



**You**

How many countries were the samples derived from?



**COVEO GPT**

> [debug] Calling HTTP endpoint

> [debug] Response received



Talked to ▾

The SARS-CoV-2 samples in the database were derived from 103 different countries.



You

How many samples derived from different countries? Give me result in table!



COVEO GPT

> [debug] Calling HTTP endpoint

> [debug] Response received



Talked to ▾

Here is a table showing the number of SARS-CoV-2 samples derived from different countries:

Country	Sample Count
United States	3,123,975
United Kingdom	2,705,542
Switzerland	137,511
Ireland	80,070
Spain	50,374
Portugal	46,800
Slovakia	40,152
Estonia	30,405
Norway	22,985
India	22,034
...	...

(Note: This table is truncated to show the top 10 countries. There are more countries in the database with varying numbers of samples.)

AI You

Give me a downloadable full table in Excel format.



COVEO GPT

I have created an Excel file containing the full table of the number of SARS-CoV-2 samples derived from different countries. You can download it using the link below:

[Download the SARS-CoV-2 Sample Data Excel File \[↗\]](#)

	A	B	C
1	Country	Sample Count	
2	United States	3123975	
3	United Kingdom	2705542	
4	Switzerland	137511	
5	Ireland	80070	
6	Spain	50374	
7	Portugal	46800	
8	Slovakia	40152	
9	Estonia	30405	
10	Norway	22985	
11	India	22034	
12	Austria	20744	
13	South Africa	20727	
14	Australia	16173	
15	Germany	14348	
16	Denmark	12464	
17	Finland	12252	



**You**

How many samples derived from Belgium?



**COVEO GPT**

> [debug] Calling HTTP endpoint

> [debug] Response received



Talked to ▾

There were 356 SARS-CoV-2 samples derived from Belgium.



You

How many samples derived from Belgium in time? Draw a graph.



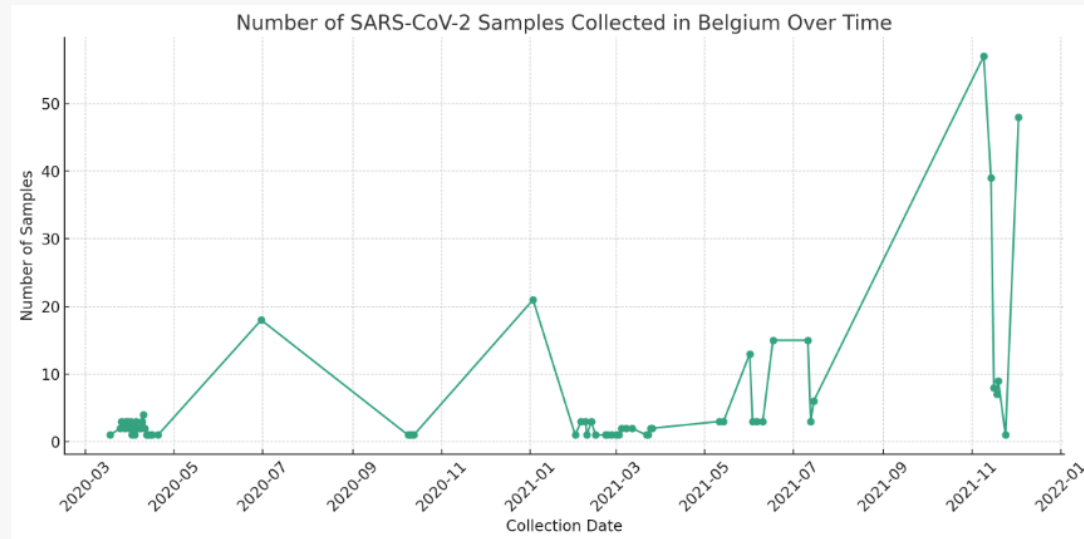
COVEO GPT

> [debug] Calling HTTP endpoint

> [debug] Response received



Talked to ▾



The graph above displays the number of SARS-CoV-2 samples collected in Belgium over time. Each point represents the count of samples collected on a specific date. As observed, there are fluctuations in the number of samples collected on different dates. [↗]

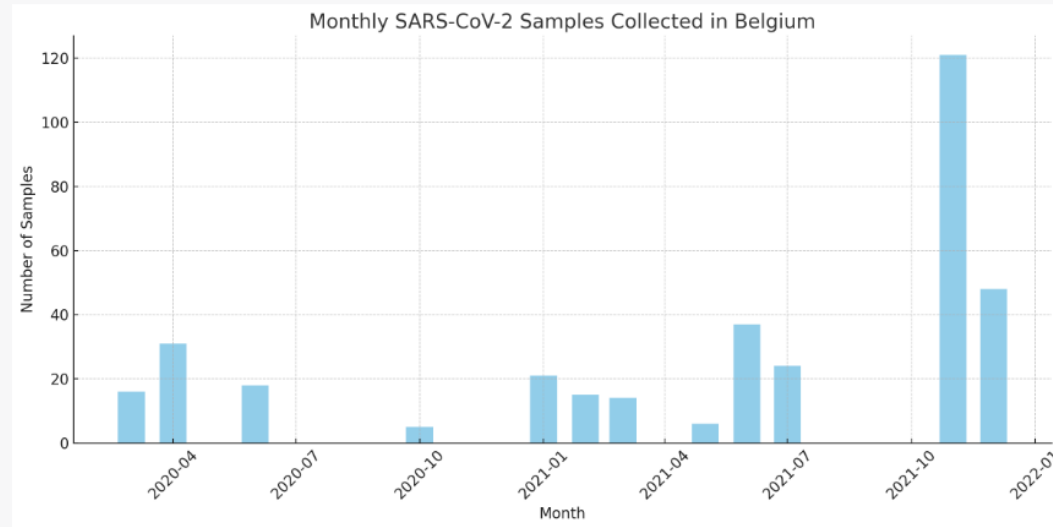


You

Please redraw the graph, use barchart, visualize the number of samples, summarize monthly the data.



COVEO GPT



The bar chart above represents the monthly summary of the number of SARS-CoV-2 samples collected in Belgium. Each bar corresponds to a specific month and illustrates the total number of samples collected during that month. This visualization helps to observe the monthly trends and variations in sample collection over time. [\[↗\]](#)



You

What was the SQL script that retrieve the data for the graph above?



COVEO GPT

To retrieve the data for the graph that summarized the monthly count of SARS-CoV-2 samples collected in Belgium, the SQL script would be structured as follows:

sql

Copy code

```
SELECT DATE_TRUNC('month', collection_date) AS collection_month, COUNT(runid) AS samp
FROM datahub_0.metadata
WHERE country_id IN (SELECT id FROM datahub_0.country WHERE country_name = 'Belgium')
GROUP BY collection_month
ORDER BY collection_month;
```

This script performs the following operations:

1. It selects the `'collection_date'` and truncates it to the month using `'DATE_TRUNC('month', collection_date)'`. This groups the dates by their respective months.
2. It counts the number of samples (`'runid'`) for each of these months using `'COUNT(runid)'`.
3. The `'WHERE'` clause filters the data to only include samples from Belgium.
4. The `'GROUP BY'` clause is used to aggregate the results by each month.
5. Finally, the results are ordered by the month (`'collection_month'`) to ensure a chronological sequence.



You

How many samples derived from Europe?



COVEO GPT

> [debug] Calling HTTP endpoint

> [debug] Response received



Talked to ▾

There were 3,205,950 SARS-CoV-2 samples derived from Europe.

The database does not store information about Europe!!!

ChatGPT first found out from internet which are the European Countries

```
SELECT COUNT(m.runid)
FROM datahub_0.metadata m
JOIN datahub_0.country c ON m.country_id = c.id
WHERE c.country_name IN ('Albania', 'Andorra', 'Armenia', 'Austria', 'Azerbaijan', 'Belarus', 'Belgium', 'Bosnia and Herzegovina',
    'Bulgaria', 'Croatia', 'Cyprus', 'Czech Republic', 'Denmark', 'Estonia', 'Finland', 'France', 'Georgia',
    'Germany', 'Greece', 'Hungary', 'Iceland', 'Ireland', 'Italy', 'Kazakhstan', 'Kosovo', 'Latvia', 'Liechtenstein',
    'Lithuania', 'Luxembourg', 'Malta', 'Moldova', 'Monaco', 'Montenegro', 'Netherlands', 'North Macedonia', 'Norway',
    'Poland', 'Portugal', 'Romania', 'Russia', 'San Marino', 'Serbia', 'Slovakia', 'Slovenia', 'Spain', 'Sweden',
    'Switzerland', 'Turkey', 'Ukraine', 'United Kingdom', 'Vatican City')
```





You

How many samples derived from EU?



COVEO GPT

> [debug] Calling HTTP endpoint

> [debug] Response received



Error talking to

There was an error in processing the request to retrieve the number of SARS-CoV-2 samples derived from the European Union (EU). The syntax of the SQL script seems to have caused an issue. I will correct it and try again.

> [debug] Calling HTTP endpoint

> [debug] Response received



Talked to ▾

There were 339,430 SARS-CoV-2 samples derived from the European Union (EU).

```
SELECT COUNT(m.runid)
FROM datahub_0.metadata m
JOIN datahub_0.country c ON m.country_id = c.id
WHERE c.country_name IN ('Austria', 'Belgium', 'Bulgaria', 'Croatia', 'Cyprus', 'Czech Republic', 'Denmark',
                          'Estonia', 'Finland', 'France', 'Germany', 'Greece', 'Hungary', 'Ireland', 'Italy',
                          'Latvia', 'Lithuania', 'Luxembourg', 'Malta', 'Netherlands', 'Poland', 'Portugal',
                          'Romania', 'Slovakia', 'Slovenia', 'Spain', 'Sweden')
```



**You**

How many "B.1" pango lineage samples are in the database?



**COVEO GPT**


> [debug] Calling HTTP endpoint

> [debug] Response received



Talked to ▾

There are 9,017 samples in the database classified under the "B.1" Pango lineage.

 **You**  
How many samples from Belgium belongs to the various Pango lineage id. Give me a table

 **COVEO GPT**

> [debug] Calling HTTP endpoint

> [debug] Response received

✓ Talked to ▾

Here is a table showing the number of SARS-CoV-2 samples from Belgium belonging to various Pango lineage IDs:

Pango Lineage	Sample Count
B.1	31
B.40	2
B	1
B.1.1	1
B.1.1.75	1
B.3	1

This table lists the different lineages and the corresponding count of samples from Belgium.



You

Give me the SQL script that generates the table above



COVEO GPT

The SQL script used to generate the table showing the number of SARS-CoV-2 samples from Belgium belonging to various Pango lineage IDs is as follows:

sql

Copy code

```
SELECT p.lineage, COUNT(*) AS sample_count
FROM datahub_0.pangolin p
JOIN datahub_0.metadata m ON p.run_id = m.runid
JOIN datahub_0.country c ON m.country_id = c.id
WHERE c.country_name = 'Belgium'
GROUP BY p.lineage
ORDER BY sample_count DESC
```

This script performs the following actions:

- Joins the `pangolin` table (`p`) with the `metadata` table (`m`) on `run\_id`, and then with the `country` table (`c`) on `country\_id`.
- Filters the data to include only those samples that are from Belgium (`c.country\_name = 'Belgium'`).
- Groups the data by the `lineage` column from the `pangolin` table.
- Counts the number of samples for each lineage.
- Orders the results in descending order based on the count of samples.