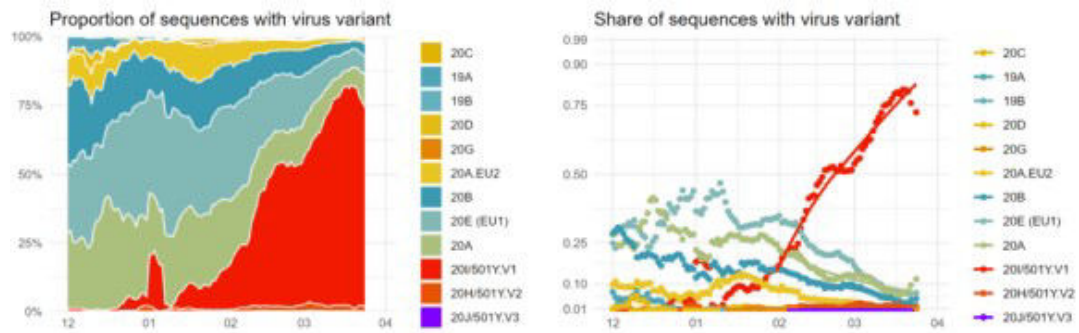


Why should I care?



Proportion and share of sequences of different SARS-CoV-2 variants in Germany. The variants with the N501Y mutation (e.g. red one) are alert variants with increased infectivity.

[Source](#).

The first quarter of 2021 has brought us common knowledge about the importance of genomic sequencing for battling COVID-19. Apparently, monitoring SARS-CoV-2 variants and mutations in countries will be crucial to epidemic spread, vaccine development, and efficiency. There is a well-written sequence of interesting news on this topic in *Nature*:

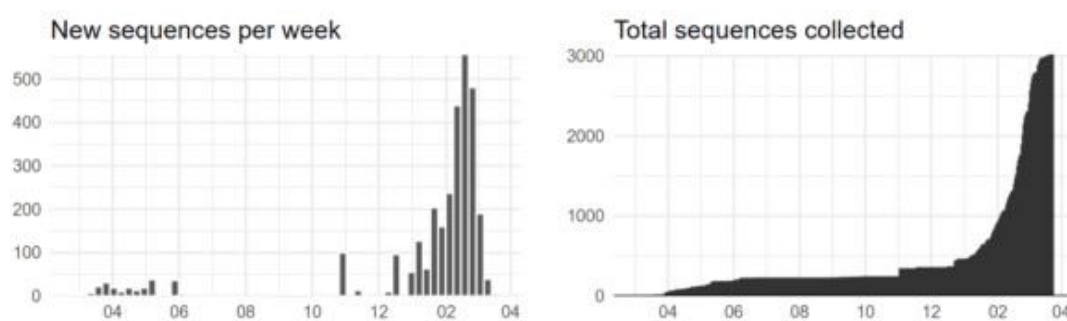
| 01–07 [Could new COVID variants undermine vaccines? Labs scramble to find out](#) |
01–15 [Alarming COVID variants show vital role of genomic surveillance](#) | 02–03
[Scientists call for fully open sharing of coronavirus genome data](#) | 03–19 [Rare COVID reactions might hold key to variant-proof vaccines](#) |

The question arises: what can we do to support the cause?

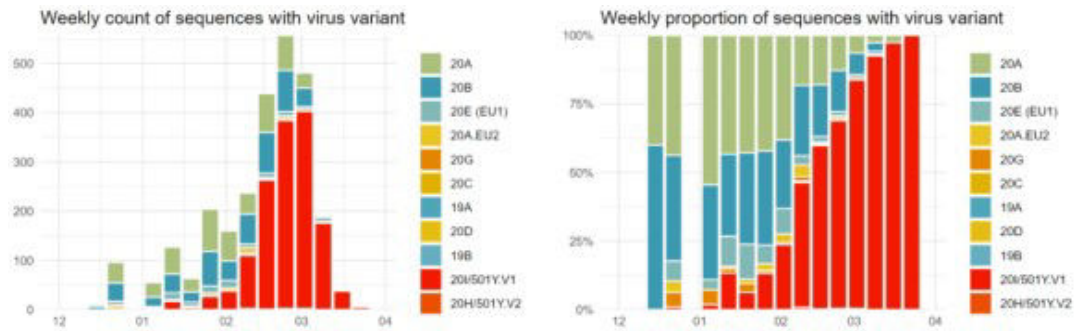
MI² against COVID

Within the [MI² against COVID](#) initiative, we developed a monitor of SARS-CoV-2 variants in Europe, based on the data from [GISAID](#). It is available at <https://monitor.mi2.ai> and updated daily.

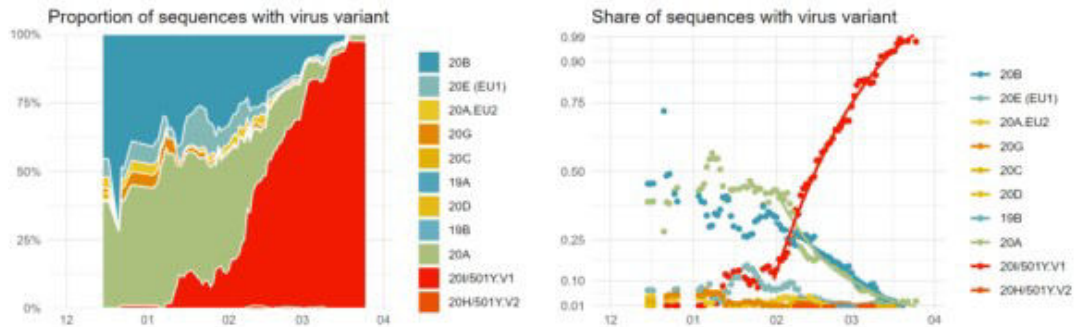
Currently, there are visualizations of sequences' metadata for 33 countries. Let's take a look at some of them on the example of Poland (2021–03–30).



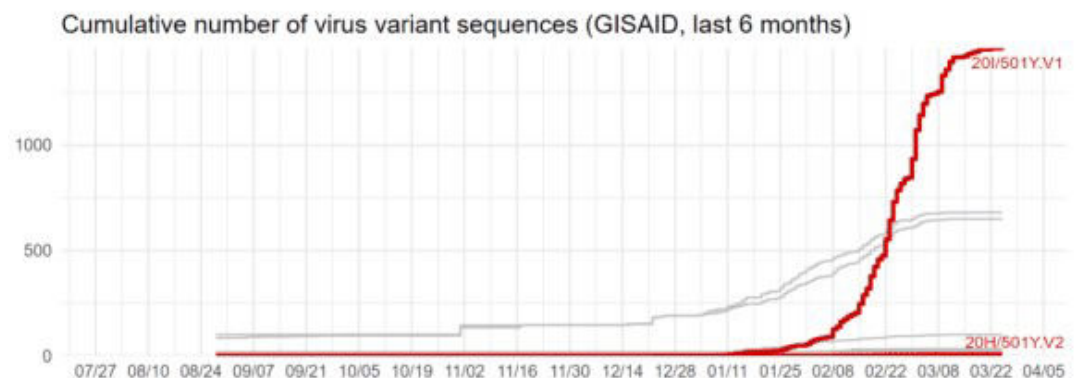
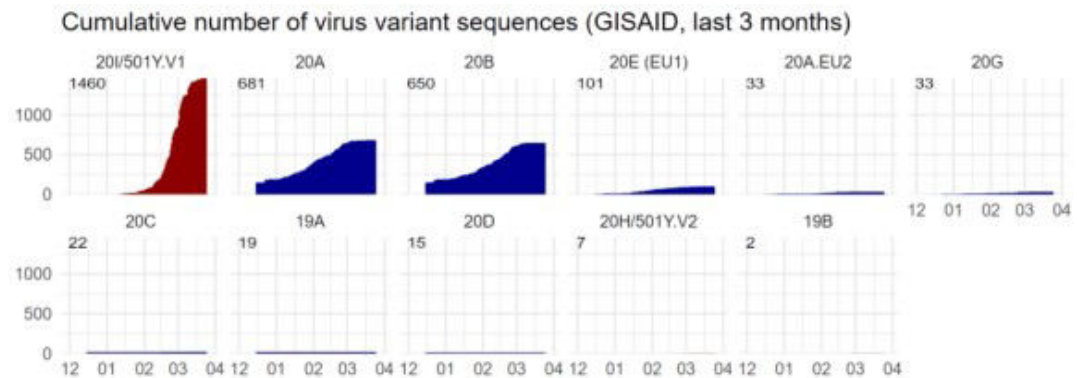
As of 2021–03–30, there are 3098 SARS-CoV-2 sequences from Poland in the GISAID database. The last sample was collected on 2021–03–25.



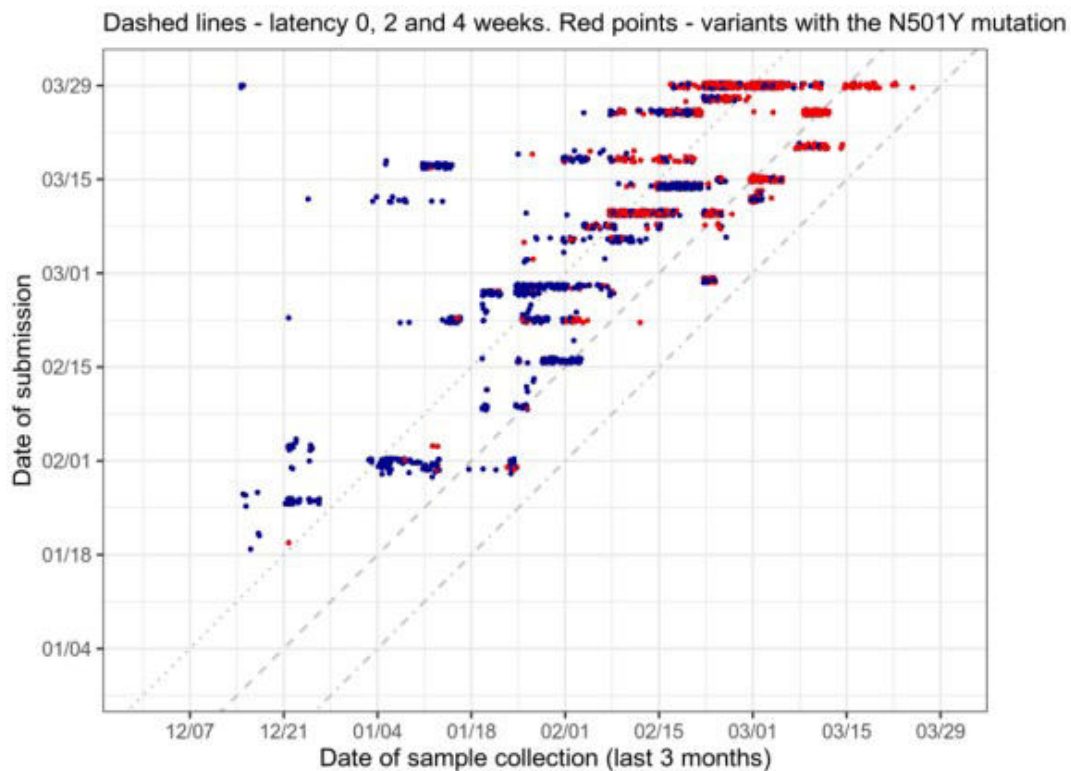
Looking at the last quarter, we see the change in proportion of the virus variants obtained from the GISAID database. The data is presented in a weekly aggregation.



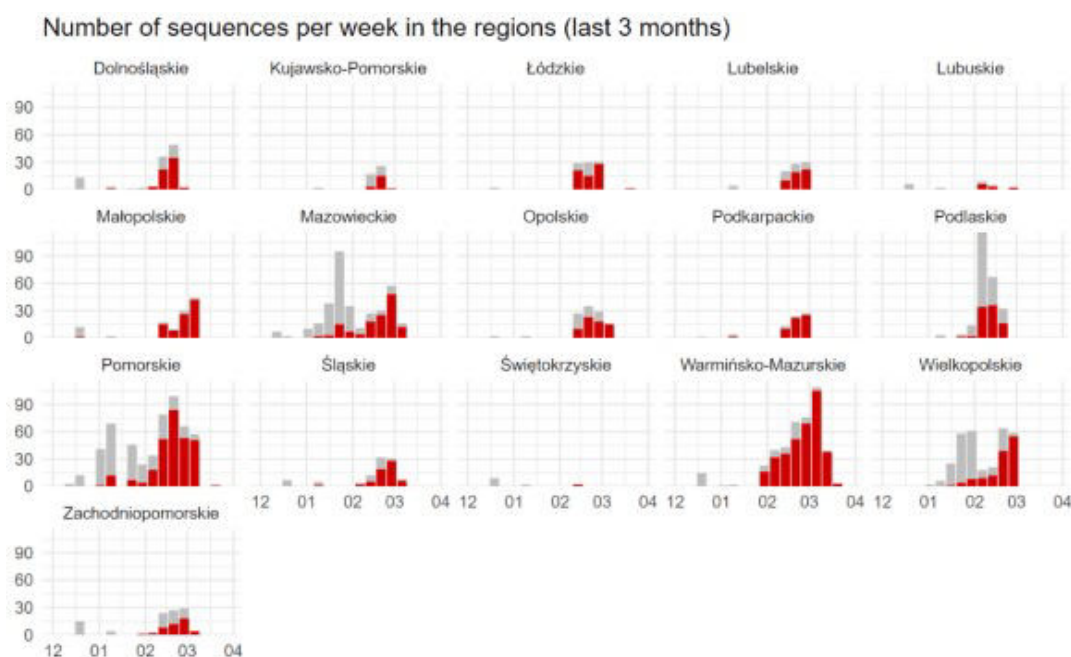
From the proportions, we obtain the share of the virus variants obtained from the GISAID database. The data is smoothed with a 7-day window.

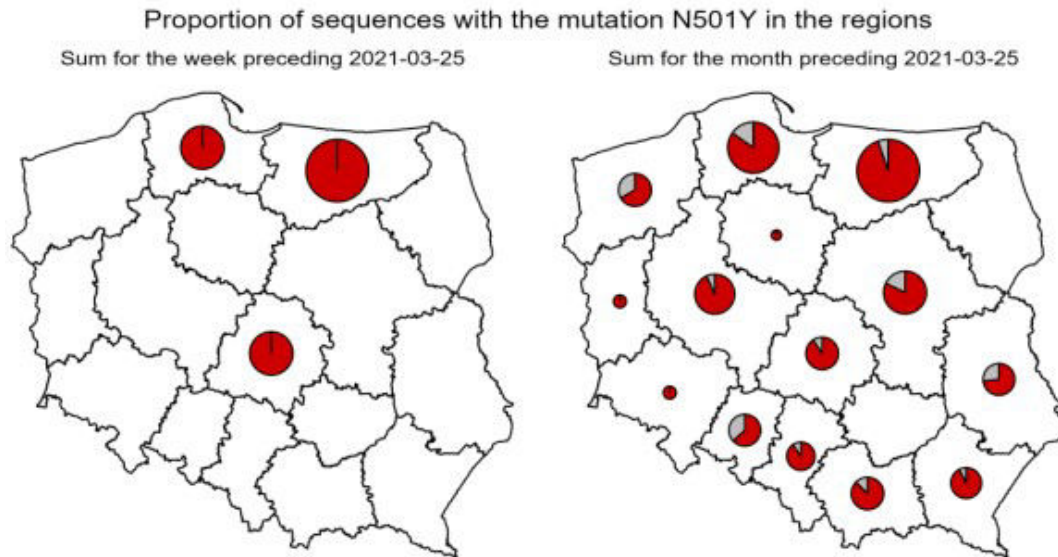


In our database, there are 11 variants of the virus tagged by the GISAID database. The most common are [20I/501Y.V1](#), [20A](#), [20B](#), [20E \(EU1\)](#), [20A.EU2](#). The variants with the N501Y mutation (highlighted in red) are alert variants with increased infectivity. Similar visualization [is available](#) for the virus lineages tagged by the Pango database.



Unfortunately for Poland, the time from the sample collection to its submission into the GISAID database is usually 2 to 5 weeks.





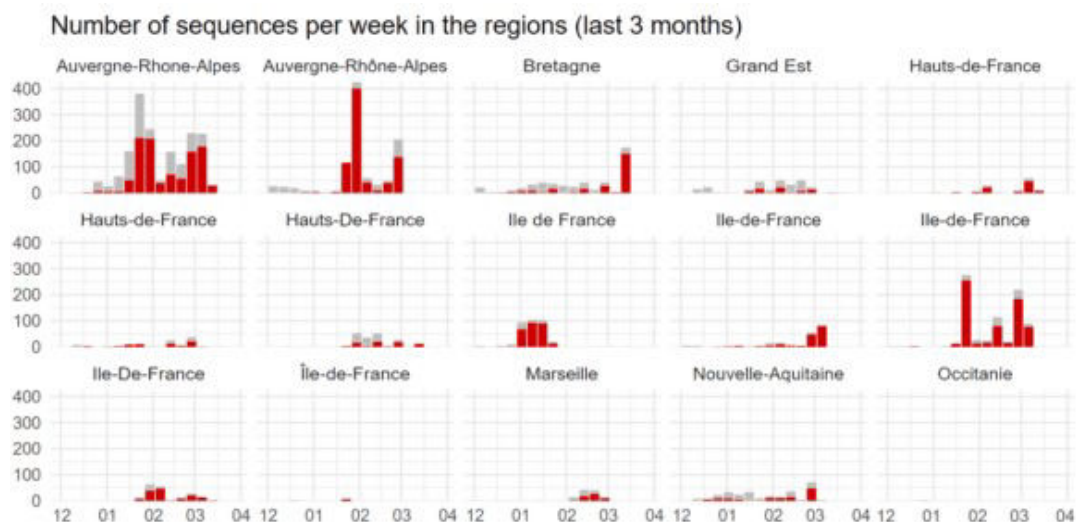
We monitor the number of variants with the N501Y mutation in the regions. More can be found at <https://monitor.mi2.ai/2021-03-30/poland/?lang=en>.

Details

The website is updated daily with reports for each European country. We automated the whole process with [AirFlow](#). Sequences and their metadata are acquired from the [GISAID](#) database. Then, they are preprocessed with [pangolin](#), [nextclade](#), and [blast](#) tools to be stored in-house. Periodically, the reports are generated using R tools from the [tidyverse](#), most notably [ggplot2](#) and [forcats](#). The map is generated using the [sf](#) and [scatterpie](#) packages. Finally, the website contents are updated on GitHub and hosted with Pages.

How to contribute?

The source code for this project is openly available at [GitHub](#). One can contribute by concatenating similarly spelled regions for any given European country. As seen below for France, there are various typos in the GISAID database, which can be corrected.



[Monitor of SARS-CoV-2 variants in France.](#)

Another way is providing a map of country regions visualized in R on the example of Poland. ...