LIFE SCIENCES
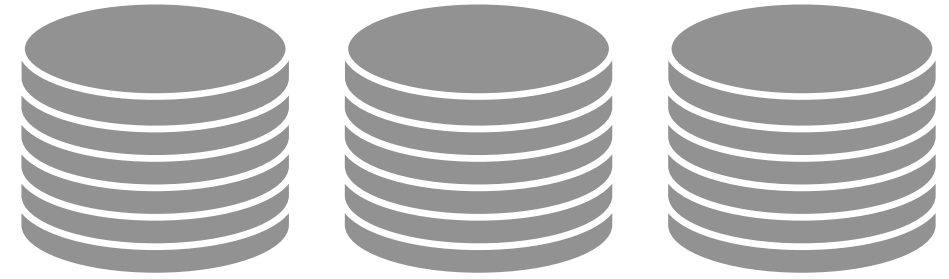PhD SCHOOL
GENEVA

UNIVERSITÉ
DE GENÈVE
FACULTY OF SCIENCE
Section of Pharmaceutical Sciences

The LOTUS Initiative: Integrating Natural Products to the Wikidata Knowledge Base

LS2 Annual Meeting 2022
Life Sciences in the 2020s:
Quantitation, Integration & Prediction
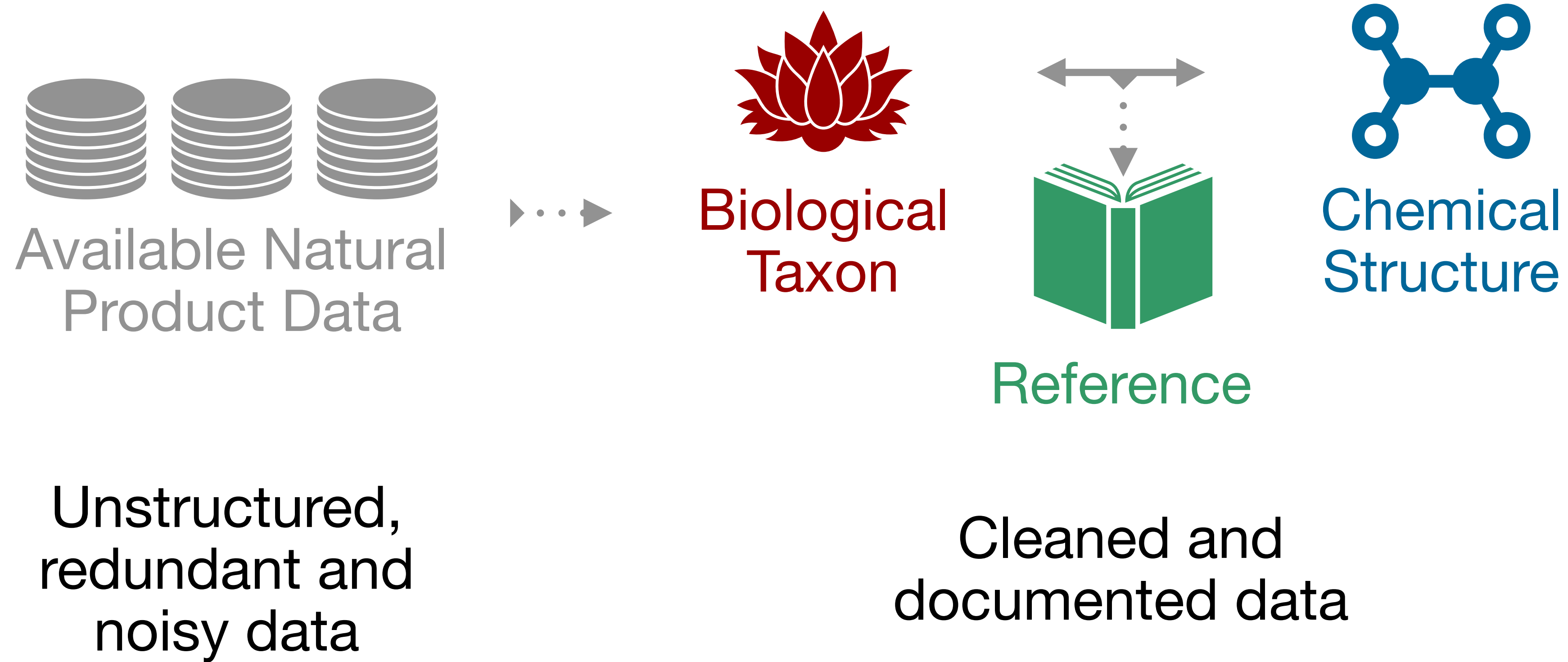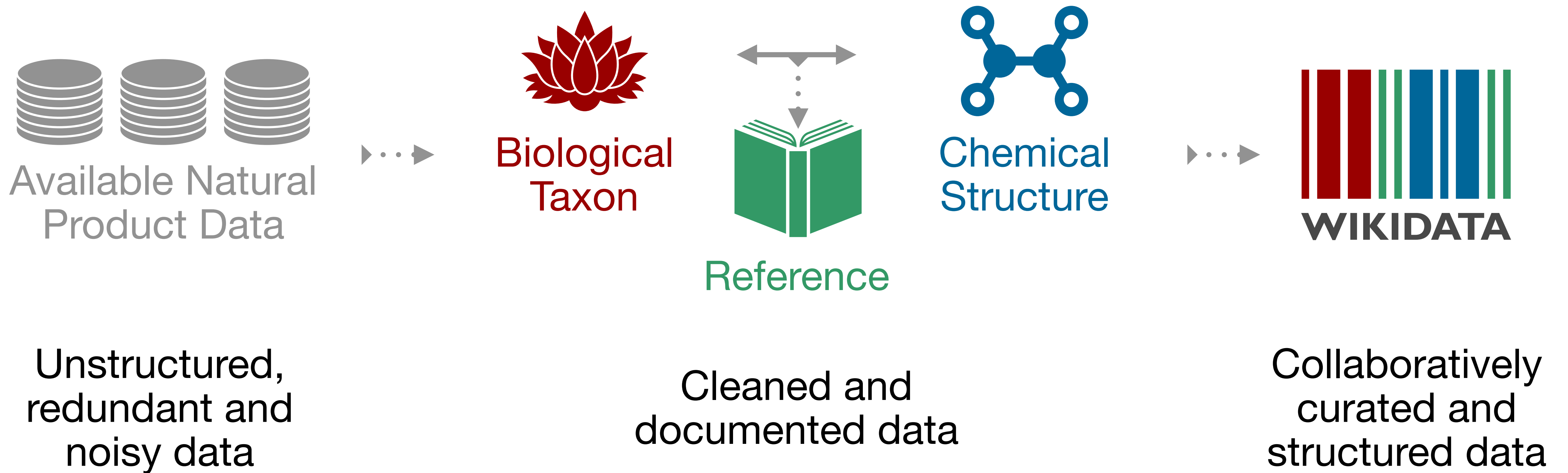
Campus Irchel, Zürich, 2022-04-22

The initiative

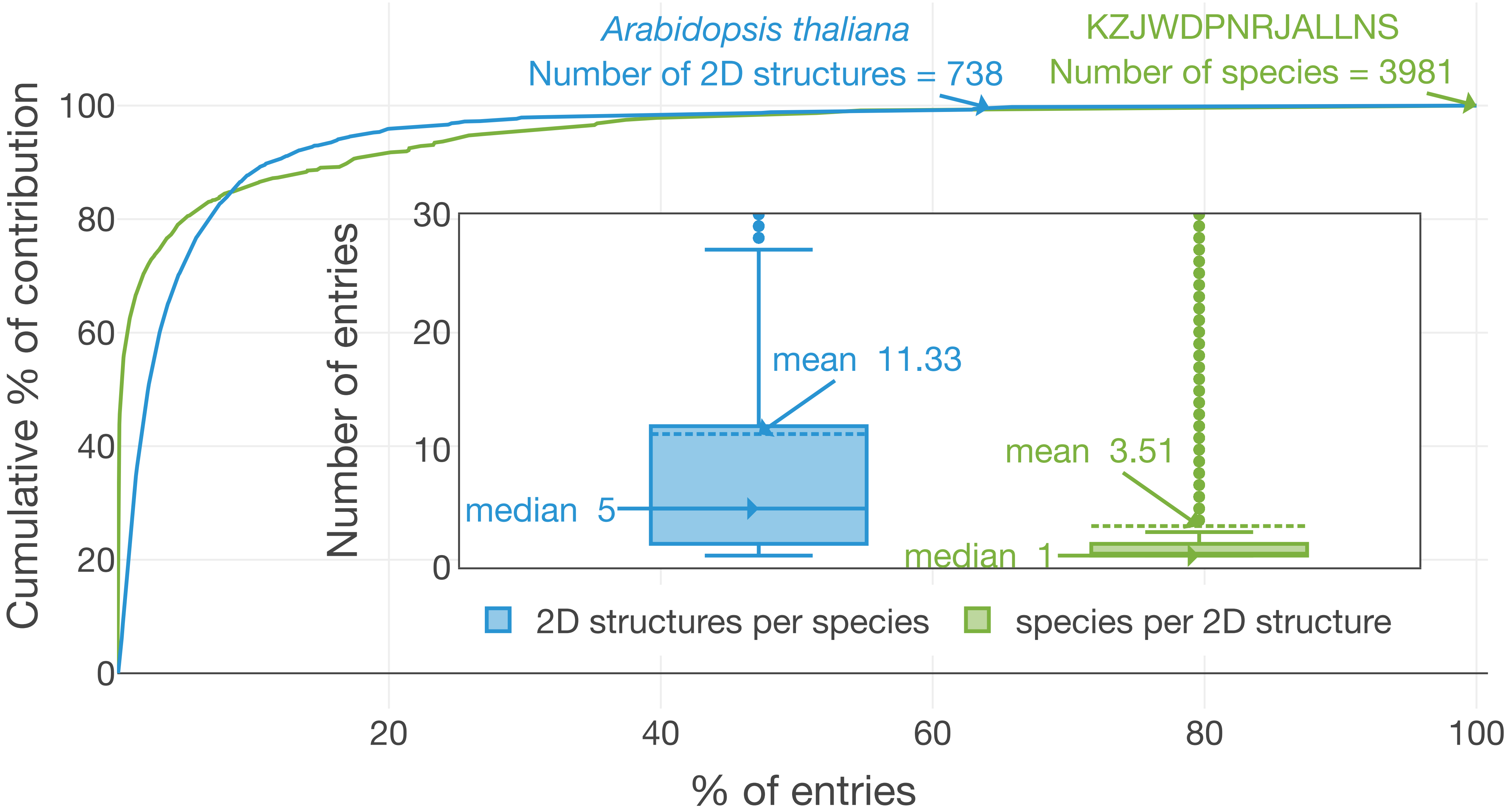Available Natural
Product Data

Unstructured,
redundant and
noisy data

The initiative

Available Natural Product Data

Biological Taxon

Reference

Chemical Structure

Unstructured, redundant and noisy data

Cleaned and documented data

The initiative

Available Natural Product Data → Biological Taxon · Reference · Chemical Structure → WIKIDATA

Unstructured, redundant and noisy data

Cleaned and documented data

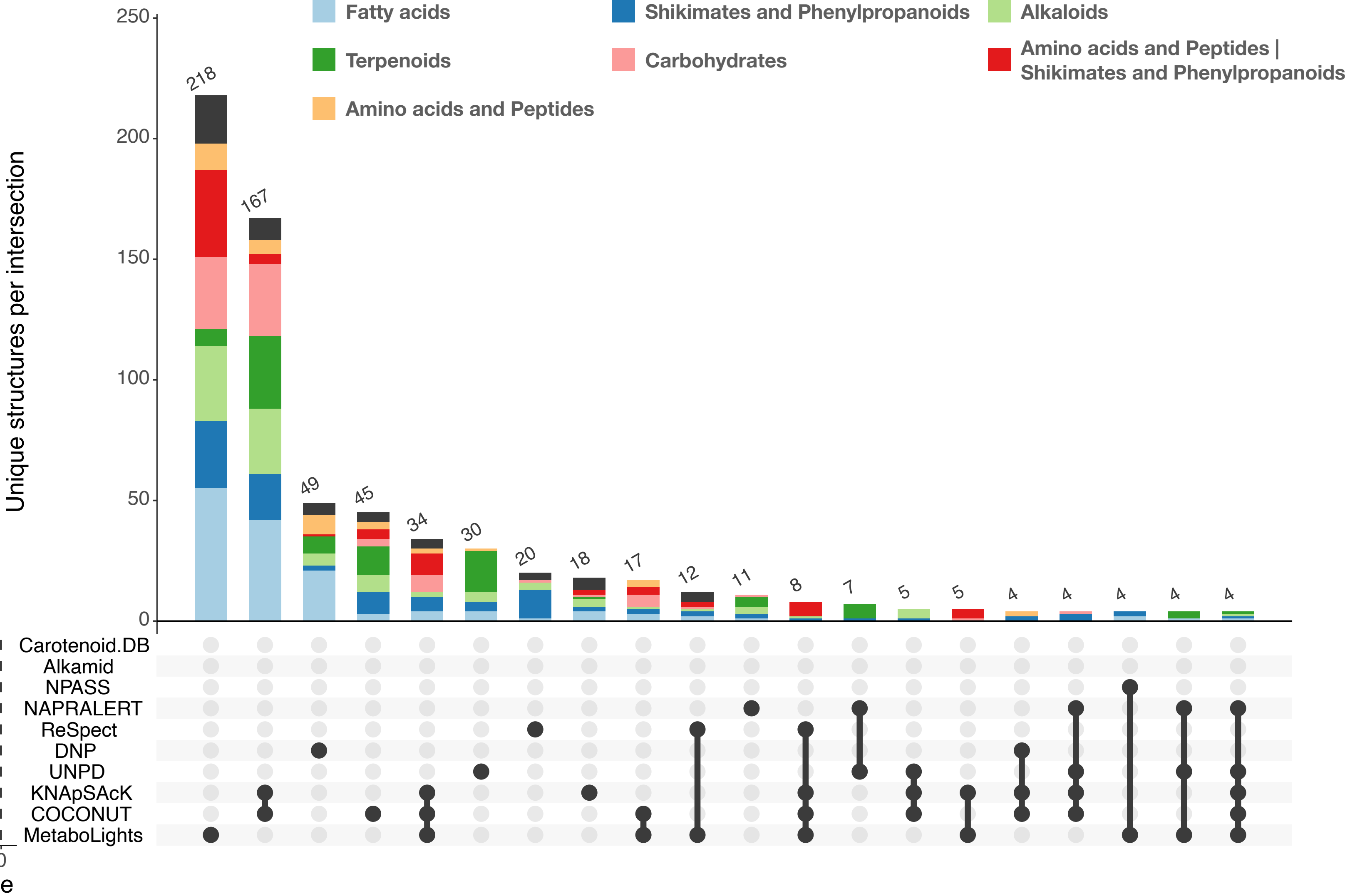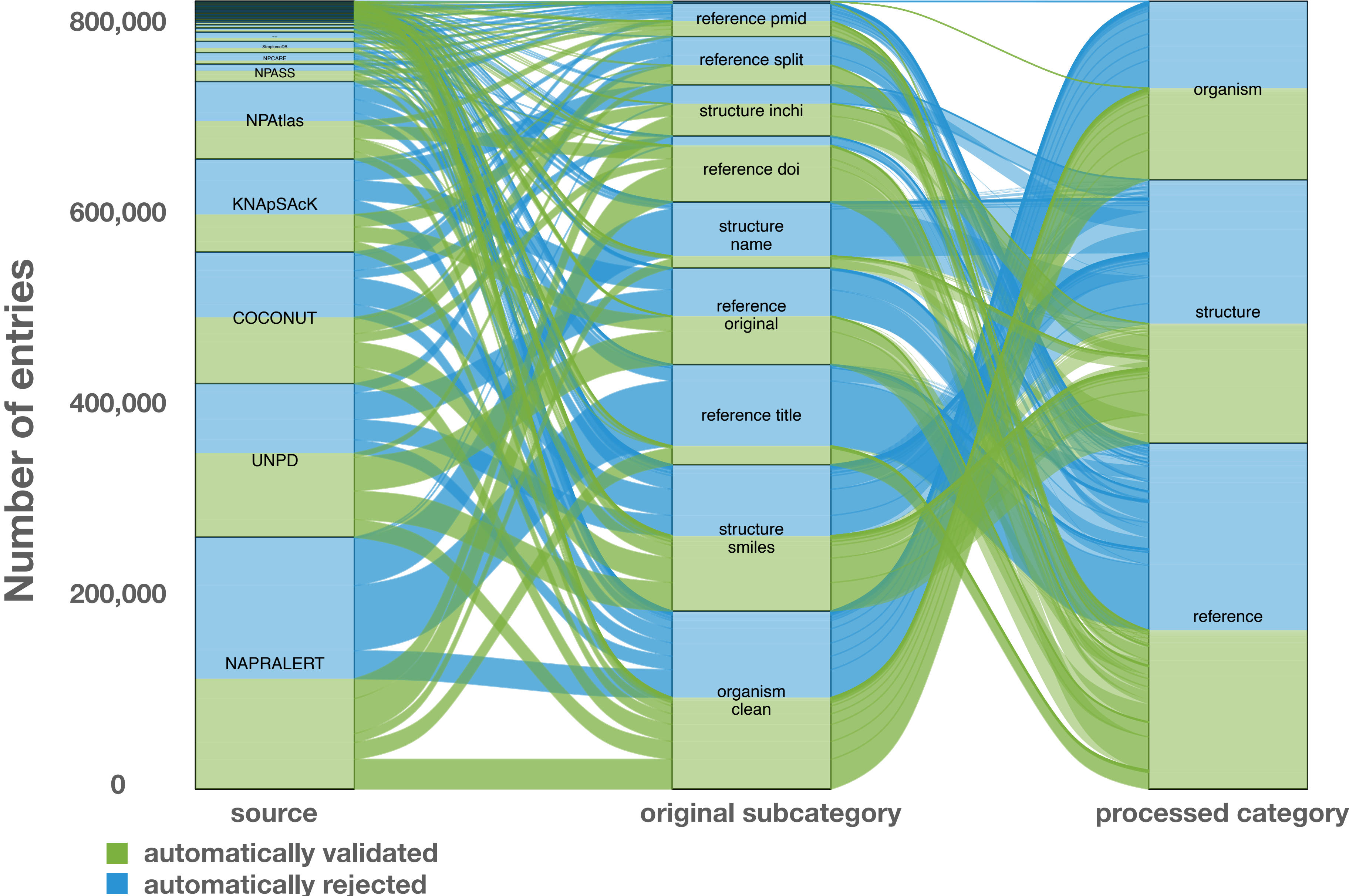Collaboratively curated and structured data

1

The initiative - *Quantitation*

The initiative - *Quantitation*

A. Structures (2D) reported in *Arabidopsis thaliana*

The initiative - *Integration*

Number of entries

source: StreptomeDB, NPCARE, NPASS, NPAtlas, KNApSAcK, COCONUT, UNPD, NAPRALERT

original subcategory: reference pmid, reference split, structure inchi, reference doi, structure name, reference original, reference title, structure smiles, organism clean

processed category: organism, structure, reference

automatically validated
automatically rejected

4

The initiative - *Integration*

1. Harmonization

2. Cleaning

3. Validation

4. Dissemination

Data Source

Data Harmonized

Structures

Organisms

References

Data Cleaned

Internal DB (Single Source of Truth)

LNPN

Open Natural Products Databases

Translation Cleaning Deduplication

Automated Validation Rules

Curated Documented Structure-Organism Pairs

WIKIDATA

The initiative - *Integration*

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

The initiative - *Integration*

« Hey Wiki, how many compounds are structurally similar to compounds labeled as antibiotics? (grouped by the parent taxon of the containing organism) »

parent_taxon_name	count
Streptomyces	412
Streptomycetaceae	206
Apis	42
Penicillium	25
Torrubiella	25
Micromonospora	22
Saccharopolyspora	21
Kitasatospora	21
Fusarium	21
Albifimbria	21
Aspergillus	20

https://w.wiki/3HMA

6

The initiative - *Integration*

« Hey Wiki, how many compounds are structurally similar to compounds labeled as antibiotics? (grouped by the parent taxon of the containing organism) »

parent_taxon_name	count
Streptomyces	412
Streptomycetaceae	206
Apis	42
Penicillium	25
Torrubiella	25
Micromonospora	22
Saccharopolyspora	21
Kitasatospora	21
Fusarium	21
Albifimbria	21
Aspergillus	20

https://w.wiki/3HMA

The initiative - *Prediction*

3D Structure

3D β-sitosterol (KZJWDPNRJALLNS-VJSFXXLFSA-N)

■ Presence   ■ Absence

The initiative - *Prediction*

■ Archaeplastida　　■ Fungi　　■ Metazoa　　■ Not attributed (Bacteria and Algae)
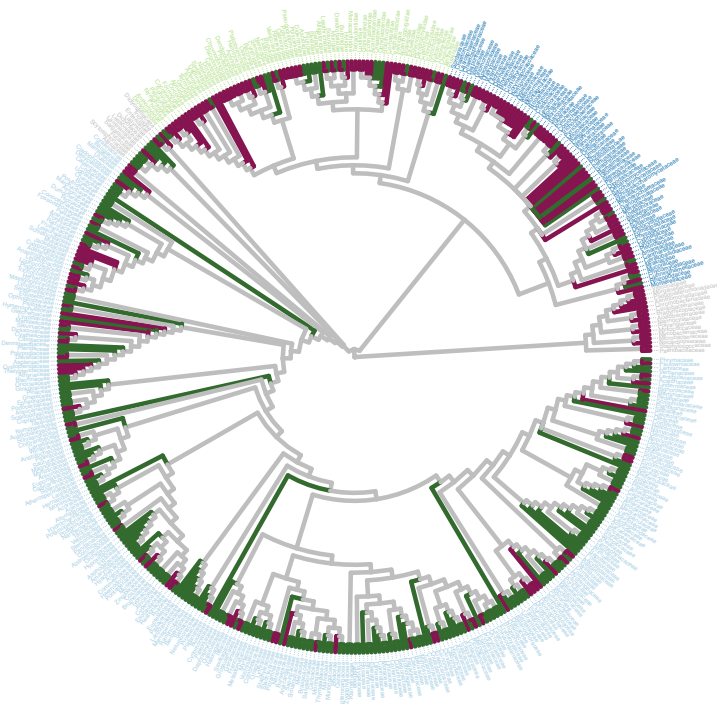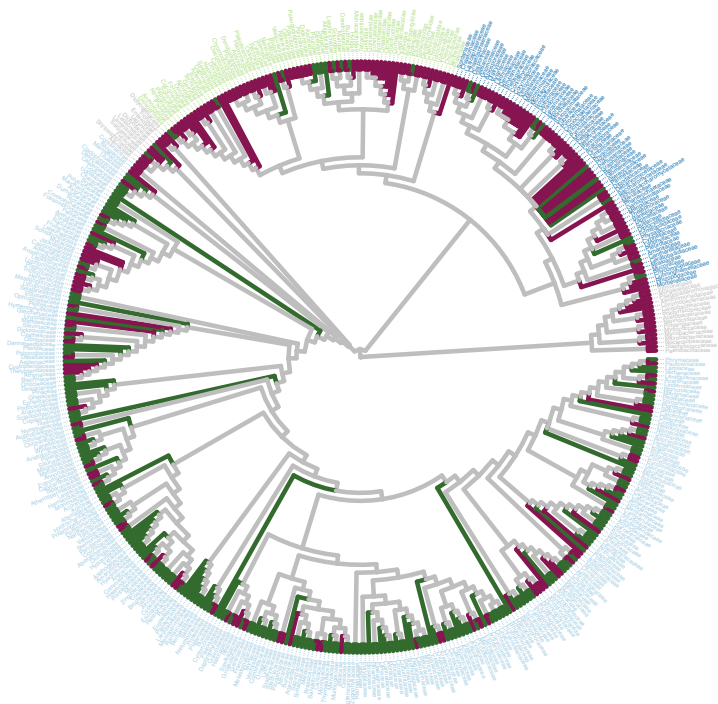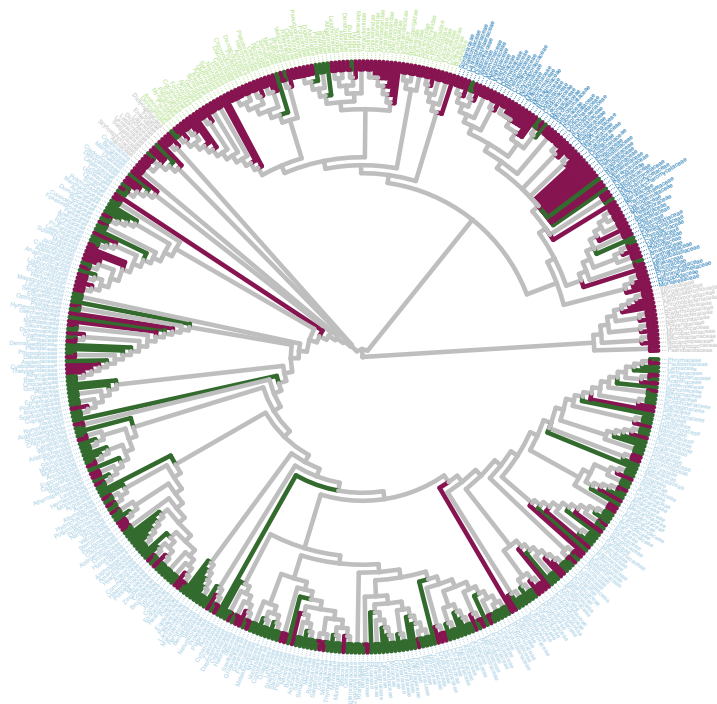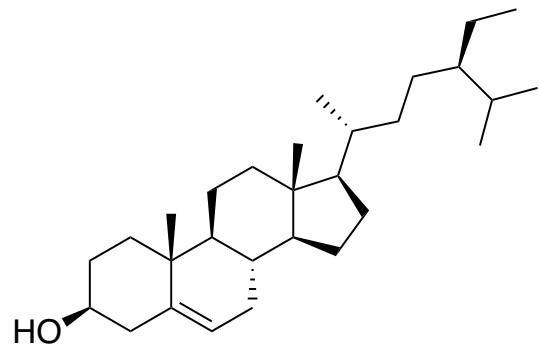
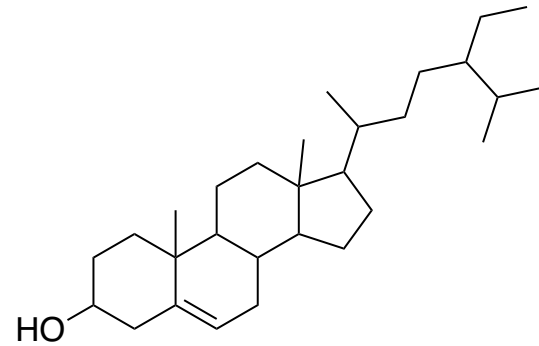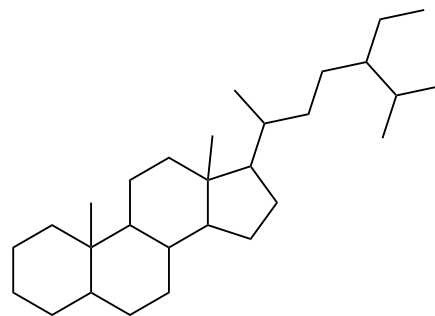3D Structure　　<　　2D Structure　　<　　Class　　<　　Superclass　　<　　Pathway
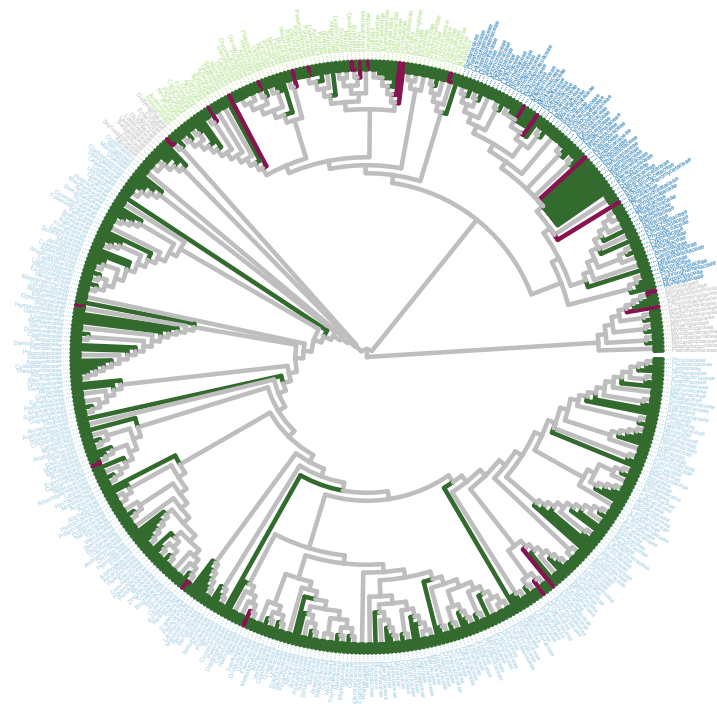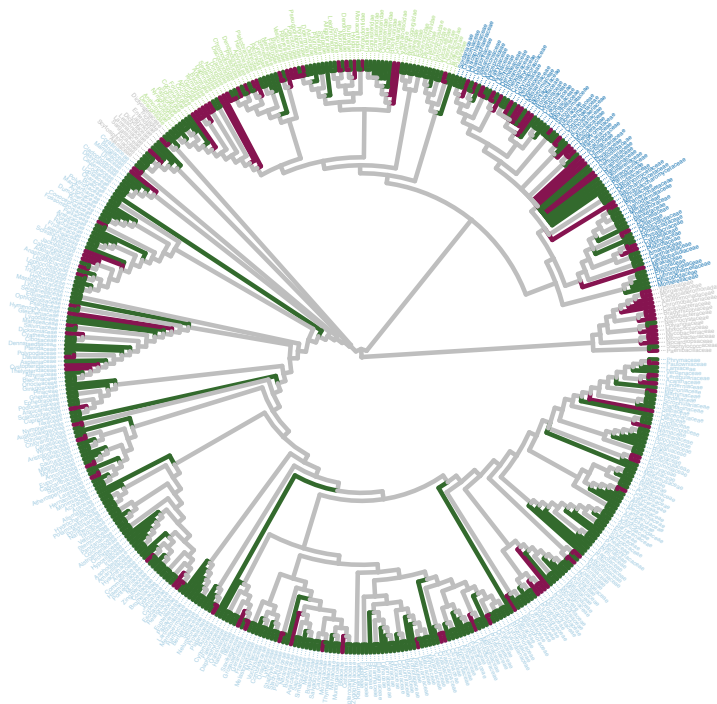
3D β-sitosterol
(KZJWDPNRJALLNS-
VJSFXXLFSA-N)
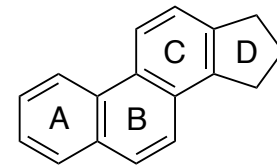
2D β-sitosterol
(KZJWDPNRJALLNS)

Stigmastanes

Steroids

Terpenoids
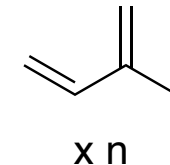
■ Presence　　■ Absence

The initiative - *Prediction*

Homo sapiens

	Similarity
Arabidopsis thaliana	651.7
Escherichia coli	581.7
Trypanosoma brucei	292.2
Vitis vinifera	209.8
Lotus corniculatus	195.0

The initiative - *Prediction*

Homo sapiens

	Similarity
Arabidopsis thaliana	651.7
Escherichia coli	581.7
Trypanosoma brucei	292.2
Vitis vinifera	209.8
Lotus corniculatus	195.0

Gentiana lutea

	Similarity
Swertia japonica	66.5
Arabidopsis thaliana	38.4
Gentiana macrophylla	37.0
Gentianella campestris	30.1
Gentiana scabra	28.4

The initiative - *Prediction*
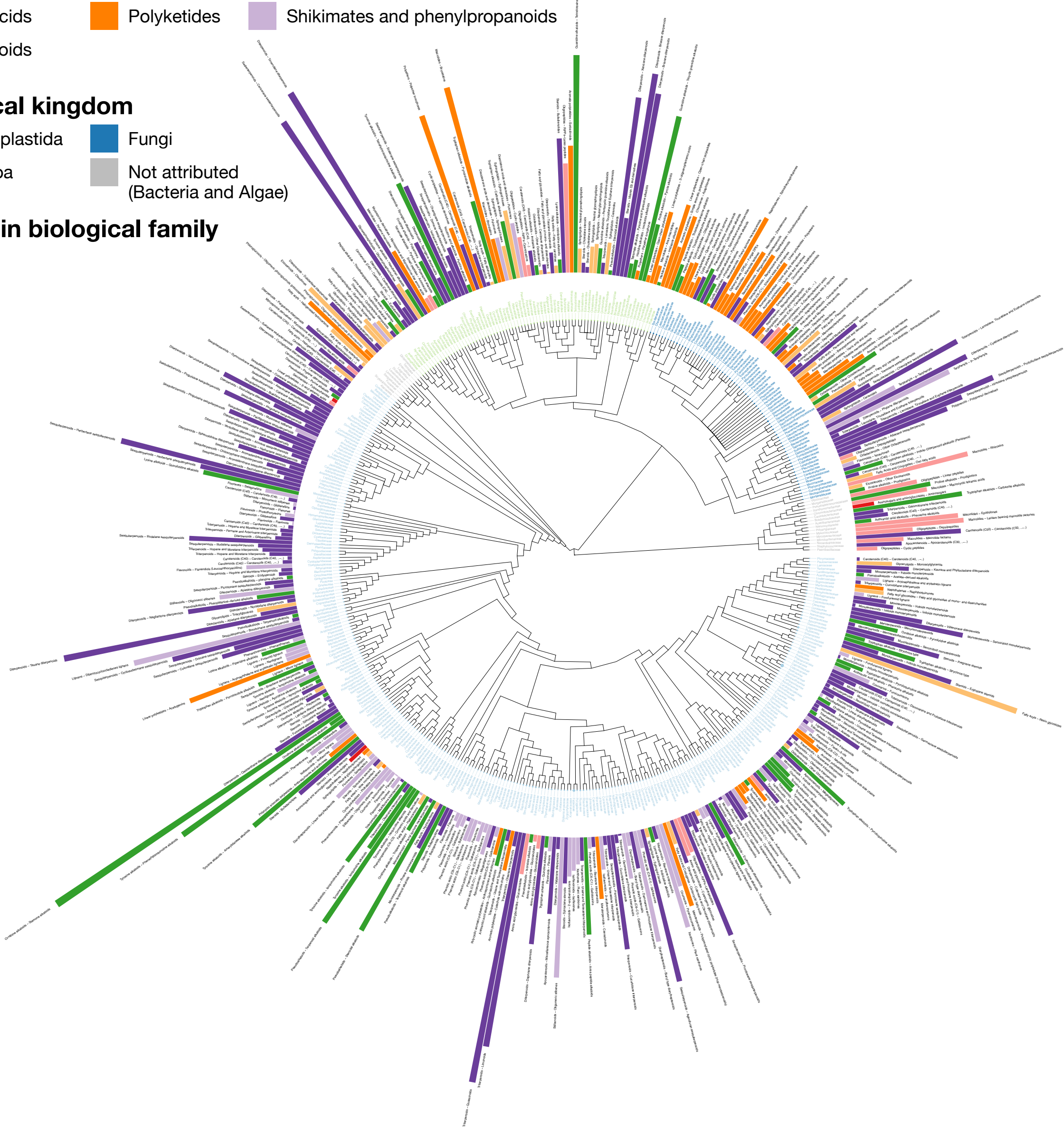

Chemical pathway
- Alkaloids
- Amino acids and peptides
- Carbohydrates
- Fatty acids
- Polyketides
- Shikimates and phenylpropanoids
- Terpenoids

Biological kingdom
- Archaeplastida
- Fungi
- Metazoa
- Not attributed (Bacteria and Algae)

Genera in biological family
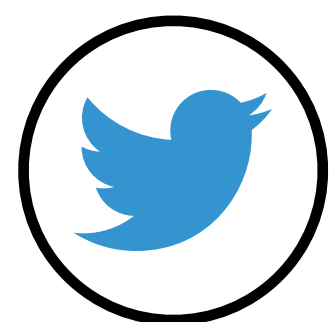- 200
- 400
- 600
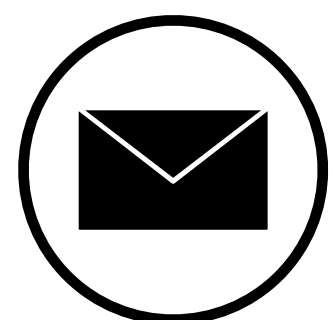
The initiative - *Contact*

lotus.naturalproducts.net/

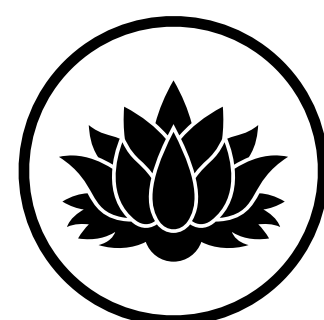wikidata.org/wiki/Wikidata:WikiProject_Chemistry/Natural_products

twitter.com/TheLOTUSInitia1

t.me/the_lotus_initiative

lotus-project@protonmail.com

lotus.nprod.net/