### Supplementary file 1. Wikidata Queries

#### Query 1 - *Arabidopsis thaliana*

This query answers to the following question:

What are the compounds found in Mouse-ear cress Arabidopsis thaliana (Q158695) or children taxa?

Link: <https://w.wiki/4Vcv>

#title: What are the compounds found in Mouse-ear cress Arabidopsis thaliana (Q158695) or children taxa?
SELECT DISTINCT ?structure ?structureLabel ?structure\_inchi WHERE {
 VALUES ?taxon {
 wd:Q158695 # You can remove the Qxxxxxx and hit Ctrl+space, type the first letters and it should autocomplete
 }
 ?children (wdt:P171\*) ?taxon. # Include children taxa
 ?structure wdt:P234 ?structure\_inchi; # Get the InChI
 (p:P703/ps:P703) ?children. # Found in given taxon/taxa
 SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}

#### Query 2 - *β*-sitosterol

This query answers to the following question:

Which organisms are known to contain β-sitosterol?

Link: <https://w.wiki/4VFn>

#title: Which organisms are known to contain Beta-Sitosterol (Q121802)?
SELECT DISTINCT ?taxon ?taxon\_name WHERE {
 VALUES ?compound {
 wd:Q121802 # You can remove the Qxxxxxx and hit Ctrl+space, type the first letters and it should autocomplete
 }
 ?compound (p:P703/ps:P703) ?taxon. # Found in taxon
 ?taxon wdt:P225 ?taxon\_name. # Get scientific name of the taxon
 SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}

#### Query 3 - *β*-sitosterol stereoisomers

This query answers to the following question:

Which organisms are known to contain which stereoisomers of β-sitosterol?

Link: <https://w.wiki/4VFq>

#title: Which organisms are known to contain stereoisomers of Beta-Sitosterol (Q121802)?
SELECT ?taxon\_name ?compound ?InChIKey
WITH {
 SELECT ?compound ?InChIKey WHERE {
 VALUES ?target {
 wd:Q121802
 }
 ?target wdt:P235 ?queryKey.
 ?compound wdt:P235 ?InChIKey.
 FILTER (REGEX(STR(?InChIKey), CONCAT("^", SUBSTR($queryKey,1,14), "-")))
 FILTER (?InChIKey != ?queryKey)
 }
} AS %compounds
WHERE {
 INCLUDE %compounds
 ?compound (wdt:P703/wdt:P225) ?taxon\_name.
 SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}

#### Query 4 - Pigments

This query answers to the following question:

Which pigments are found in which taxa, according to which reference?

Link: <https://w.wiki/4VFx>

#title: Which pigments are found in which taxa, according to which reference?
# special thanks goes to User:Lmichan for updating this information!
SELECT DISTINCT ?compound ?compoundLabel ?taxon ?taxonname ?DOI
WITH {
 SELECT ?compound WHERE {
 ?compound (wdt:P31\*/wdt:P279\*) wd:Q161179. # get pigments
 }
} AS %compounds
WITH {
 SELECT ?compound ?P703statement WHERE {
 INCLUDE %compounds
 ?compound p:P703 ?P703statement. # check for "found in taxon" statements
 }
} AS %P703statement
WITH {
 SELECT ?compound ?taxon ?DOI WHERE {
 INCLUDE %P703statement
 ?P703statement ps:P703 ?taxon ; # get the respective taxa
 prov:wasDerivedFrom / pr:P248 [ # get the reference supporting that statement
 wdt:P356 ?DOI # get the DOI for the reference
 ] .
 }
} AS %taxa
WHERE {
 {
 INCLUDE %taxa

 ?taxon wdt:P225 ?taxonname . # get the taxon name
 }
 ?compound rdfs:label ?compoundLabel . # get compound labels
 FILTER (LANG(?compoundLabel) = "en") . # filter for English
}
ORDER BY ASC(?compoundLabel)
LIMIT 10000

#### Query 5 - Sister taxon compounds

This query answers to the following question:

What are examples of organisms where compounds were found in an organism sharing the same parent taxon, but not in the organism itself?

Link: <https://w.wiki/4Wt3>

#title: What are examples of organisms where compounds were found in an organism sharing the same parent taxon, but not the organism itself?
SELECT DISTINCT ?compound ?compoundLabel ?taxonname\_with\_compound ?taxonname\_without\_compound ?parent\_taxon WITH{
 SELECT DISTINCT ?compound ?taxon\_with\_compound ?parent\_taxon
 WHERE {
 ?compound wdt:P235 ?inchikey.
 SERVICE bd:sample { ?compound wdt:P703 ?taxon\_with\_compound. bd:serviceParam bd:sample.limit 1000 }
 ?taxon\_with\_compound wdt:P171 ?parent\_taxon.
 }
} AS %taxon\_with\_compound
WITH
{
 SELECT DISTINCT ?taxon\_without\_compound ?parent\_taxon ?compound
 WHERE {
 INCLUDE %taxon\_with\_compound
 ?taxon\_without\_compound wdt:P171 ?parent\_taxon.
 FILTER (?taxon\_with\_compound != ?taxon\_without\_compound)
 }
} AS %taxon2
WHERE {
 INCLUDE %taxon\_with\_compound
 INCLUDE %taxon2
 FILTER NOT EXISTS {?compound wdt:P703 ?taxon\_without\_compound.}
 ?taxon\_with\_compound wdt:P225 ?taxonname\_with\_compound.
 ?taxon\_without\_compound wdt:P225 ?taxonname\_without\_compound.
 ?compound rdfs:label ?compoundLabel.
 FILTER(LANG(?compoundLabel) = "en").
}

#### Query 6 - *Zephyranthes* sister taxon compounds

This query answers to the following question:

Which Zephyranthes spp. lack compounds known from ≥ 2 other Zephyranthes?

Link: <https://w.wiki/4VG3>

#title: Which Zephyranthes (Q191364) spp. lack compounds known from at least two species in the genus?
PREFIX target: <http://www.wikidata.org/entity/Q191364> # Zephyranthes
SELECT DISTINCT ?compound ?compoundLabel ?taxon\_with\_compound ?another\_taxon\_with\_compound ?taxon\_without\_compound WITH {
 SELECT DISTINCT ?compound ?taxon\_YES\_1 ?taxon\_YES\_2
 WHERE {
 ?compound wdt:P703 ?taxon\_YES\_1 .
 ?compound wdt:P703 ?taxon\_YES\_2 .
 ?taxon\_YES\_1 wdt:P171 target: .
 ?taxon\_YES\_2 wdt:P171 target: .
 FILTER (?taxon\_YES\_2 != ?taxon\_YES\_1)
 }
} AS %taxa\_with\_compound
WITH
{
 SELECT DISTINCT ?taxon\_NO ?compound
 WHERE {
 INCLUDE %taxa\_with\_compound
 ?taxon\_NO wdt:P171 target: .
 FILTER (?taxon\_YES\_1 != ?taxon\_NO)
 }
} AS %taxon\_without\_compond
WHERE {
 INCLUDE %taxa\_with\_compound
 INCLUDE %taxon\_without\_compond
 FILTER NOT EXISTS { ?compound wdt:P703 ?taxon\_NO .}
 VALUES ?classes {
 wd:Q11173
 wd:Q59199015
 }
 ?taxon\_YES\_1 wdt:P225 ?taxon\_with\_compound .
 ?taxon\_YES\_2 wdt:P225 ?another\_taxon\_with\_compound .
 ?taxon\_NO wdt:P225 ?taxon\_without\_compound .
 ?compound (wdt:P31\*/wdt:P279\*) ?classes .
 ?compound rdfs:label ?compoundLabel.
 FILTER(LANG(?compoundLabel) = "en").
}

#### Query 7 - Antibiotic-like compounds

This query answers to the following question:

How many compounds are structurally similar to compounds labeled as antibiotics? Results are grouped by the parent taxon of the organism they were found in.

Link: <https://w.wiki/4VG4>

#title: How many compounds are structurally similar to compounds labeled as antibiotics? Results are grouped by the parent taxon of the organism they were found in.
PREFIX sachem: <http://bioinfo.uochb.cas.cz/rdf/v1.0/sachem#> # prefixes needed for structural similarity search
PREFIX idsm: <https://idsm.elixir-czech.cz/sparql/endpoint/>
SELECT ?parent\_taxon ?parent\_taxon\_name (COUNT(DISTINCT ?compound) AS ?count) WHERE {
 SERVICE idsm:wikidata {
 VALUES ?CUTOFF {
 "0.9"^^xsd:double
 }
 SERVICE <https://query.wikidata.org/bigdata/namespace/wdq/sparql> {
 VALUES ?MESH {
 "D000900"
 }
 ?antibiotic ((wdt:P279\*)/wdt:P2868/wdt:P486) ?MESH;
 wdt:P233 ?smiles.
 }
 ?compound sachem:similarCompoundSearch \_:b40.
 \_:b40 sachem:query ?smiles;
 sachem:cutoff ?CUTOFF.
 }
 hint:Prior hint:runFirst "true"^^xsd:boolean.
 ?compound wdt:P703 ?taxon.
 ?taxon wdt:P171 ?parent\_taxon.
 OPTIONAL { ?parent\_taxon wdt:P225 ?parent\_taxon\_name. }
 SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
GROUP BY ?parent\_taxon ?parent\_taxon\_name
ORDER BY DESC (?count)

#### Query 8 - Indolic scaffold

This query answers to the following question:

Which organisms contain indolic scaffolds? Count occurrences, group and order the results by the parent taxon.

Link: <https://w.wiki/4VG9>

#title: Which organisms contain indolic scaffolds? Count occurences, group and order the results by the parent taxon.
PREFIX sachem: <http://bioinfo.uochb.cas.cz/rdf/v1.0/sachem#> # prefixes needed for structural similarity search
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX p: <http://www.wikidata.org/prop/>
PREFIX idsm: <https://idsm.elixir-czech.cz/sparql/endpoint/>
SELECT ?parent\_taxon ?parent\_taxon\_name (COUNT(DISTINCT ?compound) AS ?count) WHERE {
 SERVICE idsm:wikidata {
 VALUES ?SUBSTRUCTURE {
 "NCCC1=CNC2=C1C=CC=C2" # indolic scaffold
 }
 ?compound sachem:substructureSearch \_:b16.
 \_:b16 sachem:query ?SUBSTRUCTURE.
 }
 hint:Prior hint:runFirst "true"^^xsd:boolean.
 ?compound p:P703 ?statement;
 wdt:P235 ?inchikey.
 ?statement ps:P703 ?taxon.
 ?taxon wdt:P171 ?parent\_taxon.
 ?parent\_taxon wdt:P225 ?parent\_taxon\_name.
 SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
GROUP BY ?parent\_taxon ?parent\_taxon\_name
ORDER BY DESC (?count)

#### Query 9 - Bioactive compounds in Actinobacteria

This query answers to the following question:

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

Link: <https://w.wiki/4VGC>

#title: Which compounds with known bioactivities were isolated from Actinobacteria (Q26262282), between 2014 and 2019, with related organisms and references?
SELECT ?organism ?organism\_name ?compound ?compound\_smiles (GROUP\_CONCAT(DISTINCT ?meshLabel; SEPARATOR = "|") AS ?bioactivities) ?isolation\_reference ?reference\_title WHERE {
 ?organism (wdt:P171\*) wd:Q26262282;
 wdt:P225 ?organism\_name.
 ?compound wdt:P235 ?compound\_id;
 wdt:P233 ?compound\_smiles;
 p:P703 ?statement;
 (wdt:P2868/wdt:P486) ?meshId.
 ?mesh wdt:P486 ?meshId;
 rdfs:label ?meshLabel.
 FILTER(LANGMATCHES(LANG(?meshLabel), "EN"))
 ?statement ps:P703 ?organism;
 prov:wasDerivedFrom ?ref.
 ?ref pr:P248 ?isolation\_reference.
 ?isolation\_reference wdt:P1476 ?reference\_title;
 wdt:P356 ?reference\_doi;
 wdt:P577 ?reference\_date.
 FILTER(((YEAR(?reference\_date)) >= 2014 ) && ((YEAR(?reference\_date)) <= 2019 ))
}
GROUP BY ?organism ?organism\_name ?compound ?compound\_smiles ?isolation\_reference ?reference\_title
LIMIT 100000

#### Query 10 - *Aspergillus* spp. terpenoids

This query answers to the following question:

Which compounds labeled as terpenoids were found in Aspergillus spp., between 2010 and 2020, with related references?

Link: <https://w.wiki/4VGD>

#title: Which compounds labelled as terpenoid (Q426694) were found in Aspergillus (Q335130) spp., between 2010 and 2020, with related references?
SELECT ?compound ?compound\_inchi (GROUP\_CONCAT(DISTINCT ?isolation\_reference; SEPARATOR = "|") AS ?isolation\_references) (GROUP\_CONCAT(DISTINCT ?reference\_title; SEPARATOR = "|") AS ?references\_titles) WHERE {
 VALUES ?taxon {
 wd:Q335130
 }
 VALUES ?chemical\_class {
 wd:Q426694
 }
 ?organism (wdt:P171\*) ?taxon.
 ?compound wdt:P235 ?compound\_id;
 wdt:P234 ?compound\_inchi;
 ((wdt:P31|wdt:P279)/(wdt:P279\*)) ?compound\_class;
 p:P703 ?statement.
 ?statement ps:P703 ?organism;
 (prov:wasDerivedFrom/pr:P248) ?isolation\_reference.
 ?isolation\_reference wdt:P1476 ?reference\_title;
 wdt:P356 ?reference\_doi;
 wdt:P577 ?reference\_date.
 FILTER(((YEAR(?reference\_date)) >= 2010 ) && ((YEAR(?reference\_date)) <= 2020 ))
 FILTER(?compound\_class = ?chemical\_class)
}
GROUP BY ?compound ?compound\_inchi

#### Query 11 - Triples

This query answers to the following question:

Which are the available referenced structure-organism pairs on Wikidata? (example limited to 1000 results)

Link: <https://w.wiki/4VFh>

#title: Which are the available referenced structure-organism pairs on Wikidata? (example limited to 1000 results)
SELECT DISTINCT ?structure ?structure\_inchikey ?taxon ?taxon\_name ?reference ?reference\_doi WHERE {
 ?structure wdt:P235 ?structure\_inchikey; # get the inchikey
 p:P703[ # statement found in taxon
 ps:P703 ?taxon; # get the taxon
 (prov:wasDerivedFrom/pr:P248) ?reference ]. # get the reference
 ?taxon wdt:P225 ?taxon\_name. # get the taxon scientific name
 ?reference wdt:P356 ?reference\_doi. # get the reference DOI
}
LIMIT 1000