

Monitor wariantów SARS-CoV-2

Hubert Baniecki, Piotr Piątyszek, Hanna Zdulska

10.05.2021

Co robimy?

1. Pozyskiwanie, przetwarzanie i wizualizacja danych o wariantach SARS-CoV-2 w Europie

2. Raportowanie o ogniskach nowych i potencjalnie groźnych wariantów w Polsce

Indyjska odmiana koronawirusa w Polsce. "Co najmniej 16 przypadków"

04.05.2021

<https://www.polsatnews.pl/wiadomosc/2021-05-04/indyjski-wariant-koronawirusa-w-polsce-konferencja-ministra-zdrowia-od-1300>



Minister zdrowia Adam Niedzielski



- Dwie osoby z rodziny polskiego dyplomaty zakażone były indyjskim wariantem COVID-19 - mówił minister zdrowia Adam Niedzielski. Jak dodał, w Polsce zlokalizowano dwa ogniska odmiany indyjskiej - w okolicach Warszawy i Katowicach. Stwierdzono dotychczas 16 przypadków zakażeń. Szef GIS poinformował, że wśród zakażonych są siostry zakonne, które brały udział w rekolekcjach w Zaborowie.

Plan

- dlaczego?
- podział wariantów SARS-CoV-2
- regularna analiza metadanych miliona sekwencji
- monitor dla całej Europy na bazie danych z GISAID **vs**
- monitor dla Polski na bazie danych lokalnych (Ministerstwa Zdrowia, PZH)
- kulisy scrapowania danych z GISAID

nature

NEWS · 07 JANUARY 2021 · UPDATE 08 JANUARY 2021 <https://www.nature.com/articles/d41586-021-00031-0>

Could new COVID variants undermine vaccines? Labs scramble to find out

Researchers race to determine why lineages identified in Britain and South Africa spread so quickly and whether they'll compromise vaccines.

NEWS · 06 MAY 2021

<https://www.nature.com/articles/d41586-021-01222-5>

Pfizer COVID vaccine protects against worrying coronavirus variants

Data from Qatar provide strongest evidence yet that COVID-19 strains thought to pose a threat to immunization efforts.

RESEARCH ARTICLE

<https://science.sciencemag.org/content/372/6538/eabg3055>

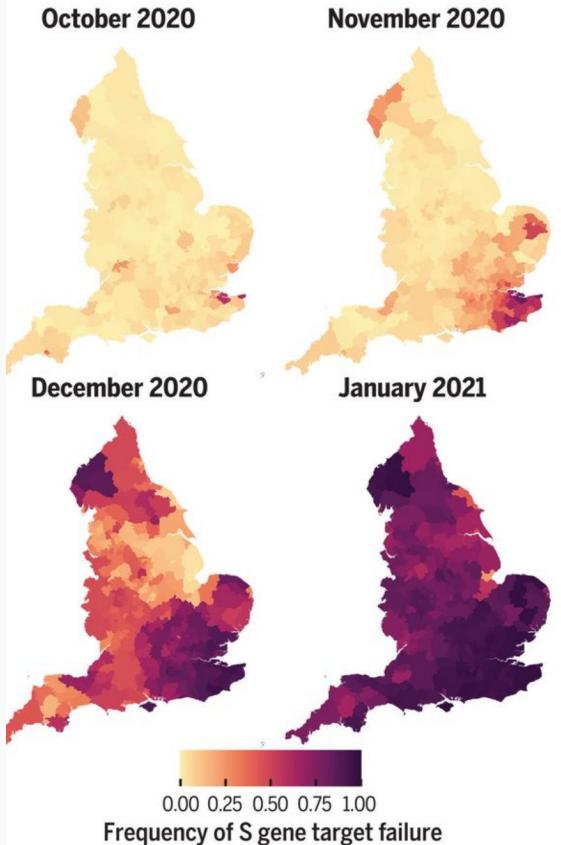
Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England

 Nicholas G. Davies^{1,*},  Sam Abbott^{1,†},  Rosanna C. Barnard^{1,†},  Christopher I. Jarvis^{1,†}, Adam J. Kucharski^{1,†},  J...

* See all authors and affiliations

Science 09 Apr 2021:
Vol. 372, Issue 6538, eabg3055
DOI: 10.1126/science.abg3055

Wariant brytyjski, afrykański, indyjski, kolejny



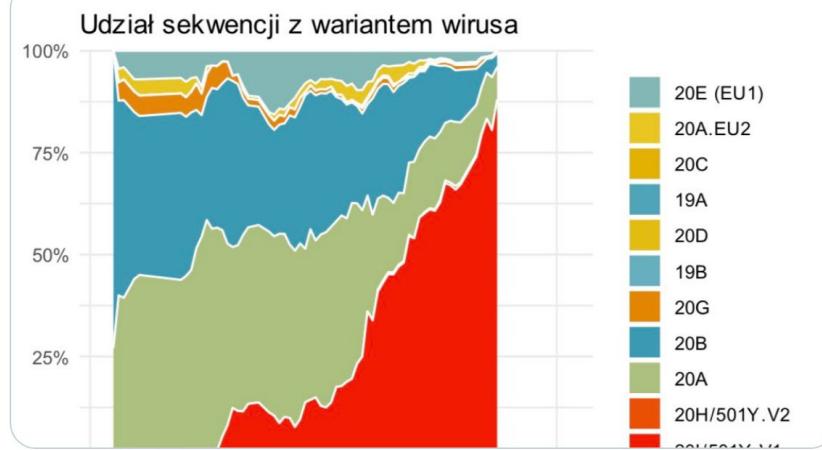
CONCLUSION

More than 98% of positive SARS-CoV-2 infections in England are now due to VOC 202012/01, and the spread of this new variant has led to a surge in COVID-19 cases and deaths. Other countries should prepare for potentially similar outcomes.



Adam Niedzielski @a_niedzielski · Mar 20

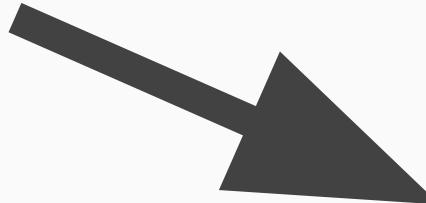
Mutacja brytyjska (20I/501Y.V1) wypiera inne warianty wirusa w Polsce. Jej udział w kolejnych badaniach genomu osiąga już wartość 80%.



Kod wirusa: ~30k literek A, T, G i C

>hCoV-19/Wuhan/WIV04/2019|EPI_ISL_402124|2019-12-30

```
ATTAAGGTTTACCTCCCAGGTAAACAAACCAACTTCGATCTCTGTAGATCTGTTCTCTAAACGAACCTTAA  
AATCTGTGGTGTCACTCGGCTGCATGTTAGTCAGTCAGCAGTATAATTAAACTAATTACTGTTGACAGG  
ACACGAGTAACCGTCTATCTCTGCAGGCTGCTTACGGTTCTGTCAGCCGATCATCAGCACATCTAGGTT  
CGTCCGGGTGACCGAAAGGTAAAGATGGAGAGCCTGTCCTGGTTCAACGAGAAAACACGTCCAACTCAGTTGC  
CTGTTTACAGGTTCGCGACGTGCTCGTACGTGGCTTGGAGACTCCGTGGAGGGAGGTCTTACAGAGGCACGTCAACAT  
CTTAAAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGCCTTGCCCTAACTGAAACAGCCCTATGTGTTCATCAA  
ACGTCGGATGCTCGAACCTCATGGTATGGTTGAGCTGGTAGCAGAACTCGAAGGCATTCACTGTC  
GTAGTGGTGAGACACTTGGTGTCTTGTCCCTCATGGGGCGAAATACCAGTGGTACCGCAAGGTTCTTCTCGTAAG  
AACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGGCCGATCTAAAGTCATTGACTTAGGCACGGCTGGCACTGA  
TCCTTATGAAGATTTCAAGAAAATGGAACACTAAACATAGCACTGGTGTACCCGTGAACCATGCGTGAGCTTAACG  
.....
```



```
| MFVFLVLLPLVSSQCVNLTTRQLPPAYTNSFRGVYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDPVLPFND  
| GYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFL  
| MDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLTPGDSSSGWTAGAAA  
| VVV/GVI DPRTEI I KVNFENGTTDAVDCAI DDI SETKCTI KSETVEKGTVOTSNERVODTESTV/REPNTTNI CDEGEV/ENATREASIVVA  
.....
```

Mutacja **S:N501Y** to zamiana jednej literki

Podział wariantów

Nextstrain Clade	Pango Lineage	Other Names
20E (EU1)	B.1.177	20A.EU1
20A .EU2	B.1.160	
20I / 501 Y .V1	B.1.1.7 ↗	VOC 202012/01
20H / 501 Y .V2	B.1.351 ↗	501Y.V2
20J / 501 Y .V3	P.1 ↗	
20C / S: 452 R	B.1.427, B.1.429	CAL.20C
20A / S: 484 K	B.1.525 ↗	
20C / S: 484 K	B.1.526	(Part of Pango lineage)
20A / S: 154 K	B.1.617.1 ↗	
20A / S: 478 K	B.1.617.2 ↗	
20A / S: 439 K	B.1.258	
20A / S: 98 F	B.1.221	
20C / S: 80 Y	B.1.367	
20B / S: 626 S	B.1.1.277	
20B / S: 1122 L	B.1.1.302	

<https://covariants.org>

Podział PANGO lineages

Article | Published: 15 July 2020

<https://www.nature.com/articles/s41564-020-0770-5>

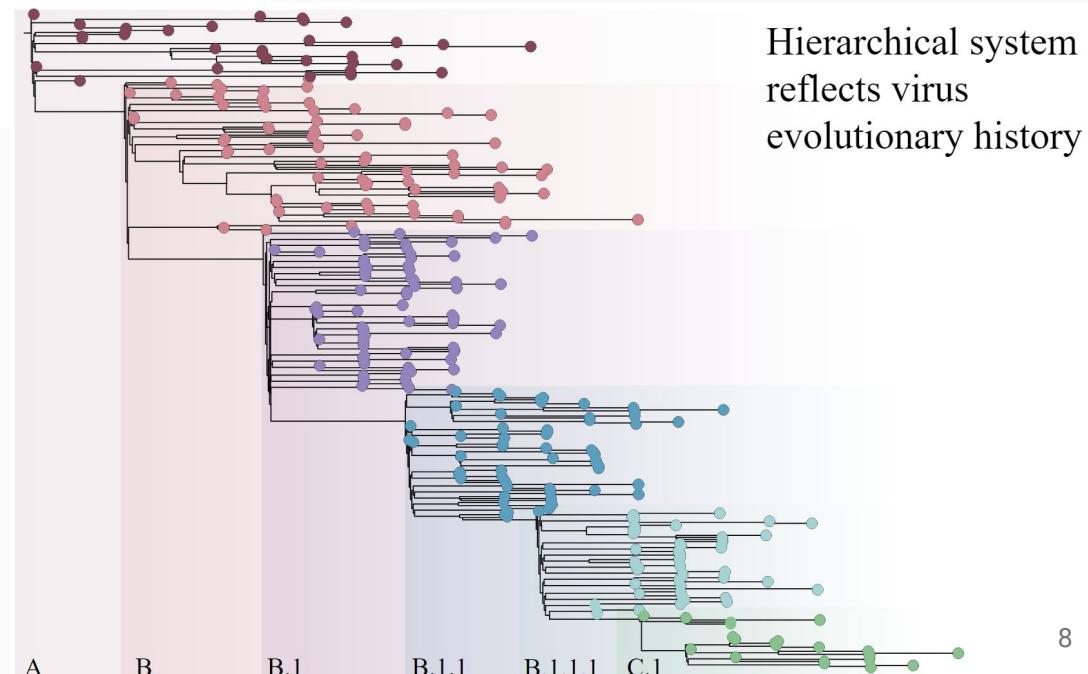
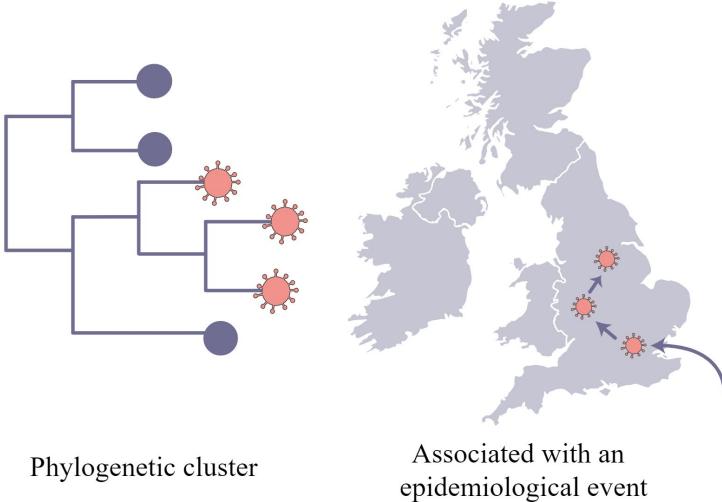
A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology

Andrew Rambaut Edward C. Holmes Áine O'Toole, Verity Hill, John T. McCrone, Christopher Ruis,
Louis du Plessis & Oliver G. Pybus

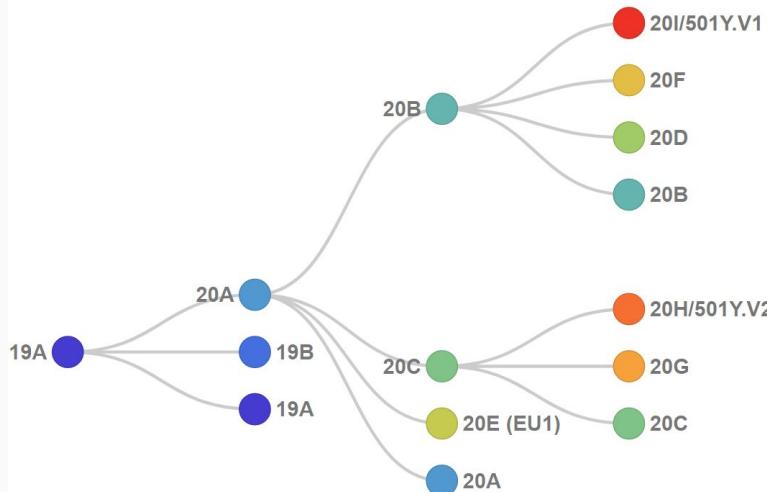
<https://cov-lineages.org>

Nature Microbiology 5, 1403–1407(2020) | Cite this article

41k Accesses | 207 Citations | 1071 Altmetric | Metrics



Podział NEXSTRAIN clades



<https://clades.nextstrain.org>

English ▾

Nextclade

Clade assignment, mutation calling, and quality control right inside your browser

Simple
No installation or setup - drop a file and see the results

Private
No remote processing - sequence data never leaves your computer

Mutation Calling
Find differences of your sequences relative to the reference in standard numbering

Clade Assignment
Find out in which Nextstrain clades your samples are from

Quality Control
Check your data against Nextstrain's QC metrics

FASTA

TXT

Drag & Drop a file or click to select

OR

PASTE SEQUENCES

Show me an Example

⚠️ Nextclade is currently under active development. Implementation details and data formats are subjects to change. The app may contain bugs. Please report any issues and leave feedback at github.com/neherlab/webclades

https://webclades-hnrba7o73.vercel.app/

What is Nextclade?

Variant of Interest (VOI)

B.1.1.7 (brytyjski). Kluczowe mutacje: N501Y, D614G.
Dostępne są dowody na zwiększoną zakaźność i dotkliwość. Dominujący wariant w Europie.

B.1.351 (afrykański). Kluczowe mutacje: K417T, E484K, N501Y, D614G. Dostępne są dowody na zwiększoną zakaźność i dotkliwość.

P.1 (brazylijski)

Variant of Concern (VOC)

B.1.617.2 (indyjski). Kluczowe mutacje: L452R, T478K, D614G.

B.1.427/B.1.429 (USA)



Obecnie dla SARS-CoV-2 nie ma zdefiniowanych Variants of High Consequence (VOHC)

Baza danych

© 2008 - 2021 | Terms of Use | Privacy Notice | Contact

You are logged in as **Hubert Baniecki** - [logou](#)

Registered Users EpiFlu™ **EpiCoV™** My profile

 [EpiCoV™](#)  [Search](#)  [Downloads](#)  [Upload](#)

Pandemic coronavirus causing COVID-19

A previously unknown human coronavirus (hCoV-19) was first detected in late 2019 in patients in the City of Wuhan, who suffered from respiratory illnesses including atypical pneumonia, an illness that has become known as coronavirus disease (COVID-19). The coronavirus originated from an animal host and is closely related to the virus responsible for the Severe Acute Respiratory Syndrome coronavirus (SARS).

On 10. January 2020, the first virus genomes and associated data were publicly shared via GISAID. The World Health Organization announced on 11. March 2020 the first coronavirus pandemic. As the pandemic progresses, scientists from around the globe are tracking the virus and its genome sequences to ensure optimal virus diagnostic tests, to track and trace the ongoing outbreak and to identify potential intervention options. Several analyses to assist with these efforts are offered here, including sequence alignments, diagnostic primer and probe coordinates, 3D protein models, drug targets, phylogenetic trees and many more.



by A*STAR Singapore

[Search](#)

hCoV-19 data sharing via GISAID

1,462,224 submissions



Emerging Variants

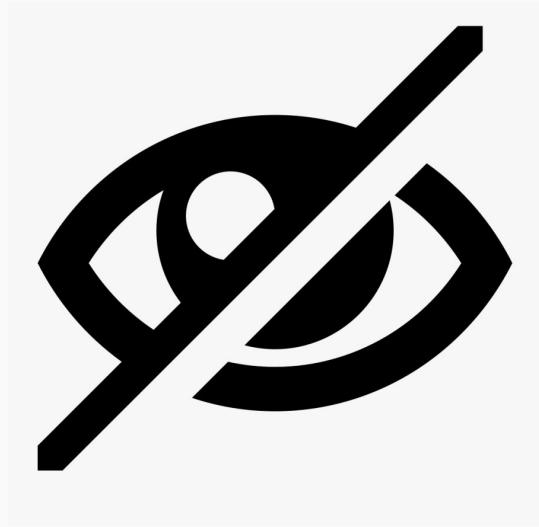


Official GISAID reference sequence



PrimerChecker

Million sekwencji w rok
- jak monitorować?



Monitor dla Europy

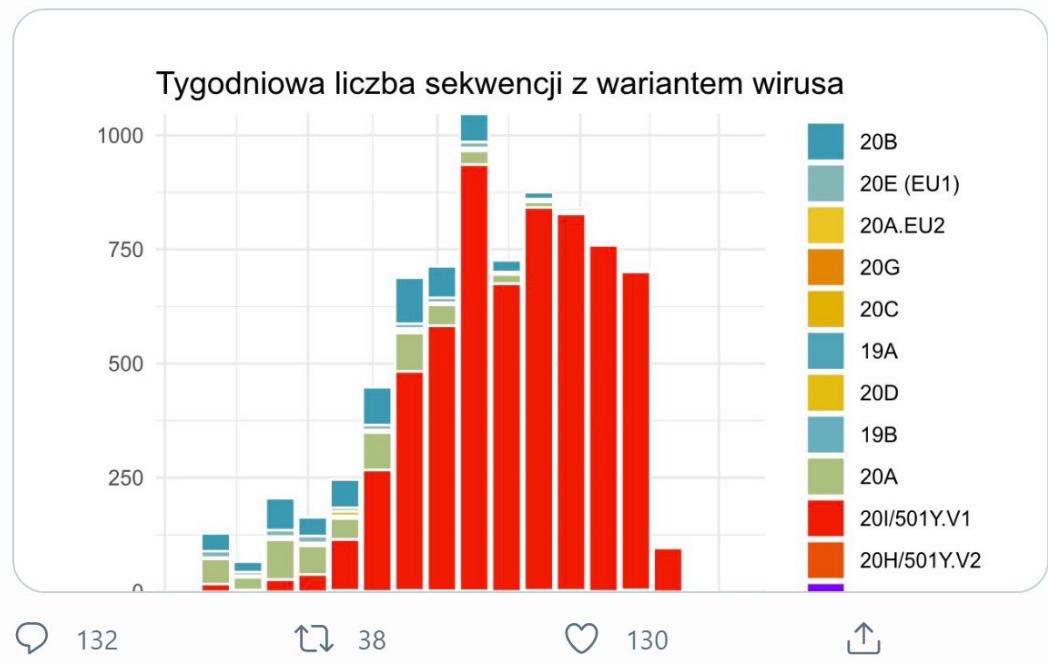
<https://monitor.mi2.ai>

Open Source:
<https://github.com/MI2DataLab/crs19-monitor>

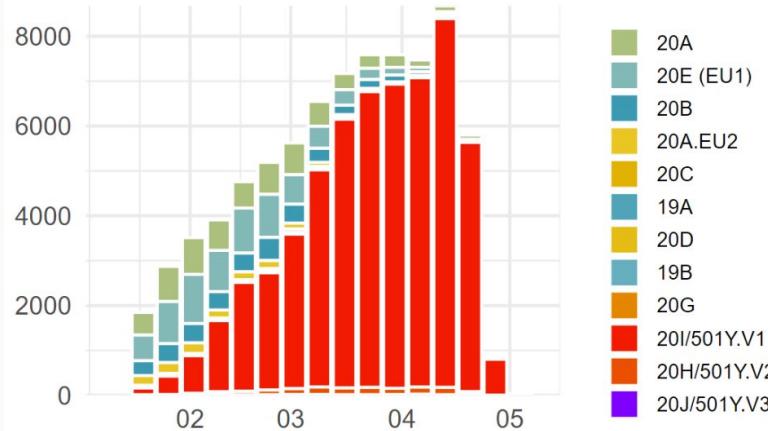


Adam Niedzielski @a_niedzielski · May 5

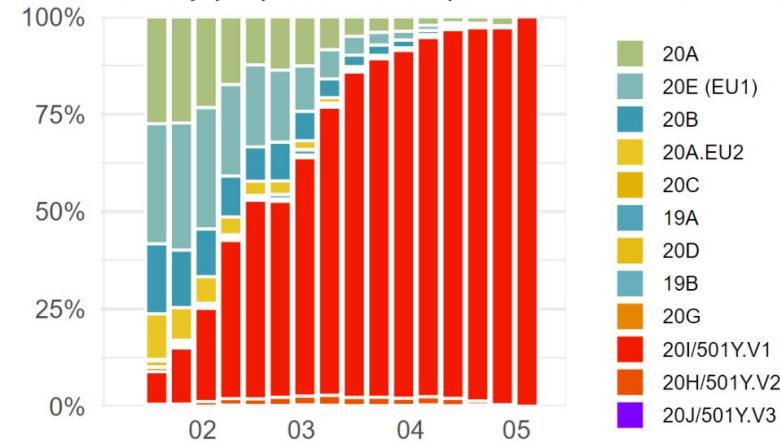
Już ponad 8,2 tys próbek sekwencjonowanych od początku roku do końca kwietnia w ramach monitorowania wariantów wirusa obecnego w Polsce. Dzięki temu projektowi wzrasta nasze bezpieczeństwo epidemiczne. Szczegóły i aktualizacje można znaleźć pod adresem:
monitor.crs19.pl/2021-05-05/pol...



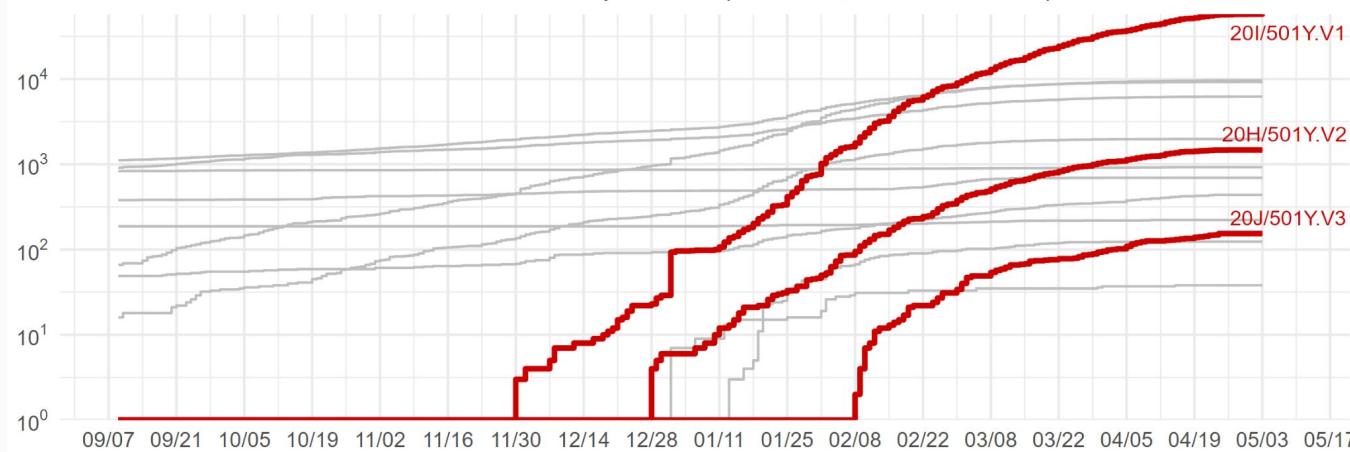
Weekly count of sequences with virus variant



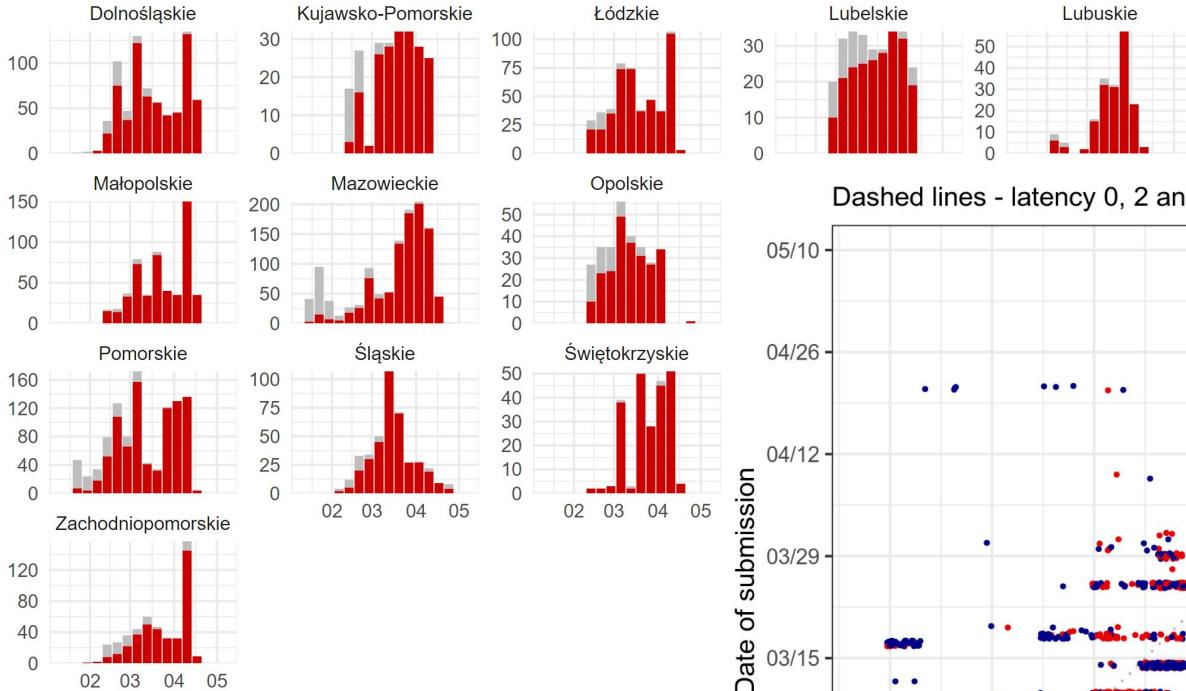
Weekly proportion of sequences with virus variant



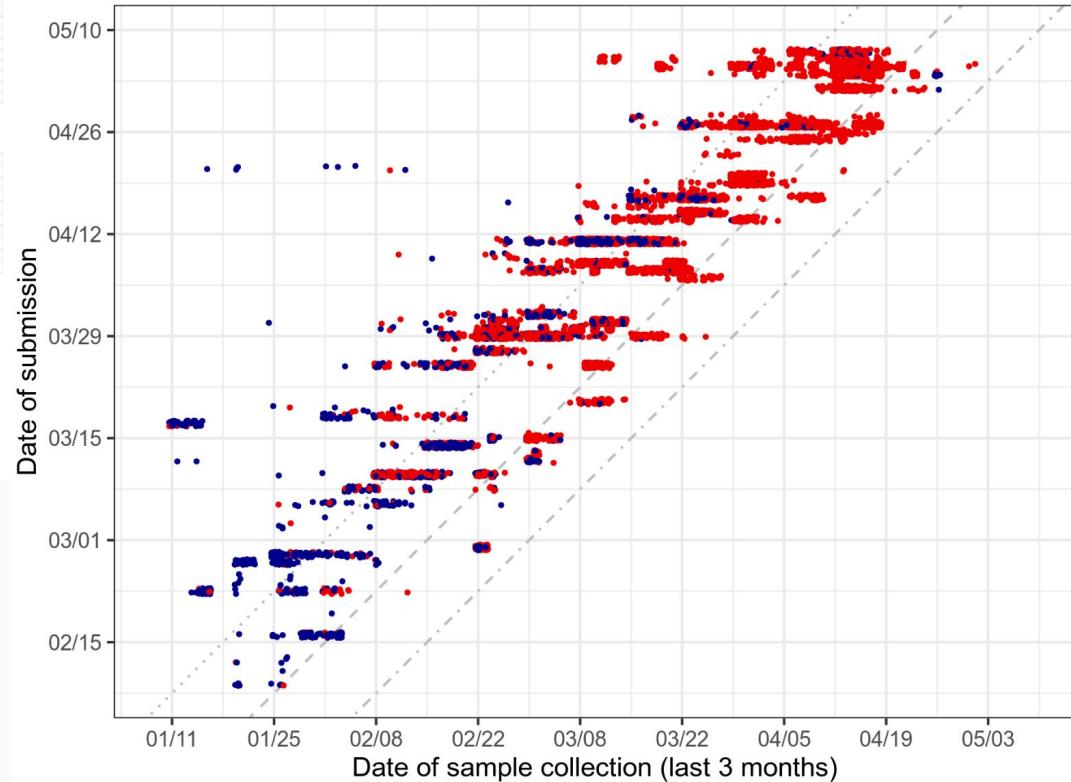
Cumulative number of virus variant sequences (GISAID, last 6 months)

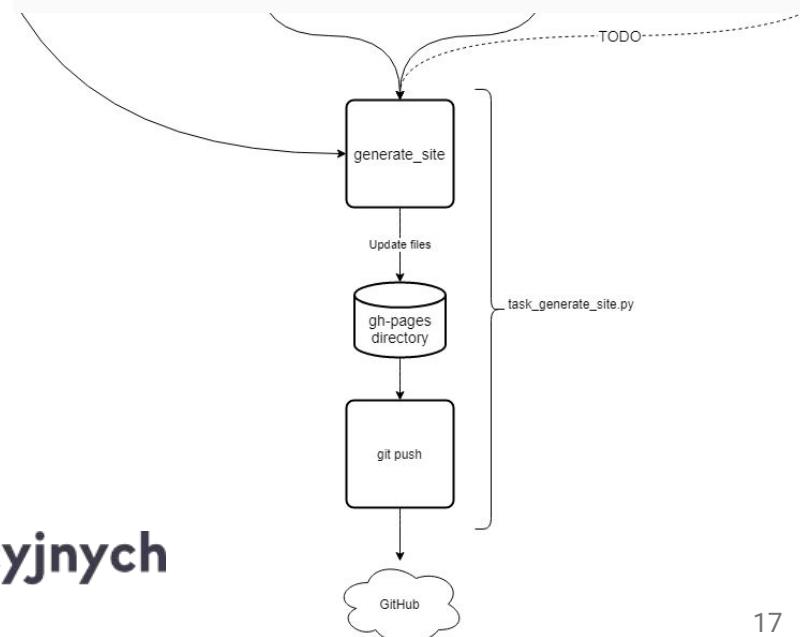
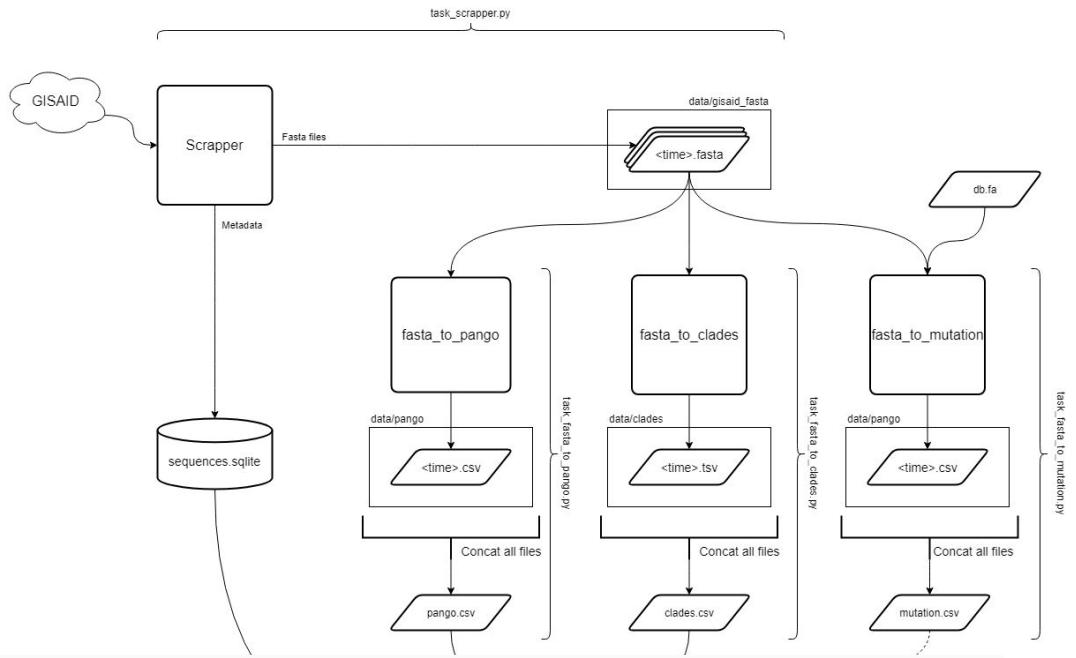


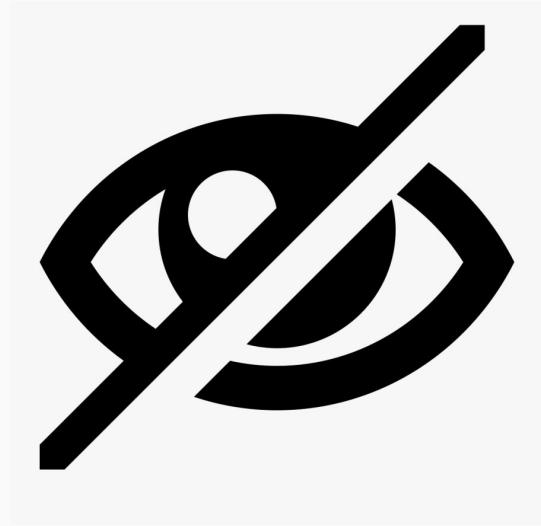
Number of sequences per week in the regions (last 3 months)

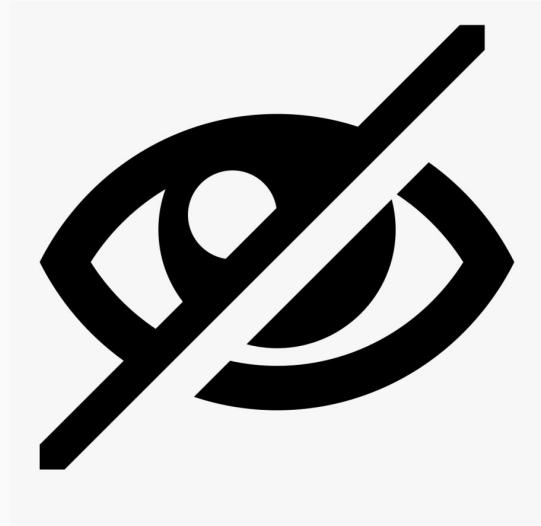


Dashed lines - latency 0, 2 and 4 weeks. Red points - variants with the N501Y mutation

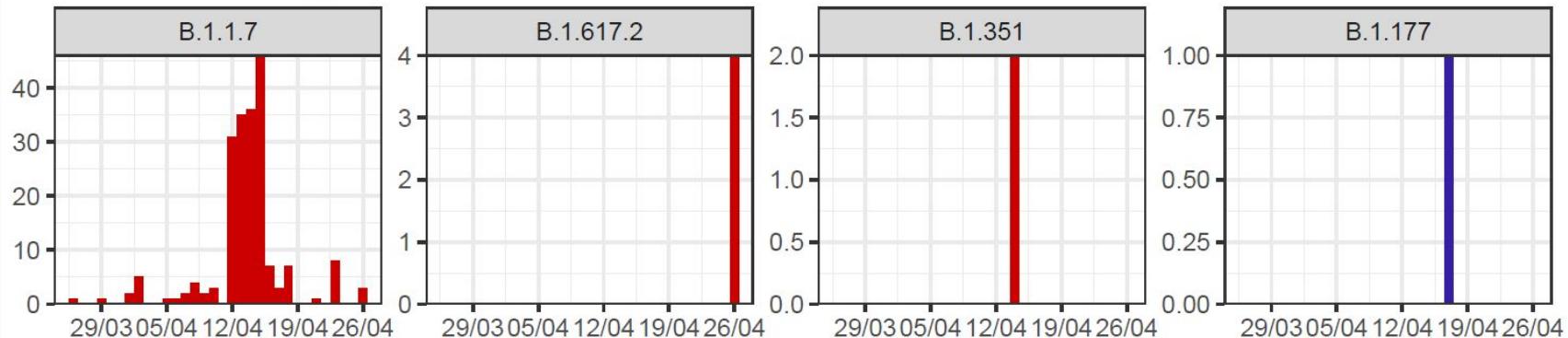




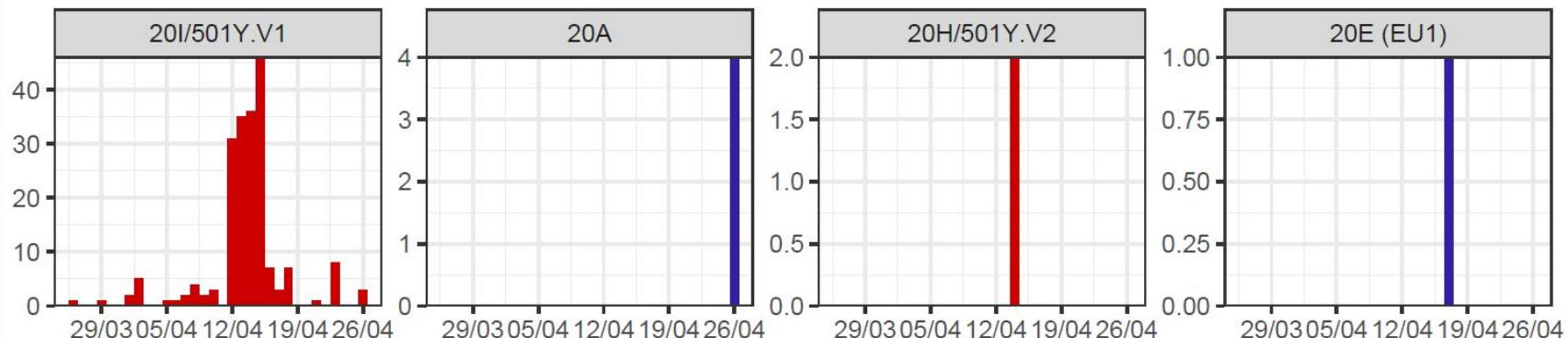




Liczba wystąpień poszczególnych wariantów SARS-CoV-2 w podziale na dni nomenklatura Pango



nomenklatura Clades





Original research article

The first SARS-CoV-2 genetic variants of concern (VOC) in Poland: The concept of a comprehensive approach to monitoring and surveillance of emerging variants

B.1.351 (known as 501Y.V2 South African variant)

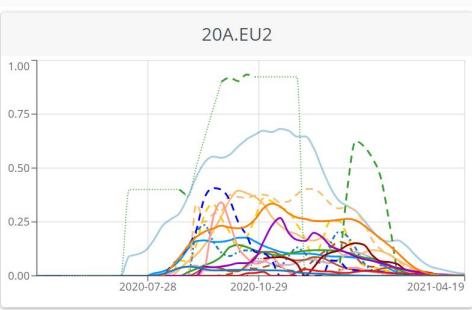
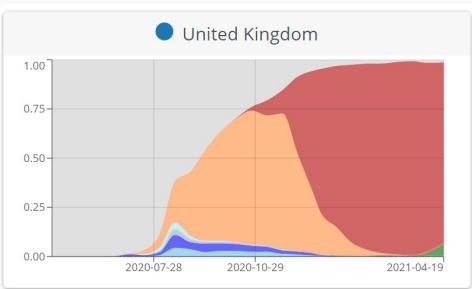
Radosław Charkiewicz ^{a, b, c, 1}, Jacek Nikliński ^{a, b, 1}, Przemysław Biecek ^d, Joanna Kiśluk ^{a, b}, Sławomir Pancewicz ^e, Anna M. Moniuszko-Malinowska ^e, Robert Flisiak ^f, Adam J. Krętowski ^g, Janusz Dzięcioł ^{b, h}, Marcin Moniuszko ⁱ, Rafał Gierczyński ^j, Grzegorz Juszczak ^k, Joanna Reszeć ^{b, l}

10.02.2021

06.03.2021

<https://www.sciencedirect.com/science/article/pii/S1896112621000201>

Jak to robią inni?



Shared mutations

Sort by: Commonness Position

20I/501Y.V1 (B.1.1.7)	20H/501Y.V2 (B.1.351)	20J/501Y.V3 (P.1)	20B/S.484K (P.2)	20C/S.452R (B.1.427/9)	20C/S.484K (B.1.526)	20A/S.484K (B.1.525)	20A/S.154K (B.1.617.1)	20A/S.478K (B.1.617.2)
Shared mutations								
S: H 69 -	S: L 18 F							
S: V 70 -								
S: Y 144 -								
	S: K 417 N	S: K 417 T						
			S: L 452 R					
	S: E 484 K	S: E 484 K	S: E 484 K	S: E 484 K	S: E 484 K	S: E 484 K	S: E 484 Q	S: L 452 R
S: N 501 Y	S: N 501 Y	S: N 501 Y						
S: D 614 G	S: D 614 G	S: D 614 G	S: D 614 G	S: D 614 G	S: D 614 G	S: D 614 G	S: D 614 G	S: D 614 G
S: P 681 H							S: P 681 R	S: P 681 R
	S: A 701 V				S: A 701 V			
		S: V 1176 F	S: V 1176 F					
Other mutations								
S: A 570 D	S: D 80 A	S: T 20 N		S: S 13 I	S: L 5 F	S: Q 52 R	S: E 154 K	S: T 19 R
S: T 716 I	S: D 215 G	S: P 26 S		S: W 152 C	S: T 95 I	S: A 67 V	S: Q 1071 H	S: E 156 -
S: S 982 A	S: L 241 -	S: D 138 Y			S: D 253 G	S: Q 677 H		S: F 157 -
S: D 1118 H	S: L 242 -	S: R 190 S				S: F 888 L		S: R 158 G
S: A 243 -	S: H 655 Y							S: T 478 K
	S: T 1027 I							S: D 950 N

Dostęp do danych - otwarty, czy nie?

NEWS · 03 FEBRUARY 2021

<https://www.nature.com/articles/d41586-021-00305-7>

nature

Scientists call for fully open sharing of coronavirus genome data

Other researchers say that restrictions at the largest SARS-CoV-2 genome platform encourage fast sharing while protecting data providers' rights.

NEWS · 05 MAY 2021

<https://www.nature.com/articles/d41586-021-01194-6>

Why some researchers oppose unrestricted sharing of coronavirus genome data

Global-south scientists say that an open-access movement led by wealthy nations deprives them of credit and undermines their efforts.

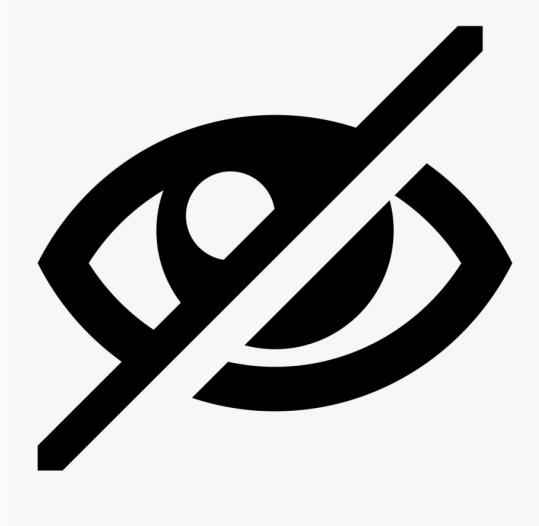
Scrapowanie GISAID czyli dynamiczne piekło

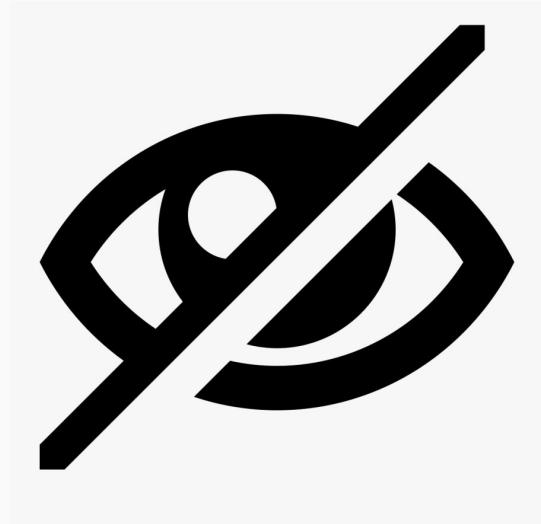
<https://www.epicov.org/epi3/frontend#291a31>



+



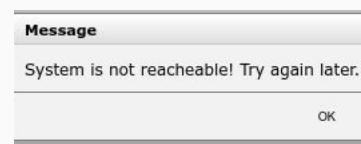
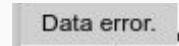
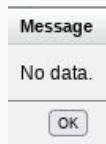




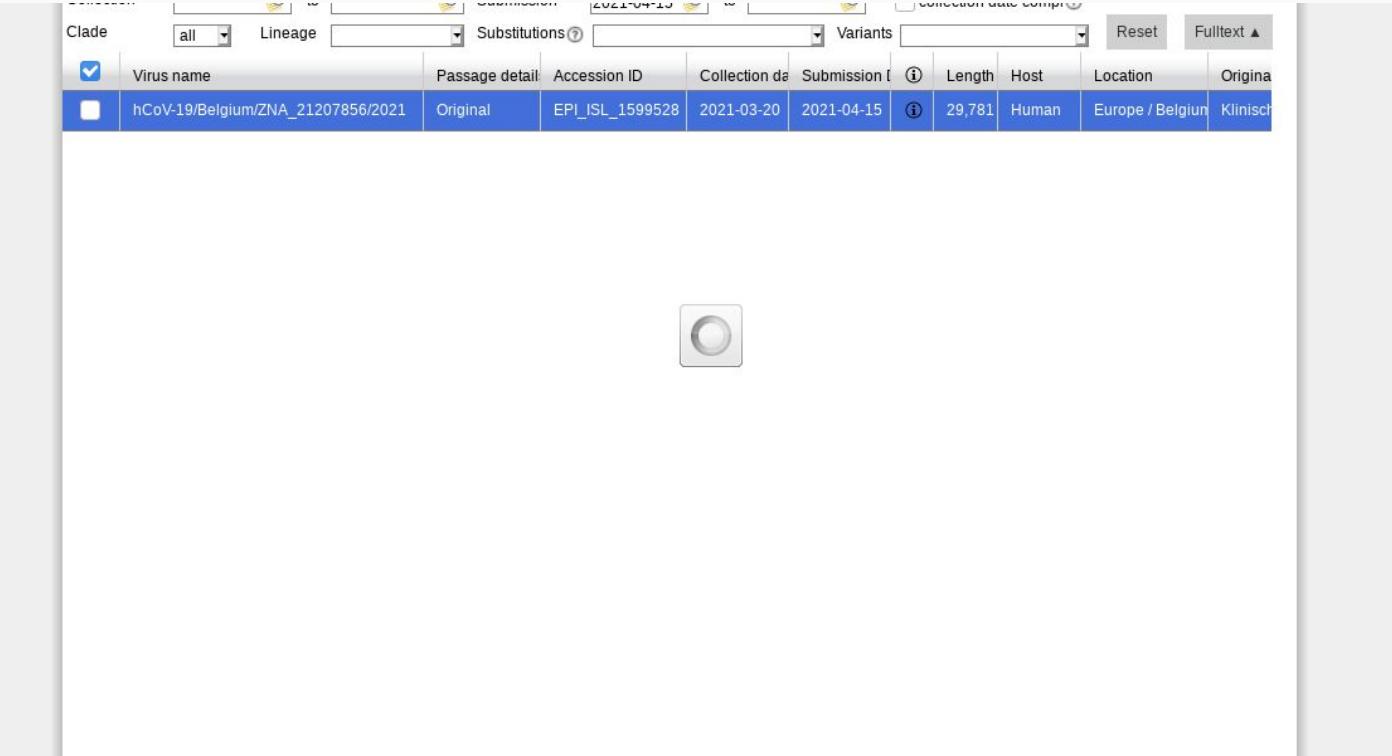
except Exception as e

```
def save_snapshot(self):  
    name = datetime.datetime.now().strftime("%Y-%m-%d_%H-%M-%S.%f")  
    self.driver.save_screenshot(self.log_dir + '/' + name + ".png")  
    with open(self.log_dir + '/' + name + '.html', 'w') as html_file:  
        html_file.write(self.driver.page_source)
```

```
except Exception as e:  
    self.save_snapshot()  
    self.close()  
    raise e
```



Spinnery i checkboxy



The screenshot shows a search interface with various filters at the top and a results table below. The filters include Clade (all), Lineage, Substitutions, Variants, Reset, and Fulltext. The results table has columns for Virus name, Passage detail, Accession ID, Collection date, Submission date, Length, Host, Location, and Original. A single row is highlighted in blue, representing hCoV-19/Belgium/ZNA_21207856/2021. The passage detail is listed as 'Original', and the accession ID is EPI_ISL_1599528. The collection date is 2021-03-20, and the submission date is 2021-04-15. The length is 29,781, and the host is Human. The location is Europe / Belgium, and the original status is Klinisch. A large, empty rectangular area is visible below the table.

Clade	Lineage	Substitutions	Variants	Reset	Fulltext			
<input checked="" type="checkbox"/> Virus name	Passage detail	Accession ID	Collection date	Submission date	Length	Host	Location	Original
<input type="checkbox"/> hCoV-19/Belgium/ZNA_21207856/2021	Original	EPI_ISL_1599528	2021-03-20	2021-04-15	29,781	Human	Europe / Belgium	Klinisch

Prove that you are not a robot

The screenshot shows a search interface for the GISAID EnFlu™ Database. The search results table lists various virus entries, including hCoV-19/Poland/PL_P1497/2021 and hCoV-19/Poland/PL_P1496/2021. A modal window titled "Prove that you are not a robot:" is displayed over the results. The modal contains a CAPTCHA image with the text "hpdm4" and a button labeled "Try again!". Below the image is a "Continue" button with a checkmark icon. At the bottom of the modal, there is a note: "Important note: In the GISAID EnFlu™ Database Access Agreement, you have accepted certain terms and conditions for viewing and using data regarding influenza viruses. To the extent the Database contains data relating to non-influenza viruses, the viewing and use of these data is subject to the same terms and conditions, and by viewing or using such data you agree to be bound by the terms of the GISAID EnFlu™ Database Access Agreement in respect of such data in the same manner as if they were data relating to influenza viruses." The modal has a "Select" checkbox at the top right. Below the modal, the search interface shows a total of 4,245 viruses, page navigation (1-5), and download links for analysis and download.

The screenshot shows the developer tools of a browser. The "Elements" tab is active, displaying a CSS rule for a component slot. The rule is: `element { inline }`. The "Layout" tab is selected. The URL bar shows the path: `/v_id="c_qr99e_u0-c_qr99e_tz" class="sys-component-slot" data-bbox="11 578 211 600"/>`

+



Jakie dane można znaleźć?

Wiek:

- 81-90
- 85+
- 942
- 4 years
- 0.92
- 51 ans
- 184d
- 14gg
- Nasopharyngeal swab
- 6Unknown
- <1
- -20
- 1899-12-31
- 'unknown', 'unkown', 'unknow'

Płeć:

- 'Male', 'M', 'Vyras', 'Ma+L9le'
- 'Female', 'F', 'Famale', 'FEMALEemale', 'Femal'
- 'Unknown', 'U', 'unkown', 'unkowne', 'Moteris', 'Inconnue', 'not provided', 'Na'
- 'Human'
- '0.0'

dwa miesiące za nami