Read Me

varSEAK Online

virSEAK
JSI's SARS-CoV-2 tool

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

2020-08-20

Table of Contents

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 variations of the SARS-CoV-2 virus genome and get an assignment to a type.

1.1 Disclaimer

This tool is intended for research use only by professionals. Details about its validation and how the types and subtypes of SARS-CoV-2 have been defined can be found in the **Development of virSEAK**[1] document. The frequency of subtypes or variants is not necessarily representative of the distribution in the world, as it is based on the sequencing data available through the NCBI Virus[2] database.

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: *[My Sequence]*, where you land on by default, and *[Overview]*.

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

On *[Overview]*, you can get an overview about the SARS-CoV-2 types and variants based on the dataset of available sequences.

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 and the **Development of virSEAK**[1] document. There also is a Back To Top function that automatically brings you back to the top of the page.

2.1.1.2 Tables

All 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").

2.2 My Sequence

2.2.1 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.

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.

2.2.2 Buttons below the graph

2.2.2.1 [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.

2.2.2.2 [Show Example]

Use the *[Show Example]* button to see what happens upon analysis of a sequence.

2.2.3 Contents below the graph

2.2.3.1 Type assignment

Directly below the buttons, the result of the assignment of your sequence to a type is displayed. Click the type to open the corresponding pop-up with more details.

Below, information concerning the type is displayed:

- The genotype patterns of the type and your sequence in detail. Differences are highlighted in dark grey with white letters
  - When hovering over a base of the pattern, the corresponding position in the graph will be highlighted in orange and vice versa. The tooltip for the position will be opened as well.
- A summary of how many matching, mismatching and missing positions your sequence's genotype pattern has as compared to the type.
- Evolution depicts the relation of your type to the reference sequence (see also below: Type tree).
- The frequency of this type is given.

2.2.3.2 Type tree

This tree structure to the right of the type assignment shows the relations between the types. The assigned type is highlighted after analysis.
Hovering over a type will open a tooltip displaying the Genotype pattern, number of isolates with this type, number of countries with this type, mismatches of this pattern to the reference sequence (root) and to the sequence of the parent (next type closer to the reference).
Click a type in the tree to open the corresponding pop-up.

2.2.3.3 Identical Isolates

To the right, the number of identical isolates in the available data is displayed together with the number of countries that this sequence has been found in. Click the number of countries to open a popup with the detailed list of countries.
Note that only completely identical sequences (no N positions) are considered.

2.2.3.4 Type (pop-up)

The type is linked to a pop-up window. In the header, the type and its respective genotype pattern is given.
On the first tab "Type", the types are displayed in a tree structure to show the relations between the types. Hovering over a type will open a tooltip displaying the Genotype pattern, number of isolates with this type, number of countries with this type, mismatches of this pattern to the reference sequence (root) and to the sequence of the parent (next type closer to the reference).
The selected type is highlighted and to the right, a corresponding table is displayed, containing three columns: Country, number of isolates and percentage of isolates. The sum for each column is given in the header of the table. By default, the table is sorted descending by number of isolates.

On the second tab, "Evolution", the same kind of table is displayed. This tab contains the tables for the selected type and its ancestors. The selected type is highlighted, the other types are linked.

Hovering will highlight the selected country in yellow in each table. Click a country to stick the highlight to it (orange).

On the third tab, "Patterns" you can view a table of the genomic positions that are considered for the type assignment. The first line lists the corresponding base of the reference sequence. The lines below that are filled with the patterns of the respective types. Only deviations are given: if the field is empty, the base is identical to the reference.
Your selected type is highlighted. Hover over the columns to open a tooltip with more details concerning the respective position and variant.
Ancestors of your selected type are highlighted in lighter shades.

Click *[Close]* or *[X]* to close the pop-up.

2.2.3.5 Isolate countries (pop-up)

The number of countries where an identical sequence has been found, is linked to a pop-up window displaying the list of countries in detail, together with the respective amount of findings in that country.

2.2.3.6 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, gene or feature, affected mat. peptides, protein id, count and frequency.
Variants that are relevant for type assignment are highlighted in orange/yellow.

2.2.3.7 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.

2.3 Overview

2.3.1 Graphical display

Here, the same graph is displayed as on My Sequence before you analyze anything. Please refer to chapter 2.2.1, page 4 for more details.

2.3.2 Contents below the graph

2.3.2.1 Variants

This table lists all variants contained in the dataset. The table contains the columns position, nucleotide change, amino acid change, gene, affected mat. peptides, protein id, count and frequency. By default, it is ordered by Count (descending).
The table is responsive to the Frequency-Filter below the legend to the right of the graph. Only variants above the selected frequency will be displayed.

2.3.2.2 Types

This tab is subdivided into two further tabs: Types and Countries.

Types is very similar to two tabs of the Type (pop-up) (see page 5). On the left, you can see the type tree, on the right, the table from the "Patterns" tab. The names of the types link to the pop-up as well.

Countries contains an overview over the countries each type has been found in. Each table contains three columns: Country, number of isolates and percentage of isolates. The sum for each column is given in the header of the table. By default, the table is sorted descending by number of isolates. The name of the type links to the type pop-up.
Hovering will highlight the selected country in yellow in each table. Click a country to stick the highlight to it (orange).

2.3.2.3 Tracks

This tab displays the Type tree (see page 5) as well as a table of all countries available in the dataset. For each country, the number of isolates and the number of types is given. You can use the search box above the table to refine the results.

When you click on a country in the table, it is highlighted in orange and the types that are present in this country are highlighted orange in the type tree. Numbers are displayed in the boxes of the types, indicating how often the respective type has been found (the sum of those numbers is the number of isolates in the table).

Additionally, a graph is displayed showing the number of findings for each type over time. By default the whole world is selected. Search for and click another country to see the data for it. The axes scale automatically depending on the amount of findings. Each type is represented in a different color. If you are uncertain, move the mouse upon the line or a dot to get a tooltip telling you the type, and for dots, the number of findings as well.
The numbers are added up from month to month.

3 References

1 Development of virSEAK (accessed 2020-05-08)
https://varseak.bio/pdf/Development_virSEAK.pdf

2 NCBI Virus (accessed 2020-05-03)
https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=SARS-CoV-2,%20taxid:2697049