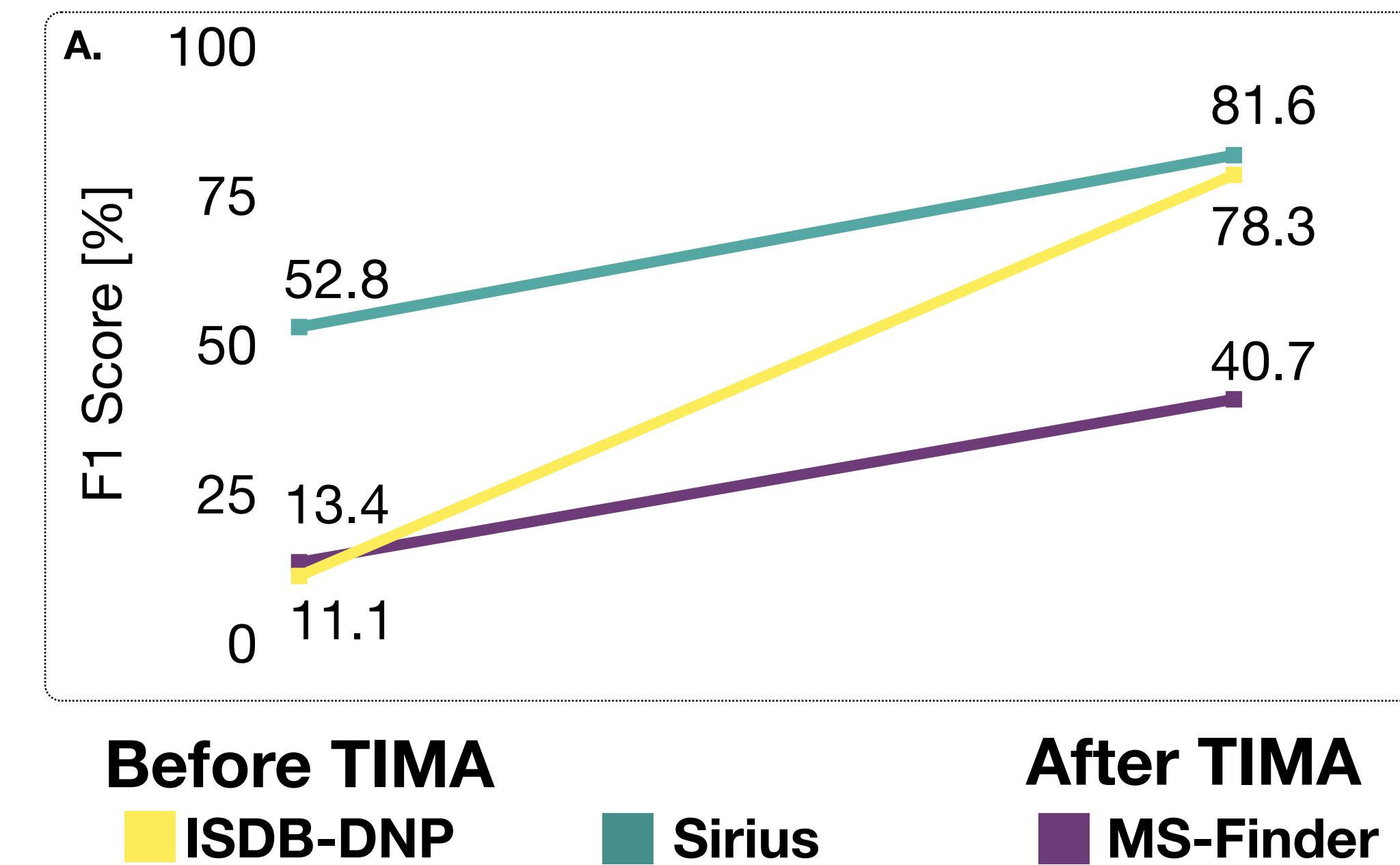
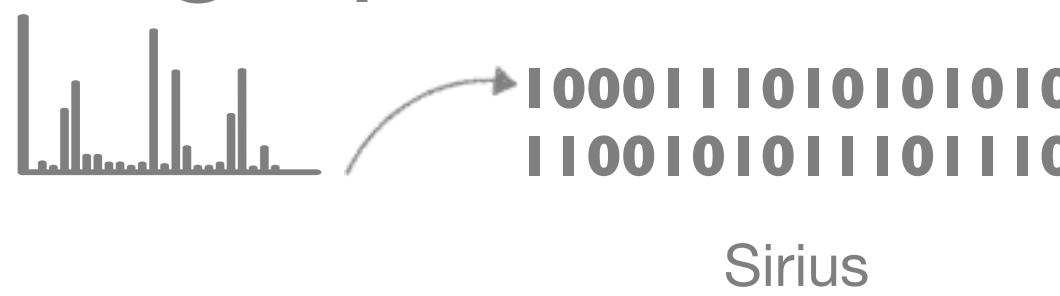


# Evaluation

Spectral similarity



Fingerprint similarity



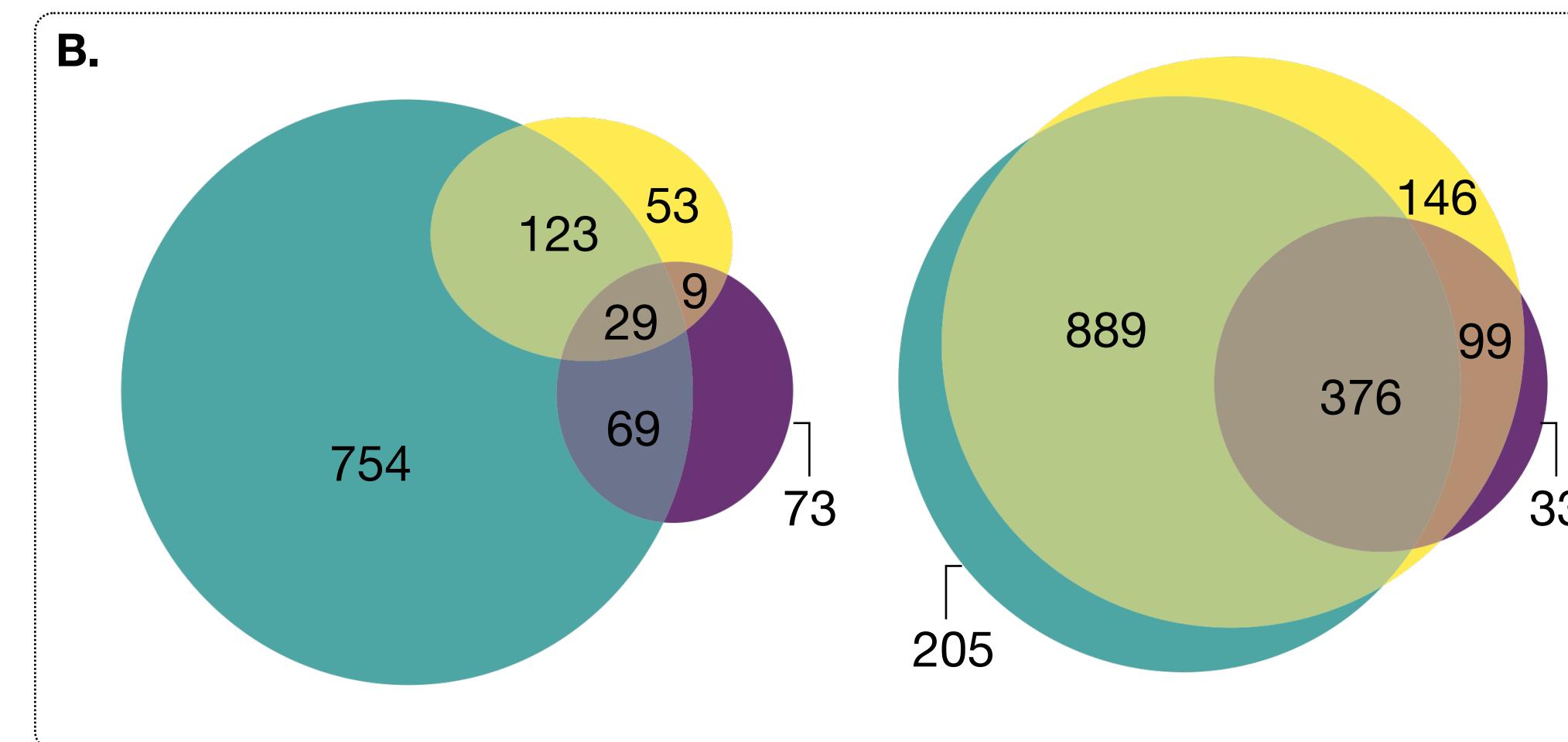
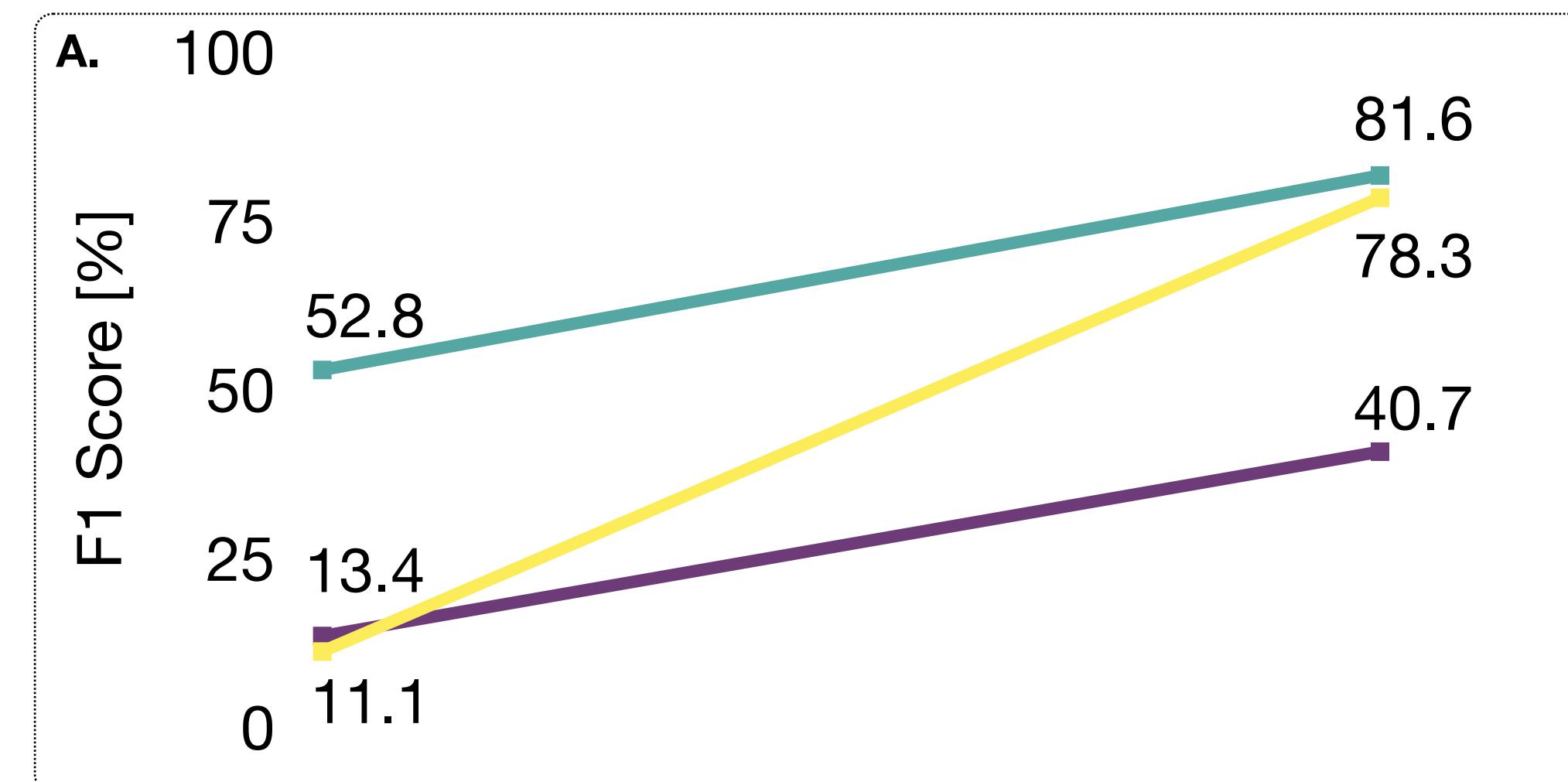
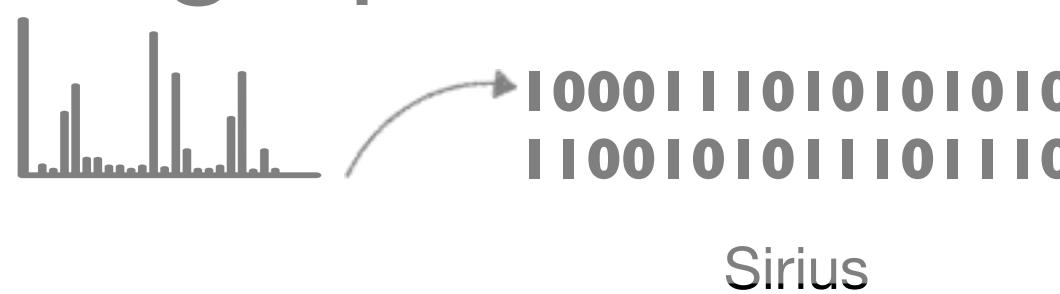
**F1 score ≈ correct candidates at rank 1**

# Evaluation

Spectral similarity

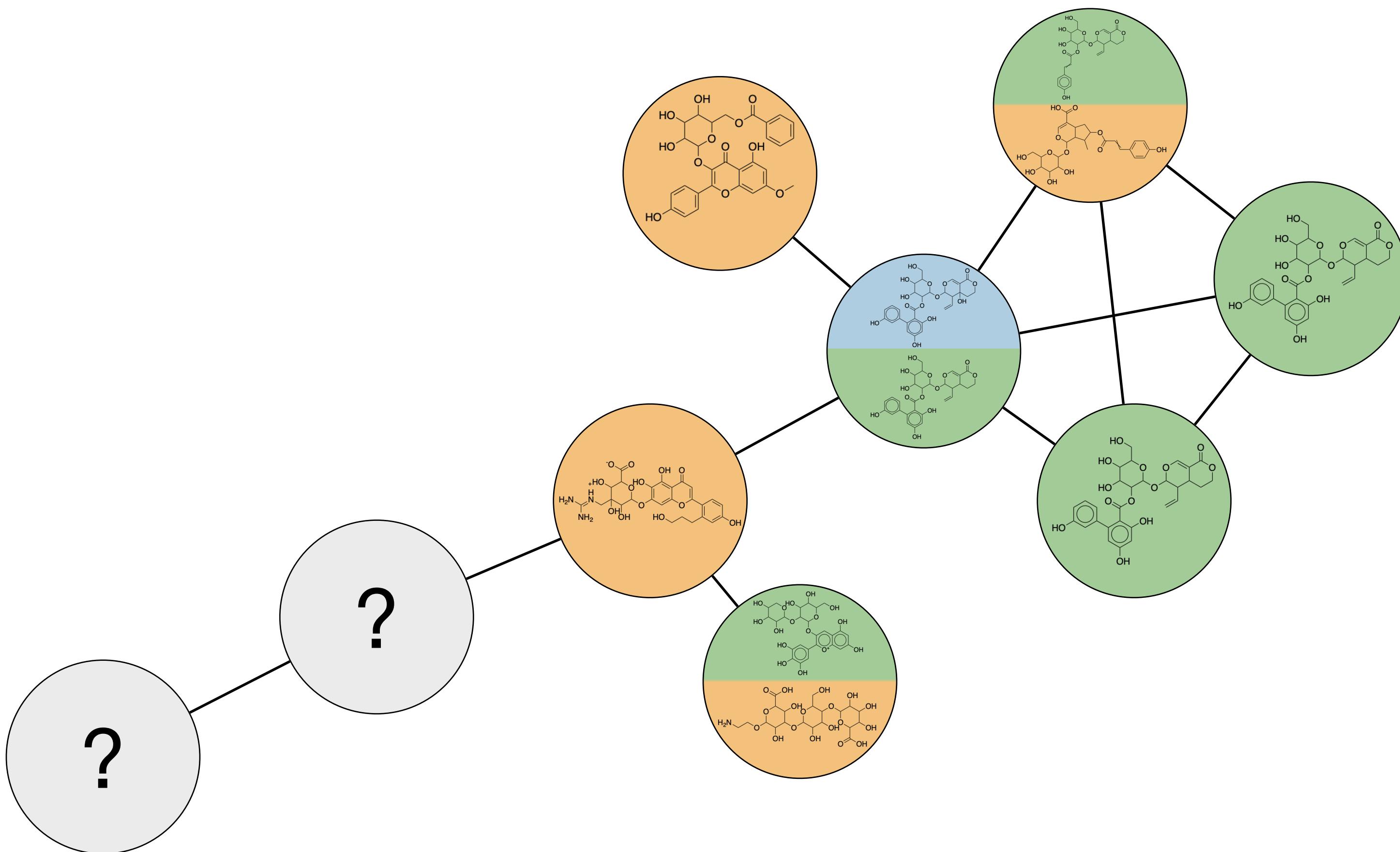


Fingerprint similarity



**F1 score ≈ correct candidates at rank 1**

# Taxonomically Informed Metabolite Annotation

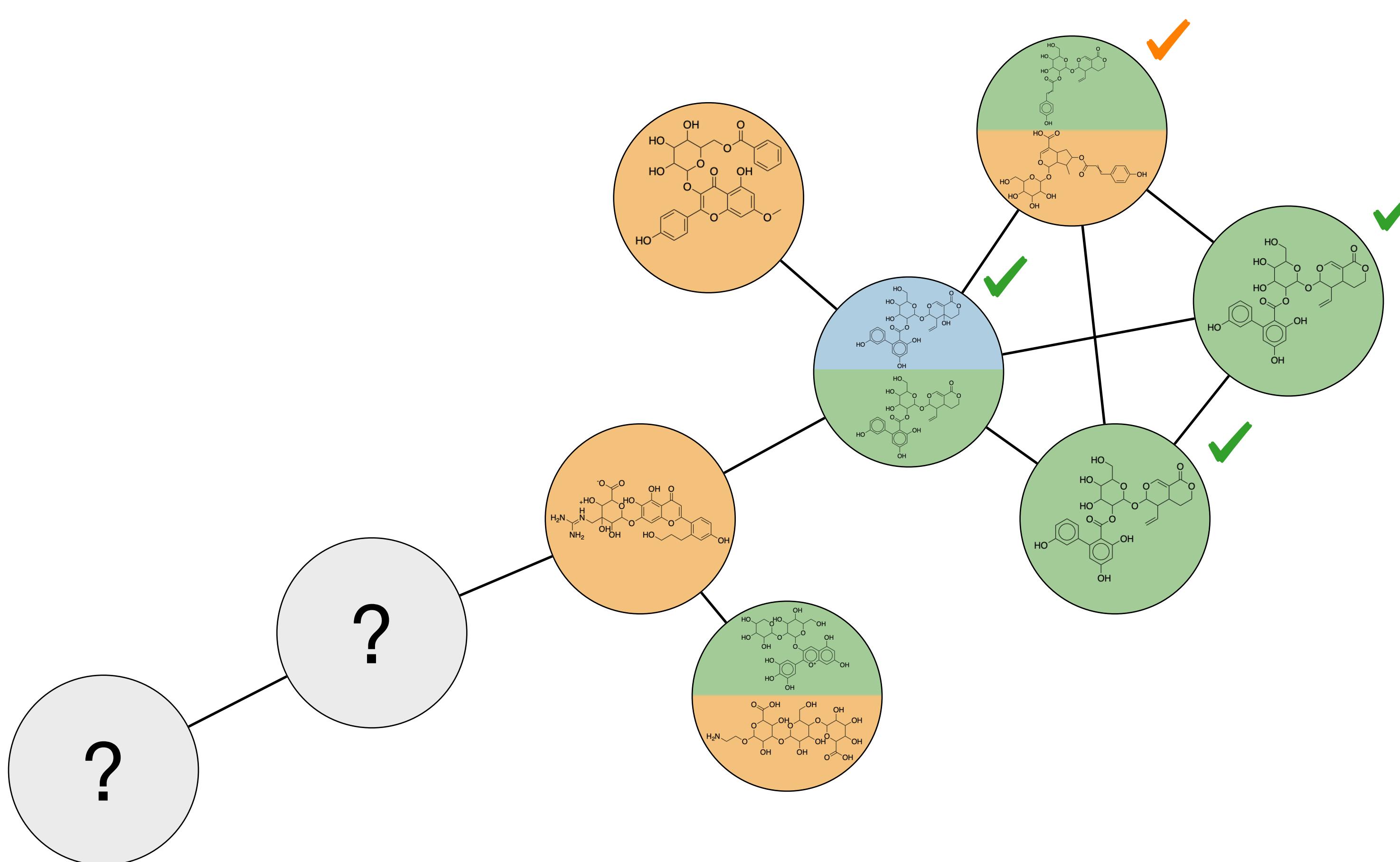


**GNPS**  
**(experimental)**

**ISDB-DNP (in silico)**

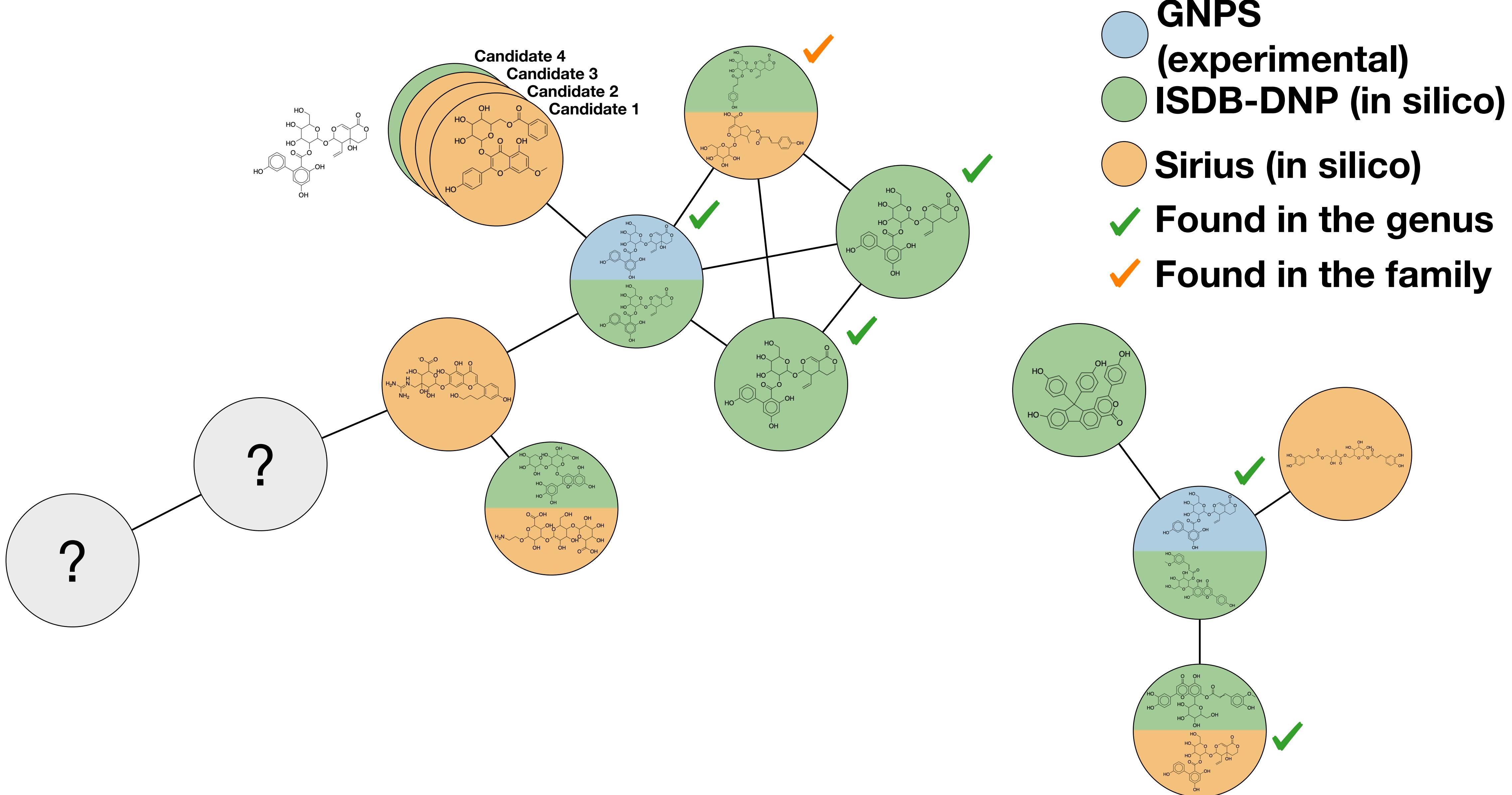
**Sirius (in silico)**

# Taxonomically Informed Metabolite Annotation

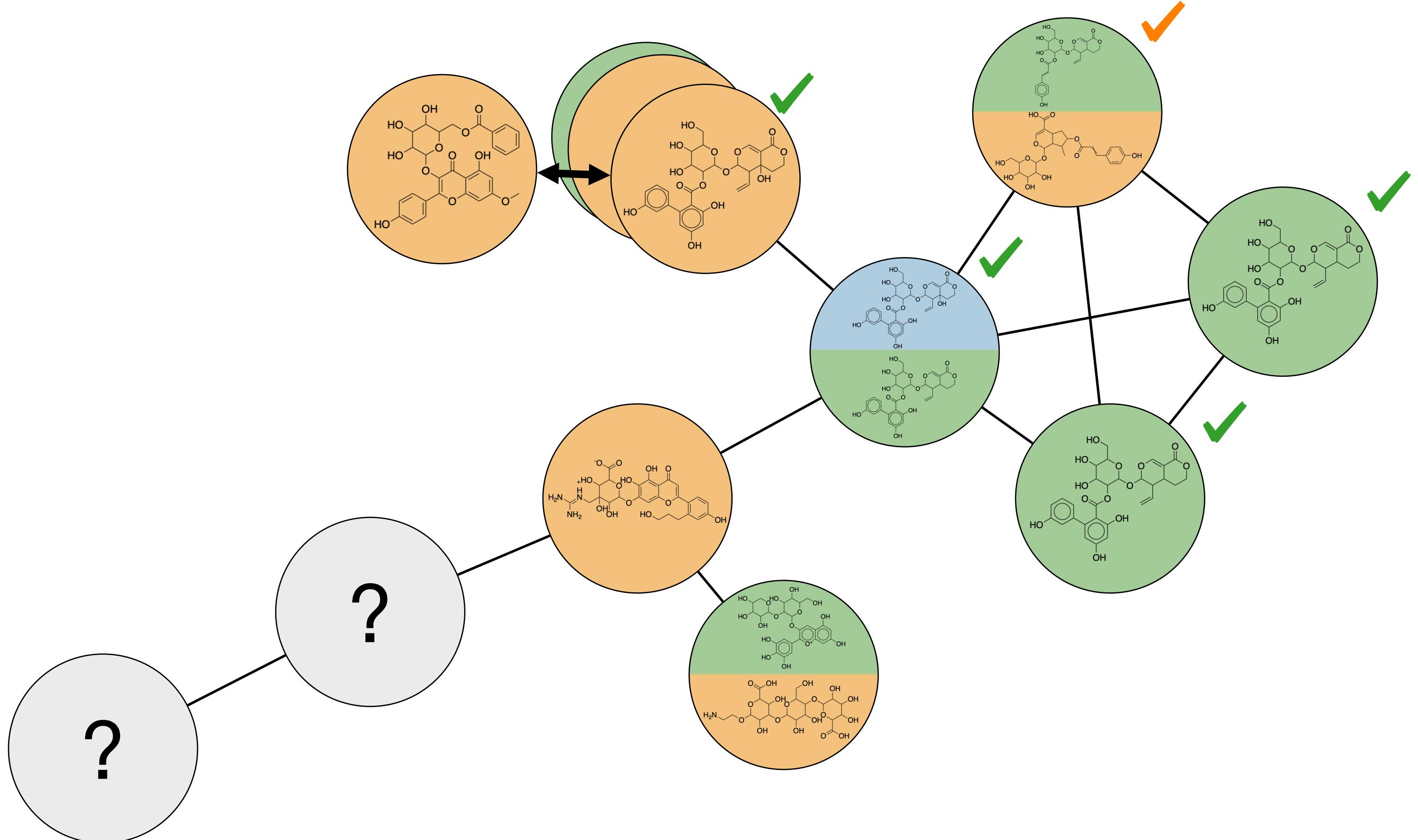


- GNPS (experimental)**
- ISDB-DNP (in silico)**
- Sirius (in silico)**
- ✓ **Found in the genus**
- ✓ **Found in the family**

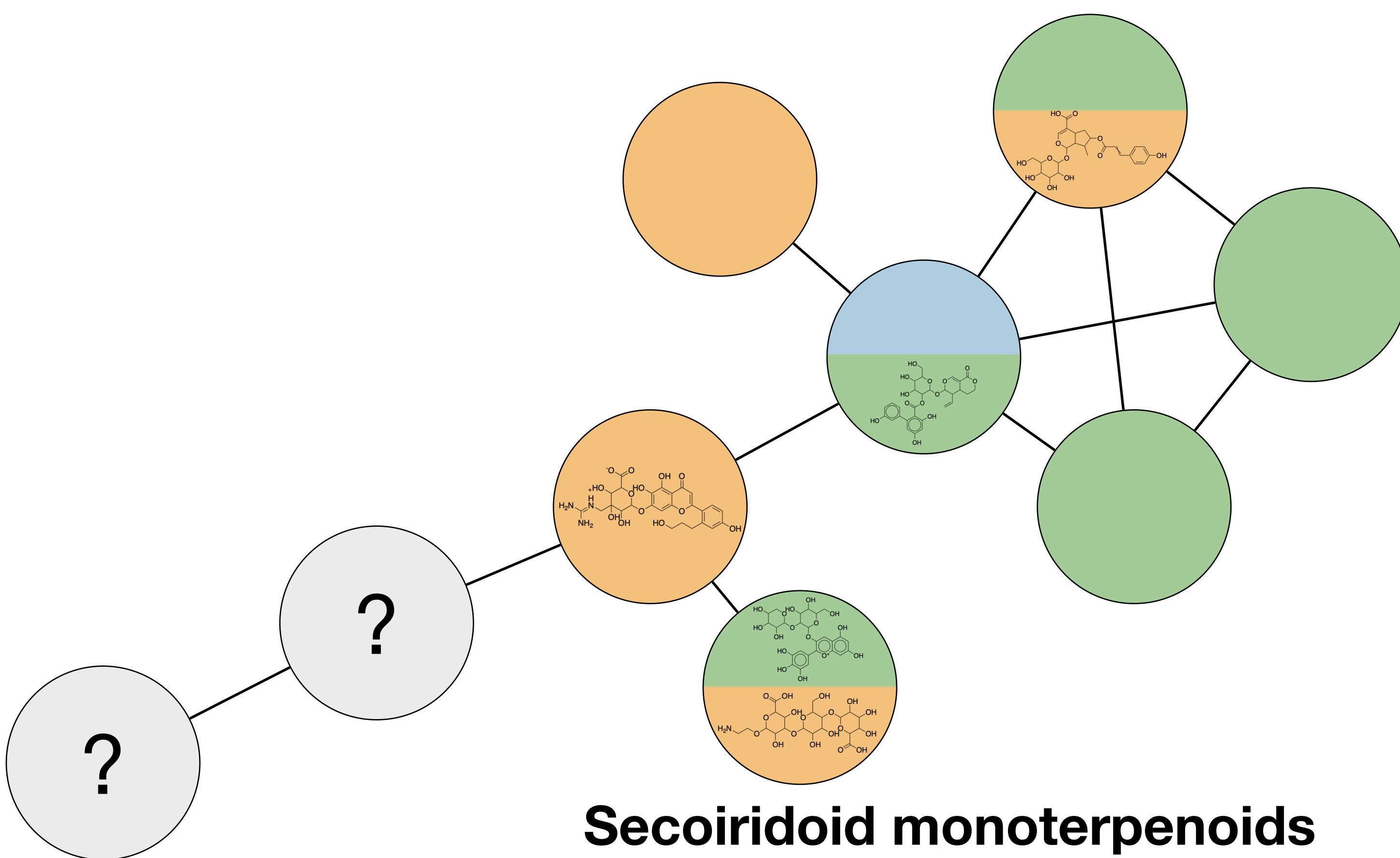
# Taxonomically Informed Metabolite Annotation



# Taxonomically Informed Metabolite Annotation

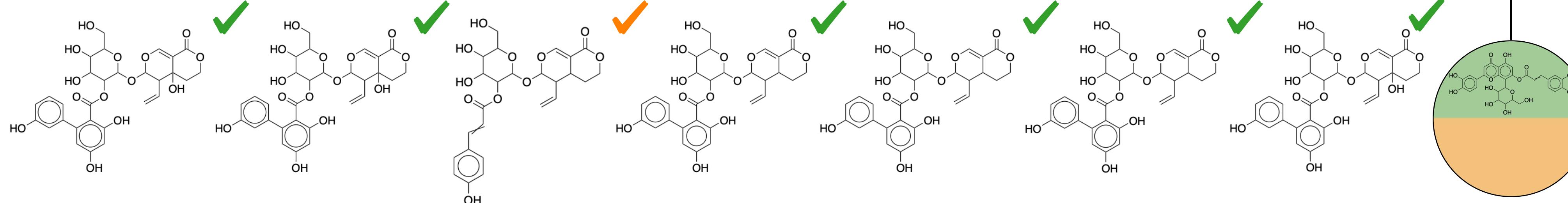


- GNPS (experimental)**
- ISDB-DNP (in silico)**
- Sirius (in silico)**
- ✓ Found in the genus**
- ✓ Found in the family**

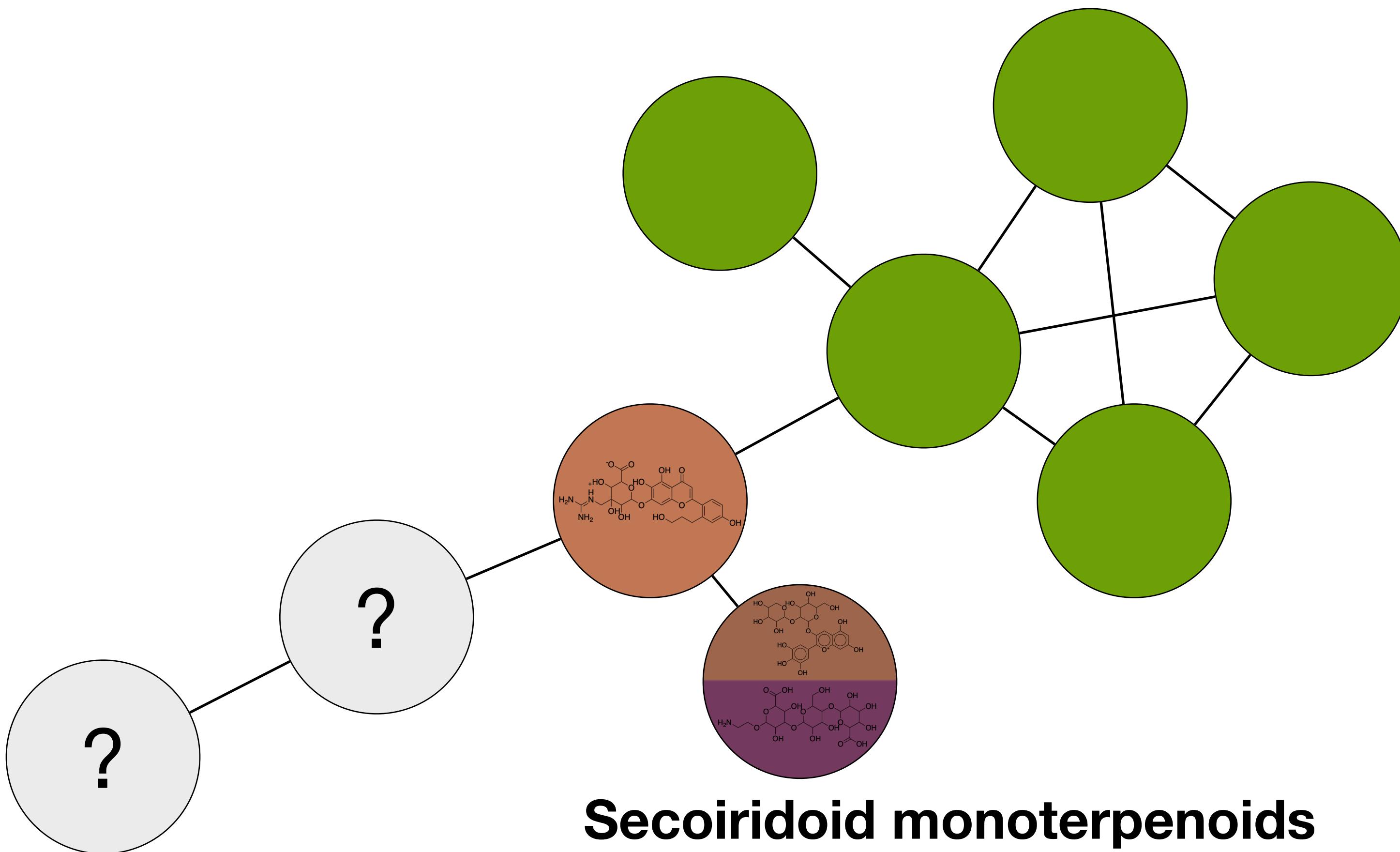


- **GNPS (experimental)**
- **ISDB-DNP (in silico)**
- **Sirius (in silico)**
- ✓ **Found in the genus**
- ✓ **Found in the family**

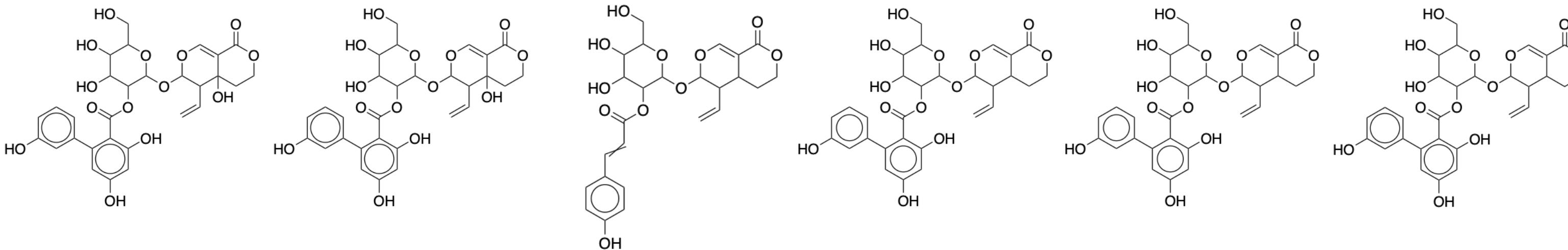
## Secoiridoid monoterpenoids



- Secoiridoid monoterpenoids
- Cinnamic acids and derivatives
- Selaginellins
- Flavones
- Anthocyanidins
- Polysaccharides

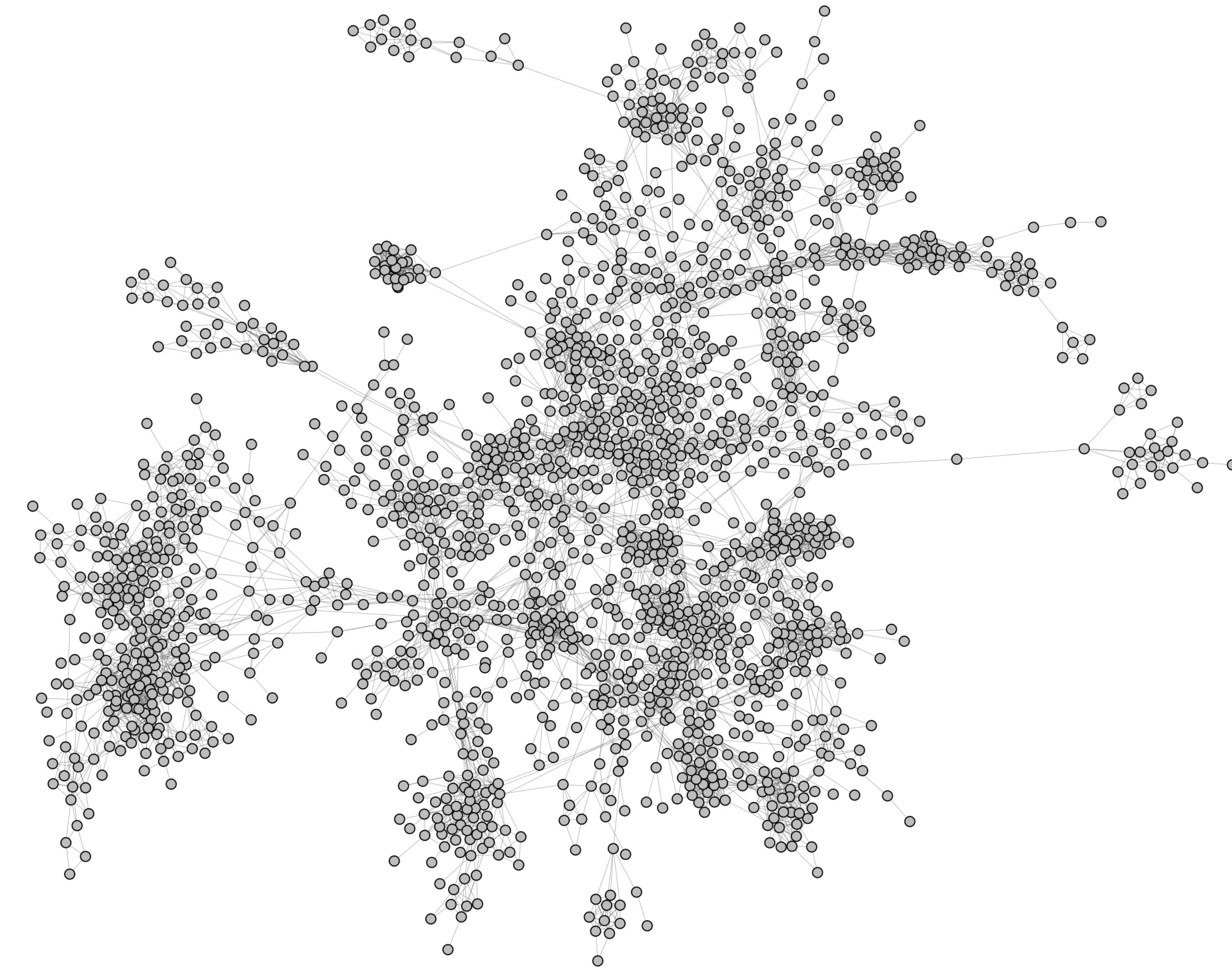


## Secoiridoid monoterpenoids



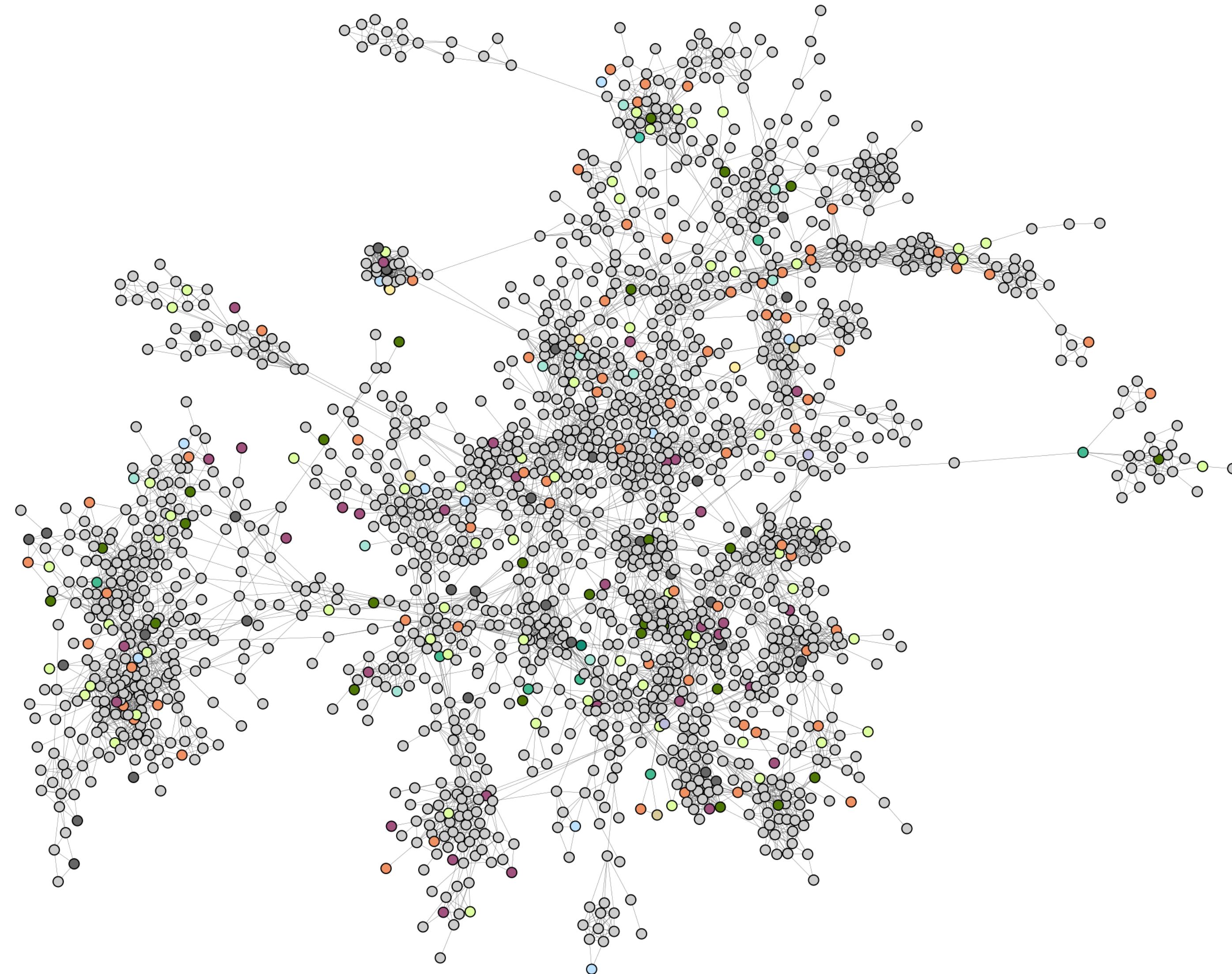
# Taxonomically Informed Metabolite Annotation

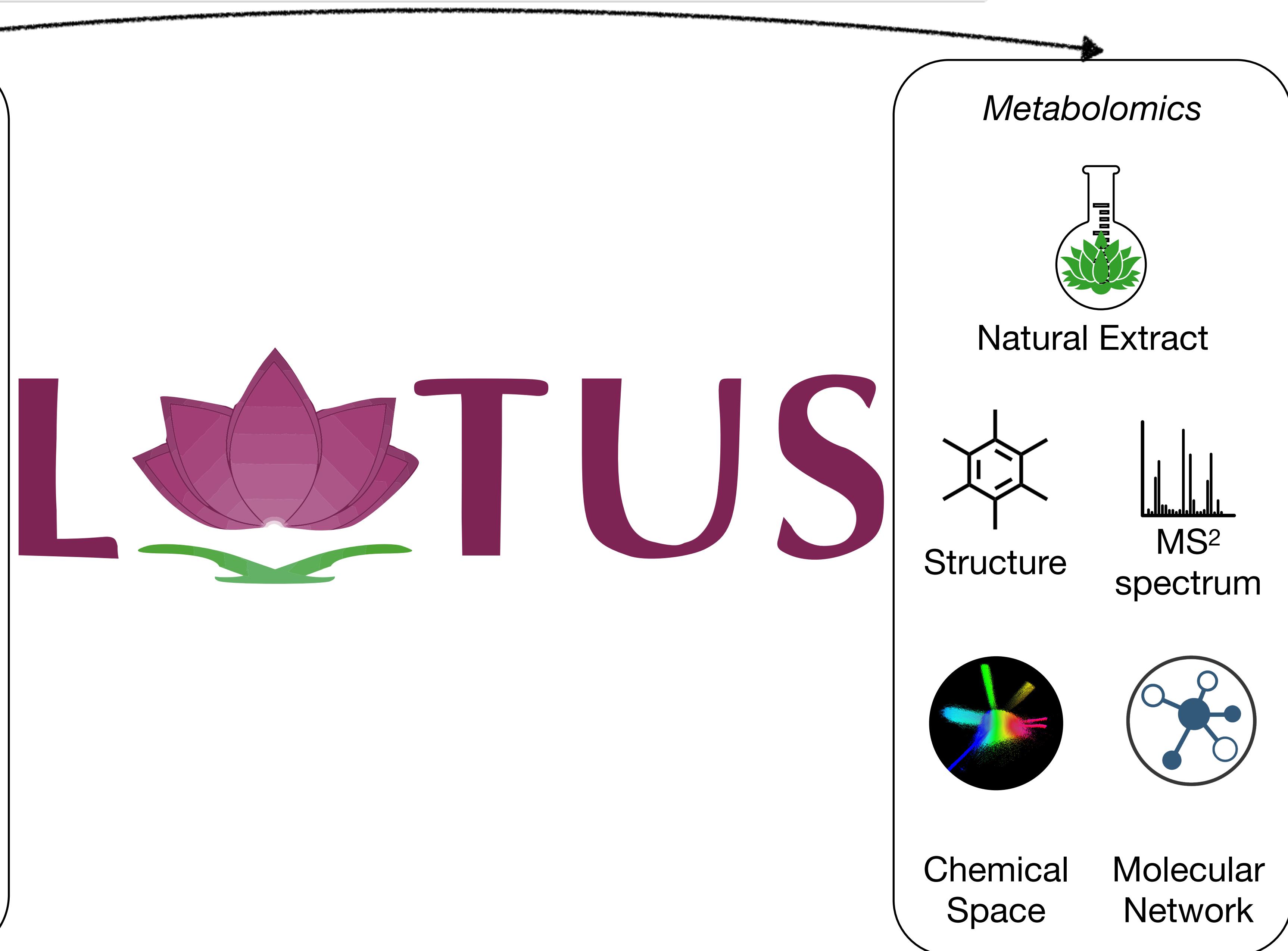
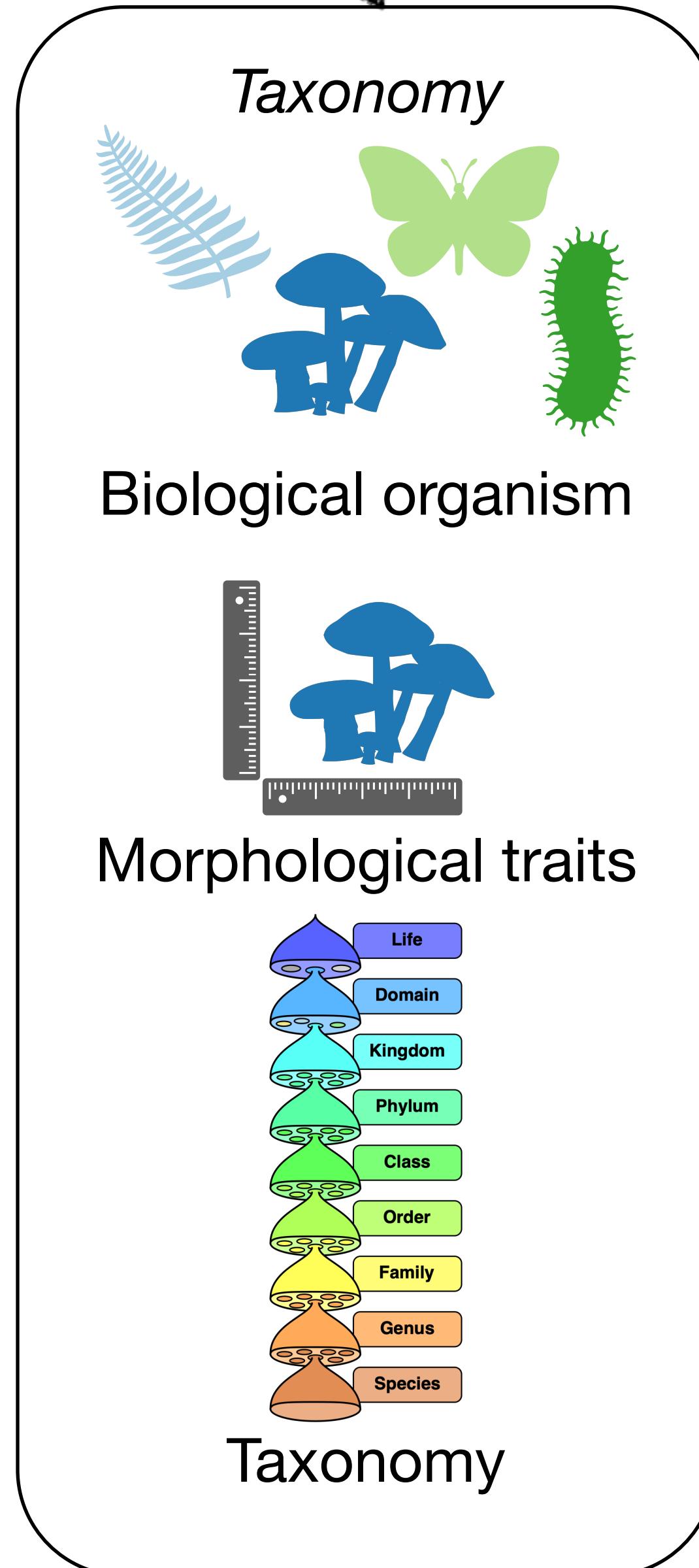
---



# Taxonomically Informed Metabolite Annotation

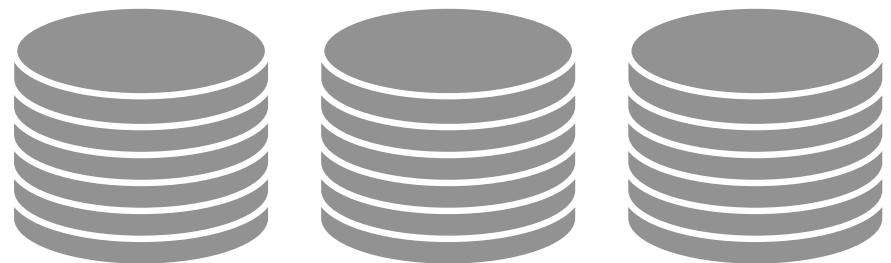
---







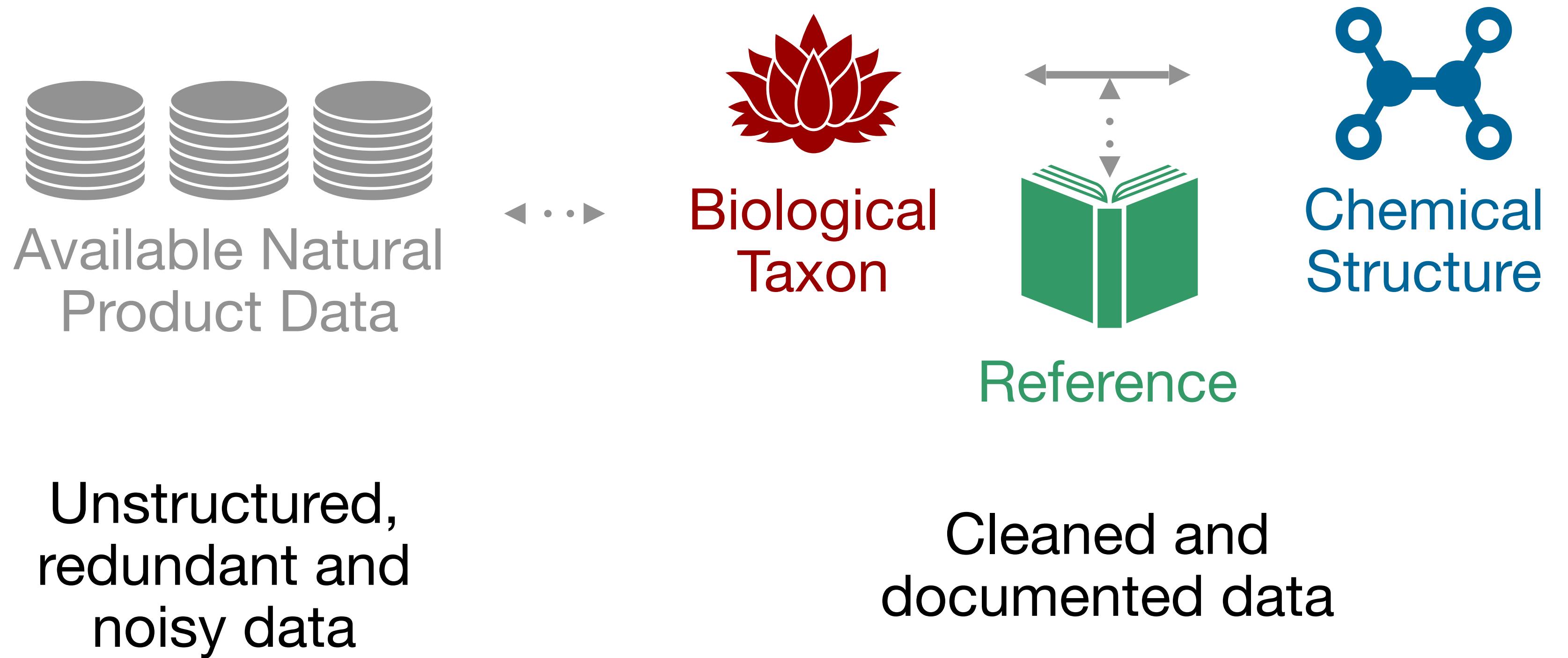
# The initiative



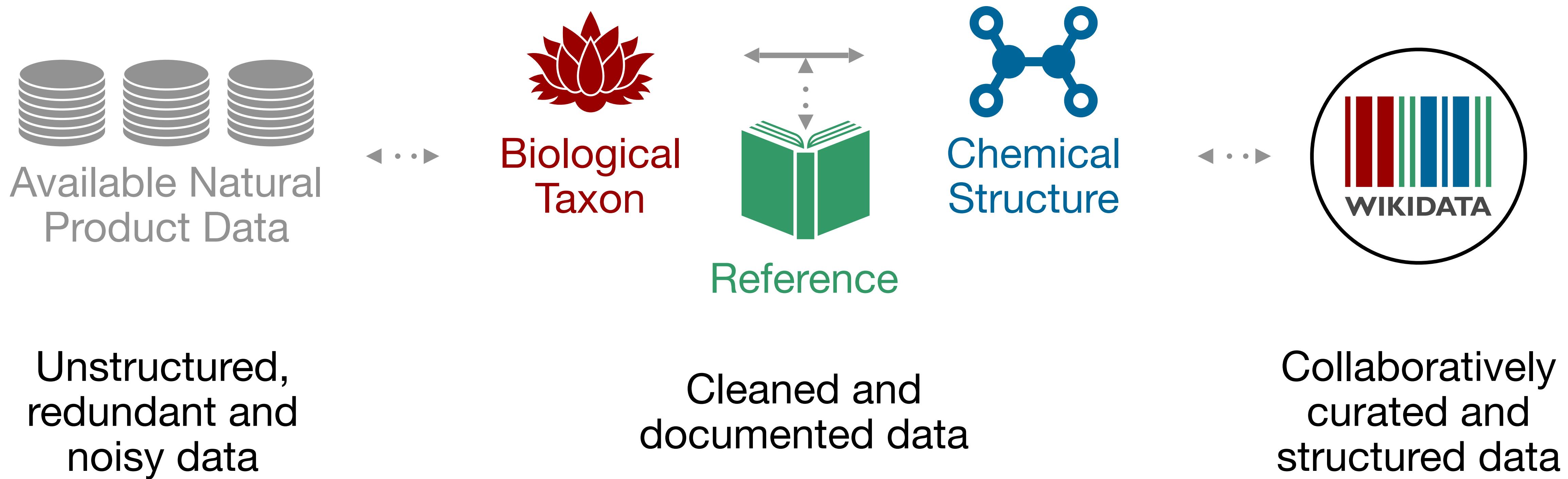
Available Natural  
Product Data

Unstructured,  
redundant and  
noisy data

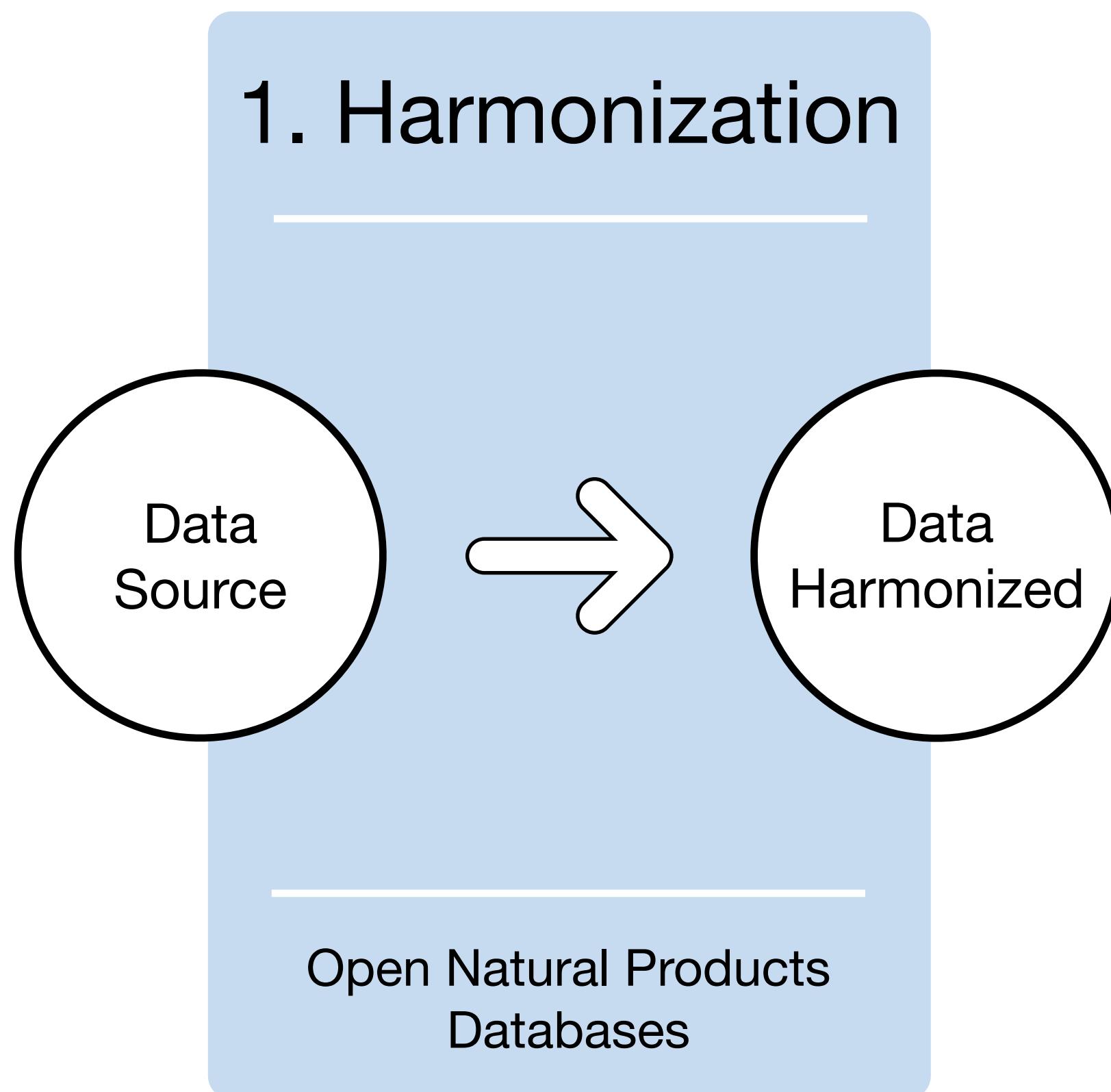
# The initiative



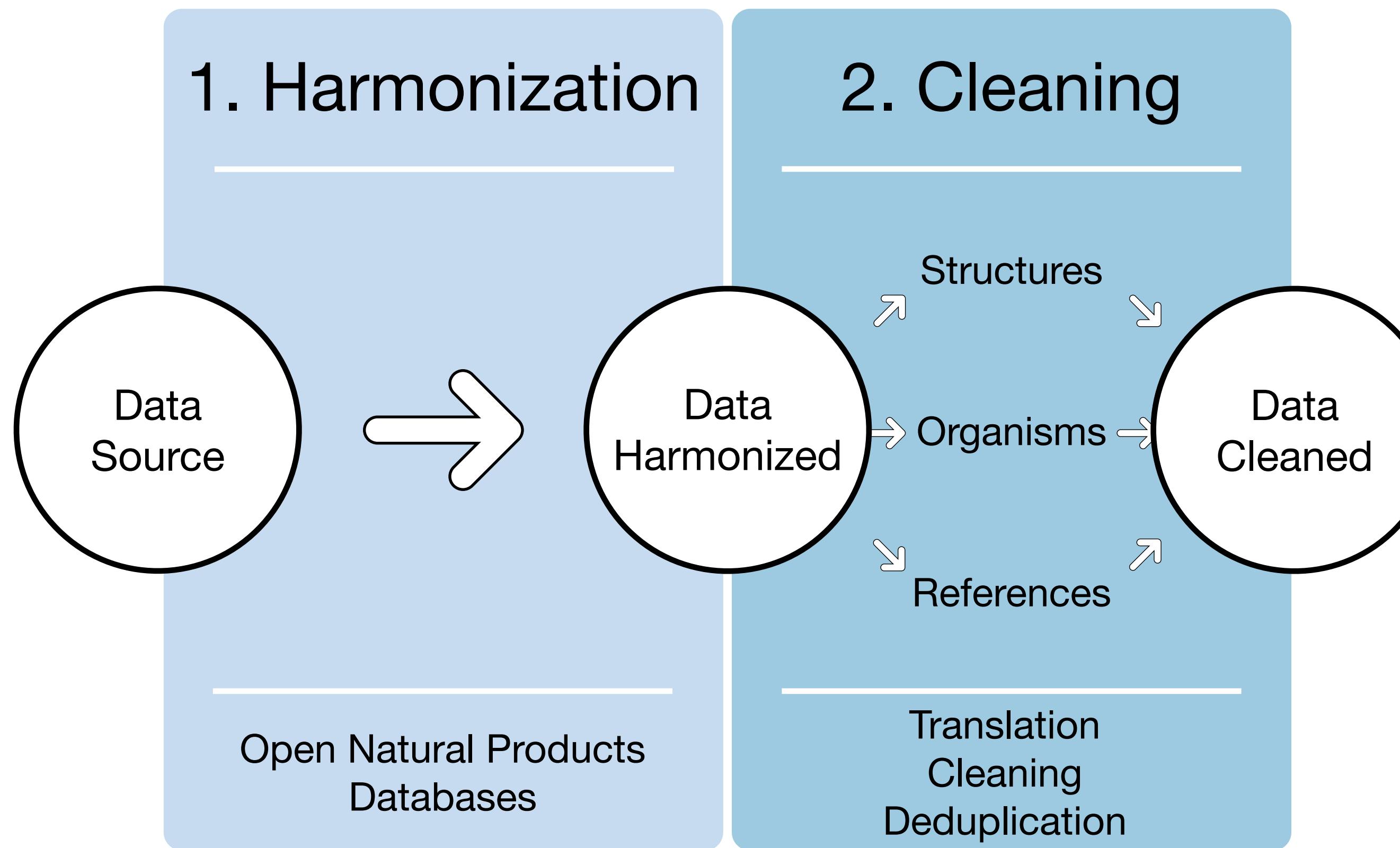
# The initiative



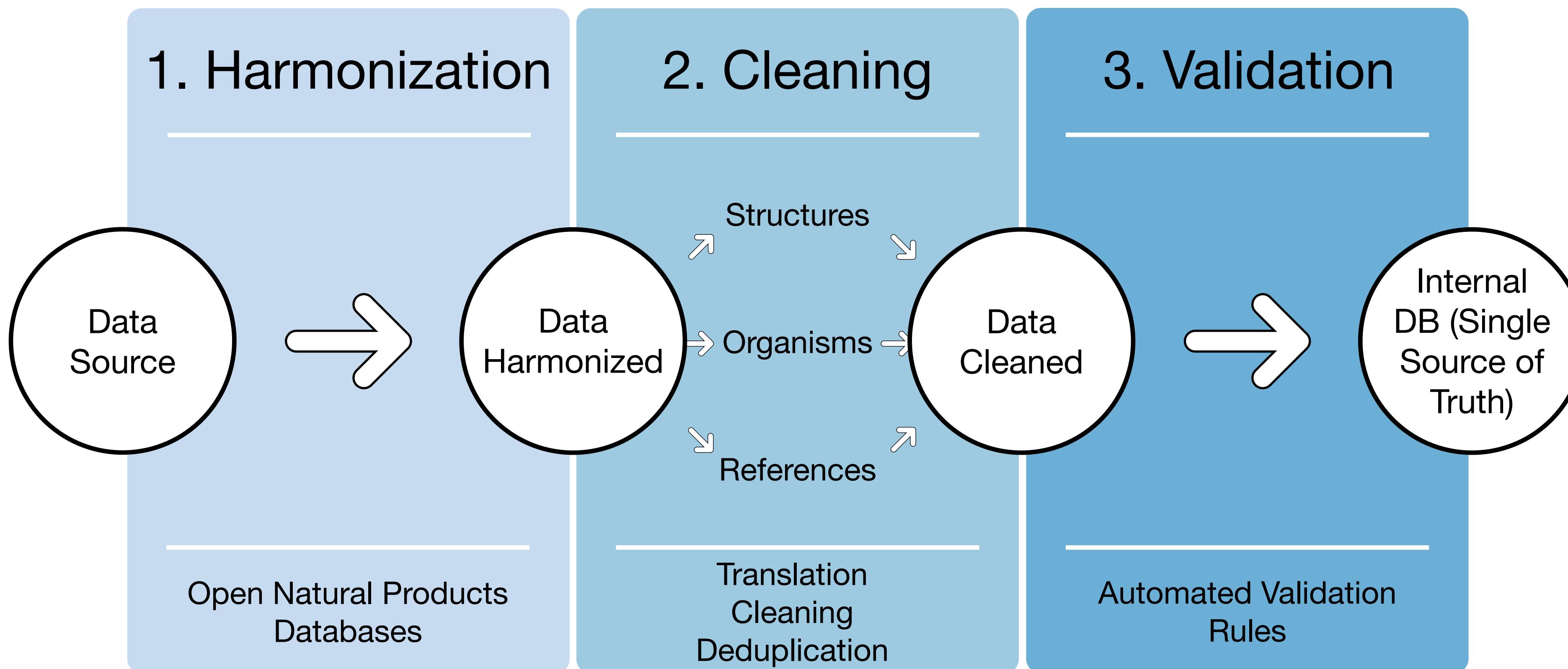
# The initiative - How?



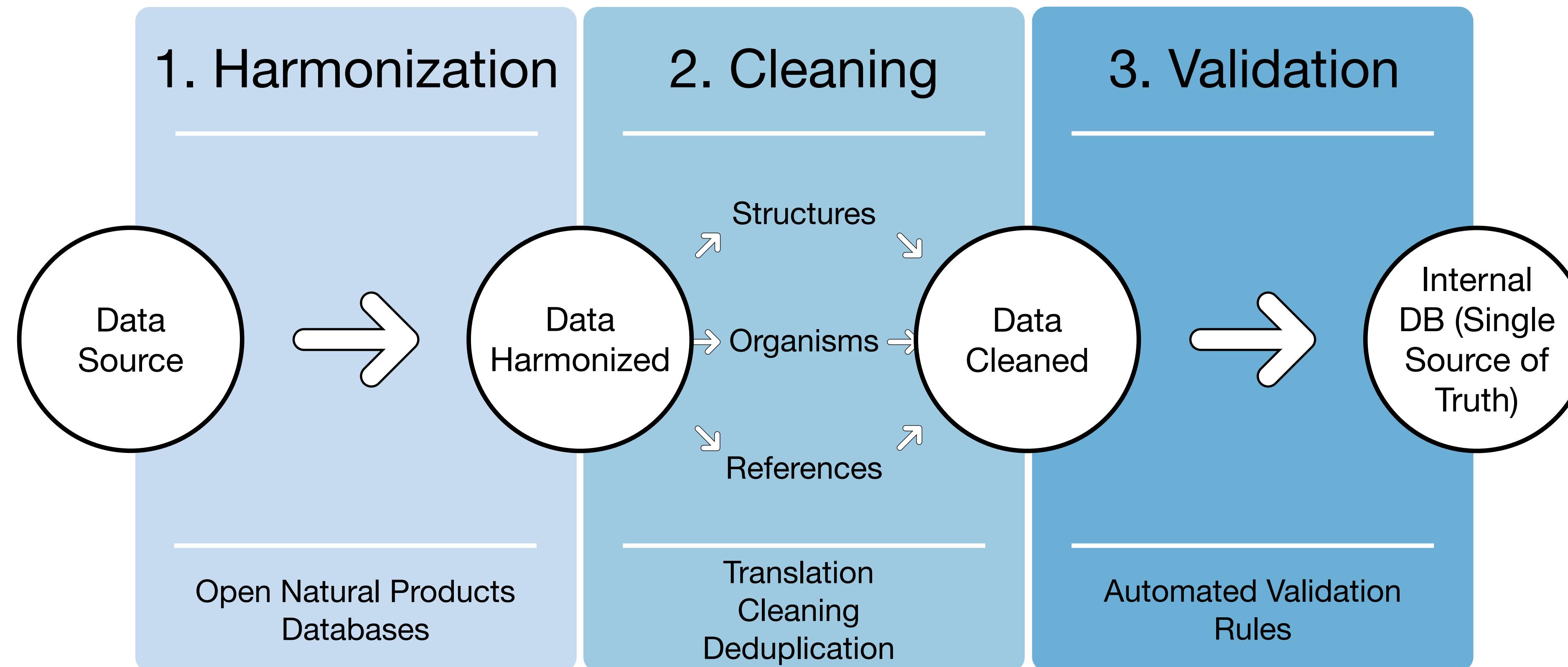
# The initiative - How?



# The initiative - How?

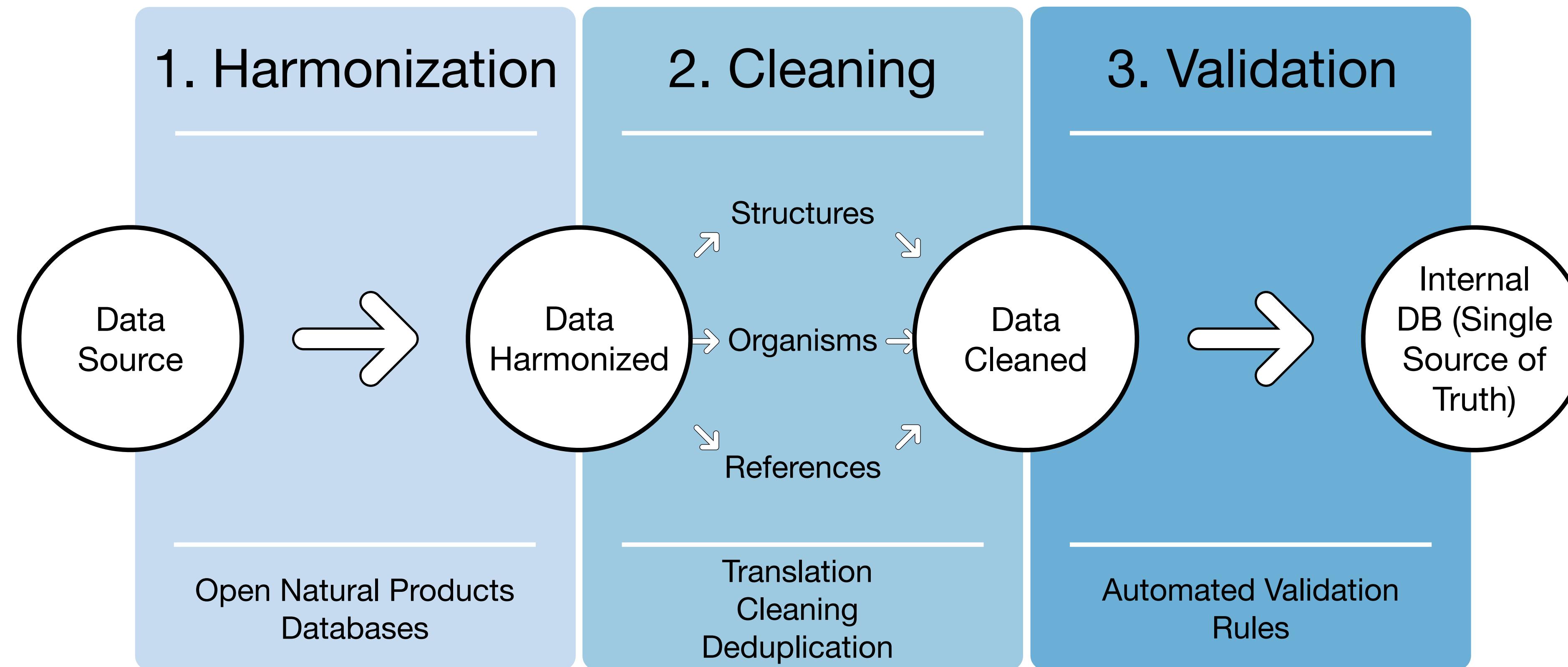


# The initiative - How?



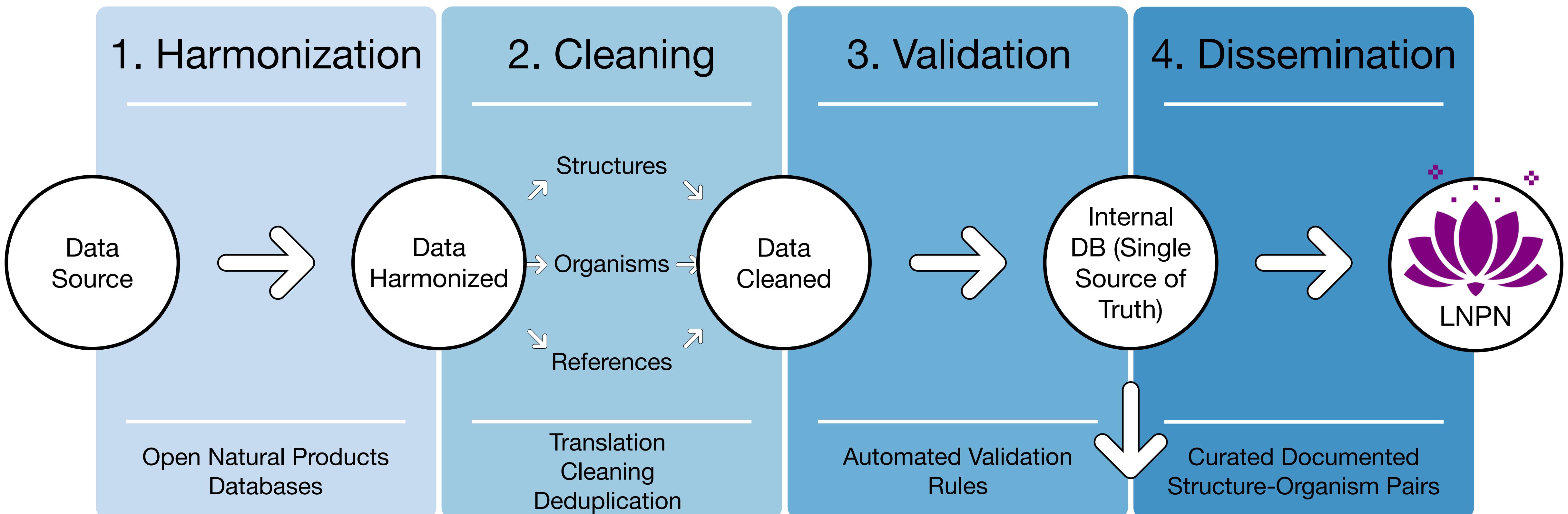
	Structure	Organism	Reference
Before curation	Cyathocaline	Stem bark of <i>Cyathocalyx zeylanica</i> CHAMP. ex HOOK. f. & THOMS. (Annonaceae)	Wijeratne E. M. K., de Silva L. B., Kikuchi T., Tezuka Y., Gunatilaka A. A. L., Kingston D. G. I., J. Nat. Prod., 58, 459-462 (1995).

# The initiative - How?

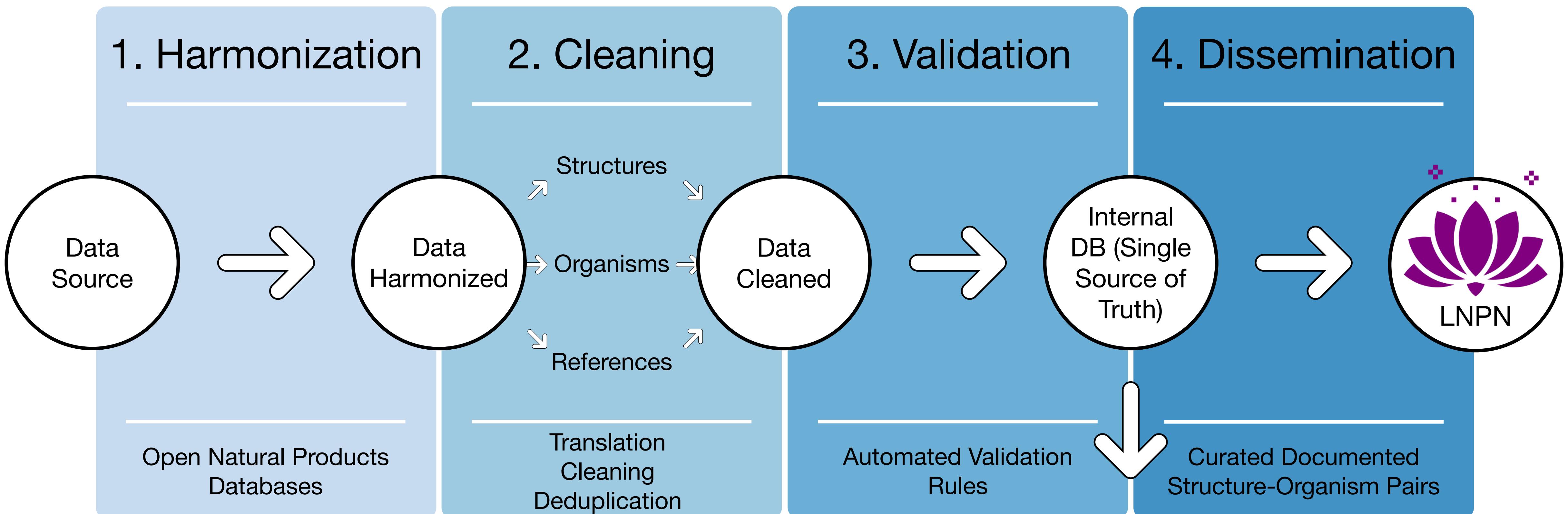


	Structure	Organism	Reference
Before curation	Cyathocaline	Stem bark of <i>Cyathocalyx zeylanica</i> CHAMP. ex HOOK. f. & THOMS. (Annonaceae)	Wijeratne E. M. K., de Silva L. B., Kikuchi T., Tezuka Y., Gunatilaka A. A. L., Kingston D. G. I., J. Nat. Prod., 58, 459-462 (1995).
After curation	VFIIVOHWCNHINZ-UHFFFAOYSA-N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

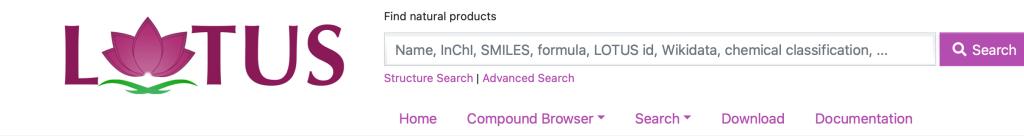
# The initiative - Where?



# The initiative - Where?



# The initiative - Where?



Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural\_products\_occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at [Wikidata](#) and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

## Component Browser

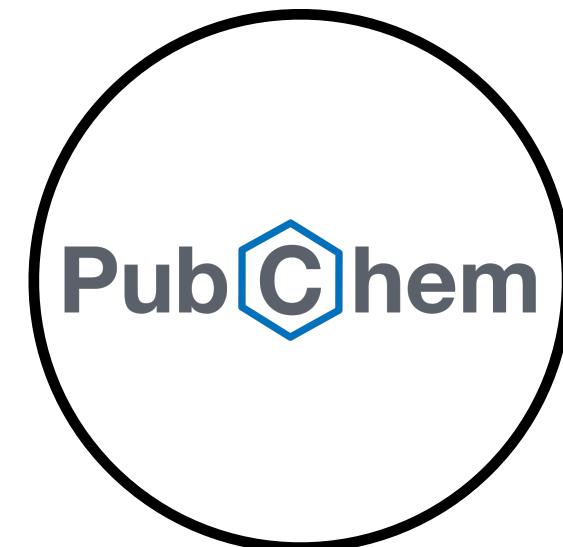
Cards Table

There are 276518 natural products in the database

« < 1 2 3 4 5 6 7 8 9 10 ... 11521 > »

<b>Q105387204</b> 3,4-dihydroxy-5-(hydroxymethyl)-5h-furan-2-one Mol. formula C5H8O5 Mol. weight 146.1 Tmp. LOTUS id LTS0249032	<b>Q27102265</b> Lysopine Mol. formula C9H18N2O4 Mol. weight 218.25 Tmp. LOTUS id LTS0160430	<b>Q105387202</b> (1r,8s,9r,10r,12r)-9,10,12-trihydro-12-(5-hydroxy-2h-furan-3-ylmethyl)-2-oxatricyclo[6.3.1.0 <sup>1,5</sup> ]dodec-3-en-3-one Mol. formula C20H26O4 Mol. weight 330.42 Tmp. LOTUS id LTS0145658	<b>Q105387201</b> (1s,2s,7s,8s,10s,11s,14s,15r,16s,17s,18s,20s,23s)-7-hydroxy-10,14,16,20-tetramethyl-10,14,16,20-azahexacyclo[12.10.0.0 <sup>1,10</sup> .0 <sup>1,12</sup> .0 <sup>1,13</sup> ]tetracos-4-en-18-yl acetate Mol. formula C29H45NO3 Mol. weight 455.67 Tmp. LOTUS id LTS0165716
<b>Q105387200</b> 7-hydroxy-10,14,16,20-tetrahydro-10,14,16,20-azahexacyclo[12.10.0.0 <sup>1,10</sup> .0 <sup>1,12</sup> .0 <sup>1,13</sup> ]tetracos-4-en-18-yl acetate Mol. formula C29H45NO3 Mol. weight 455.67 Tmp. LOTUS id LTS0137374	<b>LTS0044532</b> (4z,8o)-4,7,7-trimethyl-11-methoxy-11-oxacycloneda-4,8-dien-1-ene Mol. formula C15H22O Mol. weight 218.34 Tmp. LOTUS id LTS0044532	<b>Q105387197</b> 6-[4-(4,6-dihydroxy-7,12,16-trihydro-16,17-dihydro-17,18-octamethylpentacyclo[9.7.0.0 <sup>1,10</sup> .0 <sup>1,12</sup> .0 <sup>1,13</sup> ]tetradecan-15-yl]-2-methylhept-2-en-4-one Mol. formula C30H48O8 Mol. weight 456.7 Tmp. LOTUS id LTS0237245	<b>Q105387196</b> (6r)-6-[11s,12s,14s,15s,16s,18s,19s,20s]-11,12,16,17,18,19,20-heptamethylpentacyclo[9.7.0.0 <sup>1,10</sup> .0 <sup>1,12</sup> .0 <sup>1,13</sup> ]octadecan-15-yl-2-methylhept-2-en-4-one Mol. formula C30H48O3 Mol. weight 456.7 Tmp. LOTUS id LTS0074160

# The initiative - Where?



**LOTUS**

Find natural products  
Name, InChI, SMILES, formula, LOTUS id, Wikidata, chemical classification, ...  
Structure Search | Advanced Search

Home Compound Browser Search Download Documentation

Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural\_products occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at [Wikidata](#) and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

## Component Browser

Cards Table

There are 276518 natural products in the database

« < 1 2 3 4 5 6 7 8 9 10 ... 11521 > »

<b>Q105387204</b> 3,4-dihydroxy-5-(hydroxymethyl)-5h-furan-2-one Mol. formula C5H8O5 Mol. weight 146.1 Tmp. LOTUS id LTS0249032	<b>Q27102265</b> Lysopine Mol. formula C9H18N2O4 Mol. weight 218.25 Tmp. LOTUS id LTS0160430	<b>Q105387202</b> (1r,8s,9r,10r,12r)-9,10,12-trihydro-12-(5-hydroxy-2-furan-3-ylmethyl)-2-oxatricyclo[6.3.1.0 <sup>1,5</sup> ]dodec-1-en-3-one Mol. formula C20H26O4 Mol. weight 330.42 Tmp. LOTUS id LTS0145658	<b>Q105387201</b> (1s,2s,7s,8s,10s,11s,14s,15r,16s,17s,18s,19s,20s,23s)-7-hydroxy-10,14,16,20-tetramethyl-10,14,16,20-azahexacyclo[12.10.0.0 <sup>1,10</sup> .0 <sup>1,9</sup> .0 <sup>1,8</sup> .0 <sup>1,7</sup> ]triaconta-4-en-18-yl acetate Mol. formula C29H45NO3 Mol. weight 456.67 Tmp. LOTUS id LTS0165716
<b>Q105387200</b> 7-hydroxy-10,14,16,20-tertiary-17,18,19,20-azahexacyclo[12.10.0.0 <sup>1,10</sup> .0 <sup>1,9</sup> .0 <sup>1,8</sup> .0 <sup>1,7</sup> ]tetraconta-4-en-18-yl acetate Mol. formula C29H45NO3 Mol. weight 456.67 Tmp. LOTUS id LTS0137374	<b>LTS0044532</b>  <h2>PubChem Classification Browser</h2>		

Browse PubChem data using a classification of interest, or search for PubChem records, phenylpropionates, or Gene Ontology: DNA repair. More...

Select classification      Search

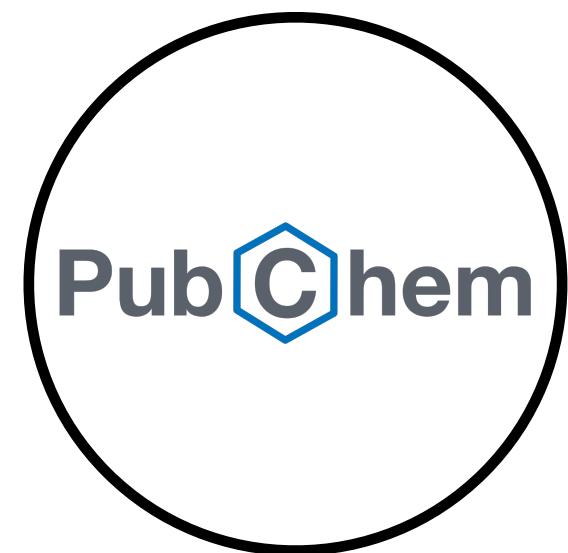
**LOTUS - the natural products occurrence database**      Key

Classification description (from LOTUS - the natural products occurrence database)  
Biological and chemical tree provided by the LOTUS (natural products occurrence database)  
Created on 09/10/2022 09:05:48 More...

Data type counts to display      Display zero count nodes?  
None Compound Taxonomy Yes No



# The initiative - Where?



**LOTUS**

Find natural products  
Name, InChI, SMILES, formula, LOTUS id, Wikidata, chemical classification, ...  
Structure Search | Advanced Search

Home Compound Browser Search Download Documentation

Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural\_products\_occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at [Wikidata](#) and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

## Component Browser

Cards Table

There are 276518 natural products in the database

« < 1 2 3 4 5 6 7 8 9 10 ... 11521 > »

Q105387204 3,4-dihydroxy-5-(hydroxymethyl)-5h-furan-2-one Mol. formula C5H8O5 Mol. weight 146.1 Tmp. LOTUS id LTS0249032	Q27102265 Lysopeine Mol. formula C9H18N2O4 Mol. weight 218.25 Tmp. LOTUS id LTS0160430	Q105387202 (1r,8s,9r,10r,12r)-9,10,12-trihydroxy-12-(5-hydroxy-2h-furan-3-ylmethyl)-2-oxatricyclo[6.3.1.0 <sup>1,7</sup> ]dodec-4-en-3-one Mol. formula C20H26O4 Mol. weight 330.42 Tmp. LOTUS id LTS0145658	Q105387201 (1s,2s,7s,8s,10s,11s,14s,15r,15s,17s,18s,19s,20s,23s)-7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[6.3.1.0 <sup>1,7</sup> ]dodec-4-en-18-yl acetate Mol. formula C29H45NO3 Mol. weight 456.67 Tmp. LOTUS id LTS0165716
Q105387200 7-hydroxy-10,14,16,20-tetrahydro-12-azahexacyclo[12.0.0.0 <sup>1,7</sup> ]tetradec-4-en-18-yl acetate Mol. formula C29H45NO3 Mol. weight 456.67 Tmp. LOTUS id LTS0137374	LTS0044532   	Q105387197 (4z,8e)-4,7,7-trimethyl-11-methoxy-11-oxacycloneda-4,8-dien-1-ene Mol. formula C15H22O Mol. weight 218.34 Tmp. LOTUS id LTS0044532	Q105387196 (6r)-6,10r,11s,12s,15s,16r,17s,18s,19s,20s,23s)-7,12,16-tetramethylpentacyclo[9.7.0.0 <sup>1,7</sup> .0 <sup>1,9</sup> .0 <sup>1,11</sup> ]octadecan-15-yl-2-methylhept-2-en-4-one Mol. formula C30H48O3 Mol. weight 456.7 Tmp. LOTUS id LTS0237245

## PubChem Classification Browser

Browse PubChem data using a classification of interest, or search for PubChem records by compound name, structure, or Gene Ontology: DNA repair. More...

Select classification

**LOTUS - the natural products occurrence database**

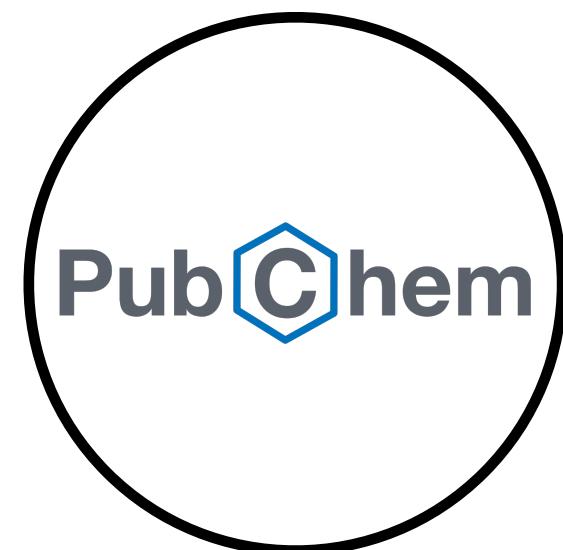
Classification description (from LOTUS - the natural products occurrence database)  
Biological and chemical tree provided by the LOTUS (natural products occurrence database)  
Created on 09/10/2022 09:05:48 More...

Data type counts to display  None Compound Taxonomy Yes No

### Browse LOTUS - the natural products occurrence database Tree

- ▼ LOTUS Tree
- ▼ Biological Tree 
  - ▶ Archaea
  - ▶ Bacteria
  - ▶ Eukaryota

# The initiative - Where?



**LOTUS**

Find natural products  
Name, InChI, SMILES, formula, LOTUS id, Wikidata, chemical classification, ...  
Structure Search | Advanced Search

Home Compound Browser Search Download Documentation

Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural\_products\_occurrence\_database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at [Wikidata](#) and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

## Component Browser

Cards Table

There are 276518 natural products in the database

« < 1 2 3 4 5 6 7 8 9 10 ... 11521 > »

Q105387204 3,4-dihydroxy-5-(hydroxymethyl)-5h-furan-2-one Mol. formula C5H8O5 Mol. weight 146.1 Tmp. LOTUS id LTS0249032	Q27102265 Lysopeine Mol. formula C9H18N2O4 Mol. weight 218.25 Tmp. LOTUS id LTS0160430	Q105387202 (1r,8s,9r,10r)-9,10,12,13-tetrahydro-12-hydroxy-2h-furan-3-ylmethoxy-2-oxatricyclo[6.3.1.0 <sup>1,5</sup> ]dodec-4-en-3-one Mol. formula C20H26O4 Mol. weight 330.42 Tmp. LOTUS id LTS0145658	Q105387201 (1r,2s,7s,8s,10r)-9,10,12,13-tetrahydro-12-hydroxy-2h-furan-3-ylmethoxy-2-oxatricyclo[6.3.1.0 <sup>1,5</sup> ]dodec-4-en-3-one Mol. formula C20H26O4 Mol. weight 330.42 Tmp. LOTUS id LTS0145658
Q105387200 7-hydroxy-10,14,16,20-terpenyl-11,15-dihydro-azahexacyclo[12.0.0.0 <sup>1,5</sup> ]tetradec-4-en-18-yl acetate Mol. formula C29H45NO3 Mol. weight 456.67 Tmp. LOTUS id LTS0137374	LTS0044532  LTS0044532	Q105387197 6-(4,6-dihydroxy-7,7,12,16-tetrahydro-1H,5H-cycloocta[9,7,6]trideca-4,8-dien-1-yl)-15-yl-2-methylhept-2-en-4-one Mol. formula C15H22O Mol. weight 218.34 Tmp. LOTUS id LTS0044532	Q105387196 (6r)-6-(4,6-dihydroxy-7,7,12,16-tetrahydro-1H,5H-cycloocta[9,7,6]trideca-4,8,11s,12s,15s,16r)-14,16-dihydroxy-7,7,12,16-tetramethylpentacyclo[9.7.0.0 <sup>1,5</sup> ]octadecan-15-yl-2-methylhept-2-en-4-one Mol. formula C30H48O3 Mol. weight 456.7 Tmp. LOTUS id LTS0237245

## PubChem Classification Browser

Browse PubChem data using a classification of interest, or search for PubChem records by name, InChI, SMILES, formula, LOTUS id, Wikidata, chemical classification, ...

Select classification

**LOTUS - the natural products occurrence database**

Classification description (from LOTUS - the natural products occurrence database)  
Biological and chemical tree provided by the LOTUS (natural products occurrence database)  
Created on 09/10/2022 09:05:48 More...

Data type counts to display  None Compound Taxonomy Yes No

### Browse LOTUS - the natural products occurrence database Tree

- ▼ LOTUS Tree
- ▼ Biological Tree 
  - ▶ Archaea
  - ▶ Bacteria
  - ▶ Eukaryota
- ▼ Chemical Tree 
  - ▶ Alkaloids
  - ▶ Amino acids and Peptides
  - ▶ Carbohydrates
  - ▶ Fatty acids
  - ▶ Polyketides
  - ▶ Shikimates and Phenylpropanoids
  - ▶ Terpenoids
- Not classified

# The initiative - Where?



**LOTUS**

Find natural products  
Name, InChI, SMILES, formula, LOTUS id, Wikidata, chemical classification, ...  
Search

Structure Search | Advanced Search

Home Compound Browser Search Download Documentation

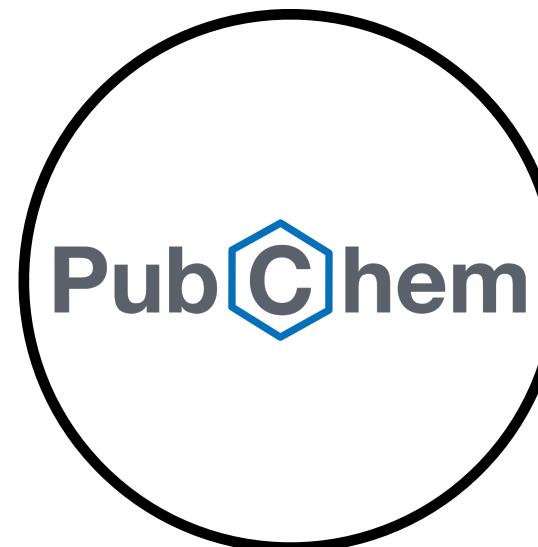
Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural\_products occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at [Wikidata](#) and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

## Component Browser

Cards Table

There are 276518 natural products in the database

Q105387204 3,4-dihydroxy-5-(hydroxymethyl)-5-hydroxyfuran-2-one Mol. formula C5H8O5 Mol. weight 146.1 Tmp. LOTUS id LTS0249032	Q27102265 Lysoine Mol. formula C9H18N2O4 Mol. weight 218.25 Tmp. LOTUS id LTS0160430	Q105387202 (1 <i>r</i> ,8 <i>s</i> ,9 <i>t</i> ,10 <i>u</i> )-9,10,12-trihydroxy-12 <i>r</i> -12 <i>s</i> -2 <i>r</i> -2 <i>s</i> -2 <i>u</i> -furan-3-ylmethoxy-2-oxatricyclo[6.3.1. <sup>0,7</sup> ]dodec-4-en-3-one Mol. formula C20H26O4 Mol. weight 330.42 Tmp. LOTUS id LTS0145658	Q105387201 (1 <i>s</i> ,2 <i>s</i> ,7 <i>s</i> ,8 <i>u</i> )-8 <i>u</i> -11 <i>s</i> -14 <i>u</i> -15 <i>r</i> -16 <i>s</i> -17 <i>u</i> -furan-3-ylmethoxy-2-oxatricyclo[6.3.1. <sup>0,7</sup> ]dodec-4-en-3-one Mol. formula C20H26O4 Mol. weight 330.42 Tmp. LOTUS id LTS0165716
Q105387200 7-hydroxy-10,14,16,20-tetrahydro-17,18-dimethylazahexacyclo[12.0.0.0 <sup>10,12</sup> ]tetradec-4-ene-18-yl acetate Mol. formula C29H45NO3 Mol. weight 456.67 Tmp. LOTUS id LTS0137374	LTS0044532 	Q105387197 (4 <sub>z</sub> ,8 <sub>o</sub> )-4,7,7-trimethyl-11-methoxy-11 <i>o</i> -cycloclaudeca-4,8-dien-1-ol Mol. formula C15H22O Mol. weight 218.34 Tmp. LOTUS id LTS0044532	Q105387196 6-(4,6-dihydroxy-7,7,12,16-tetrahydro-1 <i>o</i> -cycloocto[9.7.0.0 <sup>10,12</sup> ]octacos-4-eno-15-yl)-2-methylhept-2-en-4-one Mol. formula C30H48O3 Mol. weight 456.7 Tmp. LOTUS id LTS0237245



## PubChem Classification Browser

Browse PubChem data using a classification of interest, or search for PubChem records by phenylpropionates, or Gene Ontology: DNA repair. More...

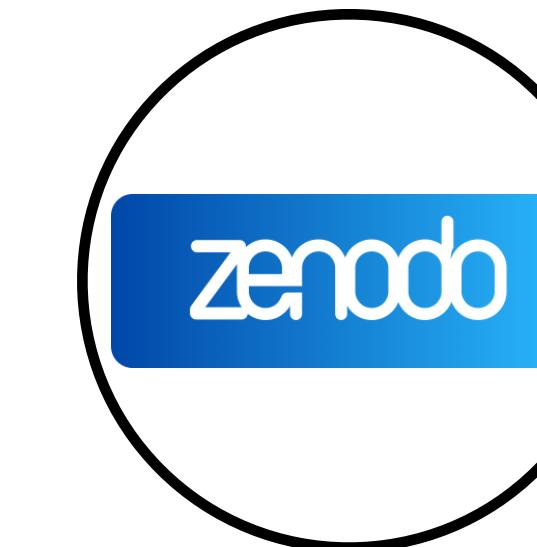
Select classification LOTUS - the natural products occurrence database Search

Classification description (from LOTUS - the natural products occurrence database)  
Biological and chemical tree provided by the LOTUS (natural products occurrence database)  
Created on 09/10/2022 09:05:48 More...

Data type counts to display Compound Display zero count nodes? Yes No

## Browse LOTUS - the natural products occurrence database Tree

- ▼ LOTUS Tree 214,481
  - ▼ Biological Tree 192,579
    - Archaea 72
    - Bacteria 17,467
    - Eukaryota 175,387
  - ▼ Chemical Tree 214,481
    - Alkaloids 28,065
    - Amino acids and Peptides 13,012
    - Carbohydrates 2,144
    - Fatty acids 10,266
    - Polyketides 27,981
    - Shikimates and Phenylpropanoids 42,177
    - Terpenoids 97,499
  - Not classified 7,328



zenodo

Search View Upload Communities Log in Sign up

## The LOTUS Initiative

### Recent uploads

Search The LOTUS Initiative View

September 16, 2022 (v4) Figure Open Access The LOTUS Initiative for Open Natural Products Research: TMAP View Rutz, Adriano; Gaudry, Amélie; TMAP of the compounds present on Wikidata curated in the frame of the LOTUS Initiative: https://doi.org/10.7554/eLife.70780 Uploaded on September 16, 2022

3 more version(s) exist for this record

September 16, 2022 (v4) Dataset Open Access The LOTUS Initiative for Open Natural Products Research: frozen dataset View Rutz, Adriano; Bisson, Jonathan; Allard, Pierre-Marie; Dataset used in the frame of the LOTUS Initiative: https://doi.org/10.7554/eLife.70780 Uploaded on September 16, 2022

5 more version(s) exist for this record

September 16, 2022 (v4) Dataset Open Access The LOTUS Initiative for Open Natural Products Research: waste to recycle View Rutz, Adriano; Bisson, Jonathan; Allard, Pierre-Marie; Dataset uploaded to Wikidata. Generated in the frame of the LOTUS Initiative: https://doi.org/10.7554/eLife.70780 Shared for further curation. Uploaded on September 16, 2022

4 more version(s) exist for this record

September 16, 2022 (v4) Dataset Open Access The LOTUS Initiative for Open Natural Products Research: biological and chemical trees View Rutz, Adriano; Bisson, Jonathan; Allard, Pierre-Marie; Biological and chemical trees made from frozen metadata (10.5281/zenodo.5794106) (for example, for PubChem) Uploaded on September 16, 2022

6 more version(s) exist for this record

September 16, 2022 (v4) Dataset Open Access The LOTUS Initiative for Open Natural Products Research: frozen dataset union wikidata (with metadata) View Rutz, Adriano; Bisson, Jonathan; Allard, Pierre-Marie; Dataset present on Wikidata used in the frame of the LOTUS Initiative: https://doi.org/10.7554/eLife.70780 Uploaded on September 16, 2022

8 more version(s) exist for this record

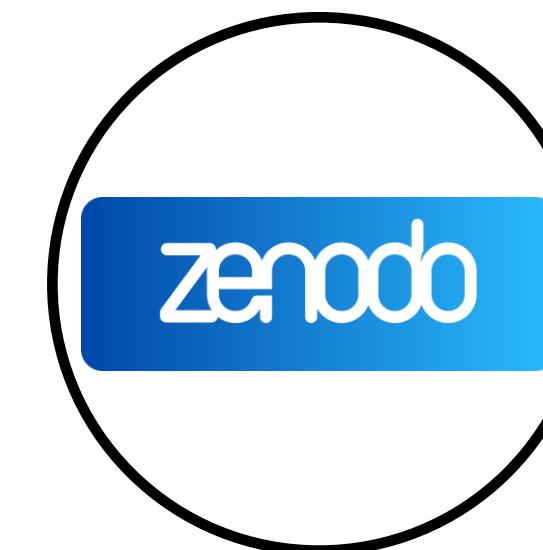
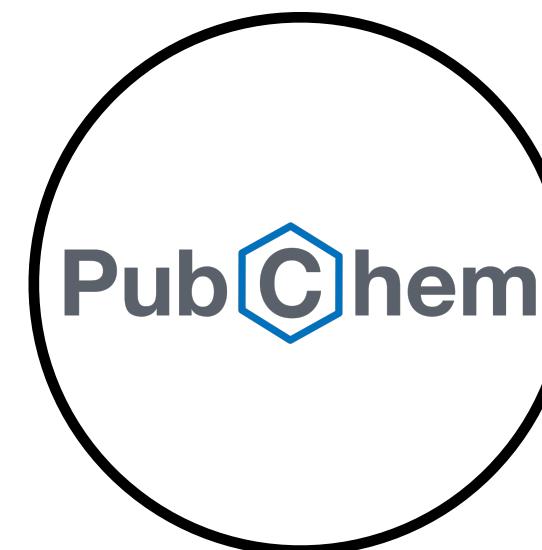
September 16, 2022 (v4) Dataset Open Access The LOTUS Initiative for Open Natural Products Research: metadata View Rutz, Adriano; Bisson, Jonathan; Allard, Pierre-Marie; Metadata of each of the three objects (structures, organisms, references) used in the frame of the LOTUS Initiative: https://doi.org/10.7554/eLife.70780 Uploaded on September 16, 2022

3 more version(s) exist for this record

Want your upload to appear in this community?  
Click the button above to upload a record directly to this community. To add one of your existing records to the community, edit the record, add this community under the "Communities" section, save, and refresh.

- The community curator will then be notified to either accept or reject your upload (see community curation policy below).
- If your upload is rejected by the curator, it will still be available on Zenodo, just not in this community.

# The initiative - Where?



**LOTUS**

Find natural products  
Name, InChI, SMILES, formula, LOTUS id, Wikidata, chemical classification, ...  
Structure Search | Advanced Search

Search

Home Compound Browser Search Download Documentation

Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural products occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at Wikidata and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

Component Browser

Cards Table

There are 276518 natural products in the database

Q105387204 3,4-dihydroxy-5-(hydroxymethyl)-5h-furan-2-one  
Mol. formula C5H8O5  
Mol. weight 146.1  
Tmp. LOTUS id LTS0249032

Q27102265 Lysopine  
Mol. formula C9H18N2O4  
Mol. weight 218.25  
Tmp. LOTUS id LTS0160430

Q105387202 3,4-dihydroxy-5-(2-hydroxy-3-oxo-2-oxazacyclo[6.3.1.0<sup>1,6</sup>]dodec-4-en-3-one)  
Mol. formula C20H26O4  
Mol. weight 330.42  
Tmp. LOTUS id LTS0145658

Q105387201 1(1a,2s,9r,10r)-2-(9,10,12-trihydroxy-12-oxo-2h-furan-3-yl)ethyl-2-oxatricyclo[6.3.1.0<sup>1,6</sup>]dodec-4-en-3-one  
Mol. formula C20H26O4  
Mol. weight 330.42  
Tmp. LOTUS id LTS0145658

Q105387200 7-hydroxy-10,14,16,20-tetrahydro-17,18-dimethyl-azahexacyclo[12.10.0.0<sup>1,6</sup>]tetra-10,14,16,20-tetra-17,18-yl acetate  
Mol. formula C29H45NO3  
Mol. weight 456.67  
Tmp. LOTUS id LTS0137374

LTS0044532  
Q105387197 6-(4,6-dihydroxy-7,7,12,16-tetrahydro-17,18-dimethylcycloneda-4,8-dien-1-yl)-15-yl-2-methylhept-2-en-4-one  
Mol. formula C15H22O  
Mol. weight 218.34  
Tmp. LOTUS id LTS0044532

Q105387196 6-(4,6-dihydroxy-7,7,12,16-tetrahydro-17,18-dimethylcycloneda-4,8-dien-1-yl)-15-yl-2-methylhept-2-en-4-one  
Mol. formula C15H22O  
Mol. weight 218.34  
Tmp. LOTUS id LTS0044532

Not classified 7,328

## PubChem Classification Browser

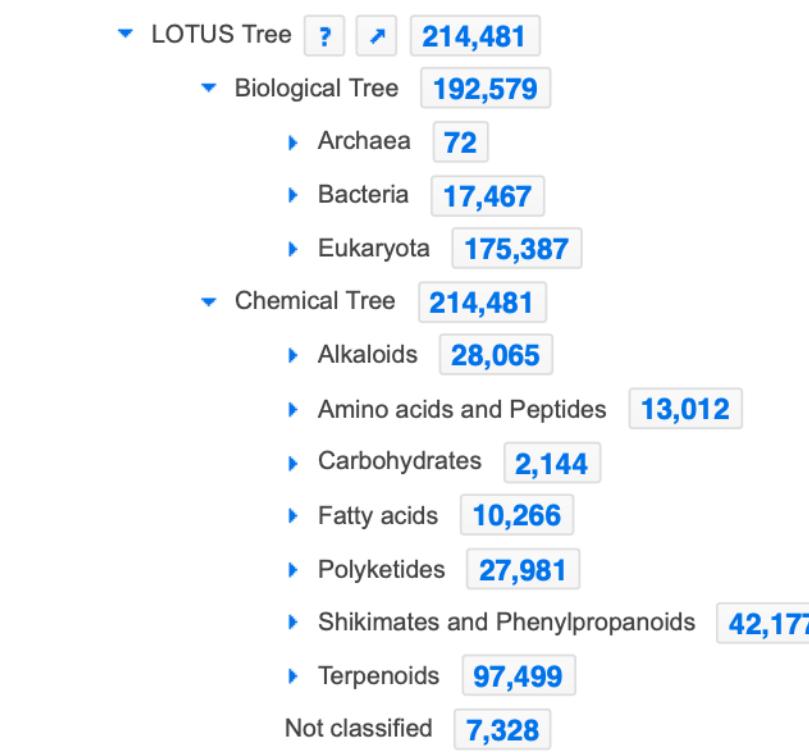
Browse PubChem data using a classification of interest, or search for PubChem records, phenylpropionates, or Gene Ontology: DNA repair. More...

Select classification

Classification description (from LOTUS - the natural products occurrence database)  
Biological and chemical tree provided by the LOTUS (natural products occurrence database)  
Created on 09/10/2022 09:05:48 More...

Data type counts to display

## Browse LOTUS - the natural products occurrence database Tree



zenodo

The LOTUS Initiative

Recent uploads

Search The LOTUS Initiative

September 16, 2022 (v1) Figure Open Access The LOTUS Initiative for Open Natural Products Research: TMAP  
Rutz, Adriano; Gaudry, Amélie; TMAP of the compounds present on Wikidata curated in the frame of the LOTUS Initiative: https://doi.org/10.7554/elife.70780  
Uploaded on September 16, 2022

3 more version(s) exist for this record

September 16, 2022 (v1) Dataset Open Access The LOTUS Initiative for Open Natural Products Research: frozen dataset  
Rutz, Adriano; Bisson, Jonathan; Allard, Pierre-Marie; Dataset used in the frame of the LOTUS Initiative: https://doi.org/10.7554/elife.70780  
Uploaded on September 16, 2022

5 more version(s) exist for this record

September 16, 2022 (v1) Dataset Open Access The LOTUS Initiative for Open Natural Products Research: waste to recycle  
Rutz, Adriano; Bisson, Jonathan; Allard, Pierre-Marie; Dataset not uploaded to Wikidata. Generated in the frame of the LOTUS Initiative: https://doi.org/10.7554/elife.70780  
Shared for further curation.  
Created: October 28, 2021  
Harvesting API: OAI-PMH interface

Want your upload to appear in this community?  
Click the button above to upload a record directly to this community. To add one of your records to the community, edit the record, add this community under the "Communities" section, save, and refresh.

The community curator will then be notified to either accept or reject your upload (see community curation policy below). If your upload is rejected by the curator, it will still be available on Zenodo, just not in this community.

September 16, 2022 (v1) Dataset Open Access The LOTUS Initiative for Open Natural Products Research: biological and chemical trees  
Rutz, Adriano; Bisson, Jonathan; Allard, Pierre-Marie; Biological and chemical trees made from frozen metadata (10.5281/zenodo.5794106) (for example, for PubChem)  
Uploaded on September 16, 2022

6 more version(s) exist for this record

September 16, 2022 (v1) Dataset Open Access The LOTUS Initiative for Open Natural Products Research: frozen dataset union wikidata (with metadata)  
Rutz, Adriano; Bisson, Jonathan; Allard, Pierre-Marie; Dataset present on Wikidata used in the frame of the LOTUS Initiative: https://doi.org/10.7554/elife.70780  
Uploaded on September 16, 2022

8 more version(s) exist for this record

September 16, 2022 (v1) Dataset Open Access The LOTUS Initiative for Open Natural Products Research: metadata  
Rutz, Adriano; Bisson, Jonathan; Allard, Pierre-Marie; Metadata of each of the three objects (structures, organisms, references) used in the frame of the LOTUS Initiative: https://doi.org/10.7554/elife.70780  
Uploaded on September 16, 2022

3 more version(s) exist for this record

## Extraction

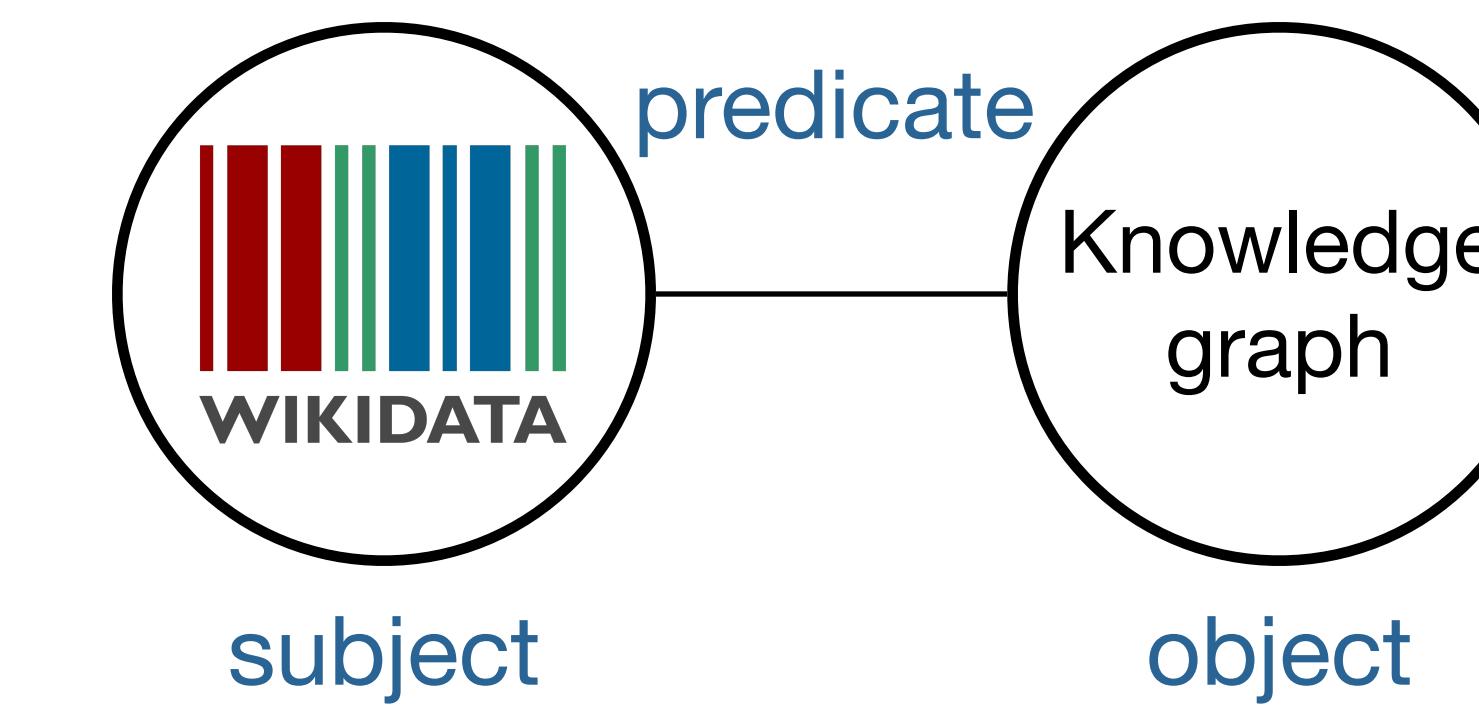
```

SELECT ?desired_items
WHERE {
  VALUES ?classes {
    Chemical compound
  }
  ?item instance of ?classes .
  ?item found in taxon ?stmt .
  ?stmt found in taxon ?taxon .
  OPTIONAL {
    ?stmt occurrence ?ref .
    ?ref stated in ?art .
  }
  SERVICE wikibase:label {
    language
    "[AUTO_LANGUAGE],en" .
  }
}
```

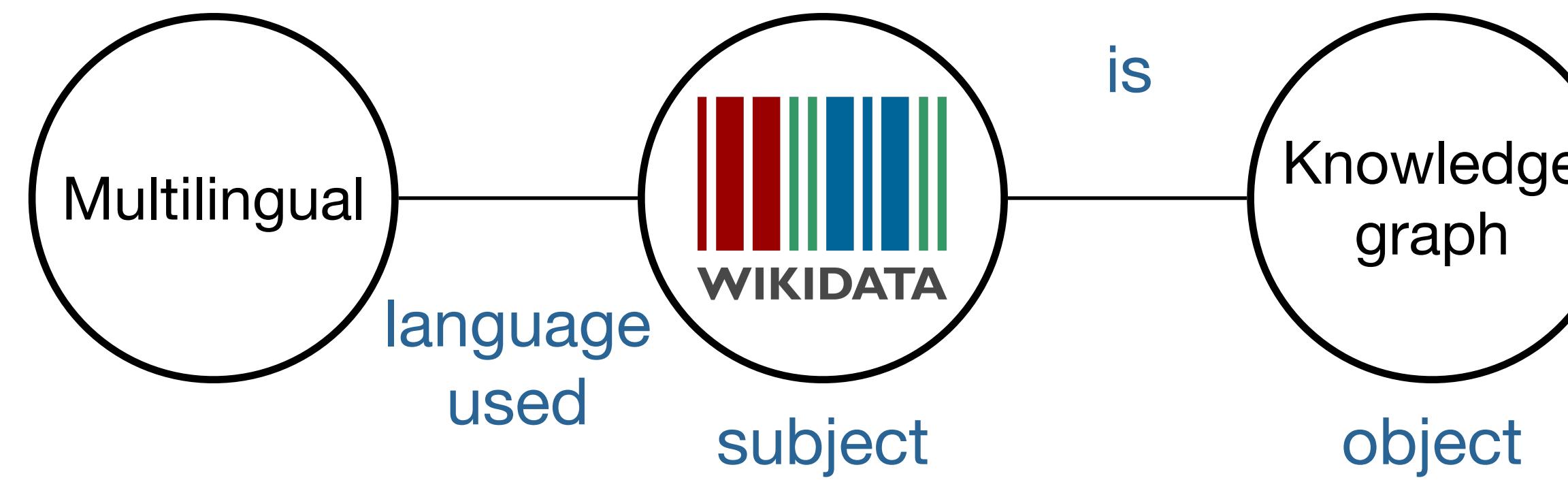
# The initiative - Wikidata



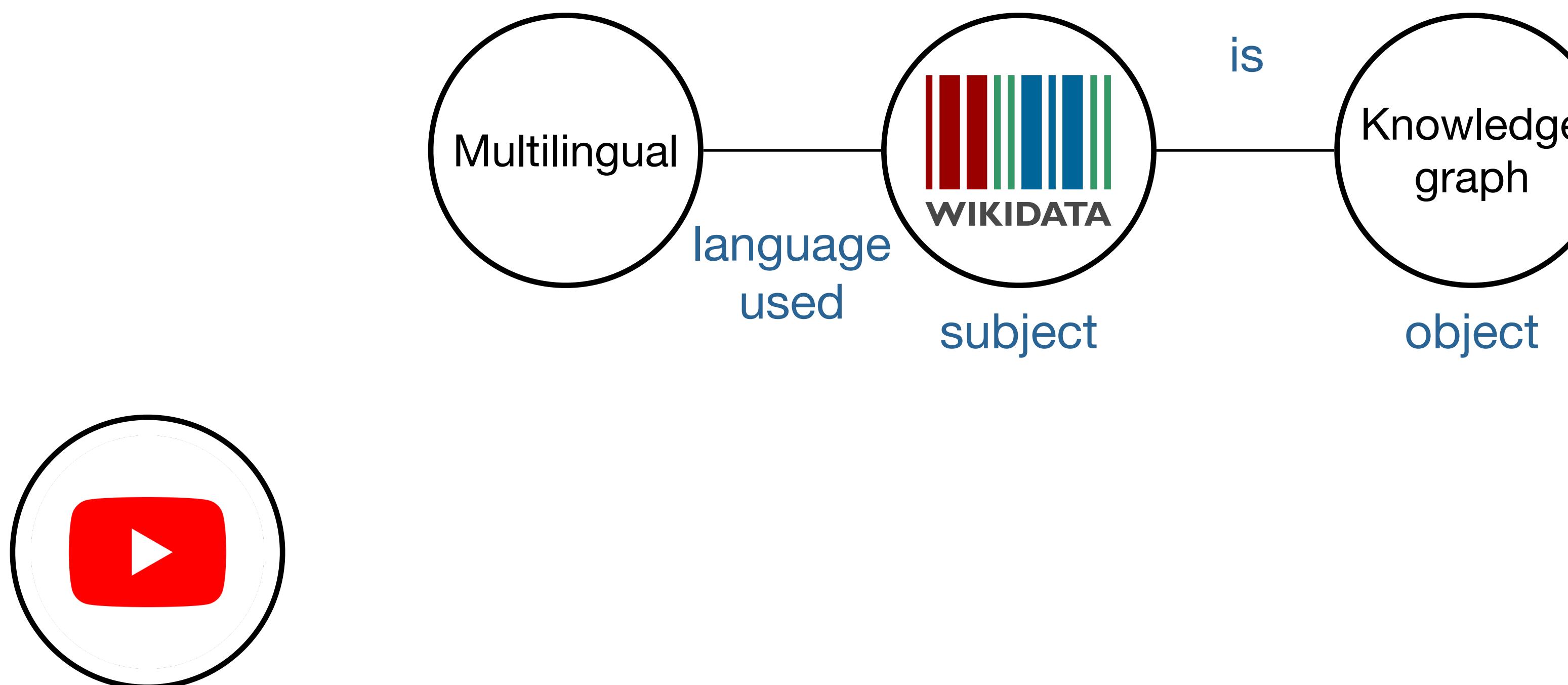
# The initiative - Wikidata



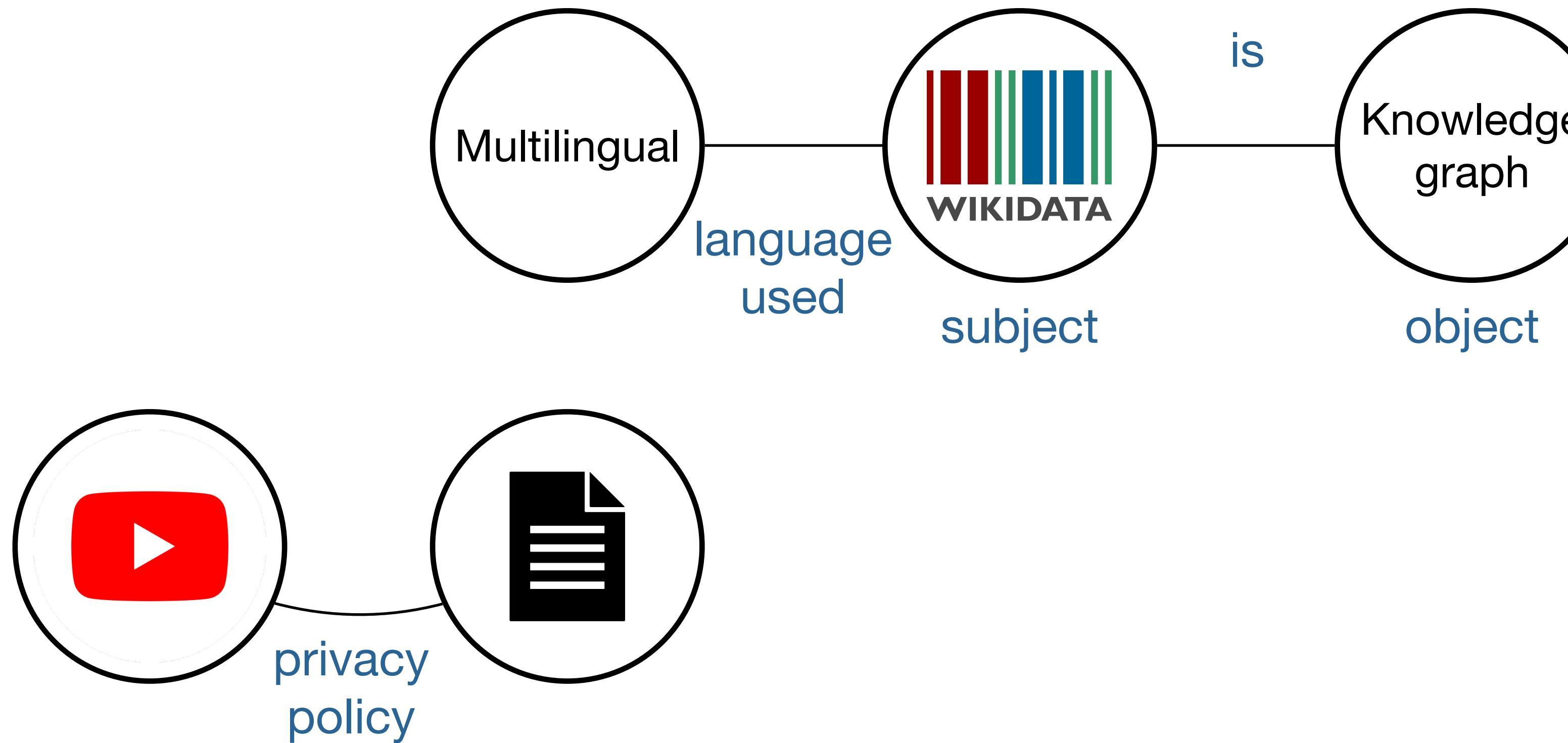
# The initiative - Wikidata



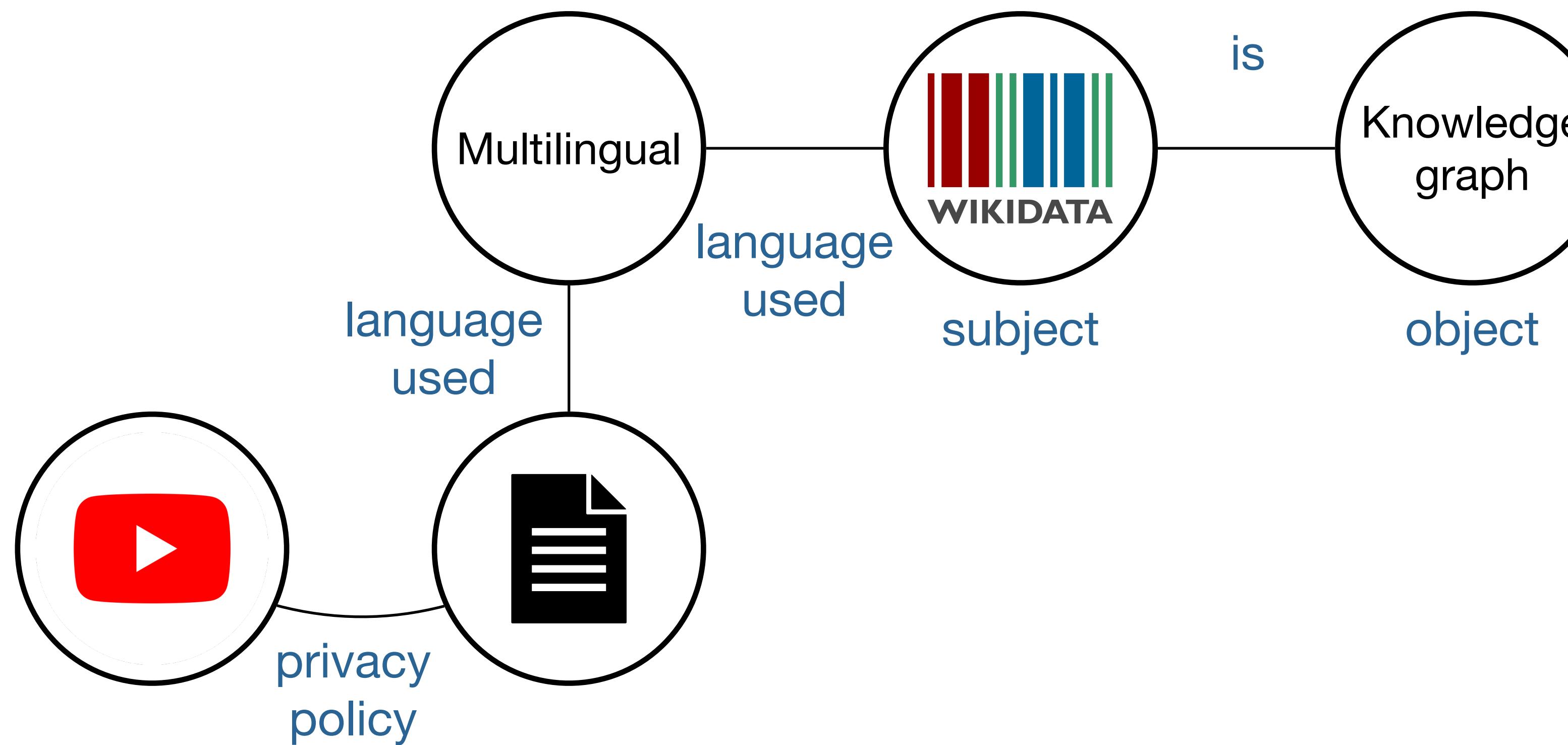
# The initiative - Wikidata



# The initiative - Wikidata



# The initiative - Wikidata



# The initiative - SPARQL

*« Hey Wiki, what are the compounds found in Swertia chirayita? »*



# The initiative - SPARQL

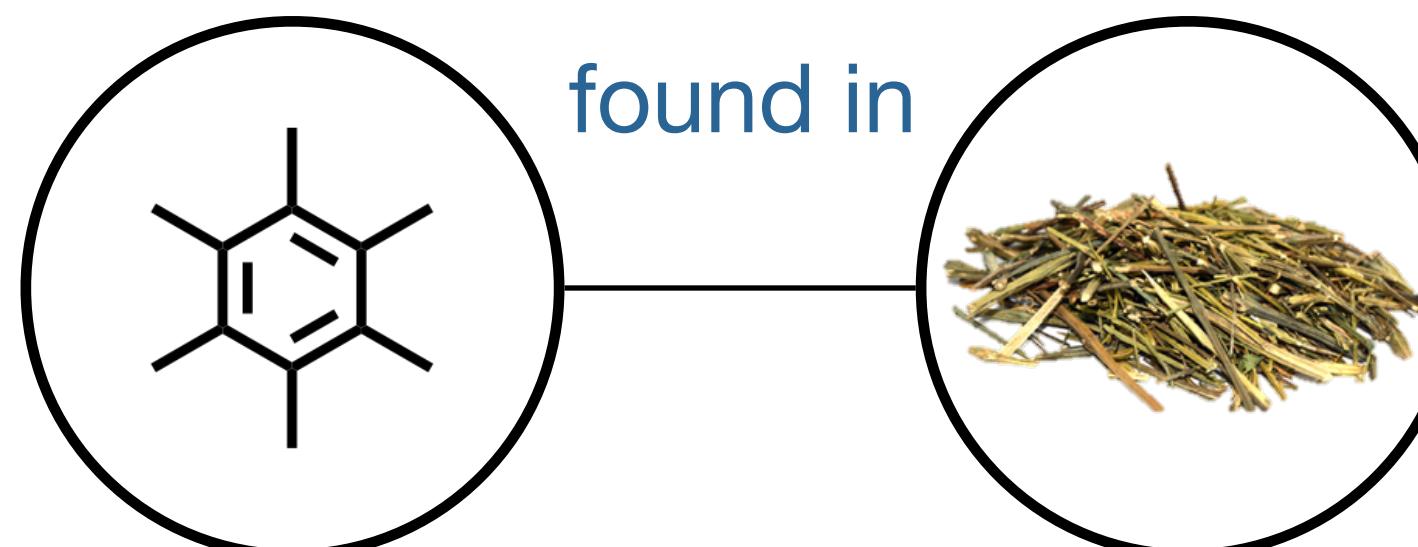
« Hey Wiki, what are the compounds found in *Swertia chirayita*? »



```
SELECT DISTINCT ?structure ?structure_smiles ?  
structure_inchikey  
WHERE {  
VALUES ?organism {  
wd:Q21318003 # Swertia chirayita  
}  
?organism_child (wdt:P171*) ?organism.  
?structure wdt:P233 ?structure_smiles;  
wdt:P235 ?structure_inchikey;  
(p:P703/ps:P703) ?organism_child.  
}
```

# The initiative - SPARQL

« Hey Wiki, what are the compounds found in *Swertia chirayita*? »



```
SELECT DISTINCT ?structure ?structure_smiles ?
  structure_inchikey
WHERE {
  VALUES ?organism {
    wd:Q21318003 # Swertia chirayita
  }
  ?organism_child (wdt:P171*) ?organism.
  ?structure wdt:P233 ?structure_smiles;
    wdt:P235 ?structure_inchikey;
    (p:P703/ps:P703) ?organism_child.
}
```

42 results in 0.4 second

row	structure	structure_smiles	structure_inchikey
1	wd:Q105165035		MIJYXULNPSFWEK-ZZAAMMQTSA-N
...	...		...
42	wd:Q1074417		AEDDIBAIWPIIBD-ZJKJAXBQSA-N

# The initiative - SPARQL

« *Hey Wiki, what are the compounds found in Gentianaceae? »*

1,004 results in 0.8 second

# The initiative - SPARQL

« Hey Wiki, what are the compounds found in Gentianaceae? »

1,004 results in 0.8 second

« Hey Wiki, what are the compounds found in Gentianales? »

10,620 results in 3.3 seconds

# The initiative - SPARQL

« Hey Wiki, what are the compounds found in Gentianaceae? »

1,004 results in 0.8 second

« Hey Wiki, what are the compounds found in Gentianales? »

10,620 results in 3.3 seconds

« Sorry Wiki, I forgot, how many species belong to Gentianales? »

43,611 results in 3.5 seconds

# The initiative - SPARQL

*« Hey Wiki, what are the compounds found in Gentianaceae? »*

1,004 results in 0.8 second

*« Hey Wiki, what are the compounds found in Gentianales? »*

10,620 results in 3.3 seconds

*« Sorry Wiki, I forgot, how many species belong to Gentianales? »*

43,611 results in 3.5 seconds

*« Ok Wiki, among the compounds found in the Gentianales how many were already described as bitter? »*

112 results in 2.7 seconds

# The initiative - SPARQL

*« Ok Wiki, which organisms contain compounds structurally similar to the ones reported as bitter in Swertia chirayita, and how many?»*

# The initiative - SPARQL

*« Ok Wiki, which organisms contain compounds structurally similar to the ones reported as bitter in Swertia chirayita, and how many?»*

```
PREFIX sachem: <http://bioinfo.uochb.cas.cz/rdf/v1.0/sachem#>
PREFIX idsm: <https://idsm.elixir-czech.cz/sparql/endpoint/>
SELECT
    ?taxon
    ?taxon_name
    (COUNT(DISTINCT ?compound) AS ?count)
WHERE {
    SERVICE idsm:wikidata {
        SERVICE <https://query.wikidata.org/bigdata/
namespace/wdq/sparql> {
            VALUES ?organism {
                wd:Q21318003 # Swertia chirayita
            }
            VALUES ?taste {
                wd:Q1517187 # Bitterness
            }
            ?organism_child (wdt:P171*) ?organism.
            ?structure wdt:P233 ?structure_smiles;
            wdt:P235 ?structure_inchikey;
            wdt:P1552 ?taste;
            (p:P703/ps:P703) ?organism_child.
        }
        ?compound sachem:similarCompoundSearch _:b40.
        _:b40 sachem:query ?structure_smiles;
        sachem:cutoff "0.9"^^xsd:double.
    }
    hint:Prior hint:runFirst "true"^^xsd:boolean.
    ?compound wdt:P703 ?taxon.
    ?taxon wdt:P225 ?taxon_name.
}
GROUP BY ?taxon ?taxon_name
ORDER BY DESC (?count)
```

# The initiative - SPARQL

« Ok Wiki, which organisms contain compounds structurally similar to the ones reported as bitter in *Swertia chirayita*, and how many?»

14 results in 2.1 seconds

row	taxon	taxon_name	count
1	wd:Q162579	<i>Gentiana purpurea</i>	3
2	wd:Q11255805	<i>Swertia japonica</i>	3
...	...	...	...
5	wd:Q158572	<i>Gentiana lutea</i>	2
6	wd:Q13859874	<i>Gentianella nitida</i>	2
...	...	...	...
14	wd:Q1074417	<i>Swertia mileensis</i>	1

```

PREFIX sachem: <http://bioinfo.uochb.cas.cz/rdf/v1.0/sachem#>
PREFIX idsm: <https://idsm.elixir-czech.cz/sparql/endpoint/>
SELECT
  ?taxon
  ?taxon_name
  (COUNT(DISTINCT ?compound) AS ?count)
WHERE {
  SERVICE idsm:wikidata {
    SERVICE <https://query.wikidata.org/bigdata/
      namespace/wdq/sparql> {
      VALUES ?organism {
        wd:Q21318003 # Swertia chirayita
      }
      VALUES ?taste {
        wd:Q1517187 # Bitterness
      }
      ?organism_child (wdt:P171*) ?organism.
      ?structure wdt:P233 ?structure_smiles;
        wdt:P235 ?structure_inchikey;
        wdt:P1552 ?taste;
        (p:P703/ps:P703) ?organism_child.
    }
    ?compound sachem:similarCompoundSearch _:b40.
    _:b40 sachem:query ?structure_smiles;
      sachem:cutoff "0.9"^^xsd:double.
  }
  hint:Prior hint:runFirst "true"^^xsd:boolean.
  ?compound wdt:P703 ?taxon.
  ?taxon wdt:P225 ?taxon_name.
}
GROUP BY ?taxon ?taxon_name
ORDER BY DESC (?count)

```

# The initiative - Community curation

English Not logged in Talk Contributions Create account Log in

Item Discussion Read View history Search Wikidata

## hyperelodione D (Q116482353)

chemical compound edit

Language	Label	Description	Also known as
English	hyperelodione D	chemical compound	

### Statements

instance of	chemical compound
-------------	-------------------

0 references

mass	548.797 dalton
------	----------------

1 reference

based on heuristic inferred from SMILES

chemical formula	$C_{36}H_{52}O_4$
------------------	-------------------

1 reference

based on heuristic inferred from SMILES

WIKIDATA

Main page  
Community portal  
Project chat  
Create a new Item  
Recent changes  
Random Item  
Query Service  
Nearby  
Help  
Donate  
  
Lexicographical data  
Create a new Lexeme  
Recent changes  
Random Lexeme  
  
Tools  
What links here  
Related changes  
Special pages  
Permanent link  
Page information  
Concept URI  
Cite this page

# The initiative - Community curation

English Not logged in Talk Contributions Create account Log in

Item Discussion Read View history Search Wikidata

## hyperelodione D (Q116482353)

chemical compound edit

Language	Label	Description	Also known as
English	hyperelodione D	chemical compound	

canonical SMILES C/C(C)=C\CC/C(C)=C/CC1=C2C3(O[H])C(C4(C/C=C(C)/CC/C=C(C)/C)C1=O)(O[H])C(C(C)=CC4)([H])CC3([H])C(C)(C)O2

reason for deprecated rank structural reassignment  
1 reference

stated in Cytotoxic polyprenylated phloroglucinol derivatives from Hypericum elodeoides Choisy modulating the transactivation of RXRa

C/C(C)=C/CC1=C2C3(O[H])C(C4(C/C=C(C)/CC/C=C(C)/C)C1=O)(O[H])C(C(C)=CC4)([H])CC3([H])C(CC/C=C(C)/C)(C)O2

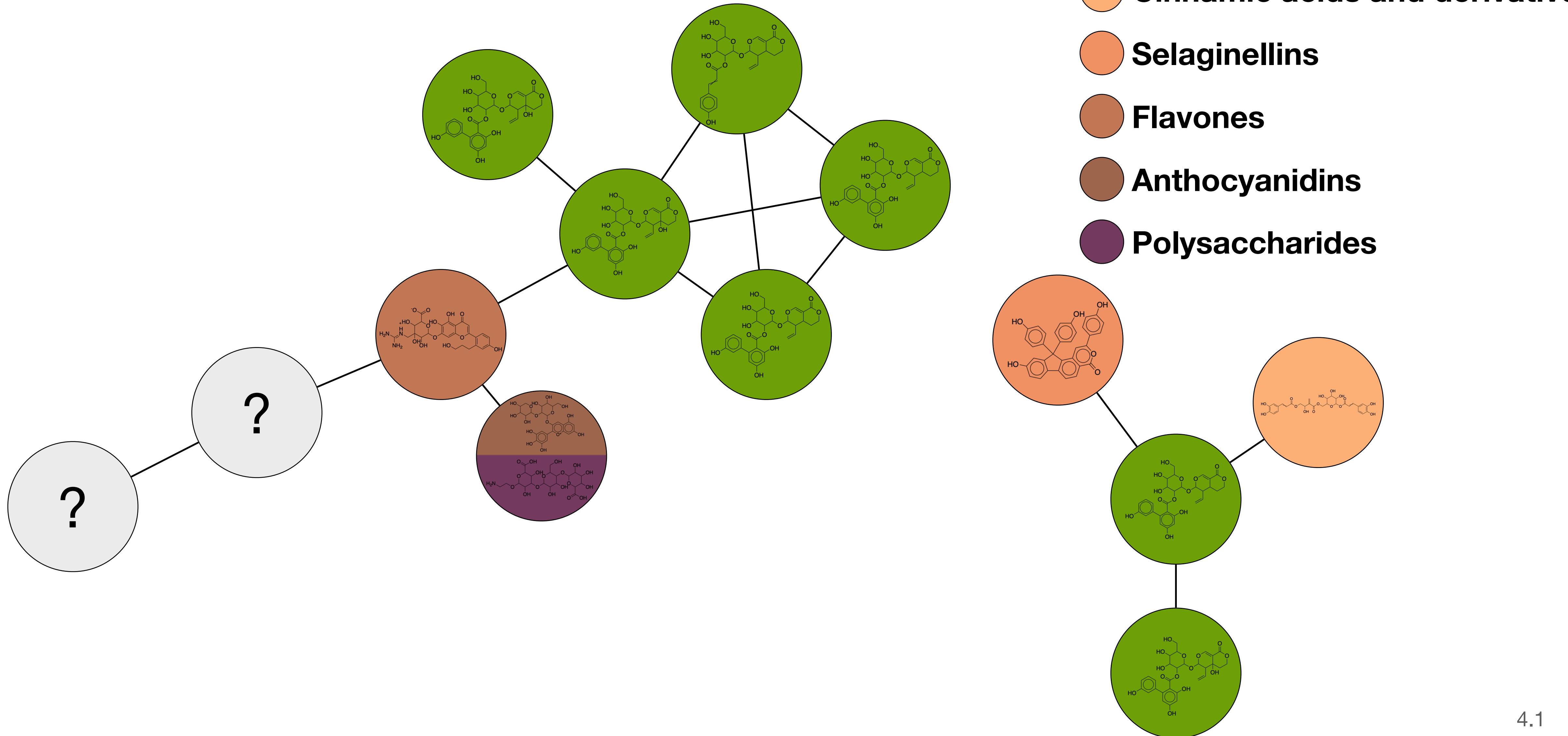
reason for preferred rank structural reassignment  
1 reference

stated in Bioinspired Total Synthesis of Erectones A and B, and the Revised Structure of Hyperelodione D

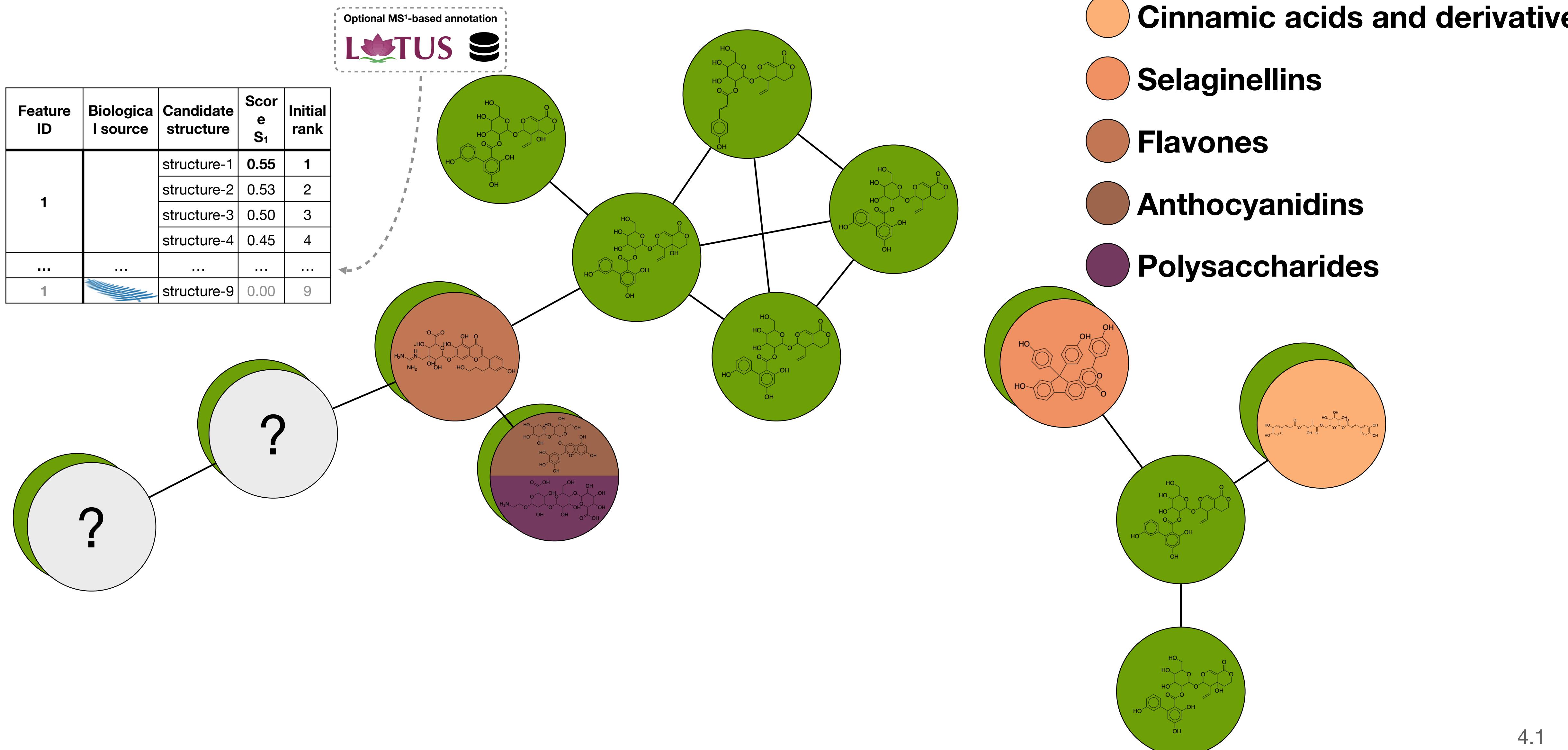
# The initiative - back to metabolites

*« But Wiki, can this help me with my metabolite annotation? »*

# MS<sup>1</sup>-based annotation

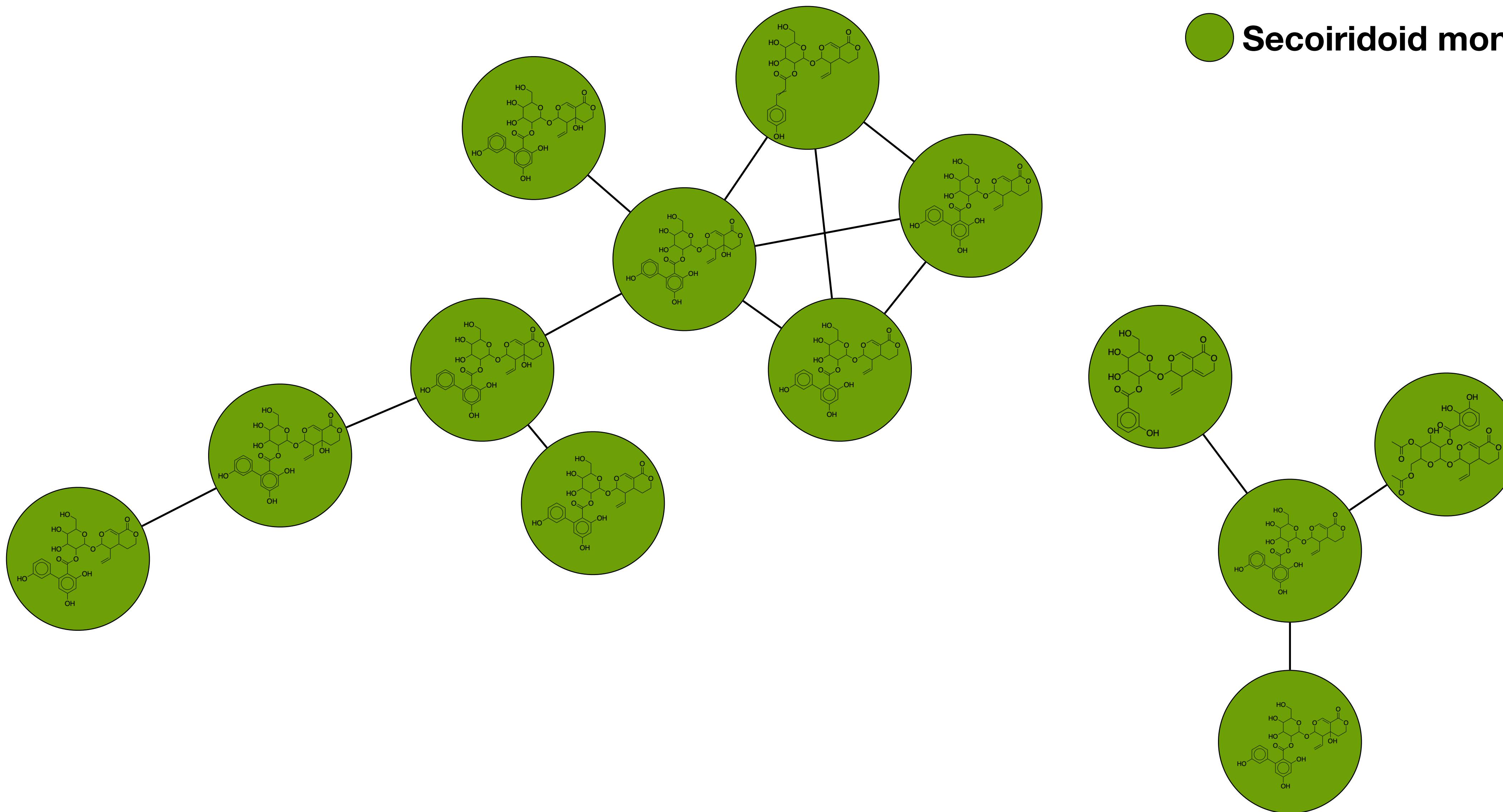


# MS<sup>1</sup>-based annotation

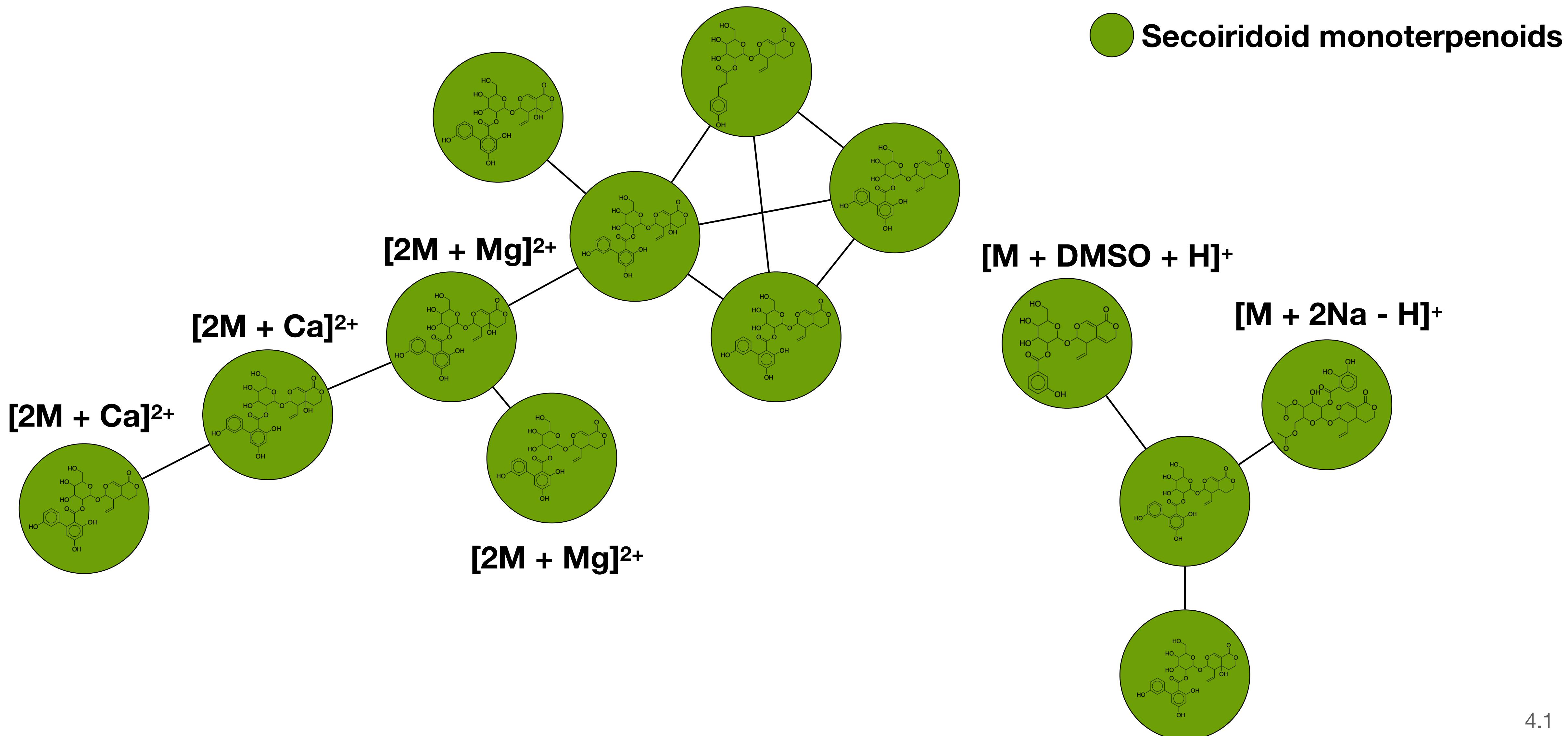


# MS<sup>1</sup>-based annotation

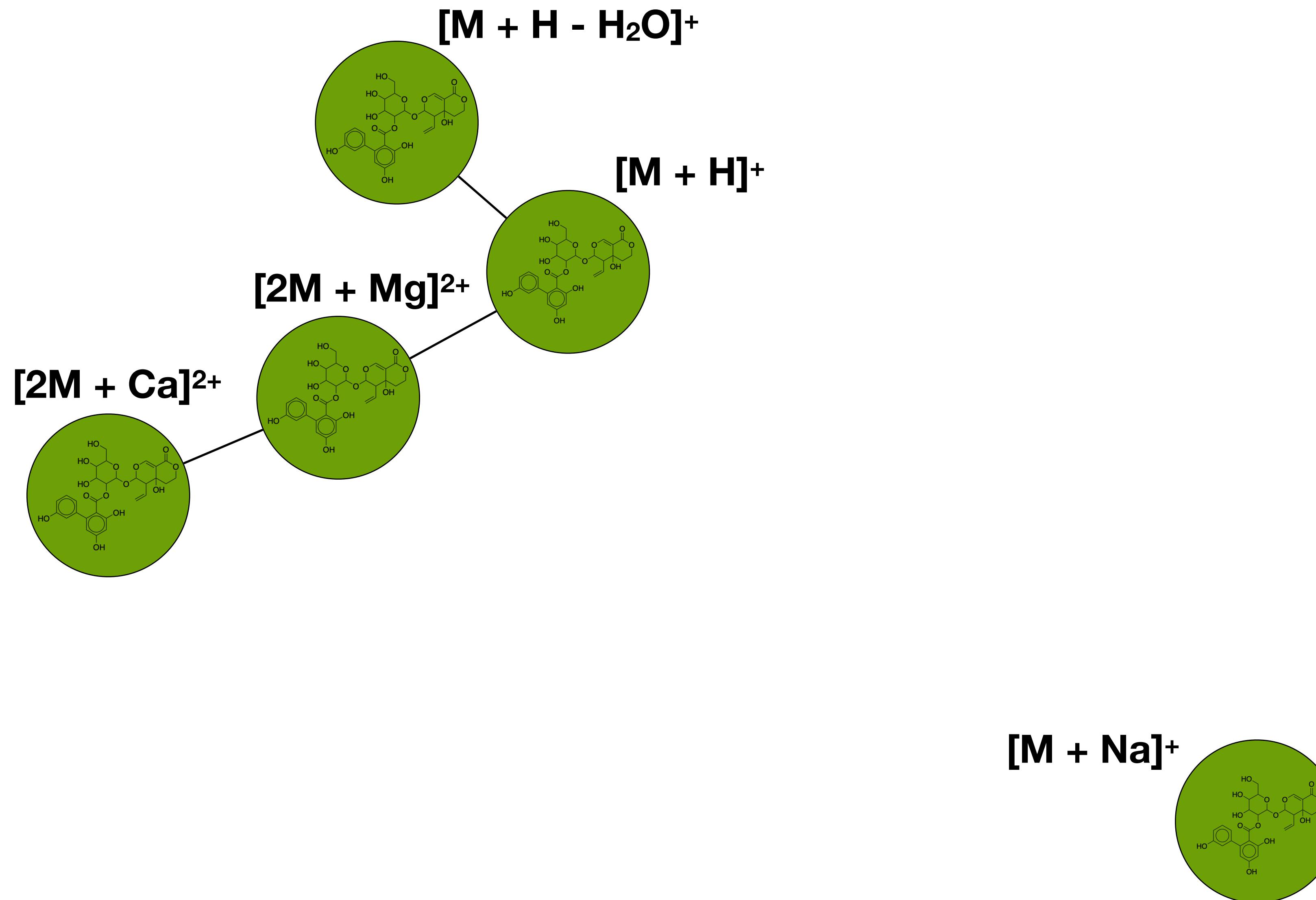
● Secoiridoid monoterpenoids



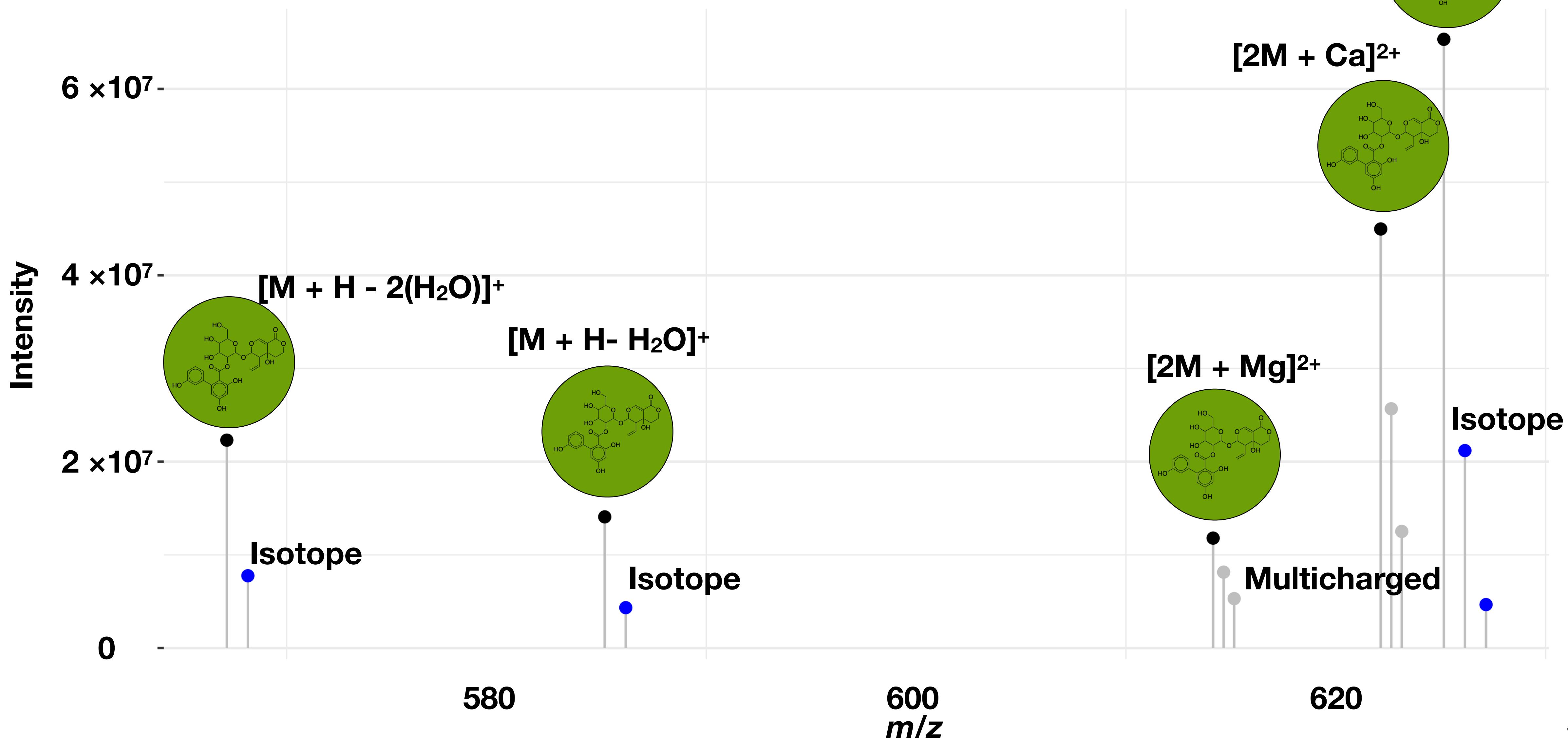
# MS<sup>1</sup>-based annotation



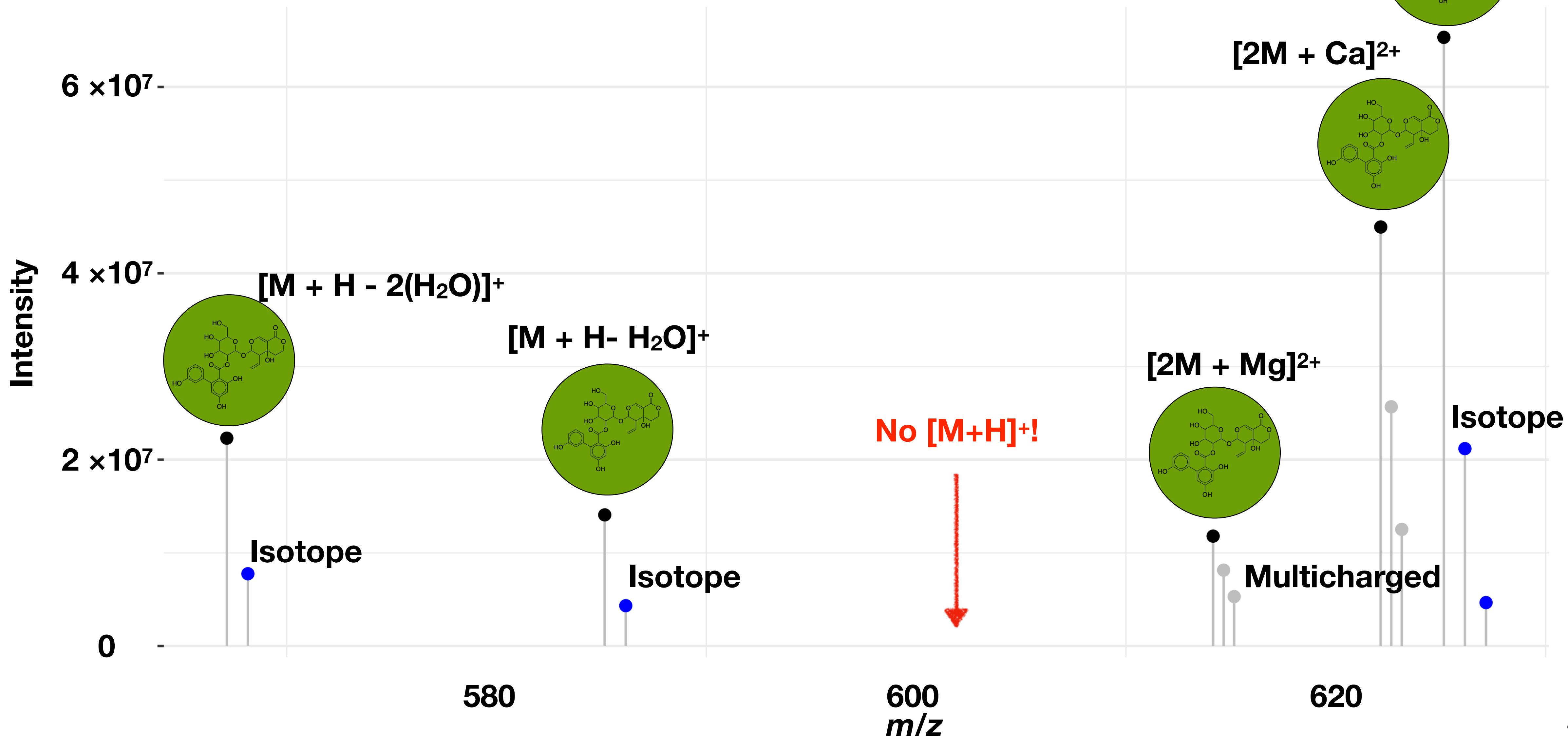
# MS<sup>1</sup>-based annotation



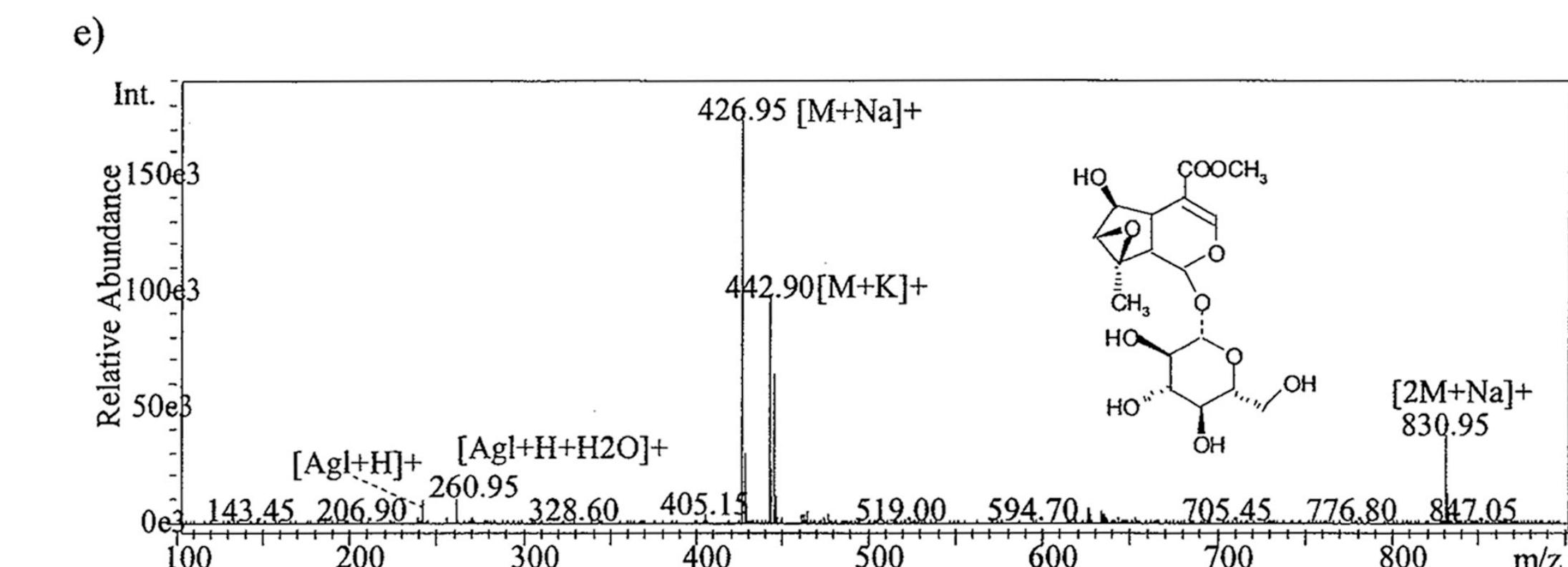
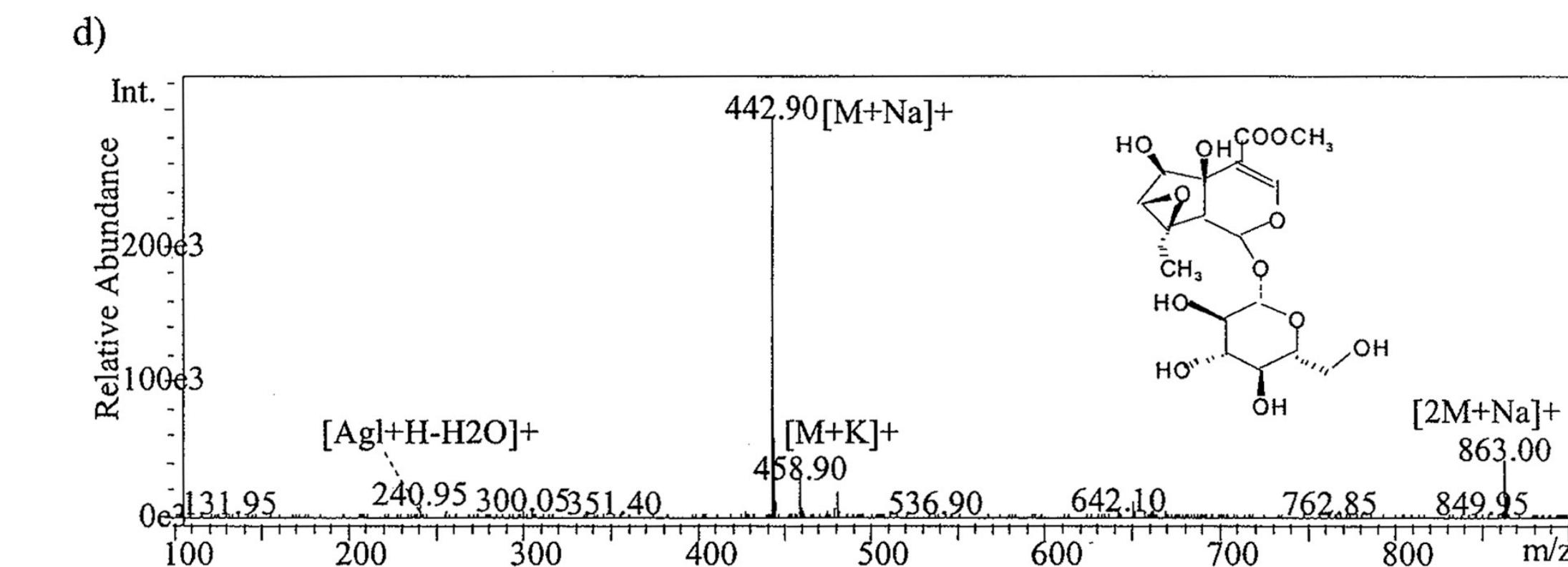
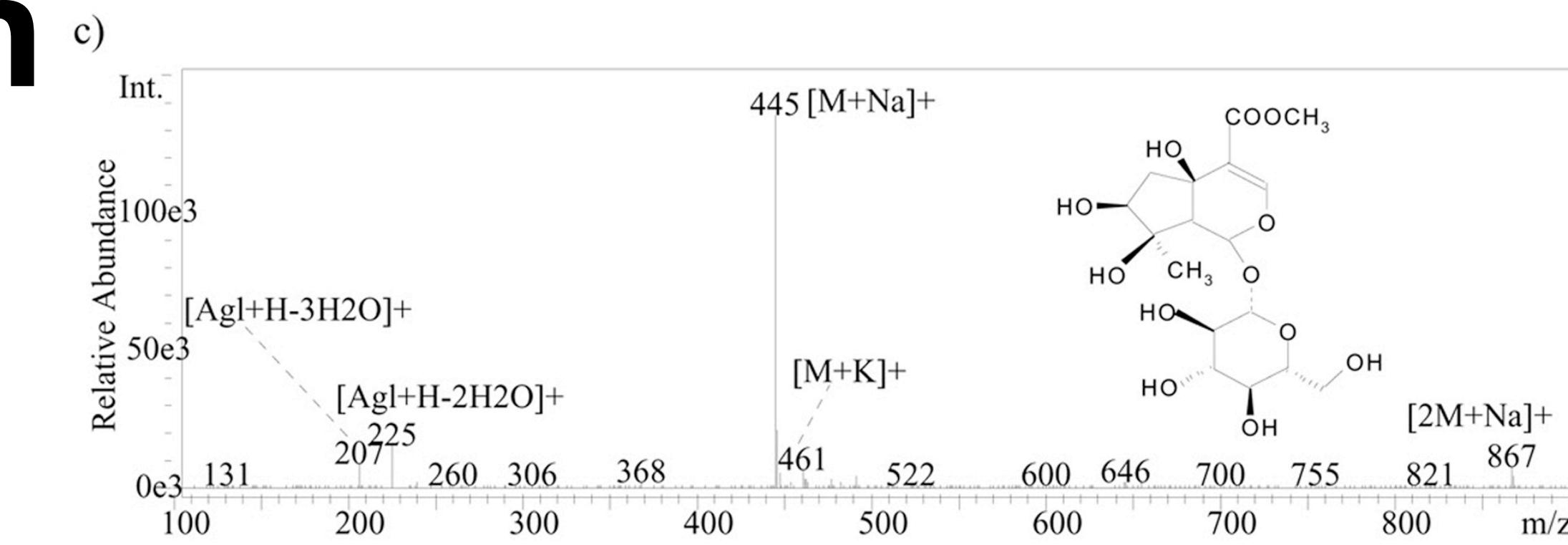
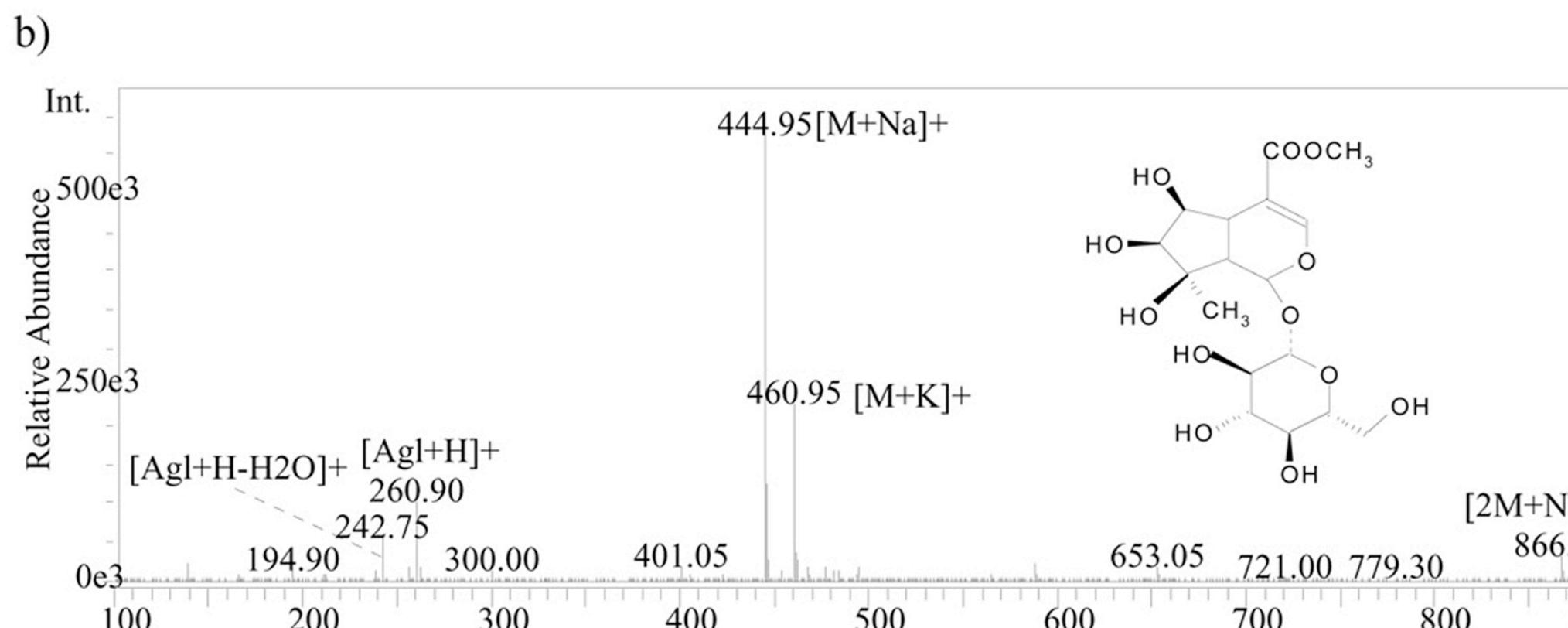
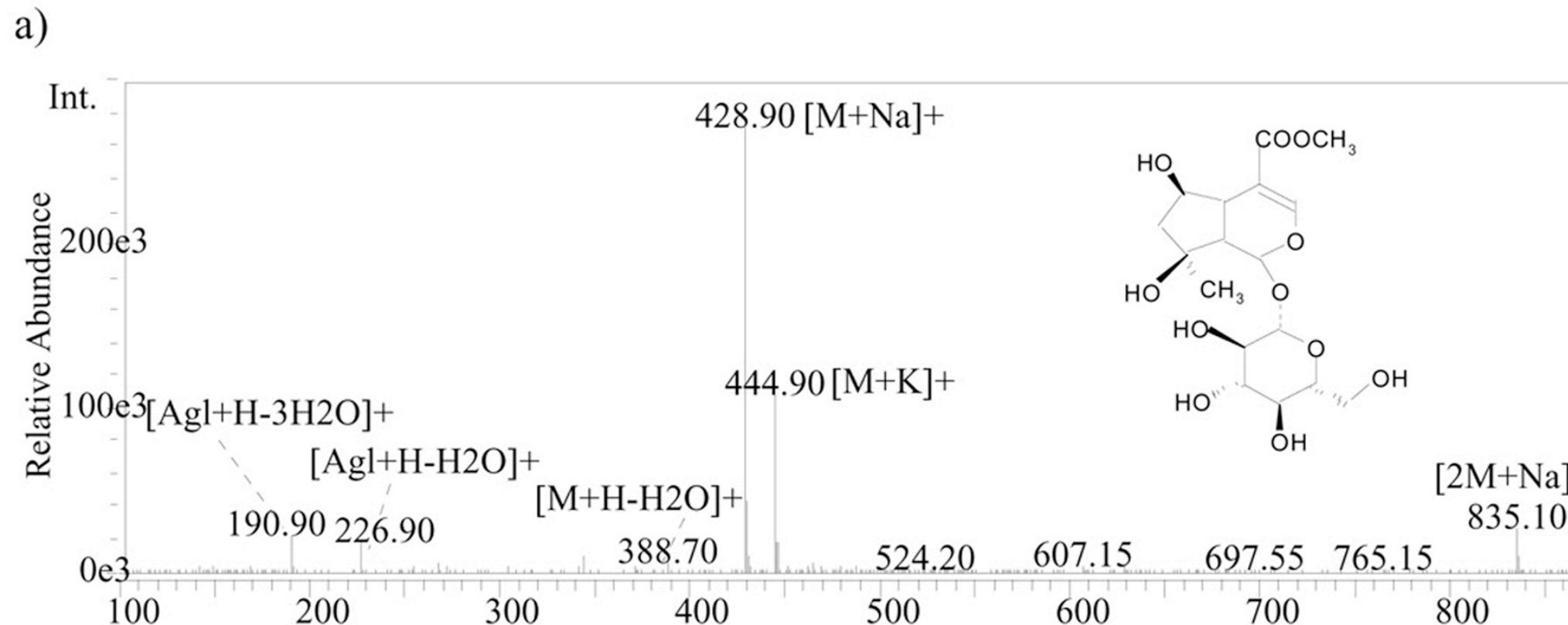
# MS<sup>1</sup>-based annotation



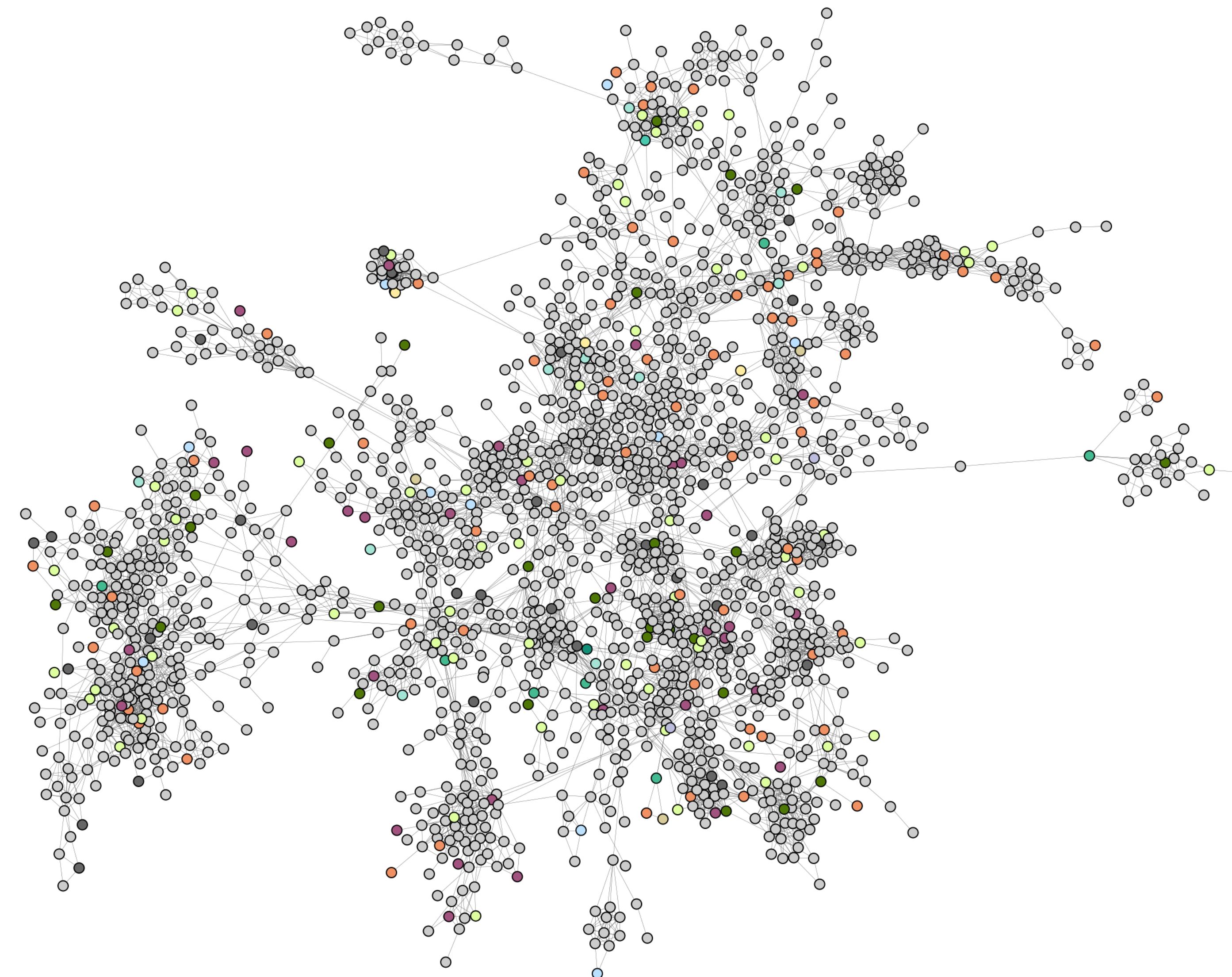
# MS<sup>1</sup>-based annotation



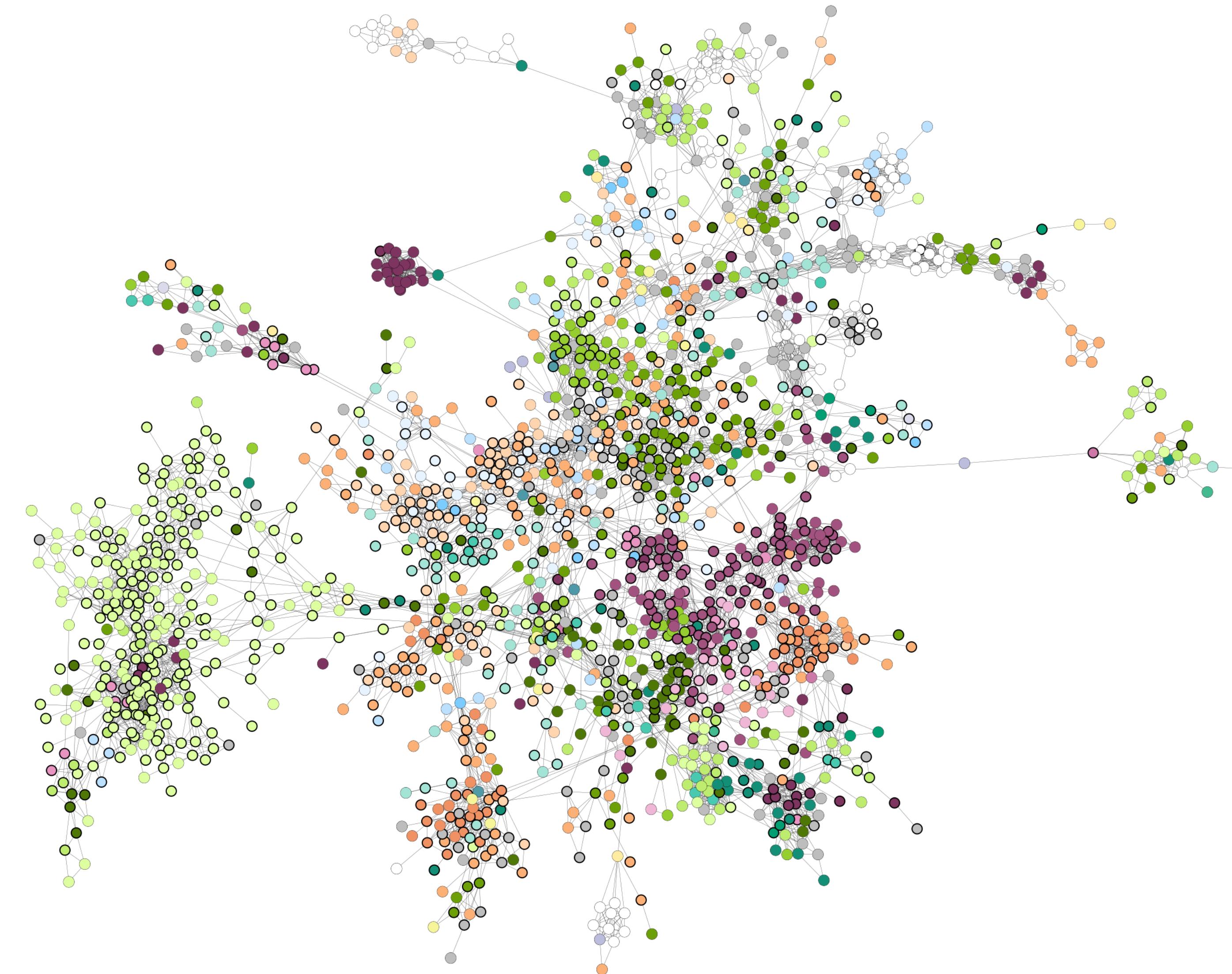
# MS<sup>1</sup>-based annotation



# Previous annotation



# Improved annotation



# Benchmark

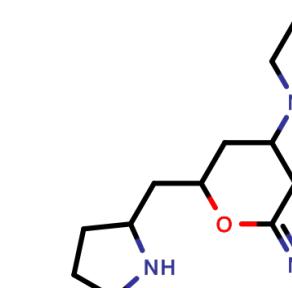
## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>  
public and third parties libraries  
66,646 experimental spectra

at least 6 fragments  
max 500 fragments  
 $100 \text{ Da} < x < 1500 \text{ Da}$   
[M+H]<sup>+</sup> adduct filtering



## STRUCTURES



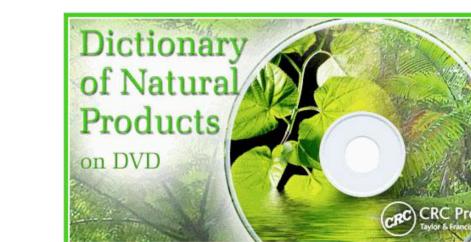
chemical translation  
and sanitization

Cleaned  
2D structures



2107 unique entries

## BIOLOGICAL SOURCES



"Alkaloid from *Brunfelsia hopeana*"



text recognition  
matching and  
resolving against the  
Catalogue of Life

Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

# Benchmark - v2

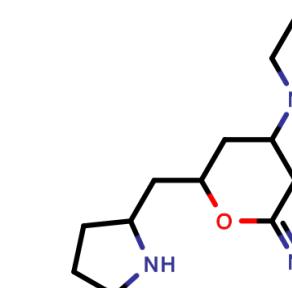
## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>  
public and third parties libraries  
~~66,646~~ experimental spectra

~~210,400~~  
at least 6 fragments  
max 500 fragments  
 $100 \text{ Da} < x < 1500 \text{ Da}$   
[M+H]<sup>+</sup> adduct filtering



## STRUCTURES

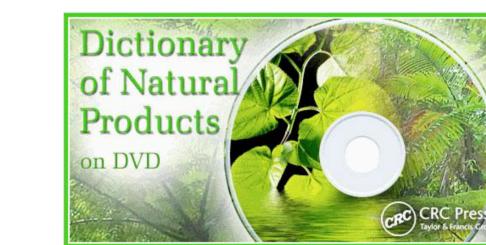


chemical translation  
and sanitization

Cleaned  
2D structures

2107 unique entries

## BIOLOGICAL SOURCES



"Alkaloid from *Brunfelsia hopeana*"



text recognition  
matching and  
resolving against the  
Catalogue of Life

Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

# Benchmark - v2

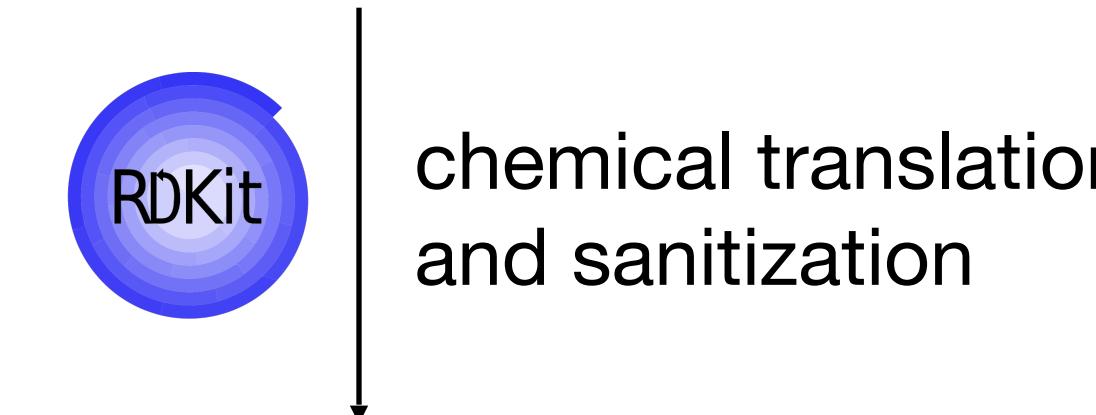
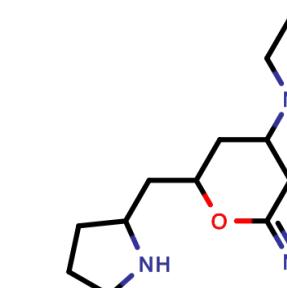
## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>  
public and third parties libraries  
~~66,646~~ experimental spectra  
~~210,400~~

at least 6 fragments  
max 500 fragments  
 $100 \text{ Da} < x < 1500 \text{ Da}$   
 $[\text{M}-\text{H}]^+$  adduct filtering



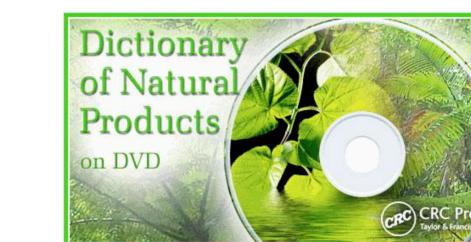
## STRUCTURES



Cleaned  
2D structures

2107 unique entries

## BIOLOGICAL SOURCES



"Alkaloid from *Brunfelsia hopeana*"



text recognition  
matching and  
resolving against the  
Catalogue of Life

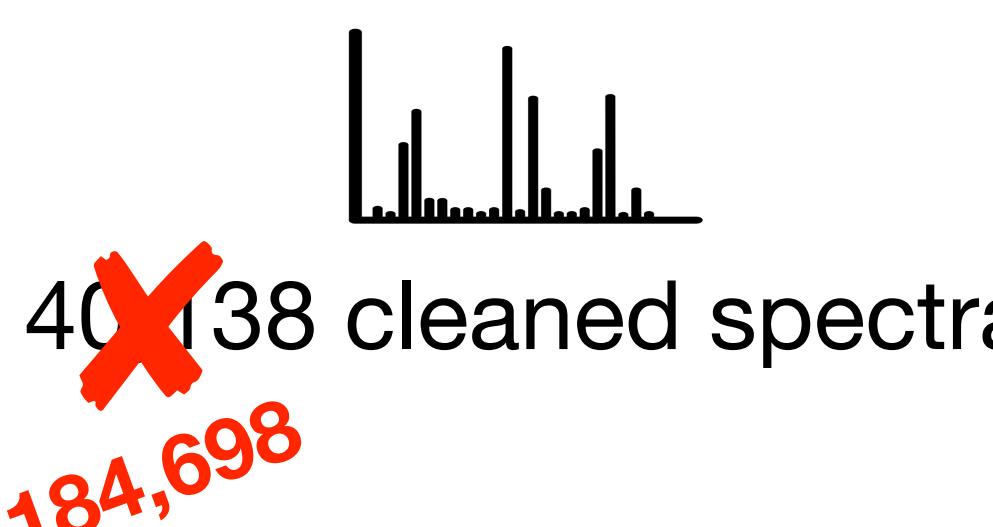
Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

# Benchmark - v2

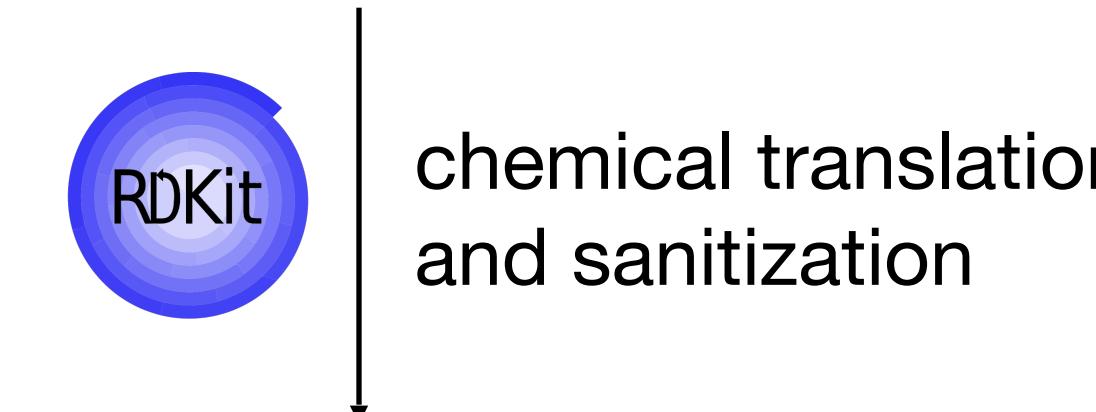
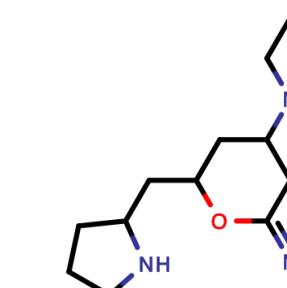
## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>  
public and third parties libraries  
~~66,646~~ experimental spectra  
~~210,400~~

at least 6 fragments  
max 500 fragments  
 $100 \text{ Da} < x < 1500 \text{ Da}$   
 $[\text{M}-\text{H}]^+$  adduct filtering

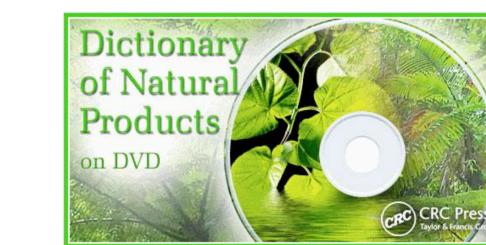


## STRUCTURES



Cleaned  
2D structures

## BIOLOGICAL SOURCES



"Alkaloid from *Brunfelsia hopeana*"



text recognition  
matching and  
resolving against the  
Catalogue of Life

Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

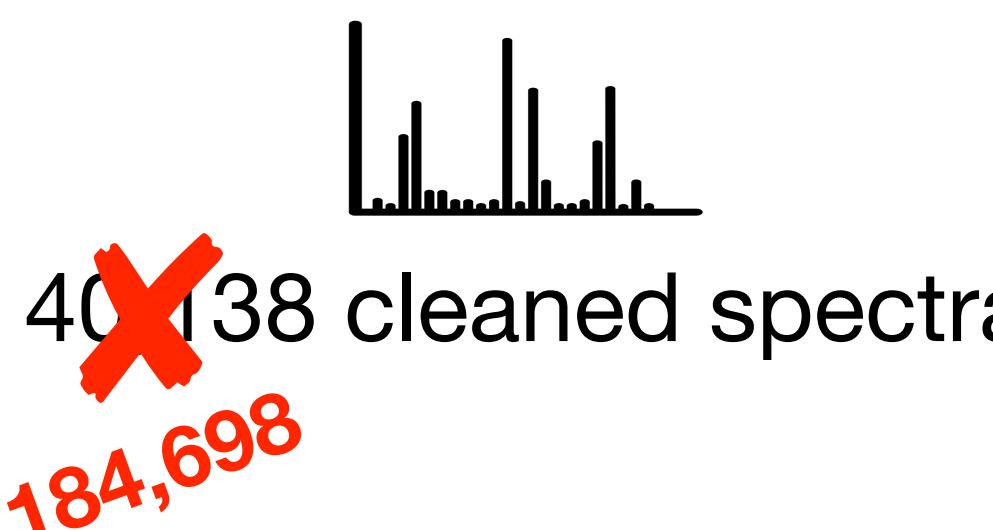
2107 unique entries

# Benchmark - v2

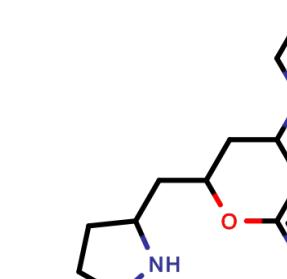
## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>  
public and third parties libraries  
~~66,646~~ experimental spectra  
~~210,400~~

at least 6 fragments  
max 500 fragments  
 $100 \text{ Da} < x < 1500 \text{ Da}$   
 $[\text{M}-\text{H}]^+$  adduct filtering



## STRUCTURES



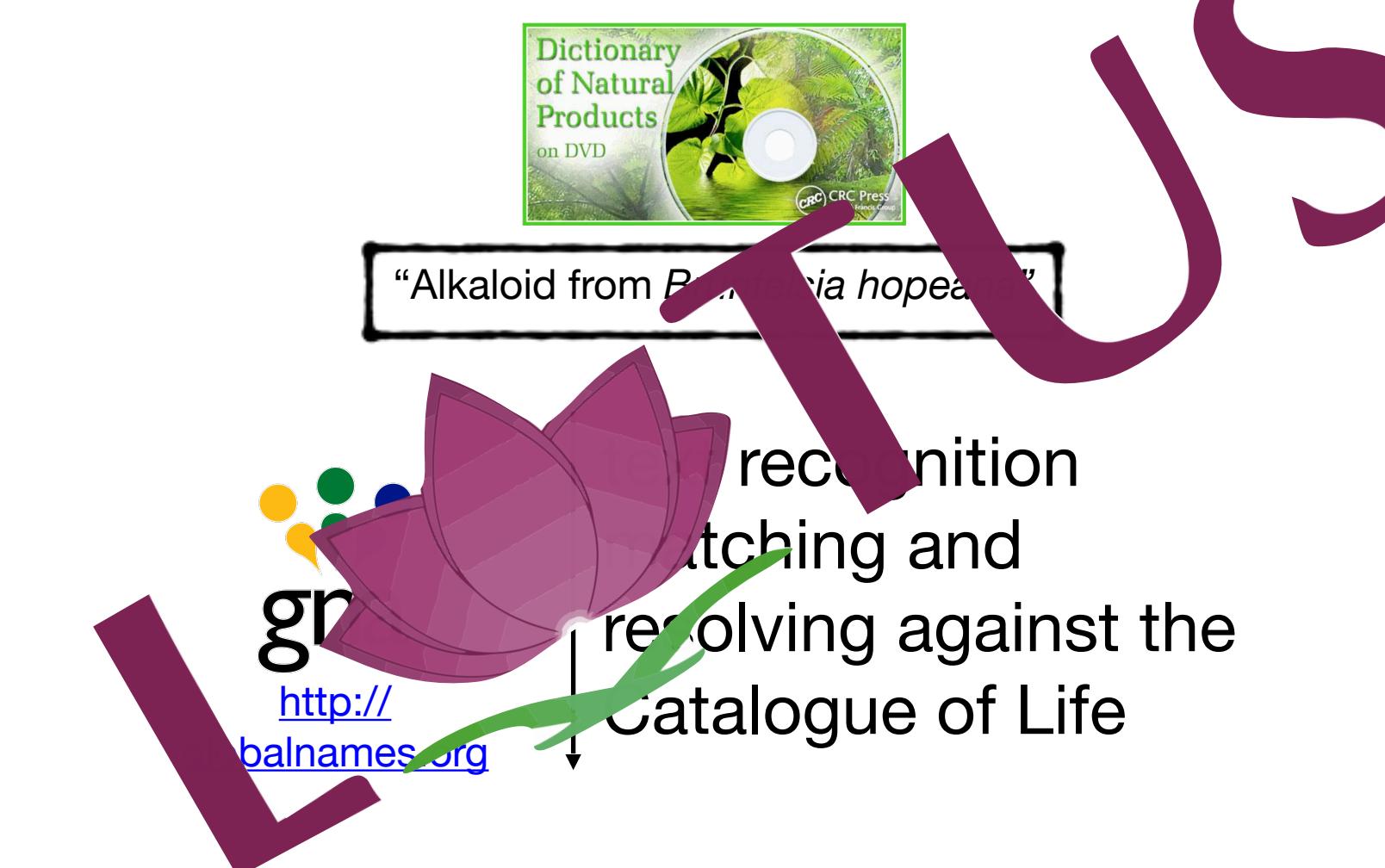
RDKit  
chemical translation  
and sanitization

Cleaned  
2D structures



2107 unique entries

## BIOLOGICAL SOURCES



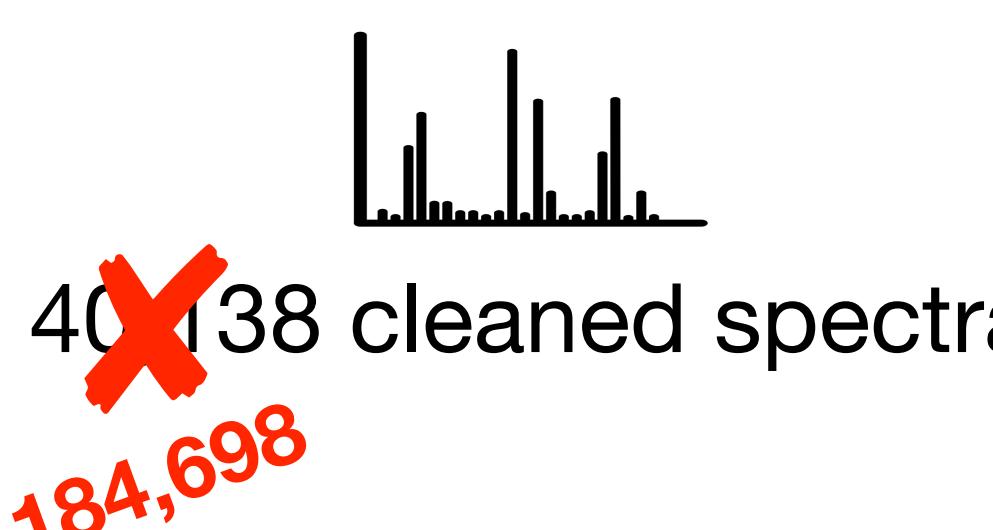
Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

# Benchmark - v2

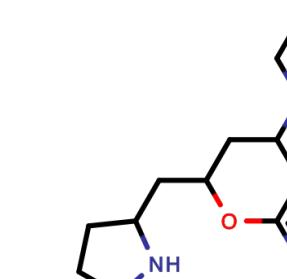
## EXPERIMENTAL SPECTRA

<https://gnps.ucsd.edu>  
public and third parties libraries  
66,646 experimental spectra  
~~210,400~~

at least 6 fragments  
max 500 fragments  
100 Da < x < 1500 Da  
[M+H]<sup>+</sup> adduct filtering



## STRUCTURES

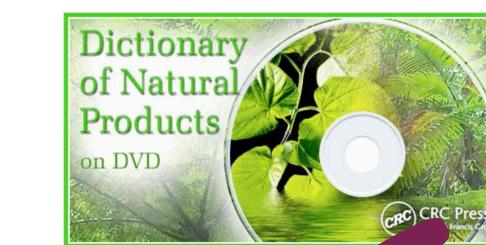


RDKit  
chemical translation  
and sanitization

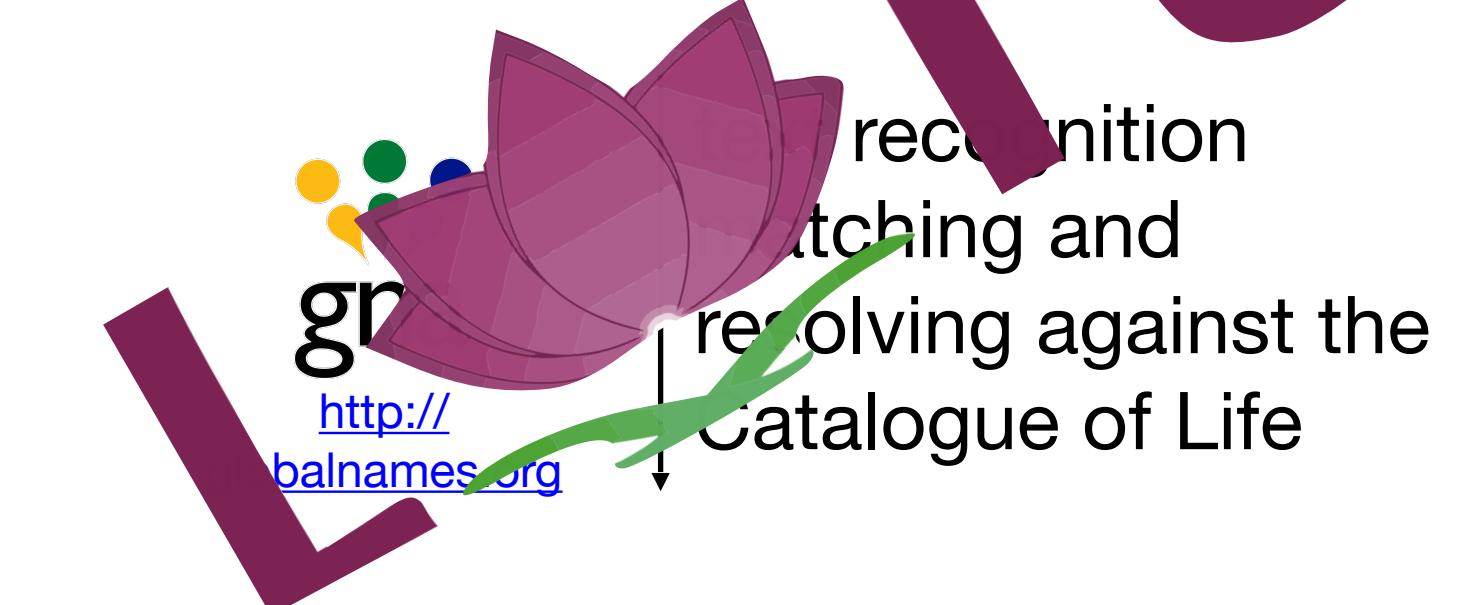
Cleaned  
2D structures



## BIOLOGICAL SOURCES



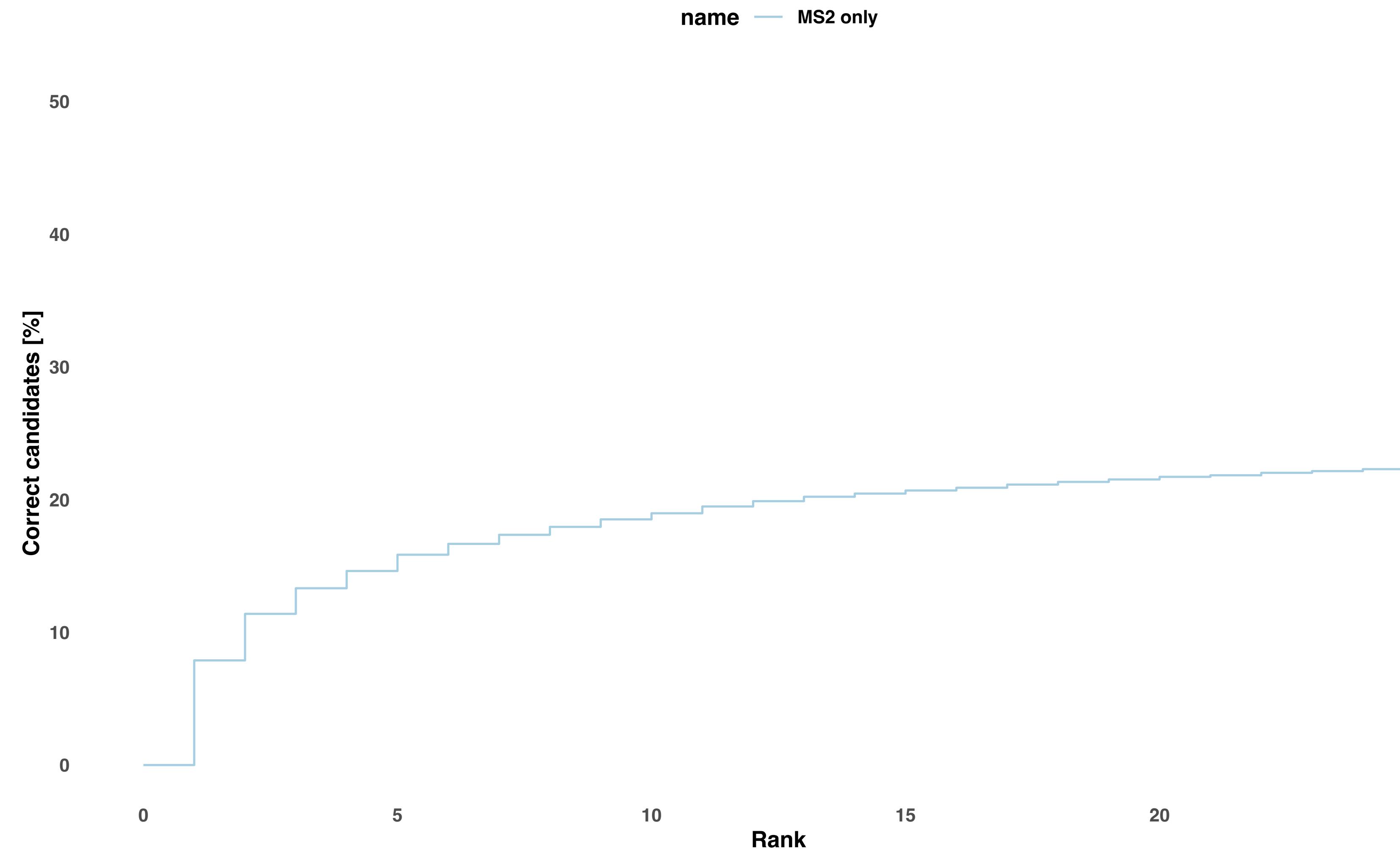
"Alkaloid from *Brunfelsia hopeana*"



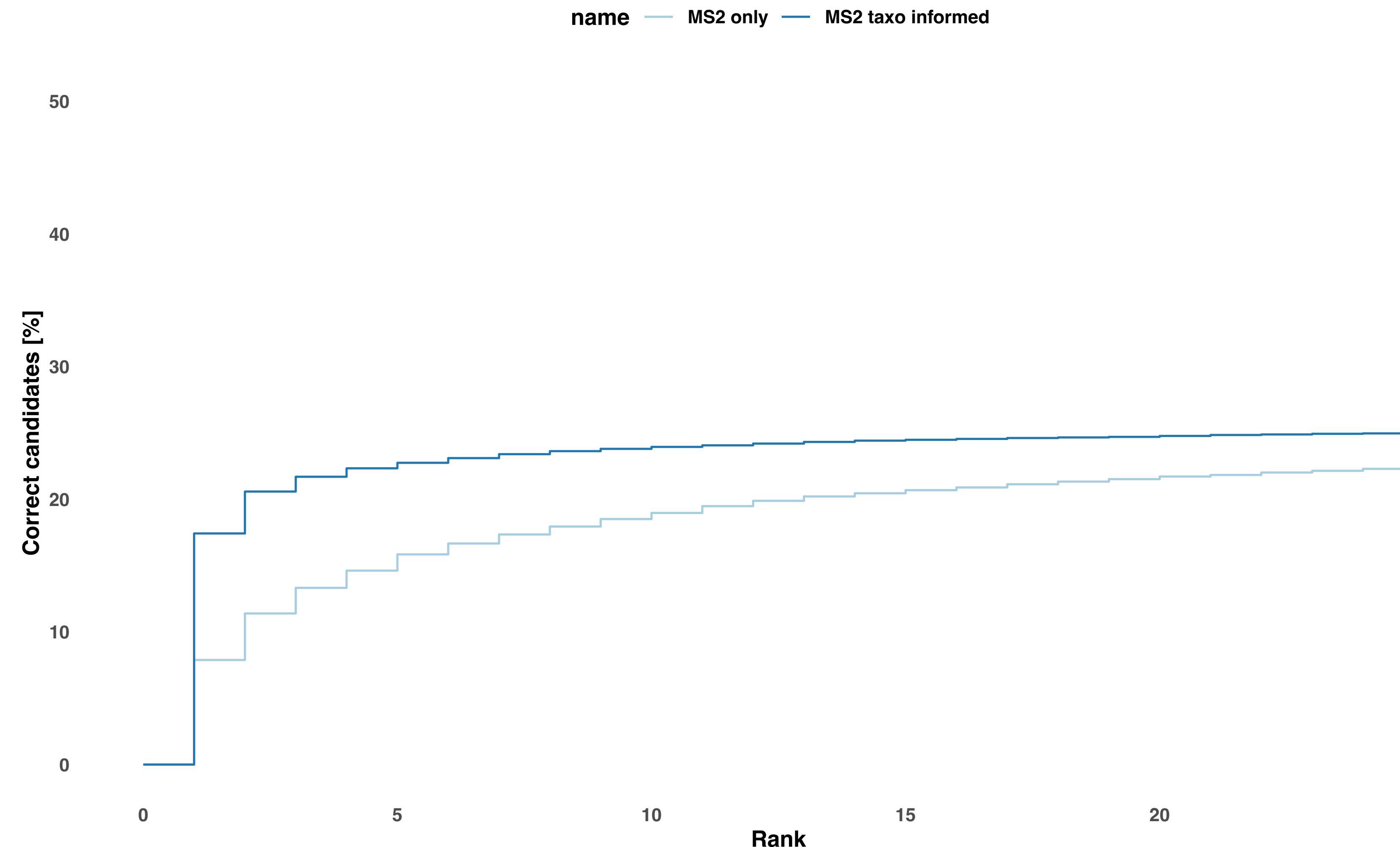
Kingdom	Order	Family	Genus	Species
Plantae	Solanales	Solanaceae	Brunfelsia	<i>Brunfelsia uniflora</i>

22,251 spectra (12,482 structures) in positive mode  
7,472 spectra (6,587 structures) in negative mode

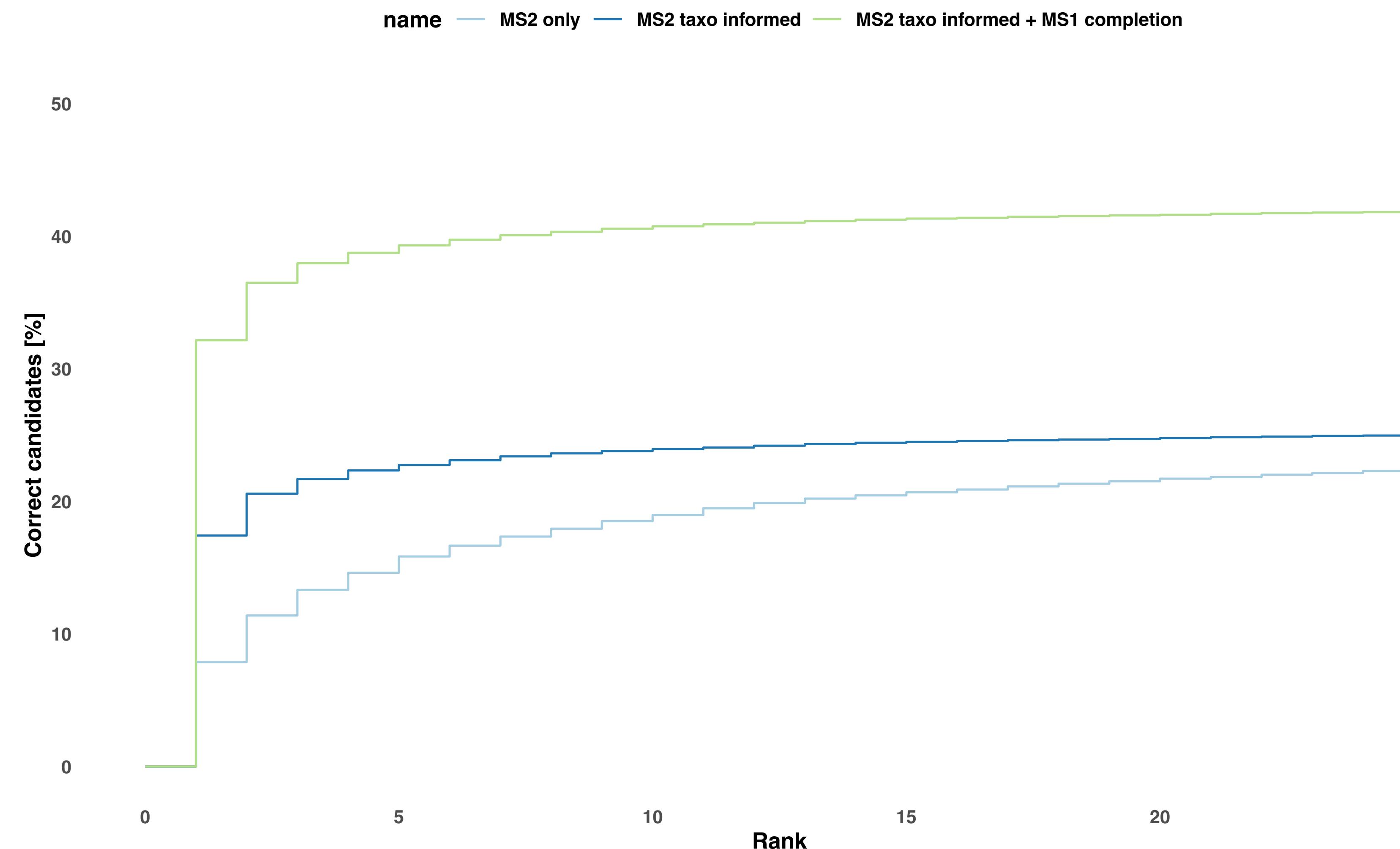
# Benchmark - v2



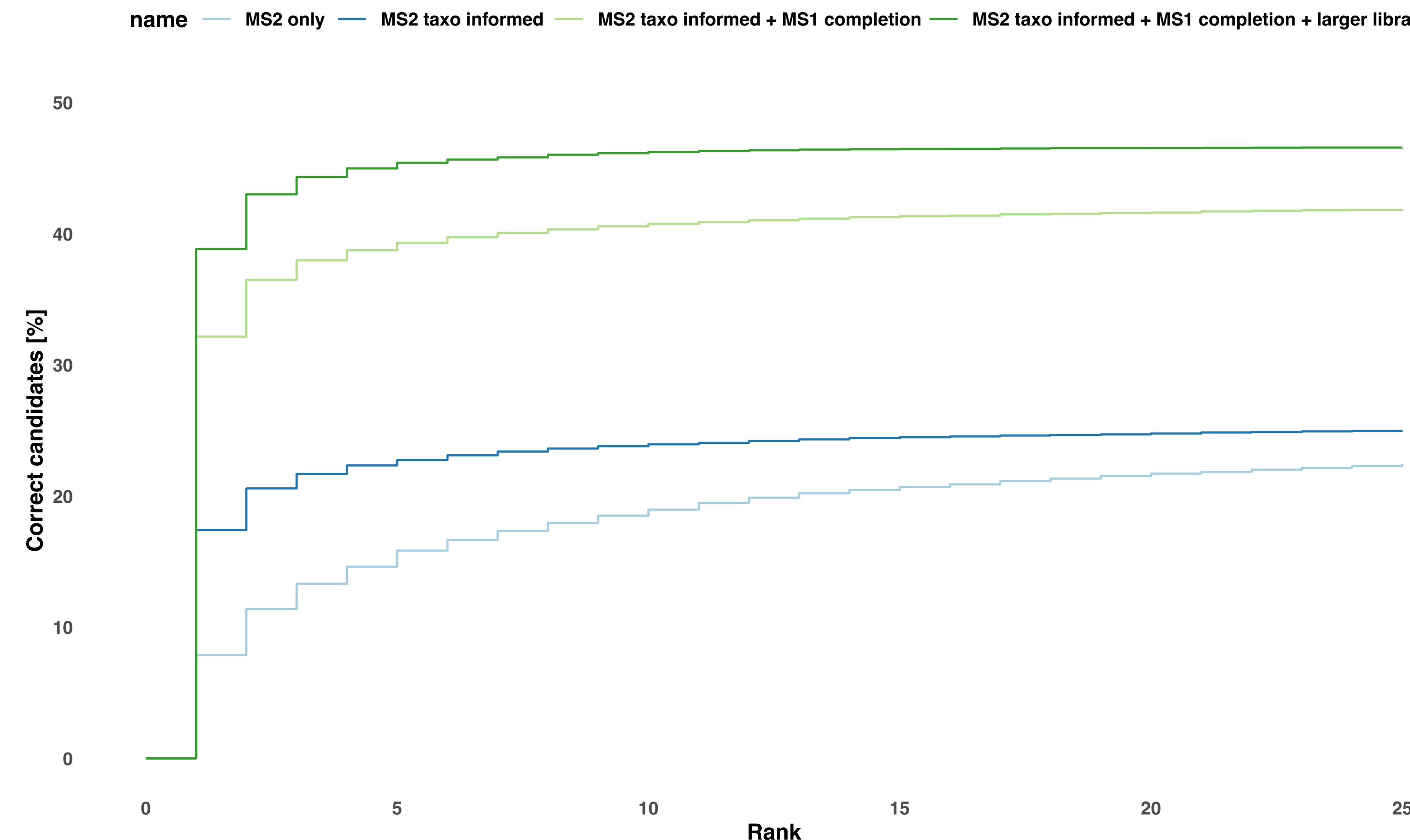
# Benchmark - v2



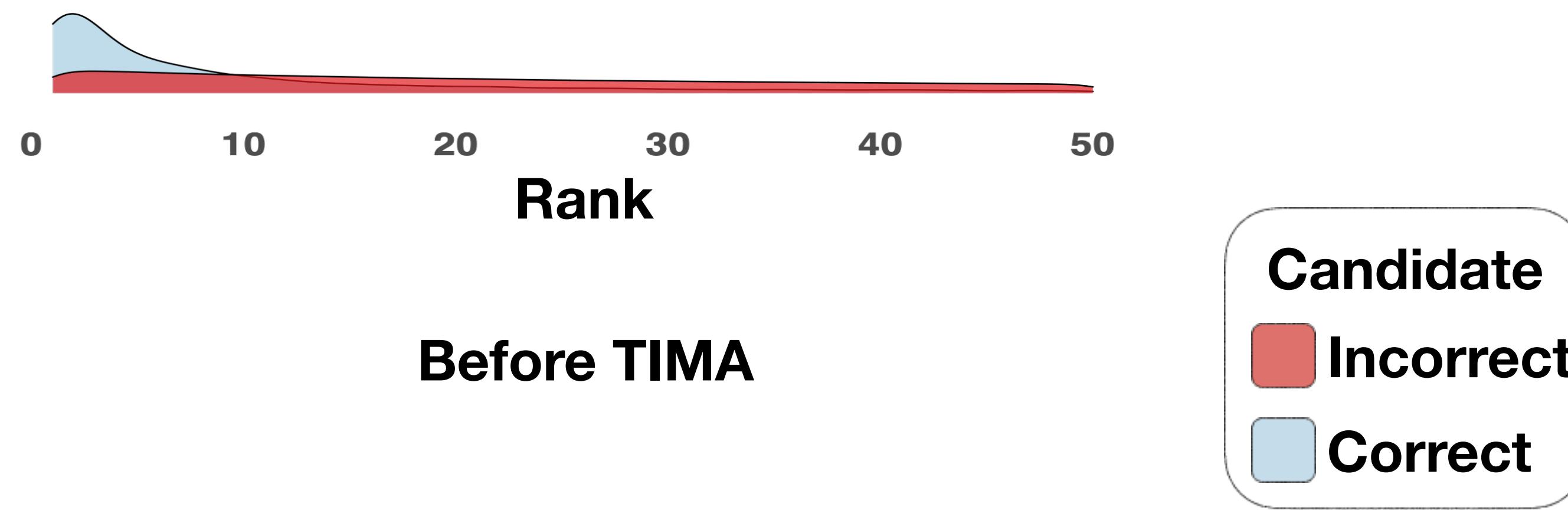
# Benchmark - v2



# Benchmark - v2



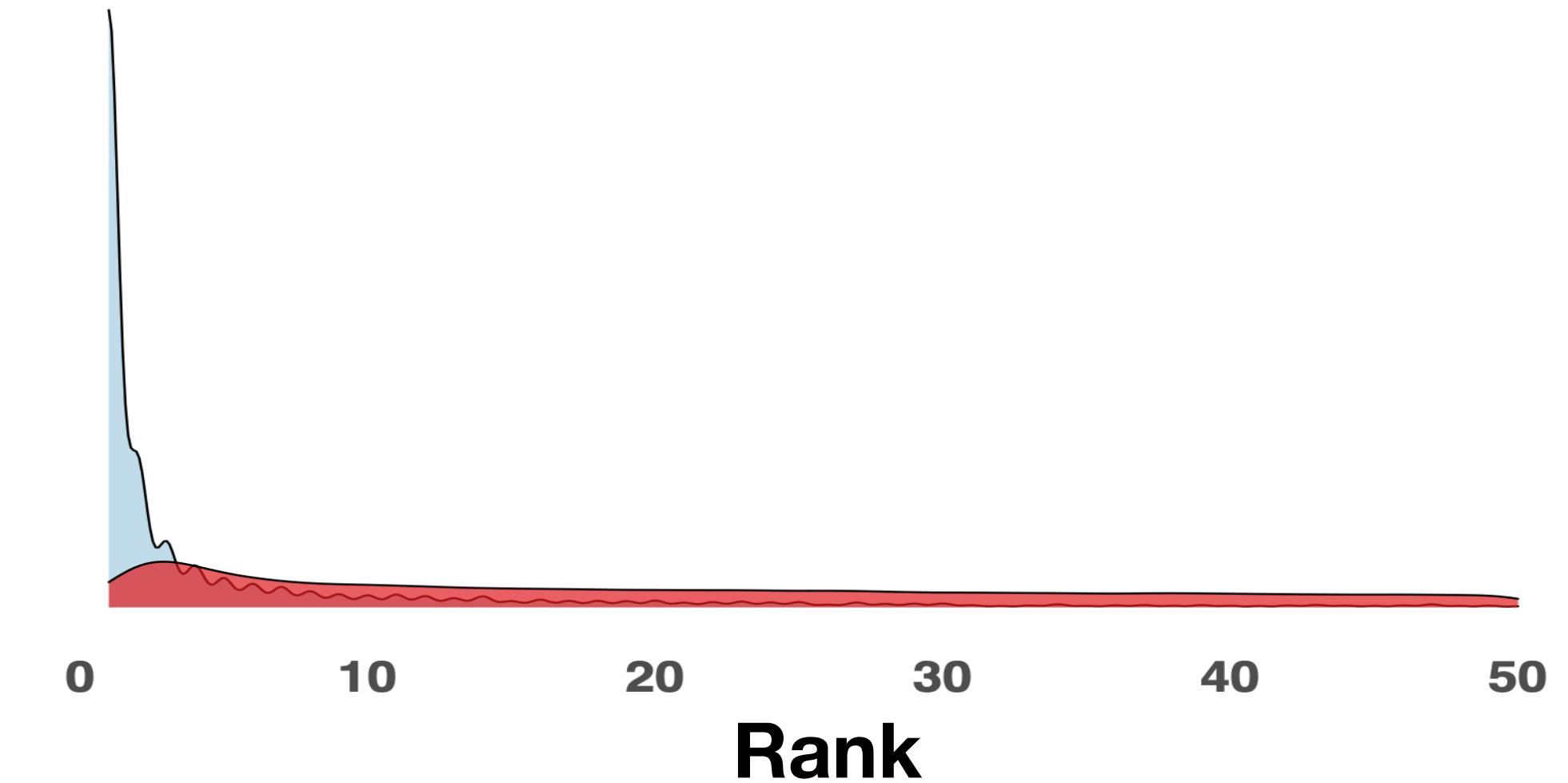
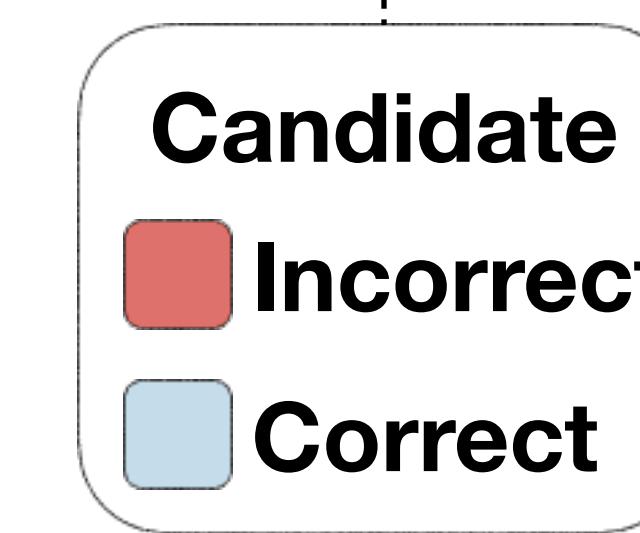
# Benchmark - v2



# Benchmark - v2



Before TIMA



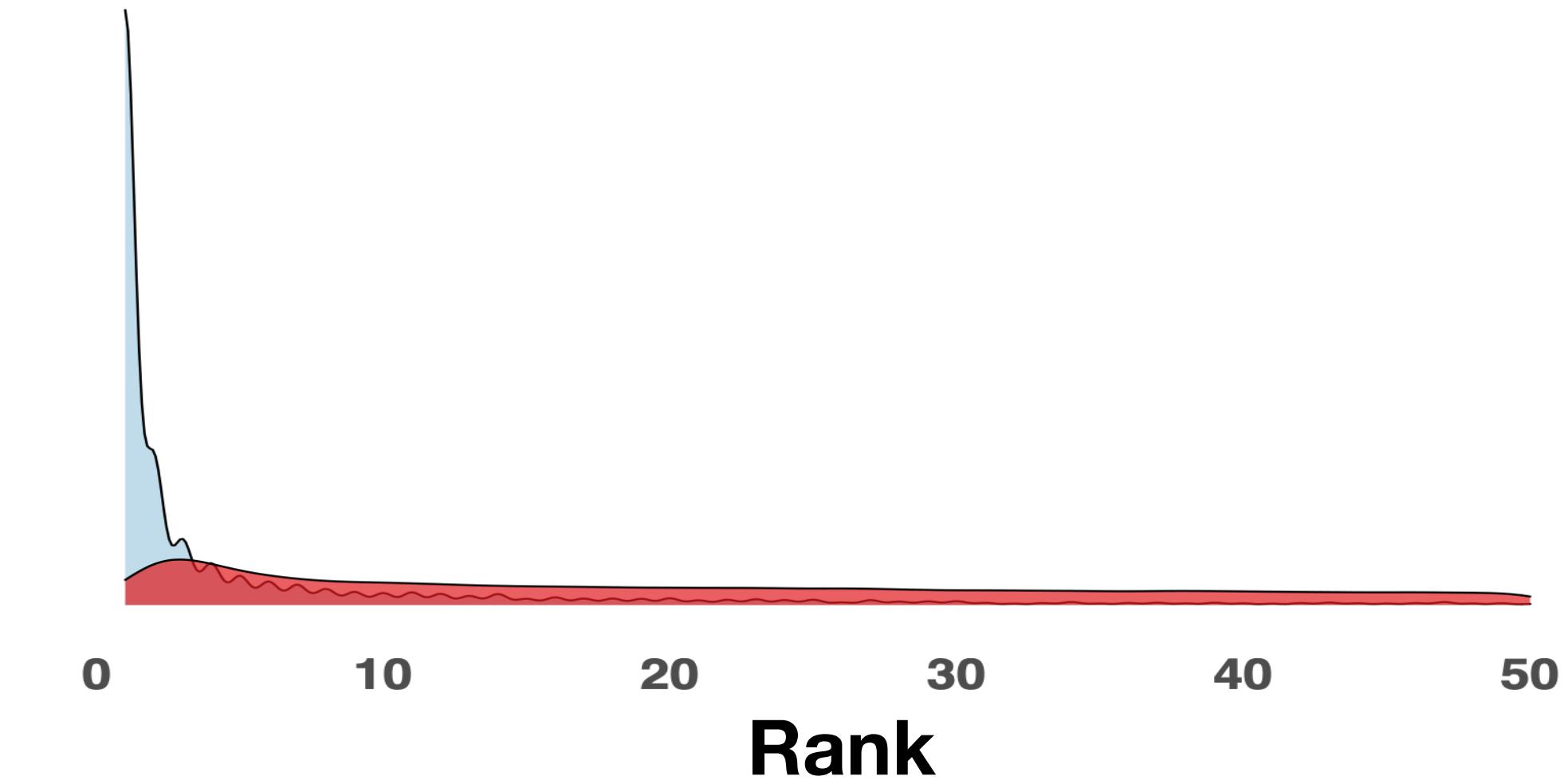
After TIMA

# Benchmark - v2

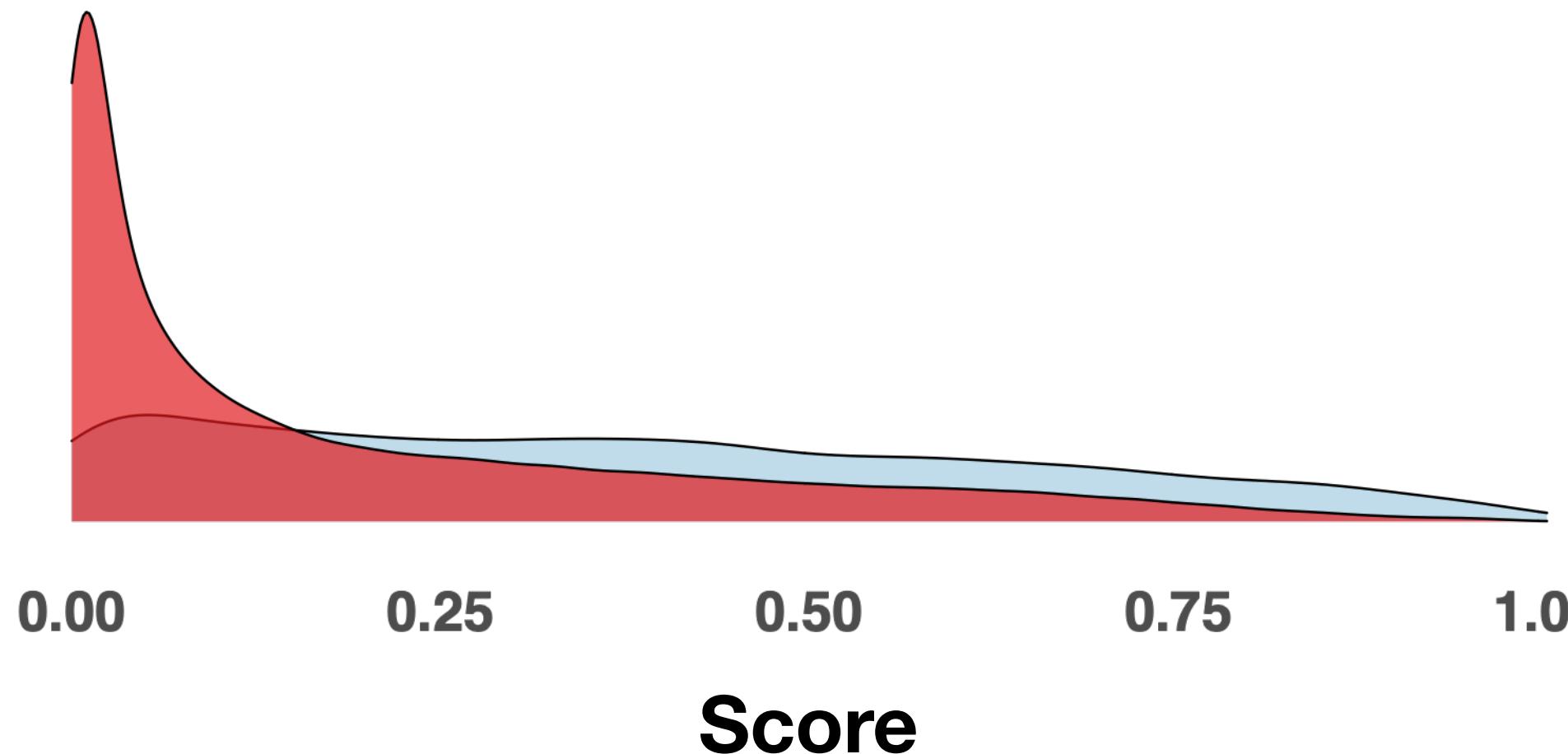


Before TIMA

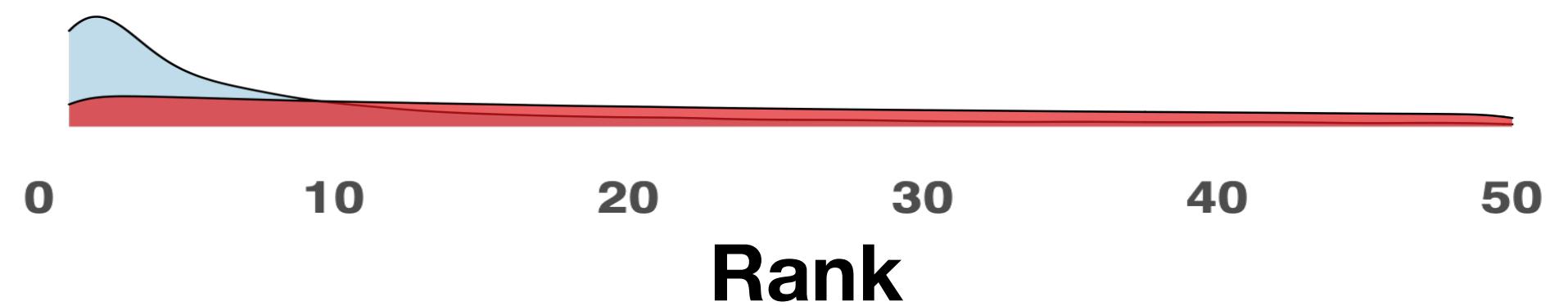
**Candidate**  
■ **Incorrect**  
■ **Correct**



After TIMA

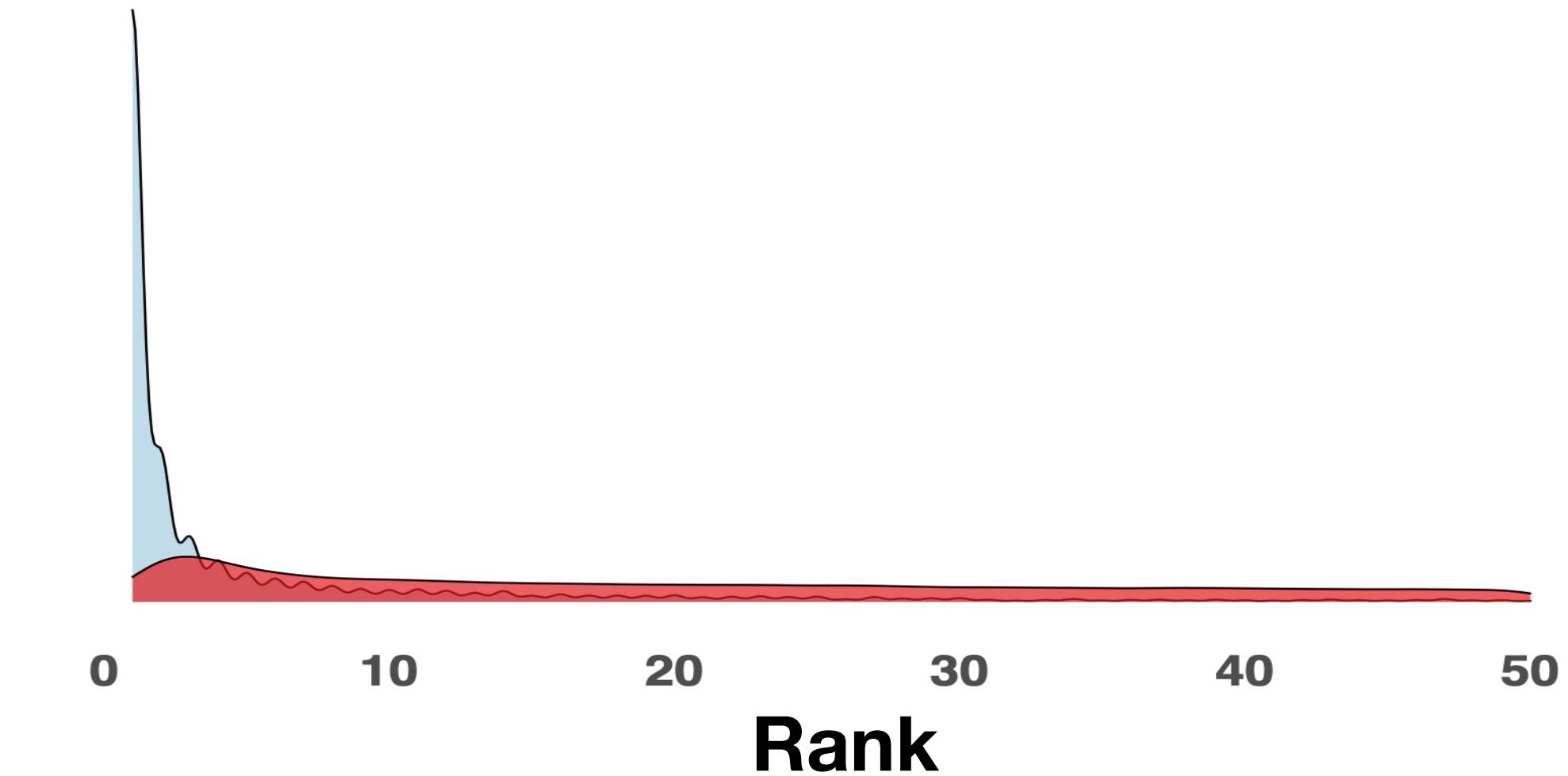


# Benchmark - v2

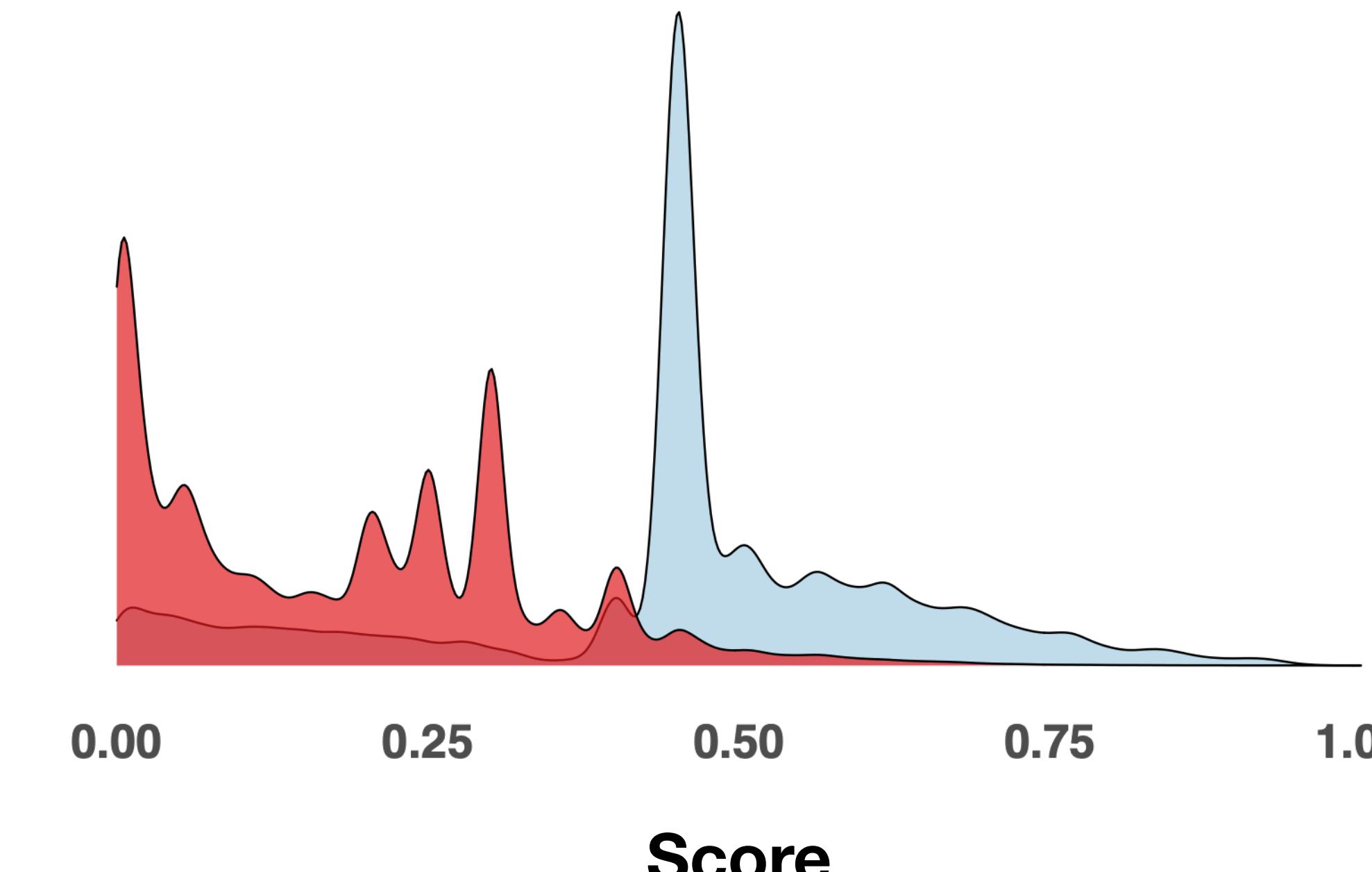
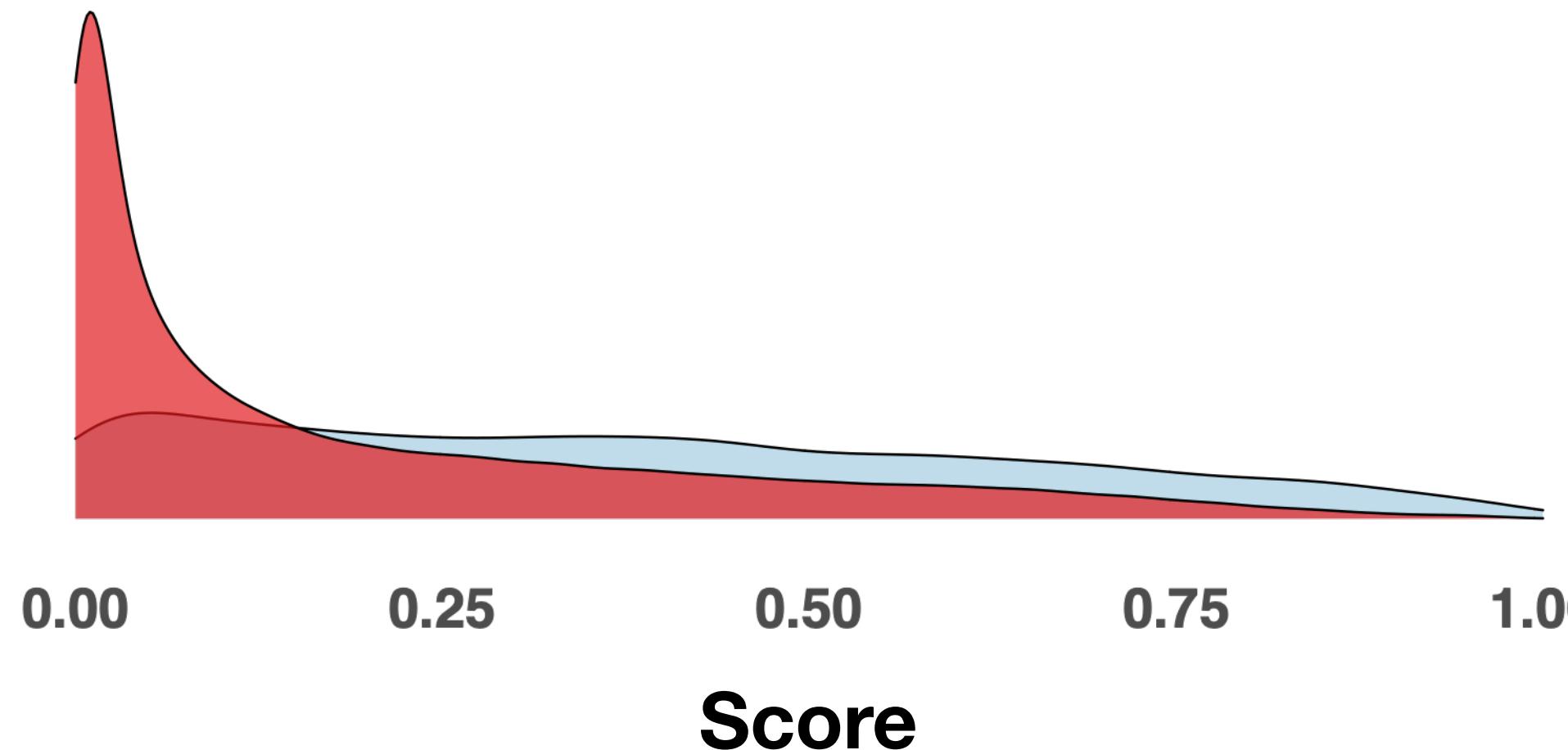


Before TIMA

**Candidate**  
■ **Incorrect**  
■ **Correct**

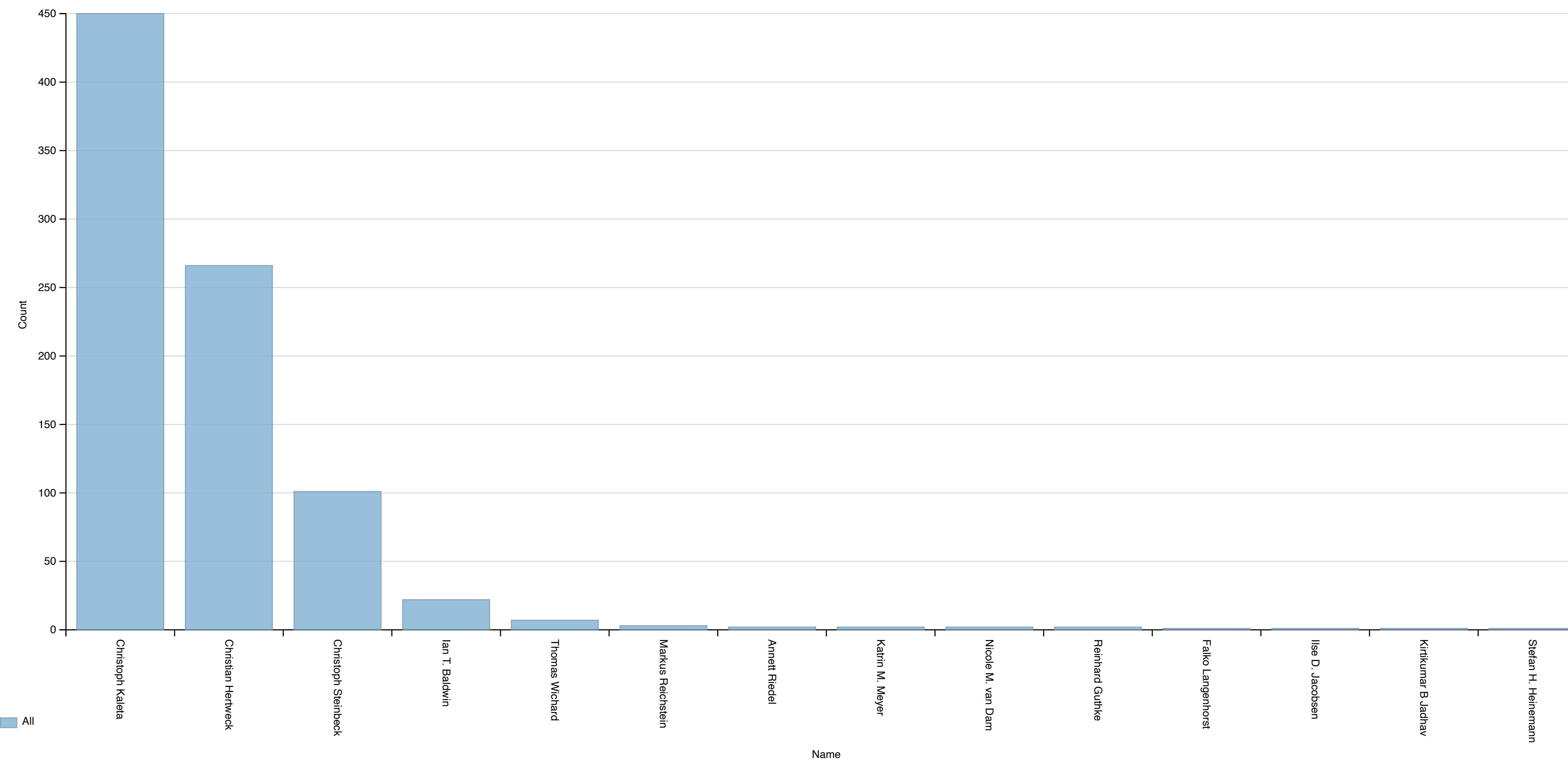


After TIMA



# From metabolites to hypotheses (and back)

« Hey Wiki, who can I contact if I want to talk about Natural Products chemistry in Jena? »



# From metabolites to hypotheses (and back)

« Hey Wiki, I saw this review<sup>1</sup> yesterday, could you do the same?»



Contents lists available at ScienceDirect

Microbiological Research

journal homepage: [www.elsevier.com/locate/micres](http://www.elsevier.com/locate/micres)



Actinobacteria in natural products research: Progress and prospects



Polpass Arul Jose <sup>a,\*</sup>,<sup>1</sup> Anjisha Maharshi <sup>a,2</sup>, Bhavanath Jha <sup>a,b,\*\*</sup>

<sup>a</sup> Marine Biotechnology and Ecology Division, CSIR- Central Salt and Marine Chemicals Research Institute, G. B. Marg, Bhavnagar, Gujarat, 364002, India

<sup>b</sup> Academy of Scientific and Innovative Research (AcSIR), CSIR, India

<sup>1</sup> <https://doi.org/10.1016/j.micres.2021.126708>

# From metabolites to hypotheses (and back)

« Hey Wiki, I saw this review<sup>1</sup> yesterday, could you do the same?»



Contents lists available at ScienceDirect

Microbiological Research

journal homepage: [www.elsevier.com/locate/micres](http://www.elsevier.com/locate/micres)



Actinobacteria in natural products research: Progress and prospects

Polpass Arul Jose <sup>a,\*</sup>,<sup>1</sup> Anjisha Maharshi <sup>a,2</sup>, Bhavanath Jha <sup>a,b,\*\*</sup>

<sup>a</sup> Marine Biotechnology and Ecology Division, CSIR- Central Salt and Marine Chemicals Research Institute, G. B. Marg, Bhavnagar, Gujarat, 364002, India

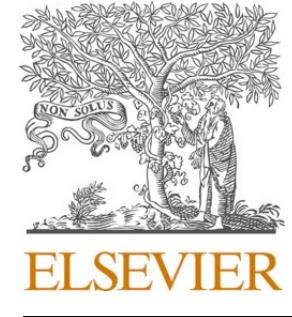
<sup>b</sup> Academy of Scientific and Innovative Research (AcSIR), CSIR, India

*Which compounds with known bioactivities were isolated from Actinobacteria (Q26262282), between 2014 and 2019, with related organisms and references?*

<sup>1</sup> <https://doi.org/10.1016/j.micres.2021.126708>

# From metabolites to hypotheses (and back)

« Hey Wiki, I saw this review<sup>1</sup> yesterday, could you do the same? »



Contents lists available at [ScienceDirect](#)

**Microbiological Research**

journal homepage: [www.elsevier.com/locate/micres](http://www.elsevier.com/locate/micres)



Actinobacteria in natural products research: Progress and prospects

Polpass Arul Jose <sup>a,\*</sup>,<sup>1</sup> Anjisha Maharshi <sup>a,2</sup>, Bhavanath Jha <sup>a,b,\*\*</sup>

<sup>a</sup> Marine Biotechnology and Ecology Division, CSIR- Central Salt and Marine Chemicals Research Institute, G. B. Marg, Bhavnagar, Gujarat, 364002, India

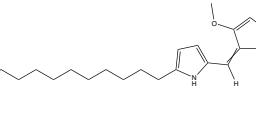
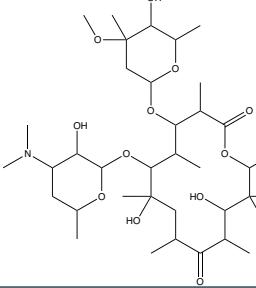
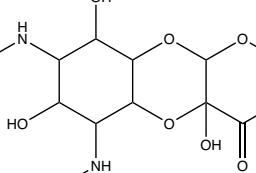
<sup>b</sup> Academy of Scientific and Innovative Research (AcSIR), CSIR, India

**Table 1**  
Novel bioactive compounds isolated from Actinobacteria during last five years (2014-2019).

Organism	Compound	Bioactivity <sup>#</sup>	Reference
<i>Arthrobacter</i> sp. PGVB1	Arthroamide	f (anti-quorum sensing)	J Nat Prod. 2015;78:2827–2831
<i>Actinoallomorus</i> sp. ID145113 and 145,206	Paramagnetoquinones A-C (1-3)	(1/2) a; (3) a	J Nat Prod. 2017;80:819–827
<i>Actinoalloteichus hymeniacidonis</i> 179DD-027	Dokdolipid B	c	Mar Drugs 2019;17:237
<i>Actinomadura atramentaria</i> NBRC 14,695**	Cinnamycin B	a	J Ind Microbiol Biotechnol. 2016;43:1159–65
<i>Actinomadura</i> sp. BCC 35,430*	Actinomadurone	b	Tetrahedron Lett. 2017;58:3223–3225
<i>Actinomadura</i> sp. K13–0306	Sagamilactam	c, f (antiparasitic)	J Antibiot. 2016;69:818–824
<i>Actinomadura</i> sp. K4S16	Nonthmicin (1) & ecteinamycin (2)	(1,2) c	Org Lett. 2017;19:1406–1409
<i>Actinomadura</i> sp. KC191*	Actinomadurol	a	J Nat Prod. 2016;79:1886–1890
<i>Actinosynnema pretiosum</i> HGF052::asm18**	Actinosynneptide A (1) & B (2)	(1,2) c	Appl Microbiol Biotechnol. 2016;101:2273–2279
<i>Amycolatopsis</i> sp. IRD-009	Pradimicin-IRD	a, c	Natural Product Res. 2019;33:1713–1720

*Which compounds with known bioactivities were isolated from Actinobacteria (Q26262282), between 2014 and 2019, with related organisms and references?*

508 results in 4.8 seconds

Taxon name	Structure	Activity	Reference
<i>Streptomyces coelicolor</i>		Antibiotic   immunosuppressive drug	<a href="#">10.1007/S11306-016-025-6</a>
<i>Saccharopolyspora erythraea</i>		Antibiotic   protein synthesis inhibitors   gastrointestinal agent	<a href="#">10.1021/ACSSYNBIO.8B00372</a>
<i>Streptomyces venezuelae</i>		Antibiotic   protein synthesis inhibitors	<a href="#">10.1016/J.JBIOTEC.2014.01.028</a>

<sup>1</sup> <https://doi.org/10.1016/j.micres.2021.126708>

# From metabolites to hypotheses (and back)

*“Hey Wiki, I saw this review<sup>1</sup> yesterday, could you do the same?”*

 ELSEVIER

Contents lists available at ScienceDirect

**Microbiological Research**

journal homepage: [www.elsevier.com/locate/micres](http://www.elsevier.com/locate/micres)

---



# Actinobacteria in natural products research: Progress and prospects

Polpass Arul Jose<sup>a,\*1</sup>, Anjisha Maharshi<sup>a,2</sup>, Bhushan R. Patel<sup>b,\*\*</sup>

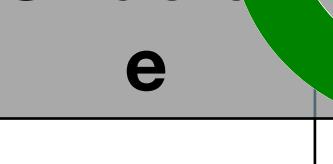
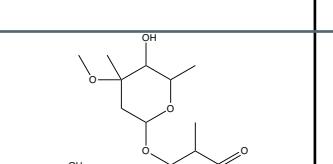
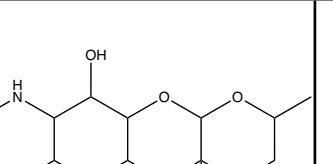
<sup>a</sup> Marine Biotechnology and Ecology Division, CSIR- Central Salt and Marine Chemicals Research Institute, Bhavnagar-364002, Gujarat, India

<sup>b</sup> Academy of Scientific and Innovative Research (AcSIR), CSIR, India

---

**Table 1**  
Novel bioactive compounds isolated from Actinobacteria during last five years (2014-2019).

*Which compounds with known bioactivities were isolated from Actinobacteria (Q26262282), between 2014 and 2019, with related organisms and references?*

Taxon name	Structure	Activity	Reference
<i>Streptomyces coelicolor</i>		Antibiotic   immunosuppressive drug	<a href="https://doi.org/10.1007/S11306-016-025-6">10.1007/S11306-016-025-6</a>
<i>Saccharopolyspora erythraea</i>		Antibiotic   protein synthesis inhibitors   gastrointestinal agent	<a href="https://doi.org/10.1021/ACSSYNBIO.8B00372">10.1021/ACSSYNBIO.8B00372</a>
<i>Streptomyces venezuelae</i>		Antibiotic   protein synthesis inhibitors	<a href="https://doi.org/10.1016/J.JBIOTEC.2014.01.028">10.1016/J.JBIOTEC.2014.01.028</a>

<sup>1</sup> <https://doi.org/10.1016/j.micres.2021.126708>

# From metabolites to hypotheses (and back)

« Hey Wiki, how many compounds are **structurally similar** to compounds labeled as **antibiotics**? Please group the results by the parent taxon of the containing organism»

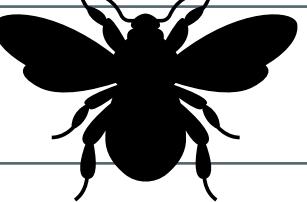
# From metabolites to hypotheses (and back)

« Hey Wiki, how many compounds are **structurally similar** to compounds labeled as **antibiotics**? Please group the results by the parent taxon of the containing organism»

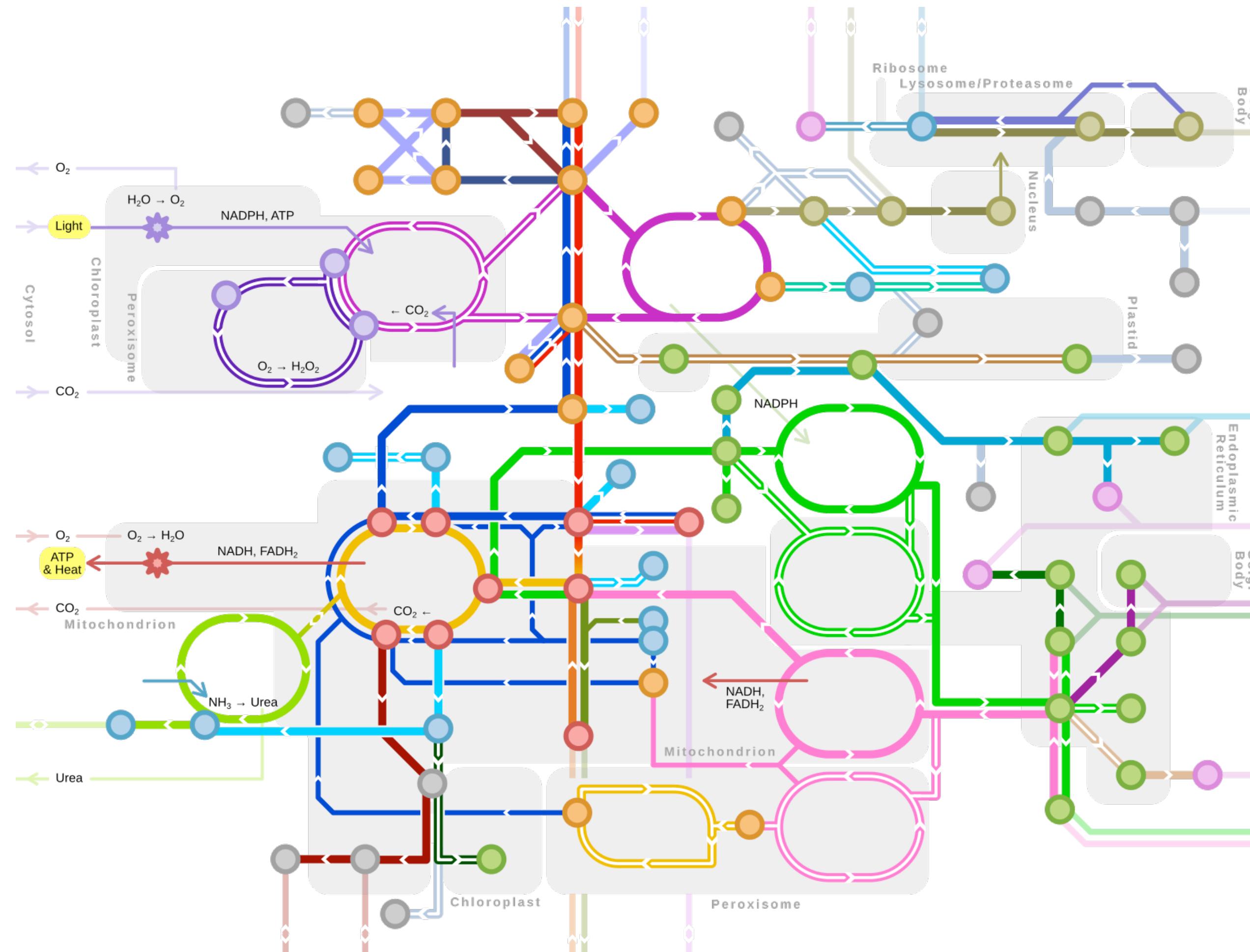
Parent taxon name	Count
<i>Streptomyces</i>	412
<i>Streptomycetaceae</i>	206
<i>Apis</i>	42
<i>Penicillium</i>	25
<i>Torrubiella</i>	25
<i>Micromonospora</i>	22
<i>Saccharopolyspora</i>	21
<i>Kitasatospora</i>	21
<i>Fusarium</i>	21
<i>Albifimbria</i>	21
<i>Aspergillus</i>	20

# From metabolites to hypotheses (and back)

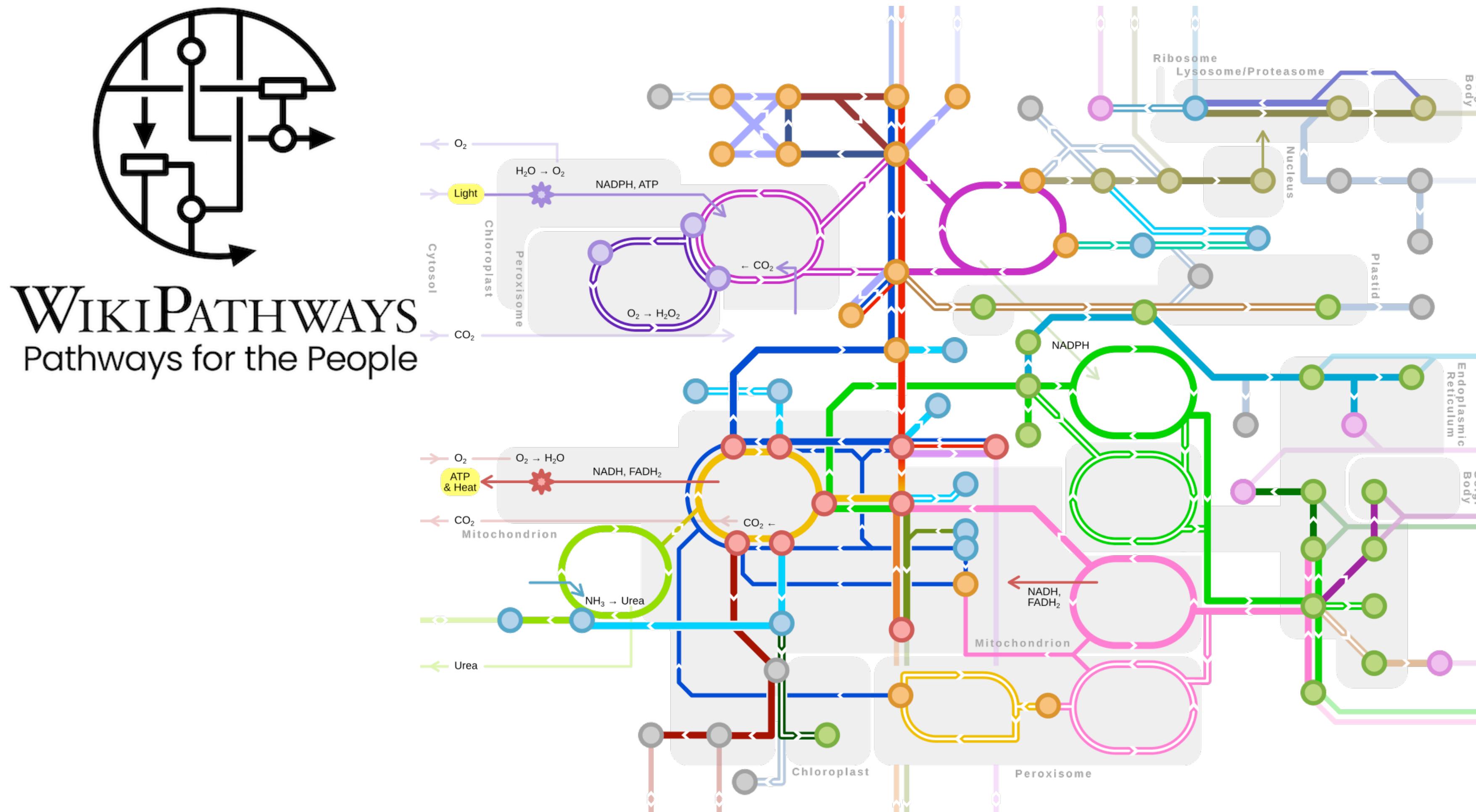
« Hey Wiki, how many compounds are **structurally similar** to compounds labeled as **antibiotics**? Please group the results by the parent taxon of the containing organism»

Parent taxon name	Count
<i>Streptomyces</i>	412
<i>Streptomycetaceae</i>	206
<i>Apis</i> 	42
<i>Penicillium</i>	25
<i>Torrubiella</i>	25
<i>Micromonospora</i>	22
<i>Saccharopolyspora</i>	21
<i>Kitasatospora</i>	21
<i>Fusarium</i>	21
<i>Albifimbria</i>	21
<i>Aspergillus</i>	20

# From metabolites to hypotheses (and back)



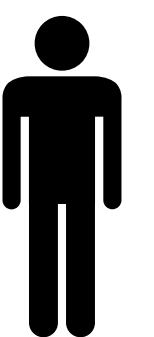
# From metabolites to hypotheses (and back)



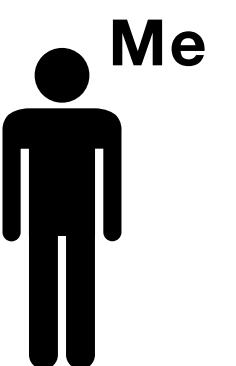
WIKIPATHWAYS  
Pathways for the People

« Hey Wiki, are the metabolites I annotated known to interact with the cholesterol biosynthesis pathway? »

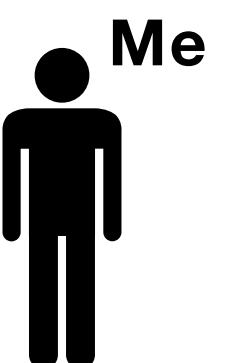
# **Sharing is caring**



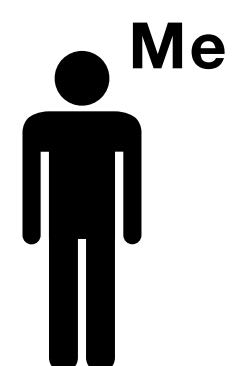
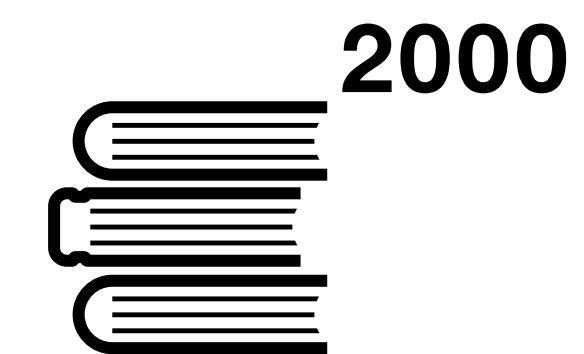
# Sharing is caring



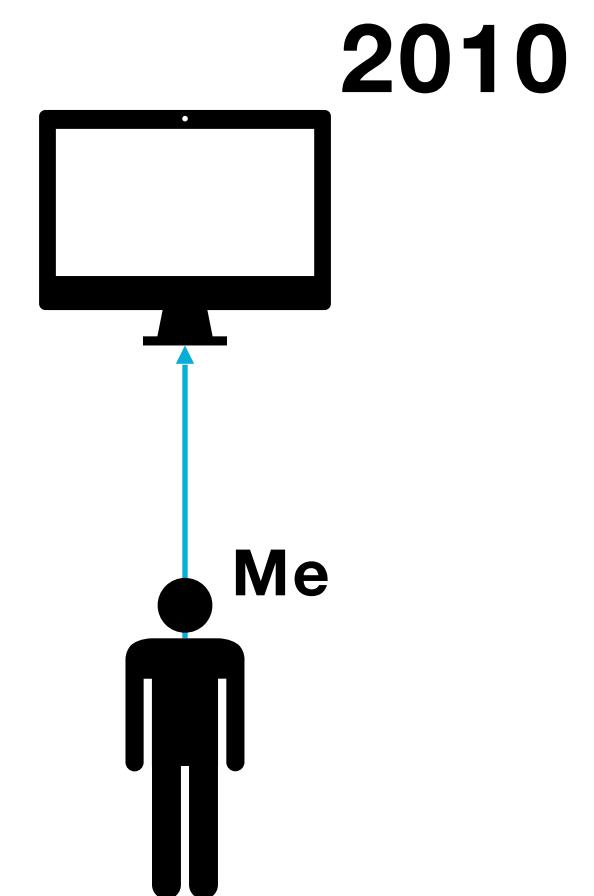
# Sharing is caring



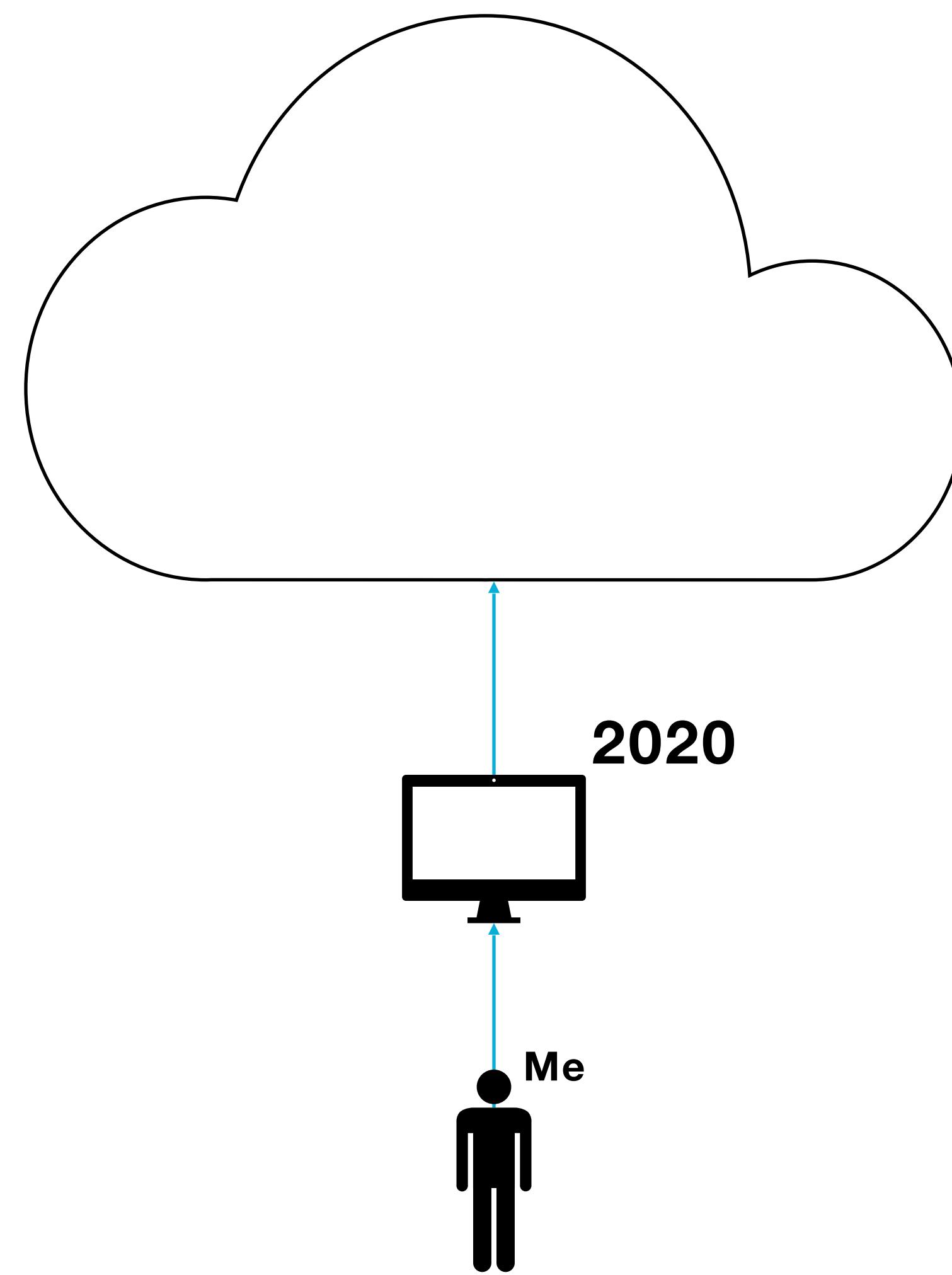
# Sharing is caring



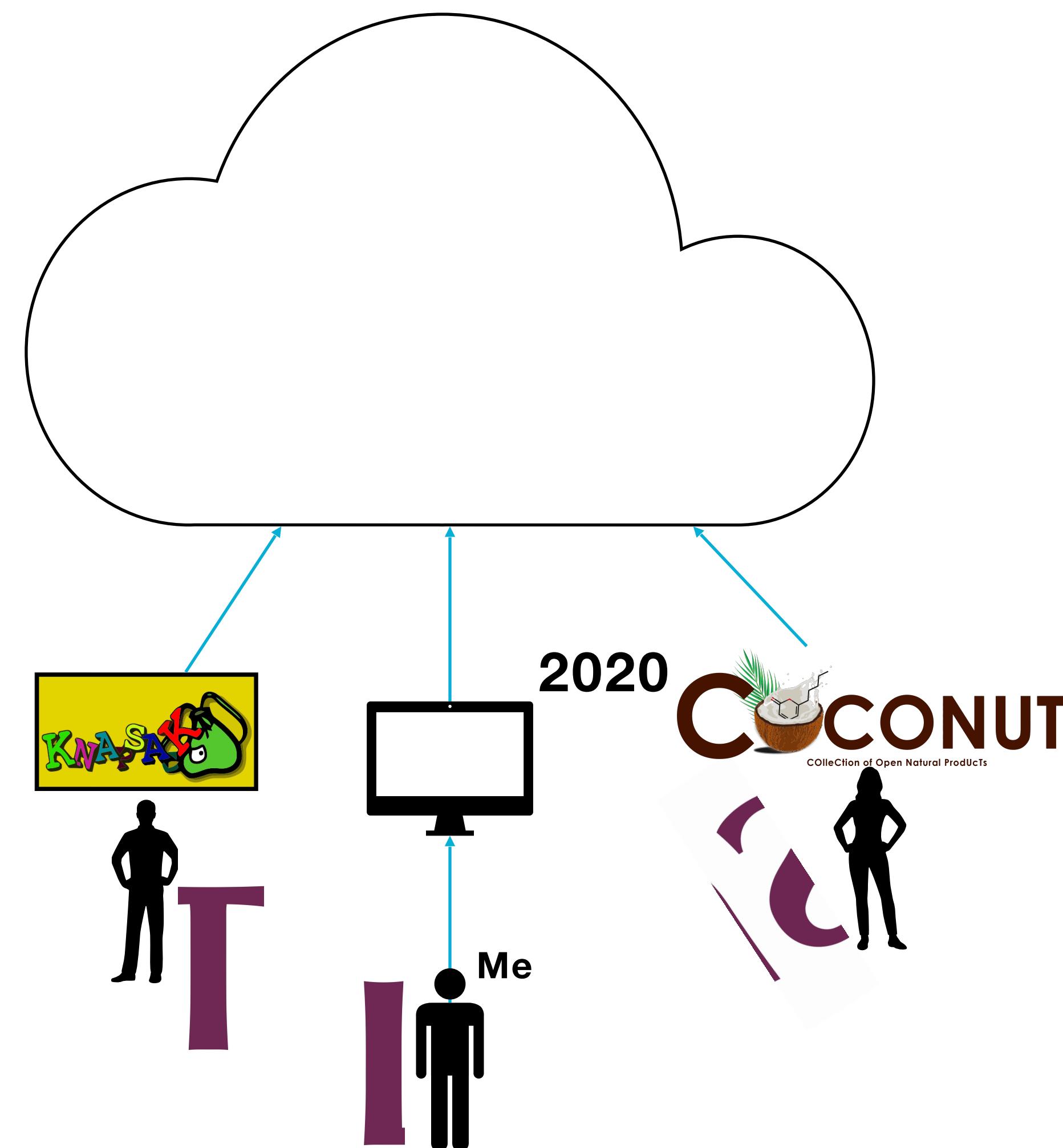
# Sharing is caring



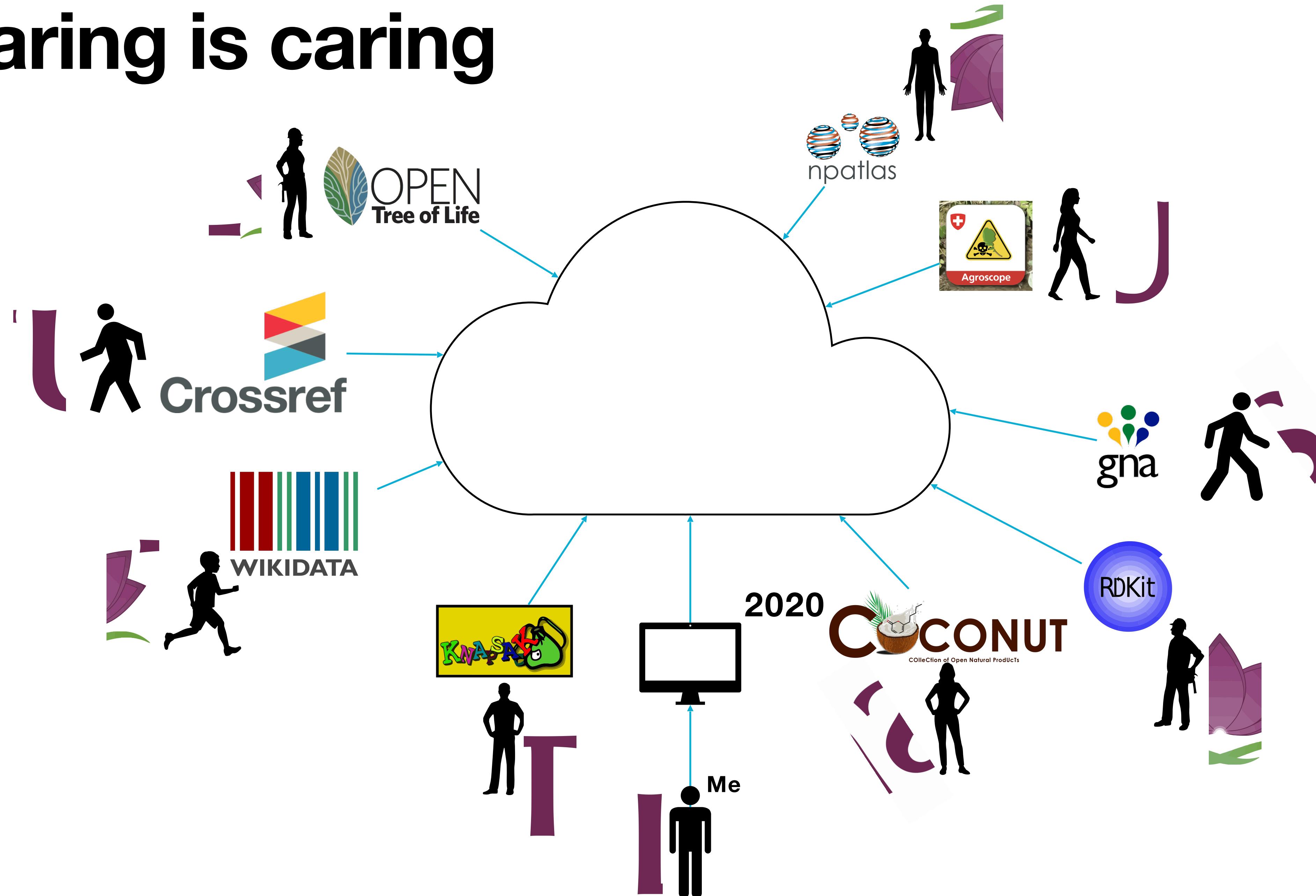
# Sharing is caring



# Sharing is caring



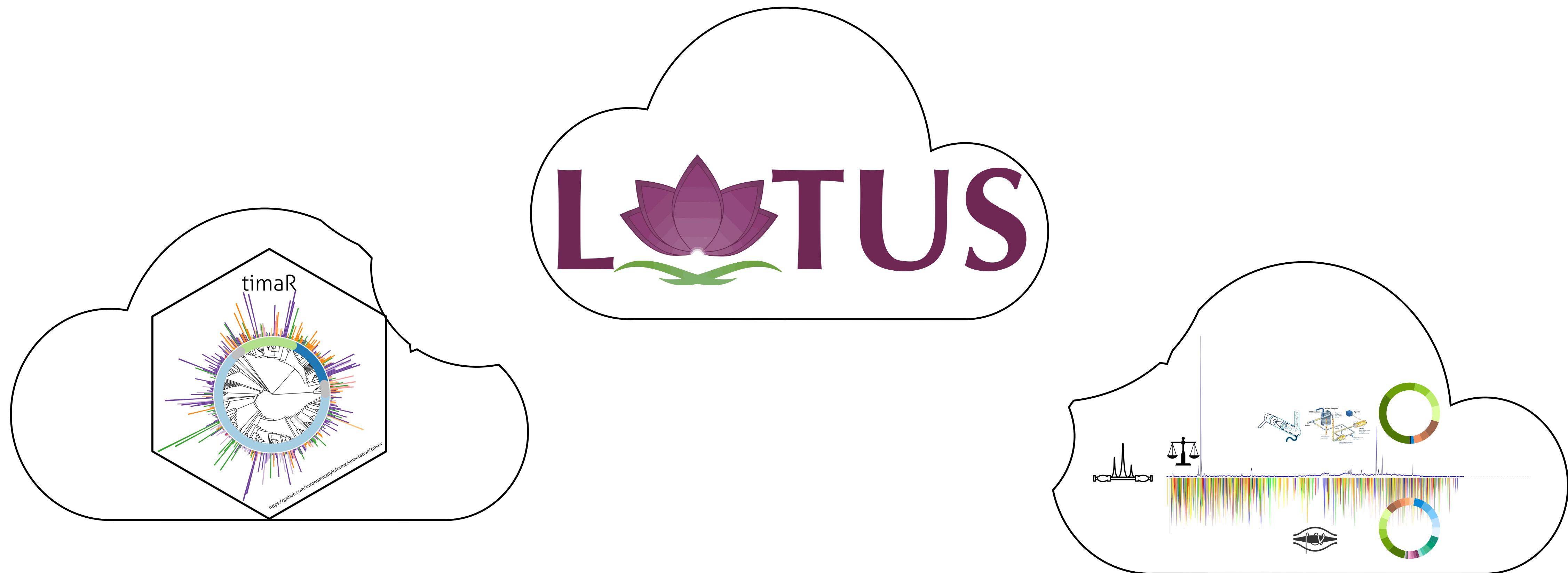
# Sharing is caring



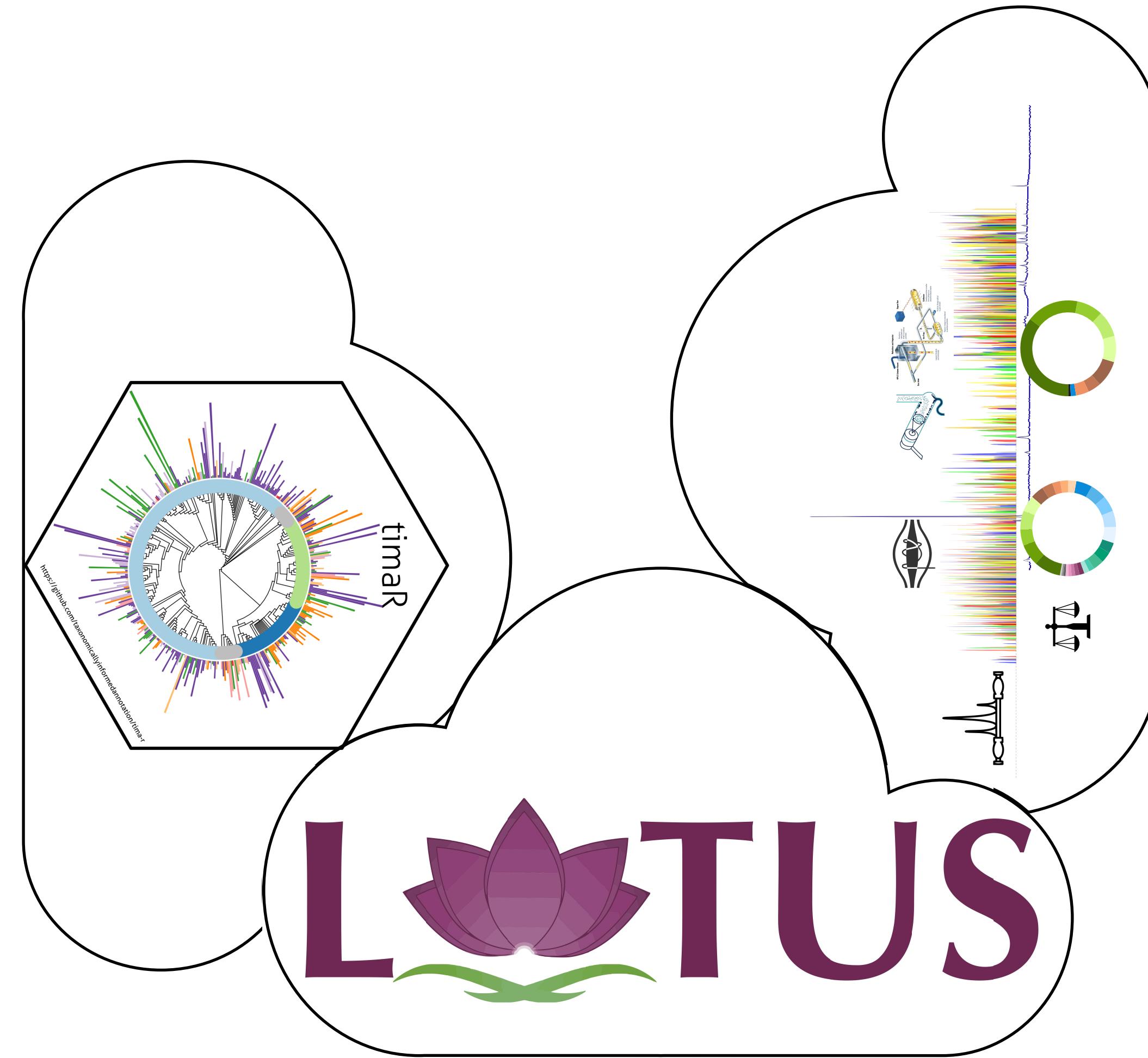
# **Swiss knife for metabolomics**



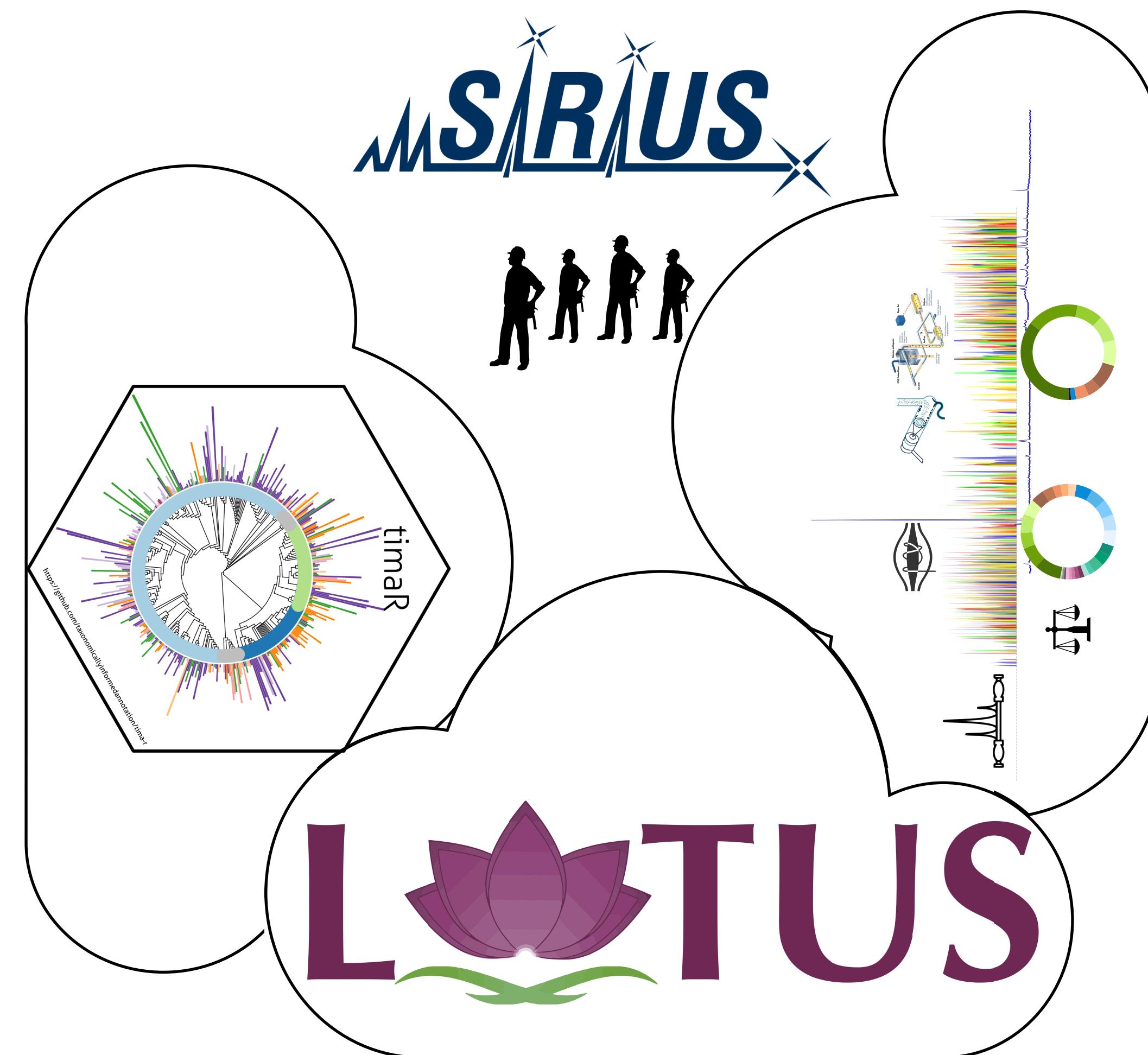
# Swiss knife for metabolomics



# Swiss knife for metabolomics

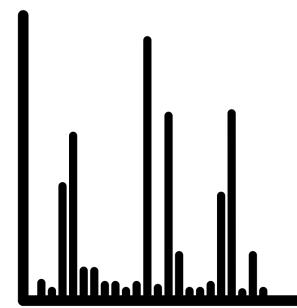


# Swiss knife for metabolomics



# Future research on the chemistry of Life

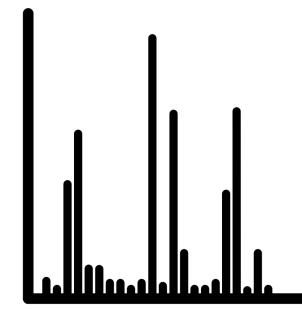
**Smarter acquisition at high throughput**



« Hey Wiki, have we  
seen this spectrum  
already? »

# Future research on the chemistry of Life

Smarter acquisition at high throughput



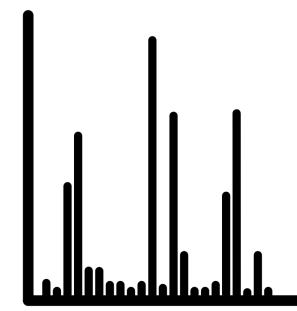
« Hey Wiki, have we  
seen this spectrum  
already? »



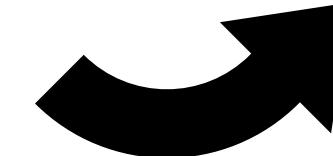
« It looks like  
something difficult to  
fragment, put some  
higher energy! »

# Future research on the chemistry of Life

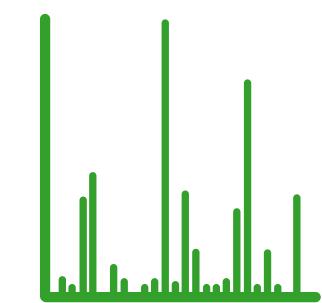
Smarter acquisition at high throughput



« Hey Wiki, have we seen this spectrum already? »



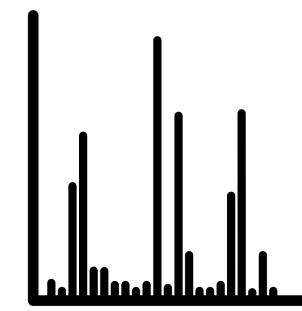
« It looks like something difficult to fragment, put some higher energy! »



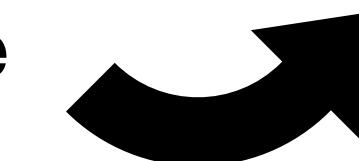
« This looks like a good fragmentation spectrum. Do we know something similar in human? »

# Future research on the chemistry of Life

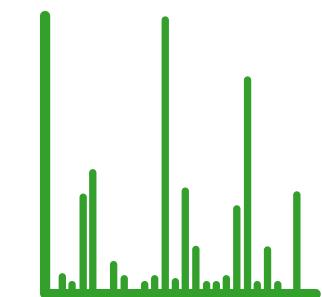
## Smarter acquisition at high throughput



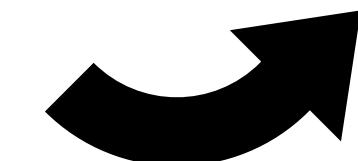
« Hey Wiki, have we seen this spectrum already? »



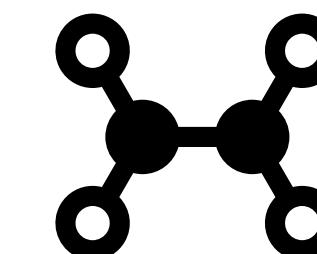
« It looks like something difficult to fragment, put some higher energy! »



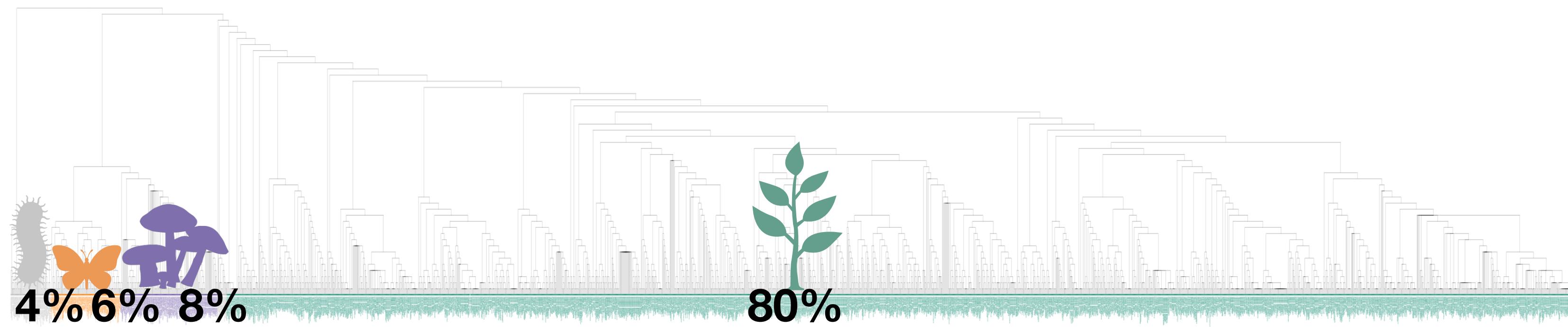
« This looks like a good fragmentation spectrum. Do we know something similar in human? »



« The closest match is a doping agent that was reported to interact with other detected metabolites also present in the sample. Maybe a new analog? »



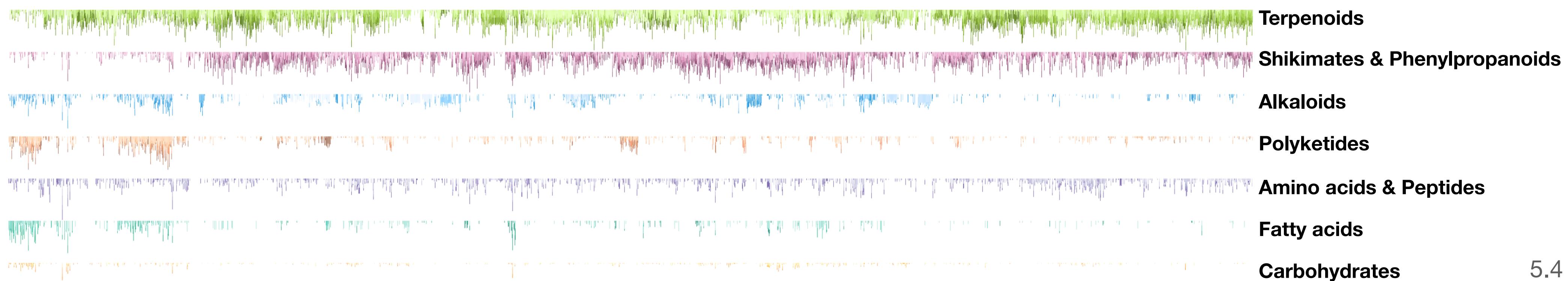
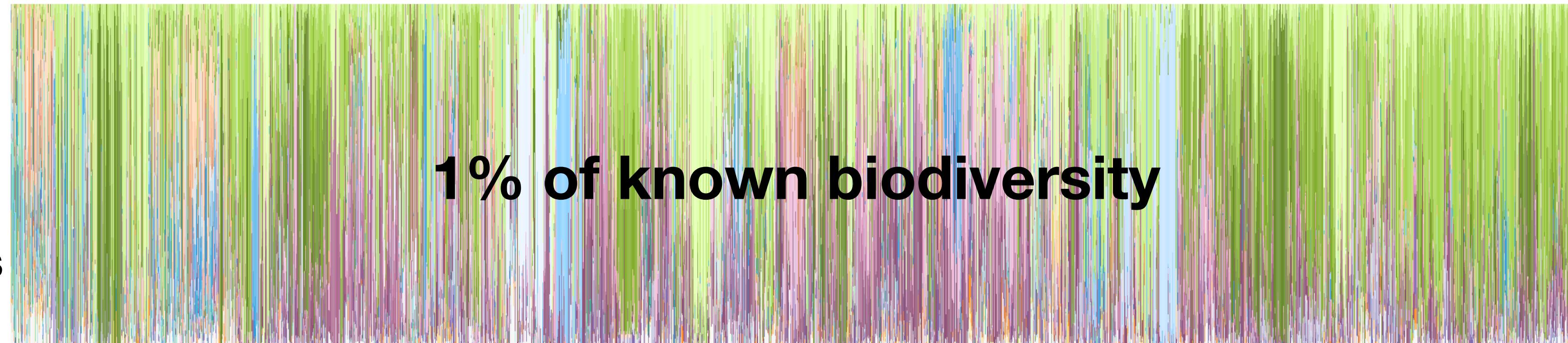
# Future research on the chemistry of Life



**7,848 species  
≥ 25 structures**

**32,907 species**

**2,956,117 species  
on Wikidata**



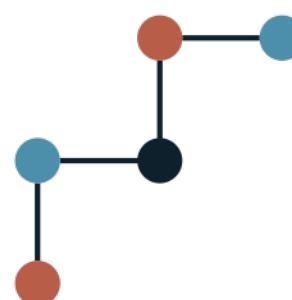
# Acknowledgements



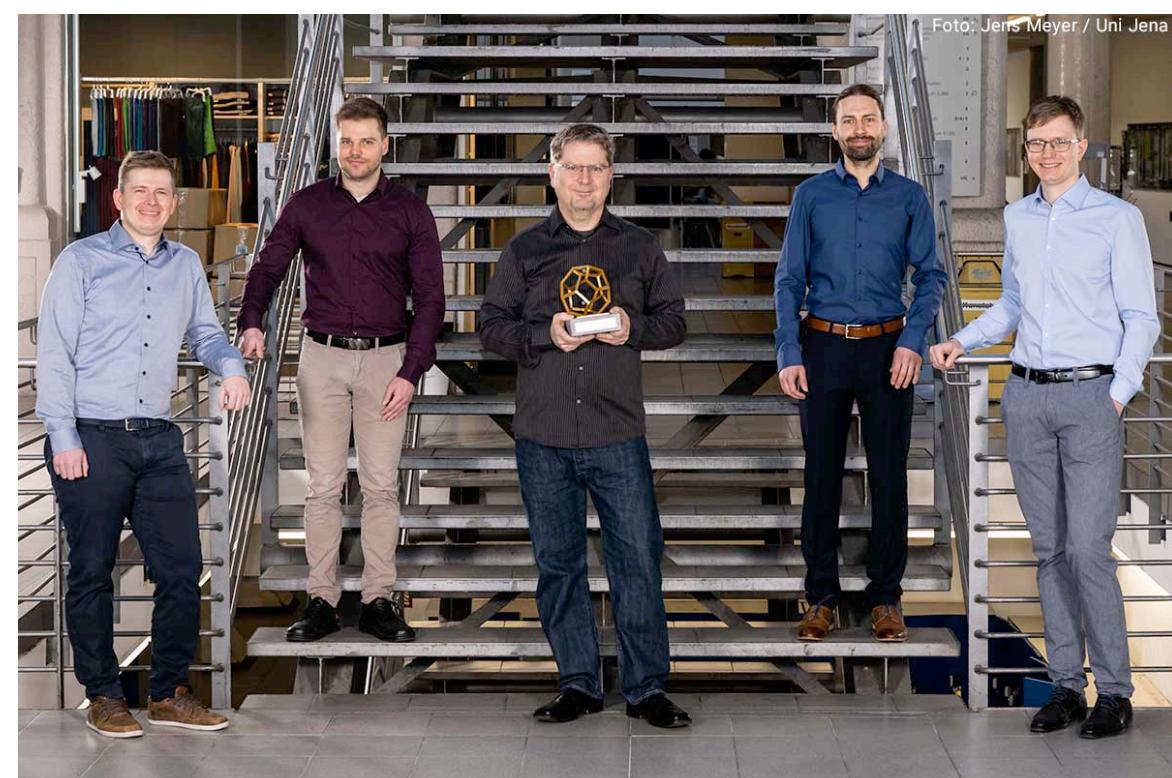
UNIVERSITÉ  
DE GENÈVE

FACULTY OF SCIENCE

Section of Pharmaceutical Sciences



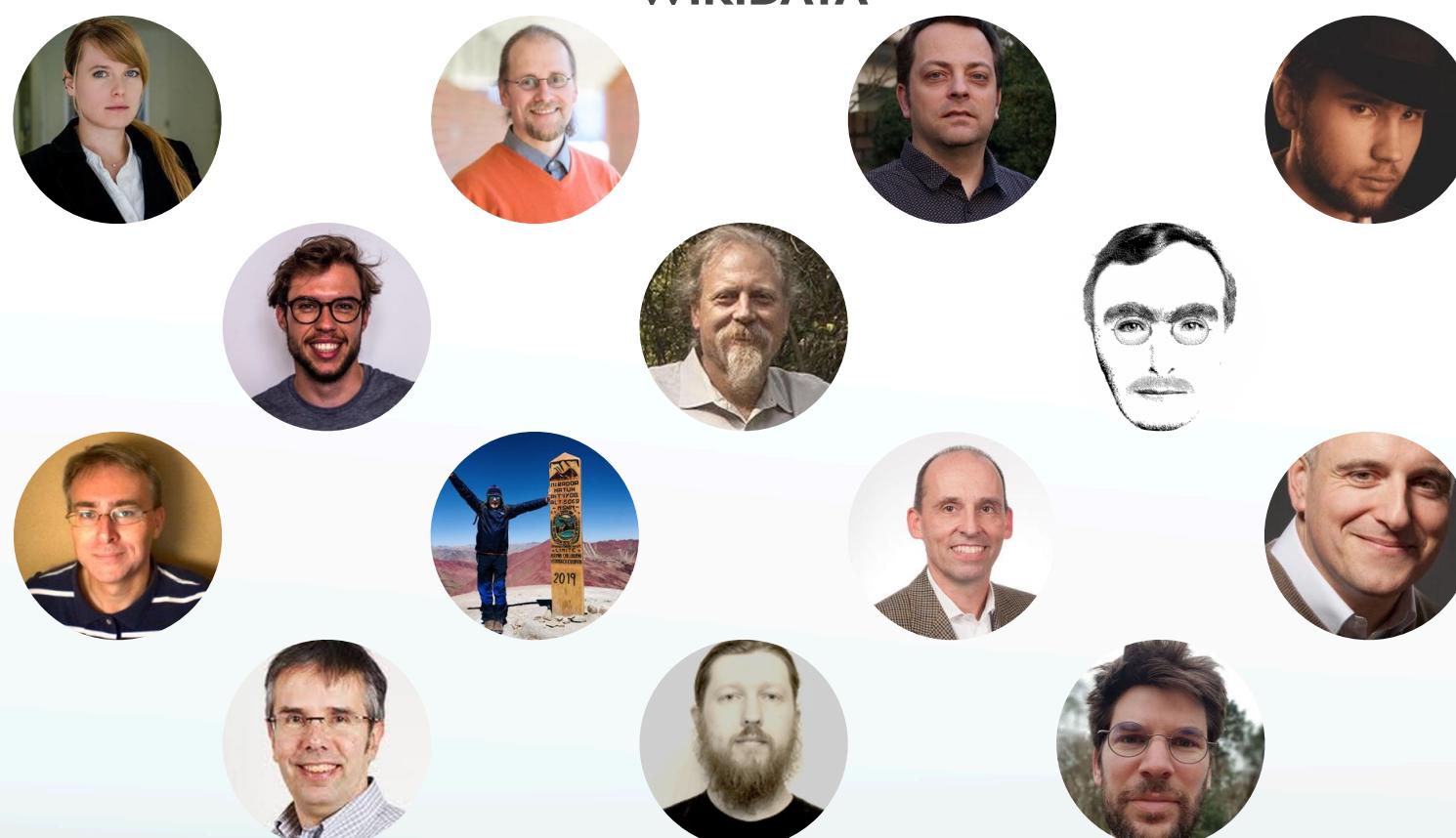
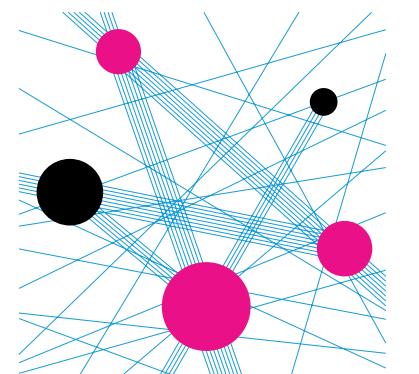
**Swiss National  
Science Foundation**



**Institute of Molecular Systems Biology**



**Prof. Nicola Zamboni**



**Contact**



[rutz@imsb.biol.ethz.ch](mailto:rutz@imsb.biol.ethz.ch)