

Read Me

virSEAK Online

virSEAK

JSI's SARS-CoV-2 tool - Analyze NGS Data / Mutation Surveillance

developed by

JSI medical systems GmbH

Tullastr. 18

77955 Ettenheim

GERMANY

phone: +49-7822/440150-21

fax: +49-7822/440150-20

email: support@varSEAK.bio

web: www.jsi-medisys.com

for research use only

2021-08-03

Table of Contents

1	Introduction.....	3
1.1	Disclaimer.....	3
2	Instructions for use.....	3
2.1	Menu.....	3
2.1.1	General features.....	3
2.2	Spreading.....	4
2.2.1	World/Country selection.....	4
2.2.2	Show Lineage Matrix.....	4
2.2.3	Lineages timeline and Lineages data.....	4
2.2.4	Variants timeline and Variants data.....	4
2.2.5	Isolates per month timeline and Isolates per month data.....	4
2.3	My Sequence.....	5
2.3.1	Graphical display.....	5
2.3.2	Buttons below the graph.....	5
2.3.3	Contents below the graph.....	6

1 Introduction

virSEAK is **JSI medical systems'** contribution to improve research related to the COVID-19 crisis. With this tool, you can look at the worldwide spread of the SARS-CoV-2 virus lineages and analyse your own fasta sequence to obtain mutations and a lineage.

We also have an [installable version](#) available to analyze high-throughput NGS sequencing data of SARS-CoV-2 samples with automatic upload of virus sequences to a central database server for mutation surveillance. Feel free to directly [contact us](#) in case you would like to know more or get a free trial license!

1.1 Disclaimer

This tool is intended for research use only by professionals.

2 Instructions for use

To display this web page we recommend the web browsers [FireFox](#) or [Chrome](#). Note: Internet Explorer is not supported.

2.1 Menu

Click the red button [*virSEAK*] or the SARS-CoV-2 image to navigate to the **virSEAK** page. At the top, two pages are offered: [*Spreading*], where you land on by default, and [*My Sequence*].

On [*Spreading*], you can get an overview about the SARS-CoV-2 lineages spreading worldwide and in different countries based on the dataset of available sequences.

On [*My Sequence*], you can analyze your own sequence.

The button [*Analyze NGS Data / Mutation Surveillance*] transfers you to the **JSI medical systems** homepage for more detailed information on how to analyze NGS data with **virSEAK**. You are of course also welcome to directly [contact us](#).

2.1.1 General features

2.1.1.1 Footer

At the bottom of the page, you can find general information, as well as this Read Me. There also is a Back To Top function that automatically brings you back to the top of the page.

2.1.1.2 Tables

Most tables have the same features: columns can be sorted ascending as well as descending and there is a search function at the top right corner. In the top left corner you can set how many entries should be displayed per page. Pages can be navigated in the lower right corner. The lower

left corner displays a navigation (e.g. "Showing 11-20 of 30 entries"). Alternatively, you can scroll through the whole table without any pagination.

2.2 Spreading

2.2.1 World/Country selection

On the left side, you can select for which country the data should be displayed. By default, the World is selected (and hence highlighted in orange).

There also is information available how many isolates are in the dataset for each country, whether they have found known lineages of concern (B.1.1.7, B.1.351, B.1.617.2 and P.1 at this timepoint), new rising lineages and how many variants (i.e. mutations) are contained.

2.2.2 Show Lineage Matrix

This pop-up displays all mutations in rows that are present in any of the samples that were assigned to the respective lineage. The frequency of this mutation for this lineage is given in the respective column.

2.2.3 Lineages timeline and Lineages data

This display is responsive to the country selection.

You can see how the amount of isolates for each known lineage develops over time. Known lineages will be displayed in color, other lineages in light grey. If there are lineages that were rising but are now decreasing in amount, they will be listed in other shades of grey.

The table to the right of the display show the highest value each known lineage has ever had (in percentage for that month), together with the absolute count of samples with this lineage in the respective month.

Click the line of a lineage to display another table to the right with a list of countries and their counts where this lineage is also present.

2.2.4 Variants timeline and Variants data

This display is responsive to the country selection.

By default, all variants listed in the variants data table are displayed in the graph. Use the checkboxes to change the selection.

Select a row and the corresponding line in the display will be highlighted in red.

Additionally, another table will be opened to the right, listing the temporal data of this variant, i.e. in how many isolates it was present each month (in absolute and percentage values).

2.2.5 Isolates per month timeline and Isolates per month data

This display is responsive to the country selection.

Here you can view how many samples (isolates) have been submitted each month per country.

2.3 My Sequence

2.3.1 Graphical display

The graph at the top of the page is headed by basic information about the virus and the current dataset (date of the dataset, number of sequences, source, reference, sequence length). The graph displays the features of the SARS-CoV-2 virus genome (horizontal bars with different colors depending on the type of feature). Below that, the positions of the variants are marked with vertical bars. The height of the bars indicates how often the variant has occurred in the dataset. The most common variants with over 98 % frequency are not displayed in this graph.

On the right, a legend indicates the meanings of the respective color.

You can interact with the graph: when hovering over either scale (horizontal or vertical), a line will appear across the graph for you to compare the position.

Hovering over features will open a tooltip with the name, position and, if available, protein id of the feature.

Hovering over the vertical bars will open a tooltip with the position, gene or region, the variant (nucleotide change) and, if applicable, amino acid change, as well as the frequency of the variant.

Below the legend on the right, you can change the filter for the frequency being displayed. By default, only variants above 4.0 % frequency are displayed. You can select to display all variants or all variants above 1.0, 2.0, 3.0, 4.0 or 5.0 %, respectively.

When a sequence has been analyzed, the graph will contain an additional line in the middle, displaying the variants of the analyzed sequence as dots. Again, a tooltip appears when hovering over a dot, displaying the position, gene or region, variant (nucleotide change), and, if applicable, the amino acid change.

2.3.2 Buttons below the graph

2.3.2.1 *[Analyze My Sequence]*

This button opens a pop-up window where you can either select a .fasta file from your computer (maximum size 100 kb) or paste a nucleotide sequence into the input field (mind the requirements listed below the field for successful analysis).

Upon pressing the button *[Analyze]*, **virSEAK** will align your sequence to the reference and compare the found variants to the dataset. You will then be able to see the variants in the graph as well as in the table below. Between the buttons and the table, your sequence will be assigned to a type. Please refer to the corresponding chapters for more details.

Press *[Reset]* to unload a file, *[Clear]* to empty the input field, and *[Cancel]* or *[X]* to close the pop-up.

2.3.2.2 *Examples:*

After the heading "Examples:" you will find several buttons with example sequences. Use them to see what happens upon analysis of a sequence. There are the following buttons:

- normal: this is a sample without a N501Y or E484K mutation, which therefore does not trigger the Alert function. It has 9 variants and is assigned to the B.1.319 lineage by Pangolin.
- Alpha: this is an example containing the mutations required to be assigned to the lineage B.1.1.7, resulting in the N501Y alert.
- Beta: this is an example containing the mutations required to be assigned to the lineage B.1.351, also resulting in the N501Y + E484K alert.
- Gamma: this is an example containing 15 of the 17 the possible mutations to be assigned to the lineage P.1, also resulting in the N501Y + E484K alert.
- N501Y: this is an example containing just the N501Y mutation which is associated with increased transmissibility. The alert function warns that it is present, while Pangolin assigns the lineage B.1.36.10.

2.3.3 Contents below the graph

2.3.3.1 Alert

The alert will warn if any of the following variants is present in the sample: L452R, E484K, E484Q, N501Y or P681R.

The list of variants may be updated automatically if new important variants arise.

If no alert is displayed, none of these variants could be found in this sample.

2.3.3.2 Pangolin lineage

To the right of the alert, the Pangolin lineage is shown (may take a few seconds to calculate). A linkout from the heading transfers you to <https://cov-lineages.org/>. Below, the following Pangolin fields are given: Lineage (in red and a bigger font size), Conflicts, Ambiguity Score, Scorpio Call, Scorpio Support, Scorpio Conflict, Status, Note, Pangolin version, PangoLEARN version and Pango version.

2.3.3.3 GISAID Clade

To the right, the assignment of the GISAID Clade is given. The GISAID Clade is calculated according to <https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/> (last accessed 2021-07-16) with the following marker variants:

S: C8782T,T28144C includes NS8-L84S
L: C241,C3037,A23403,C8782,G11083,G26144,T28144 (early clade markers in WIV04-reference sequence)
V: G11083T,G26144T NSP6-L37F + NS3-G251V
G: C241T,C3037T,A23403G includes S-D614G
GH: C241T,C3037T,A23403G,G25563T includes S-D614G + NS3-Q57H
GR: C241T,C3037T,A23403G,G28882A includes S-D614G + N-G204R
GV: C241T,C3037T,A23403G,C22227T includes S-D614G + S-A222V
GRY: C241T,C3037T,21765-21770del,21991-21993del,A23063T,A23403G,G28882A includes S-H69del, S-V70del, S-Y144del, S-N501Y + S-D614G + N-G204R

2.3.3.4 Variants

Here, a table lists the variants of your sequence, by default sorted by position (ascending). Furthermore, you can see the nucleotide and amino acid (AA) change, variant type (e.g. SNP or Del) gene or feature, affected mat. peptides, protein id, count and frequency. Variants that are relevant for the alert are highlighted in orange/yellow.

Note: As our company comes from a human genetics background, we are used to call genetic variation a "variant", not a mutation.

2.3.3.5 Alignment

On this tab, you can see the alignment of your sequence to the reference sequence. At the top, you can use the button *[Toggle Show All / Only Variants]* to fold out or fold in the lines with no deviations to the reference sequence.

Also, the length of your sequence, the number of variants and the number of the bases A, G, T, C and Others is displayed.

Below that, the alignment is given. The upper line is the reference (ref), the lower line is your sequence (My Seq). Missing bases (e.g. in the beginning of the sequence) are given as a period ("."), for deletions and insertions, a minus is used ("-"). The blue number is the position of the first base in this line. Between the sequences, colon (":") indicates identical bases whilst an asterisk ("*") highlights differences.