

1 Question 1: How many protein records are in UniProt?

```
In [1]: %endpoint https://sparql.uniprot.org/sparql
        %log debug
        %show 20
        %outfile query.log

Endpoint set to: https://sparql.uniprot.org/sparql
Logging set to DEBUG
Result maximum size: 20
Output file: /home/osboxes/Downloads/query.log

In [3]: # this is distinct version

PREFIX up: <http://purl.uniprot.org/core/>

SELECT (COUNT(DISTINCT ?protein) AS ?count)
WHERE
{
    ?protein a up:Protein .
}
```

count
322278757

Total: 1, Shown: 1

```
In [3]: PREFIX up:<http://purl.uniprot.org/core/>
        SELECT (COUNT(DISTINCT ?protein) AS ?count) # Count is for obtain the number and distinct to count each entry just once
        WHERE
        {
            ?protein a up:Protein #a = rdf:type. up:Protein is the core class in Uniprot
        }
```

count
322278757

Total: 1, Shown: 1

2 Question 2: How many Arabidopsis thaliana protein records are in UniProt?

```
In [4]: PREFIX up:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>

SELECT (COUNT(DISTINCT ?protein) AS ?count)
WHERE
{
  ?protein a up:Protein .
  ?protein up:organism taxon:3702 # up:organism links the protein with the organism.
}
```

count
121693

Total: 1, Shown: 1

Question 3: retrieve pictures of Arabidopsis thaliana from UniProt?

In [29]: *# retrieve pictures from proteins and Arabidopsis thaliana*

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX taxon: <http://purl.uniprot.org/taxonomy/>
PREFIX up: <http://purl.uniprot.org/core/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

```
SELECT ?picture
WHERE
{
  ?x a up:Protein .
  ?x up:organism taxon:3702 .
  ?picture up:height ?height .
  ?picture up:width ?width
}
```

	picture	height	width
	https://upload.wikimedia.org/wikipedia/commons/3/3a/Bee-sting-abeille-dard-2.jpg	443	390
	https://upload.wikimedia.org/wikipedia/commons/5/50/Panorpa.alpina.jpg	425	640
	https://upload.wikimedia.org/wikipedia/commons/5/55/Bombus.sylvorum-.lindsey.jpg	600	400
	https://upload.wikimedia.org/wikipedia/commons/5/55/Lasioglossum_laticeps.jpg	588	800
	https://upload.wikimedia.org/wikipedia/commons/9/92/Andrena_wilkella_f.jpg	574	800
	https://upload.wikimedia.org/wikipedia/commons/a/ac/Tuinhommel_op_vingerhoedskruid_closeup%281%29.jpg	348	478
	https://upload.wikimedia.org/wikipedia/commons/c/c4/Drawf_Honey_Bee_%28India%29.jpg	500	800
	https://upload.wikimedia.org/wikipedia/commons/c/cc/Phthorimaea_operculella_dorsal.jpg	540	720
	https://upload.wikimedia.org/wikipedia/commons/d/d9/Bombus.subterraneus-.lindsey.jpg	426	640
	https://upload.wikimedia.org/wikipedia/commons/d/db/Halictus.rubicundus-.lindsey.jpg	426	640

Total: 10, Shown: 10

Question 4: What is the description of the enzyme activity of UniProt Protein Q9SZZ8?

the description of the enzyme activity of UniProt Protein Q9SZZ8 is Beta-carotene + 4 reduced ferredoxin [iron-sulfur] cluster + 2 H(+) + 2 O(2) = zeaxanthin + 4 oxidized ferredoxin [iron-sulfur] cluster + 2 H(2)O.

```
In [8]: PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX up: <http://purl.uniprot.org/core/>
```

```
SELECT ?protein ?enzyme ?activity ?description
```

```
WHERE
```

```
{
  ?protein a up:Protein .
  ?protein up:mnemonic "BCH1_ARATH" .
  ?protein up:enzyme ?enzyme .
  ?enzyme up:activity ?activity .
  ?activity rdfs:label ?description .
}
```

protein	enzyme	activity	description
http://purl.uniprot.org/uniprot/Q9SZZ8	http://purl.uniprot.org/enzyme/1.14.15.24	http://purl.uniprot.org/enzyme/1.14.15.24#SIPF8A63F68B2741FFE	Beta-carotene + 4 reduced ferredoxin [iron-sulfur] cluster + 2 H(+) + 2 O(2) = zeaxanthin + 4 oxidized ferredoxin [iron-sulfur] cluster + 2 H(2)O.

Total: 1, Shown: 1

Question 5: Retrieve the proteins ids, and date of submission, for proteins that have been added to UniProt this year ¶

we only retrieve 10 records.

```
In [2]: PREFIX up:<http://purl.uniprot.org/core/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>

SELECT ?protein ?created
WHERE
{
  ?protein a up:Protein .
  ?protein up:created ?created .
  FILTER (?created >= "2020-01-01"^^xsd:date) .
}
LIMIT 10
```

protein	created
http://purl.uniprot.org/uniprot/A0A5D4QTF6	2020-02-26
http://purl.uniprot.org/uniprot/A0A5D4QTF6	2020-02-26
http://purl.uniprot.org/uniprot/A0A5D4SCR7	2020-02-26
http://purl.uniprot.org/uniprot/A0A5D4SCR7	2020-02-26
http://purl.uniprot.org/uniprot/A0A5D4SQH5	2020-02-26
http://purl.uniprot.org/uniprot/A0A5D4SQH5	2020-02-26
http://purl.uniprot.org/uniprot/A0A5N5U371	2020-02-26
http://purl.uniprot.org/uniprot/A0A5N5U371	2020-02-26
http://purl.uniprot.org/uniprot/A0A5N5UEH0	2020-02-26
http://purl.uniprot.org/uniprot/A0A5N5UEH0	2020-02-26

Total: 10, Shown: 10

Question 6: How many species are in the UniProt taxonomy?

There are 2615376 species in the UniProt taxonomy.

In [18]: **PREFIX** up: <http://purl.uniprot.org/core/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>

```
SELECT (count(distinct ?taxon) as ?count)
FROM <http://sparql.uniprot.org/taxonomy>
WHERE
{
    ?taxon a up:Taxon .
    #?taxon up:rank $species .
}
limit 10
```

count
2615376

Total: 1, Shown: 1

Question 7: How many species have at least one protein record? (this might take a long time to execute, so do this one last!) ¶

Query run so slowly.

```
In [3]: PREFIX up: <http://purl.uniprot.org/core/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>

SELECT count(?taxon)
WHERE
{
    ?taxon a up:Taxon .
    ?protein up:organism ?taxon .
}
GROUP BY ?taxon
HAVING(count(?protein) > 0)
```

result in next page →

callret-0
1
3
1
3
3
1
10
3
2
16796
3
115
3
3
9
18
12
24
46
6

Total: 1410946, Shown: 20

Question 8: find the AGI codes and gene names for all Arabidopsis thaliana proteins that have a protein function annotation description that mentions “pattern formation” ¶

I only query 10 result.

```
In [2]: #AGI is Arabidopsis Gene Id
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX up: <http://purl.uniprot.org/core/>
PREFIX taxon: <http://purl.uniprot.org/taxonomy/>

SELECT ?geneName ?AGI
WHERE
{
    ?protein a up:Protein .
    ?protein up:organism taxon:3702 .

    ?protein up:alternativeName ?name .
    ?protein up:encodedBy ?gene .
    ?gene skos:prefLabel ?geneName .
    ?gene up:locusName ?AGI .
    ?protein up:annotation ?annotation .
    ?annotation a up:Function_Annotation .
    ?annotation rdfs:comment ?text .
    FILTER regex(?text, "pattern formation", "i")
}
limit 10
```

result in next page →

geneName	AGI
RPK2	At3g02130
RPK2	At3g02130
RPK2	At3g02130
RPK2	At3g02130
ROPGAP3	At2g46710
ROPGAP3	At2g46710
SWEET8	At5g40260
SWEET8	At5g40260
SWEET8	At5g40260
SWEET8	At5g40260

Total: 10, Shown: 10

Question 9: what is the MetaNetX Reaction identifier (starts with “mnxr”) for the UniProt Protein uniprotkb:Q18A79? ¶

mnxr145046c3

In [14]: %endpoint <https://rdf.metanetx.org/sparql>

```
PREFIX uniprotkb: <http://purl.uniprot.org/uniprot/>
PREFIX mnx: <https://rdf.metanetx.org/schema/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```
select ?pept ?MNXR ?MNXR_ID
where
{
  ?pept mnx:peptXref uniprotkb:Q18A79 .
  ?cata mnx:cata ?pept ;
    rdfs:label ?cata_label .
  ?gpr mnx:gpr ?cata ;
    mnx:reac ?reac .
  ?reac rdfs:label ?reac_label ;
    rdfs:comment ?reac_eq .
  ?mnet mnx:mnet ?gpr ;
    rdfs:label ?mnet_label .
  ?reac mnx:mnxr ?MNXR ;
    rdfs:label ?MNXR_ID .
}
```

Endpoint set to: <https://rdf.metanetx.org/sparql>

pept	MNXR	MNXR_ID
https://rdf.metanetx.org/pept/GLGA_CLOD6	https://rdf.metanetx.org/reac/MNXR145046	mnxr145046c3

Total: 1, Shown: 1

10 Question 10: What is the official Gene ID (UniProt calls this a “mnemonic”) and the MetaNetX Reaction identifier (mnxr.....) for the protein that has “Starch synthase” catalytic activity in *Clostridium difficile* (taxon 272563).

n [8]:

```
1 PREFIX uniprotkb: <http://purl.uniprot.org/uniprot/>
2 PREFIX up: <http://purl.uniprot.org/core/>
3 PREFIX mnx: <https://rdf.metanetx.org/schema/>
4 PREFIX owl: <http://www.w3.org/2002/07/owl#>
5 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
6 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
7 PREFIX taxon: <http://purl.uniprot.org/taxonomy/>
8 PREFIX GO: <http://purl.obolibrary.org/obo/GO_>
9
10 select ?protein ?geneId ?enzyme ?MNXR_ID
11 where {
12     service <https://sparql.uniprot.org/sparql> {
13         ?protein a up:Protein ;
14                 up:organism taxon:272563 ;
15                 up:mnemonic ?geneId ;
16                 up:classifiedWith|(up:classifiedWith/rdfs:subClassOf) GO:0009011 ;
17                 up:enzyme ?enzyme .|
18     }
19
20     ?pept mnx:peptXref ?protein .
21     ?cata mnx:pept ?pept ;
22         rdfs:label ?cata_label .
23     ?gpr mnx:cata ?cata ;
24         mnx:reac ?reac .
25     ?reac rdfs:label ?reac_label ;
26         rdfs:comment ?reac_eq .
27     ?mnet mnx:gpr ?gpr ;
28         rdfs:label ?mnet_label .
29     ?reac mnx:mnxr ?MNXR ;
30         rdfs:label ?MNXR_ID .
31 }
32 limit 10
```

protein	geneId	enzyme	MNXR_ID
http://purl.uniprot.org/uniprot/Q18A79	GLGA_CLOD6	http://purl.uniprot.org/enzyme/2.4.1.21	mnxr145046c3
http://purl.uniprot.org/uniprot/Q18A79	GLGA_CLOD6	http://purl.uniprot.org/enzyme/2.4.1.21	mnxr145046c3

Total: 2, Shown: 2