

The SARS-CoV-2 Real-time Tracking and Early Warning System for SARS-CoV-2 Variants and Lineages of Concern (VoCs/LoCs)

COVID-19 is currently a worldwide epidemic, with viral mutations creating new, time-critical concerns. The BRCs play an essential role in supporting infectious disease research and must respond to emerging diseases in a manner that best enables researchers working on countermeasures. Thus, a major goal of the system enhancement is to facilitate the rapid, near real-time, identification of these variants/lineages of concern (VoCs/LoCs):

- Variants affecting transmissibility
- Variants affecting virulence
- Variants affecting antibody neutralization (immune evasion)
- Variants affecting detection

The SARS-CoV-2 resource can be accessed through banners on the ViPR home page and *Coronaviridae* and SARS-CoV-2 portal pages:

← → ↻ viprbrc.org/brc/home.spg?decorator=vipr ☆ 🌐 🔍 📄 📁 📧 📧 📧

ViPR
Virus Pathogen Resource

About Us Community Announcements Links Resources Support

 **Latest Variant of Concern for SARS-CoV-2**

Search → **Analyze** → **Save to Workbench**

Search our comprehensive database for:

- Sequences & strains
- Immune epitopes
- 3D protein structures
- Host Factor Data
- Antiviral Drugs
- Plasmid Data

[Browse All Search Types](#)

Analyze data online:

- Sequence Alignment
- Phylogenetic Tree
- Sequence Variation (SNP)
- Metadata-driven Comparative Analysis
- BLAST
- VIGOR4 Genome Annotator

[Browse All Tools](#)

Sign up for a workbench to:

- Store and share data
- Combine working sets
- Integrate your data with ViPR data
- Store and share analyses
- Custom search alert

[Sign In](#)

There is an additional way to access this information through the beta site for BV-BRC (<https://beta.bv-brc.org/>). To drill down and see more information about SARS-CoV-2 lineages and variants of concern, click on the SARS-CoV-2 Variants of Concern hyperlink that can be found at the top of the page.

BV-BRC BETA WORKSPACES ABOUT CONTACT TEAM ANNOUNCEMENTS PUBLICATIONS CITATION RELATED RESOURCES

NEW: Check out our **SARS-CoV-2 Variants and Lineages of Concern** and Genome Assembly and Annotation Service.

BACTERIAL AND VIRAL BIOINFORMATICS RESOURCE CENTER

The Bacterial and Viral Bioinformatics Resource Center (BV-BRC) is an information system designed to support research on bacterial and viral infectious diseases via integration of vital pathogen information with rich data and analysis tools. BV-BRC combines two long-running centers: PATRIC, the bacterial system, and IRD/VIPR, the viral systems.

PATRIC

The Pathosystems Resource Integration Center (PATRIC) provides an extensive collection of integrated bacterial data, tools, and visualizations for genomic analysis to aid researchers in discovery and characterization of mechanisms of pathogenicity, supporting development of diagnostics, therapeutics, and vaccines.

Quickstart Video
User Guides
Tutorials
Common Tasks
CLI Tutorial
Instructional Videos
Contact Us / Provide Feedback

IRD AND VIPR

The Influenza Research Database (IRD) and Virus Pathogen Resource (ViPR) provide a wide range of data, analysis and visualization tools, and personal workbenches for the virus research communities that facilitate understanding of viral pathogens and how they interact with their host, leading to new diagnostics, treatments and preventions.

IRD Help Manual
IRD Tutorials & Training Methods
Contact IRD
ViPR Help Manual
ViPR Tutorials & Training Materials
Contact ViPR

NEWS & ANNOUNCEMENTS

[New PATRIC Online Bacterial Bioinformatics Course](#)

We have released a new online Bacterial Bioinformatics course, freely available through Coursera. ... [read more](#)

Bacterial Bioinformatics using PATRIC

Tweets by @BVBR_DB

BVBR Retweeted
VEuPathDB @veupathdb

Variants of Concern Overview

Clicking on the hyperlink will take you to an Overview page that shows summary information about the initial five lineages of concern (e.g., PANGO lineage information, key amino acid substitutions, functional impact). This page also contains a series of tabs that provide links to additional information about sequence prevalence, overlap with sequence feature through a genome browser, etc.

BV-BRC BETA WORKSPACES ABOUT CONTACT TEAM ANNOUNCEMENTS PUBLICATIONS CITATION RELATED RESOURCES

SARS-COV-2 VARIANTS AND LINEAGES OF CONCERN

Tabs with additional information

Overview | Lineages of Concern | Lineage Prevalence | Variant Prevalence | Genome Browser | Protein Structure | Phylogeny | Resources

Welcome to BV-BRC SARS-CoV-2 Real-time Tracking and Early Warning System for SARS-CoV-2 Variants and Lineages of Concern (VOCs)

The SARS-CoV-2 Variants and Lineages of Concern resource

- Identifies and tracks emerging variants and lineages through daily processing of publicly available SARS-CoV-2 sequences
- Performs risk assessment on the variants to identify candidate VOCs by leveraging a growing knowledgebase of sequence features, including protein domains, functional regions, and immune epitopes
- Provides detail information about each VOC
- Provides their sequence prevalence in various countries and regions over time using interactive dashboards and charts
- Provides integrated view of the VOCs and importance sequence features using genome browser, protein structure viewers, and phylogenetic trees

| Lineage Of Concern | PANGO lineage | Nextstrain lineage | Other synonyms | Emergence date | Emergence location | Key AA substitutions in spike protein | Impact |
|--------------------|---------------|--------------------|--------------------------|----------------|----------------------------------|--|---|
| B.1.1.7 | B.1.1.7 | 2019Y1Y1 | VOC 20201301, UK variant | September 2020 | Southeast England | H69, V70, N501Y, D614G, P681H | Increased transmissibility; S gene target failure (SGTF) |
| B.1.351 | B.1.351 | 20H501YV2 | South African variant | October 2020 | Nelson Mandela Bay, South Africa | L241, L242, A243, K417N, E484K, N501Y, D614G | loss of serum antibody neutralization |
| P.1 | B.1.1.28 | 20J501YV3 | Brazilian variant | July 2020 | Brazil | K417T, E484K, N501Y, D614G | Increased transmissibility; loss of serum antibody neutralization |
| CAL.20C | B.1.429 | | | July 2020 | Southern California, USA | W152C, L452R, D614G | loss of monoclonal antibody binding |
| B.1.375 | B.1.375 | | | September 2020 | Massachusetts, USA | H69, V70, D614G | S gene target failure (SGTF) |

Links to PANGO information

Links to recent publications

Recent PubMed Articles

- 2021 Mar-Apr
BRI m24373 SNP and COVID-19 mortality.
Laher S and Rheinlein PH
World Acad Sci J
- 2021 Jan 28
Will the emergent SARS-CoV-2 B.1.1.7 lineage affect molecular diagnosis of COVID-19?
Rimmer JD et al.
J Med Virol
- 2021 Jan 27
Spike ACE2 as a potential therapy for COVID-19.
Kishoremurthy D et al.
Ann J Physiol Cell Physiol
- 2021 Jan 20
Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein.
Supradene N et al.
bioRxiv
- 2021 Jan 8
A novel variant and multiple introductions of SARS-CoV-2 initiated the COVID-19 epidemic in Greece.
Spanaki M et al.
J Med Virol

show more >>

Lineage of Concern

To see more information on the five lineages of concern, click on that tab.

SARS-COV-2 VARIANT AND LINEAGES OF CONCERN

Overview | **Lineages of Concern** | Lineage Prevalence | Variant Prevalence | Genome Browser | Protein Structure | Phylogeny | Resources

This will open a page that shows detailed information on these lineages. Although the default shows the B.1.1.7 lineage, clicking on the down arrow below that box will allow you to view the detailed information about any of the defined LoCs. The information provided on these pages includes:

- The name of the LoC
- The PANGO lineage
- The NextStrain lineage
- Other synonyms VOC
- Emergence location: The place where the lineage/variant was first identified
- Emergence date: The date when the lineage/variant was first identified
- Amino Acid substitutions in the Spike protein: The mutations are defined by comparing a consensus sequence derived from available lineage isolates to the reference strain (see SOP) and are depicted as follows: Wuhan amino acid residue_Location in the Spike Protein_Amino acid replacement in the LoC.
 - D614G means that a D in the Wuhan strain was replaced by a G in the B.1.1.7 lineage.
 - H69- means that the H in the Wuhan strain was lost in the B.1.1.7 lineage
 - -69H would mean that an insertion of an H at site 69 would be found in the B.1.1.7 lineage.
 - An * means that the mutation is found in 25% of the variants
 - A ** means that mutations are found in more than one variant of concern.
- Amino acid substitutions in non-spike protein and whole genome nucleotide substitutions are also provided.
- Links to consensus sequences and representative strains.
- Reference to relevant citations regarding their functional impact.

| SARS-COV-2 VARIANTS AND LINEAGES OF CONCERN | |
|---|---|
| Overview | Lineages of Concern |
| Lineage Prevalence | Variant Prevalence |
| Genome Browser | Protein Structure |
| Phylogeny | Resources |
| Lineages Of Concern | |
| Select Lineage of Concern (LoC): | B.1.1.7 |
| LoC name | B.1.1.7 |
| PANGO lineage | B.1.1.7 |
| NextStrain lineage | 20I/501Y.V1 |
| Other synonyms VOC | 202012/01, UK variant |
| Emergence location | Southeast England |
| Emergence date | September 2020 |
| Amino acid substitutions vs Wuhan-Hu-1: Spike | H69-**, V70-**, N501Y**, A570D, D614G**, P681H, T716I, S982A, D1118H |
| Spike Short Peptide Search | 69/70 deletion: NVTWFHAISGTNGTKRFD (AND) P681H: TQTNSPRRARS |
| Amino acid substitutions vs Wuhan-Hu-1: Non-Spike | nsp3: T183I, A890D, I1412T; nsp6: S106-, G107-, F108-; RNA-dependent RNA polymerase: P323L; helicase: K460R; ORF8: Q27stop |
| Nucleotide substitutions vs Wuhan-Hu-1 | C241T, C913T, C3037T, C3267T, C5388A, C5986T, T6954C, T11288-, C11289-, T11290-, G11291-, G11292-, T11293-, T11294-, T11295-, T11296-, C14408T, C14676T, C15279T, T16176C, A17615G, T21765-, A21766-, C21767-, A21768-, T21769-, G21770-, T21991-, T21992-, A21993-, A23063T, A23271C, A23403G, A23604C, T23709C, G24506T, C24914G, T27972C, G28048T, G28111A, A28271-, G28280C, A28281T, T28282A, G28881A, G28882A, G28883C, C28977T |
| Impact | Increased transmissibility; S gene target failure (SGTF) |
| SF overlap | TBD |
| ViPR representative strain link | SARS-CoV-2/human/USA/FL-CDC-STM-P012/2020 |
| GISAID representative strain | hCoV-19/England/MILK-9E05B3/2020 EPI_ISL_601443 |
| Consensus genome sequence | B.1.1.7_North America_genome_consensus.fasta |
| Consensus Spike protein | sequence B.1.1.7_North America_S protein_consensus.fasta |
| Consensus all protein | sequences B.1.1.7_North America_all proteins_consensus.fasta |
| Relevant publications | 1. CDC Emerging Variants 2. Virological.org |

* denotes mutations found in at least >25% variants, but not in all sequences
 ** denotes mutations found in more than one VoC indicating possible positive selection, shared ancestry and/or convergent evolution

To see information on additional lineages of concern, click on the drop-down box at the end of the lineage name. This will open the list of lineages, and after clicking one, the page will reload to show that data.

Select Lineage of Concern (LoC):

B.1.1.7

B.1.1.7

B.1.351

P.1

CAL.20C

B.1.375

Lineage Prevalence

BV-BRC is currently focusing on five lineages of concern (B.1.1.7, B1.351, P.1, CAL.20C and B.1.375). The sequence characteristics that define these lineages are defined on the Lineages of Concern tab, but there is additional information available about these lineages that can be

seen by clicking on the Lineage Prevalence tab.



This will open a table that shows information about all non-synonymous substitutions identified in the Spike protein. The table has three inner tabs (Table, By County Chart, and By Lineage Chart). The table has a dynamic filter at the top, followed by the table that has the following columns:

- Lineage: Information on all unique amino acid substitutions identified in the spike glycoprotein is provided.
- Sequence Features: These include protein domains, secondary structure elements, functional regions (e.g., receptor binding residues, enzyme active sites), sites of post-translational modifications, protein-protein interaction sites, site in which substitutions have been shown to alter virus characteristics, antibody and T cell epitopes.
- Country: The country where the lineage is found
- Region: The region of the country where the lineage is found
- Month: The year and month that the lineage has been found.
- Total Isolates: The number of isolates found that have the variant. The number will change based on the dynamic filter.
- Lineage Count: The number of isolate genomes in this lineage. This number will vary depending upon the filtering.
- Sequence Prevalence: The number in this column represents the total isolates of this lineage from the indicated country or region and month divided by the total number of all isolate sequences from that same country or region and month.
- Growth Rate: The growth rate is computed month over month. This number in this column is the prevalence of this variant for the current month divided by the past month.

A summary of the available data is provided at the bottom left of the table. This number changes when the filter is used.

SARS-COV-2 VARIANTS AND LINEAGES OF CONCERN

Overview | Lineages of Concern | **Lineage Prevalence** | Variant Prevalence | Genome Browser | Protein Structure | Phylogeny | Resources

Table | By Country Chart | By Lineage Chart

Dynamic filter

keyword: Lineage: Any Sequence Features: Any

Country: Any Region: Any Month: Any Total Isolates >= 0 Lineage Count >= 0 Prevalence >= 0 Growth Rate >= 0 Filter Reset Filter

| Lineage | Sequence Features | Country | Region | Month | Total Isolates | Lineage Count | Sequence Prevalence | Growth Rate |
|--------------------------------------|-------------------|---------|---------------------|--------|----------------|---------------|---------------------|-------------|
| <input type="checkbox"/> D614G | | Albania | All | All | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Albania | All | 202009 | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Albania | Moerfelden-Walldorf | All | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Albania | Moerfelden-Walldorf | 202009 | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Algeria | Adrar | All | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Algeria | Adrar | 202006 | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Algeria | Alger | All | 3 | 1 | 0.3333 | 1 |
| <input type="checkbox"/> D614G,A892S | | Algeria | Alger | All | 3 | 1 | 0.3333 | 1 |
| <input type="checkbox"/> D614G,G769V | | Algeria | Alger | All | 3 | 1 | 0.3333 | 1 |
| <input type="checkbox"/> D614G | | Algeria | Alger | 202006 | 3 | 1 | 0.3333 | 1 |
| <input type="checkbox"/> D614G,A892S | | Algeria | Alger | 202006 | 3 | 1 | 0.3333 | 1 |
| <input type="checkbox"/> D614G,G769V | | Algeria | Alger | 202006 | 3 | 1 | 0.3333 | 1 |
| <input type="checkbox"/> D614G | | Algeria | All | All | 18 | 16 | 0.8889 | 1 |
| <input type="checkbox"/> D614G,A892S | | Algeria | All | All | 18 | 1 | 0.0556 | 1 |
| <input type="checkbox"/> D614G,G769V | | Algeria | All | All | 18 | 1 | 0.0556 | 1 |
| <input type="checkbox"/> D614G | | Algeria | All | 202003 | 3 | 3 | 1 | 1 |
| <input type="checkbox"/> D614G | | Algeria | All | 202006 | 14 | 12 | 0.8571 | 1 |
| <input type="checkbox"/> D614G,A892S | | Algeria | All | 202006 | 14 | 1 | 0.0714 | 1 |
| <input type="checkbox"/> D614G,G769V | | Algeria | All | 202006 | 14 | 1 | 0.0714 | 1 |
| <input type="checkbox"/> D614G | | Algeria | All | 202007 | 1 | 1 | 1 | 1.1667 |

1 - 200 of 25000 results

Available results

Filters are provided for each of the columns in the table. An additional keyword filter is also provided and can be used for any term.

keyword: Lineage: Any Sequence Features: Any

Country: All Region: All Month: All Total Isolates >= 10 Lineage Count >= 5 Prevalence >= 0.05 Growth Rate >= 1 Filter Reset Filter

Text can be entered into the filter box. Click on the down arrow at the end of any of the other filters will open a drop-down box where selections can be made.

Overview | Lineages of Concern | **Lineage Prevalence** | Variant Prevalence | Genome Browser | Protein Structure | Phylogeny | Resources

Table | By Country Chart | By Lineage Chart

keyword: Lineage: Any Sequence Features: Any

Country: All Region: All Month: Any Prevalence >= 0.05 Growth Rate >= 1 Filter Reset Filter

Lineage: Any

- A222V,A262S,P272L,D614G
- A222V,D614G
- A222V,D614G,Q675H
- A222V,E583D,D614G
- A222V,P272L,D614G
- A27S,D614G
- A27V,D614G
- A520S,D614G
- A522S,D614G
- A67V,D614G
- D138H,D614G
- D138Y,D614G
- D215Y,D614G

Any or all of the filtering facets can be used to generate Boolean AND operations. Simply enter text in the text box (if desired) and choose any other value, and then click on the Filter button.

keyword: Lineage: Any Sequence Features: Any

Country: USA Region: Any Month: Any Total Isolates >= 0 Lineage Count >= 0 Prevalence >= 0 Growth Rate >= 0 Filter Reset Filter

This will reload the table to show the results based on that selection.

keyword Lineage: Any Sequence Features: Any

Country: USA Region: All Month: All Total Isolates >= 100 Lineage Count >= 10 Prevalence >= 0 Growth Rate >= 1 Filter Reset Filter

| Lineage | Sequence Features | Country | Region | Month | Total Isolates | Lineage Count | Sequence Prevalence | Growth Rate |
|--|-------------------|---------|--------|-------|----------------|---------------|---------------------|-------------|
| <input type="checkbox"/> A222V,D614G | | USA | All | All | 78243 | 56 | 0.0007 | 1 |
| <input type="checkbox"/> A222V,L229F,D614G | | USA | All | All | 78243 | 11 | 0.0001 | 1 |
| <input type="checkbox"/> A243S,D614G | | USA | All | All | 78243 | 62 | 0.0008 | 1 |
| <input type="checkbox"/> A262S,D614G | | USA | All | All | 78243 | 14 | 0.0002 | 1 |
| <input type="checkbox"/> A263S,D614G | | USA | All | All | 78243 | 37 | 0.0005 | 1 |
| <input type="checkbox"/> A27S,D614G | | USA | All | All | 78243 | 27 | 0.0003 | 1 |
| <input type="checkbox"/> A27S,D614G,G1085A | | USA | All | All | 78243 | 117 | 0.0015 | 1 |
| <input type="checkbox"/> A27V,D614G | | USA | All | All | 78243 | 39 | 0.0005 | 1 |
| <input type="checkbox"/> A344S,D614G | RBD | USA | All | All | 78243 | 29 | 0.0004 | 1 |
| <input type="checkbox"/> A475S,D614G | RBD,RBD ACE2 | USA | All | All | 78243 | 63 | 0.0008 | 1 |
| <input type="checkbox"/> A520S,D614G | RBD | USA | All | All | 78243 | 85 | 0.0011 | 1 |
| <input type="checkbox"/> A522S,D614G | RBD | USA | All | All | 78243 | 20 | 0.0003 | 1 |
| <input type="checkbox"/> A522S,D614G,E780Q | RBD | USA | All | All | 78243 | 43 | 0.0005 | 1 |
| <input type="checkbox"/> A570S,D614G | | USA | All | All | 78243 | 12 | 0.0002 | 1 |

Nothing s
Select one o
on the left t
details and possi

The filter can be reset to the default values by clicking on the Reset Filter button.

keyword Lineage: Any Sequence Features: Any

Country: USA Region: Any Month: Any Total Isolates >= 0 Lineage Count >= 0 Prevalence >= 0 Growth Rate >= 0 Filter Reset Filter

Variant Prevalence

While Lineages of Concern describe constellations of mutations that share a common evolutionary ancestor, Variants of Concern track single amino acid substitutions that may share a common ancestor or have been generated through convergent evolution. The Variant Prevalence information is organized in a similar fashion to the LoC Prevalence described above. All variants are defined based on the reference Wuhan-hu-1 genome. To find information on the individual SARS-CoV-2 variants and their prevalence, which currently only include variations in the Spike protein (S), click on the Variant Prevalence tab.



This will open a table that shows all the variants in the spike protein. The table has three inner tabs (Table, By County Chart, and By Lineage Chart). The table has a dynamic filter at the top, followed by the table that has the following columns:

- **AA Variant:** The amino acid variant (indicated by the amino acid called in the Wuhan spike protein, the location of that amino acid in the protein sequence, followed by the amino acid seen in the variant).
- **Sequence Features:** These include protein sequence features, and include protein domains, secondary structure elements, functional regions (e.g., receptor binding residues, enzyme active sites), sites of post-translational modifications, protein-protein interaction sites, site in which substitutions have been shown to alter virus characteristics, antibody and T cell epitopes.
- **Country:** The country where the variant is found
- **Region:** The region of the country where the variant is found
- **Month:** The year and month that the variant has been found
- **Total Isolates:** The number of isolates found that have the variant. The number will change based on the dynamic filter.

- Lineage Count: The number of lineages that has the variant.
- Sequence Prevalence: The number in this column represents the isolates with a given aa variant from the indicated country or region and month divided by the total number of all isolate sequences from that same country or region and month.
- Growth Rate: The growth rate is computed month over month. This number in this column is the prevalence of this variant for the current month divided by the past month.

A summary of the available data is provided at the bottom left of the table. This number changes when the filter is used.

SARS-COV-2 VARIANTS AND LINEAGES OF CONCERN

Overview | Lineages of Concern | Lineage Prevalence | **Variant Prevalence** | Genome Browser | Protein Structure | Phylogeny | Resources

Table | By Country Chart | By Lineage Chart

keyword: Sequence Features: Any Country: Any Region: Any Month: Any

Total Isolates >= 0 Lineage Count >= 0 Prevalence >= 0 Growth Rate >= 0 **Filter** **Reset Filter**

| AA Variant | Sequence Features | Country | Region | Month | Total Isolates | Lineage Count | Sequence Prevalence | Growth Rate |
|--------------------------------|-------------------|---------|--------------------|--------|----------------|---------------|---------------------|-------------|
| <input type="checkbox"/> D614G | | Albania | All | All | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Albania | All | 202009 | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Albania | Moerfelden-Waldorf | All | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Albania | Moerfelden-Waldorf | 202009 | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Algeria | Adrar | All | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Algeria | Adrar | 202006 | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> A892S | | Algeria | Alger | All | 3 | 1 | 0.3333 | 1 |
| <input type="checkbox"/> D614G | | Algeria | Alger | All | 3 | 3 | 1 | 1 |
| <input type="checkbox"/> G769V | | Algeria | Alger | All | 3 | 1 | 0.3333 | 1 |
| <input type="checkbox"/> A892S | | Algeria | Alger | 202006 | 3 | 1 | 0.3333 | 1 |
| <input type="checkbox"/> D614G | | Algeria | Alger | 202006 | 3 | 3 | 1 | 1 |
| <input type="checkbox"/> G769V | | Algeria | Alger | 202006 | 3 | 1 | 0.3333 | 1 |
| <input type="checkbox"/> A892S | | Algeria | All | All | 18 | 1 | 0.0556 | 1 |
| <input type="checkbox"/> D614G | | Algeria | All | All | 18 | 18 | 1 | 1 |
| <input type="checkbox"/> G769V | | Algeria | All | All | 18 | 1 | 0.0556 | 1 |
| <input type="checkbox"/> D614G | | Algeria | All | 202003 | 3 | 3 | 1 | 1 |
| <input type="checkbox"/> A892S | | Algeria | All | 202006 | 14 | 1 | 0.0714 | 1 |
| <input type="checkbox"/> D614G | | Algeria | All | 202006 | 14 | 14 | 1 | 1 |
| <input type="checkbox"/> G769V | | Algeria | All | 202006 | 14 | 1 | 0.0714 | 1 |
| <input type="checkbox"/> D614G | | Algeria | All | 202007 | 1 | 1 | 1 | 1 |
| <input type="checkbox"/> D614G | | Algeria | Rileta | All | 1 | 1 | 1 | 1 |

1 - 200 of 25000 results

Filters are provided for each of the columns in the table. An additional keyword filter is also provided and can be used for any term.

keyword: Sequence Features: Any Country: Any Region: Any Month: Any

Total Isolates >= 0 Lineage Count >= 0 Prevalence >= 0 Growth Rate >= 0 **Filter** **Reset Filter**

Text can be entered into the filter box. Click on the down arrow at the end of any of the filters will open a drop-down box where selections can be made.

keyword: Sequence Features: Any

Total Isolates >= 0 Lineage Count >= 0 Prevalence >= 0

| AA Variant | Sequence Features | Country |
|--------------------------------|-------------------|---------|
| <input type="checkbox"/> D614G | | Albania |
| <input type="checkbox"/> D614G | | Albania |
| <input type="checkbox"/> D614G | | Albania |
| <input type="checkbox"/> D614G | | Albania |
| <input type="checkbox"/> D614G | | Algeria |
| <input type="checkbox"/> D614G | | Algeria |
| <input type="checkbox"/> A892S | | Algeria |
| <input type="checkbox"/> D614G | | Algeria |
| <input type="checkbox"/> G769V | | Algeria |

Sequence Features: Any

- Any
- Cleavage site
- Fusion peptide 1
- Fusion peptide 2
- Heptad repeat 1
- Heptad repeat 2
- KxHox motif
- Mutagenesis
- RBD
- RBD ACE2 Motif
- Signal peptide
- Transmembrane

Any or all of the filtering facets can be used. Simply enter text in the text box (if desired) and choose any other value, and then click on the Filter button. This will reload the table to show the results based on that selection.

BV-BRC BETA WORKSPACES ABOUT CONTACT TEAM ANNOUNCEMENTS PUBLICATIONS CITATION RELATE

SARS-COV-2 VARIANTS AND LINEAGES OF CONCERN

Overview Lineages of Concern Lineage Prevalence **Variant Prevalence** Genome Browser Protein Structure Phylogeny Resources

Table By Country Chart By Lineage Chart

China Sequence Features: Any Country: Any Region: Any Month: Any

Total Isolates >= 0 Lineage Count >= 0 Prevalence >= 0 Growth Rate >= 5 Filter Reset Filter

| AA Variant | Sequence Features | Country | Region | Month | Total Isolates | Lineage Count | Sequence Prevalence | Growth Rate |
|--------------------------------|-------------------|---------|----------|--------|----------------|---------------|---------------------|-------------|
| <input type="checkbox"/> T307I | | China | All | 202002 | 345 | 5 | 0.0145 | 5 |
| <input type="checkbox"/> S247R | | China | Hangzhou | 202002 | 4 | 1 | 0.25 | 7.5075 |
| <input type="checkbox"/> D614G | | China | All | 202003 | 165 | 44 | 0.2667 | 16.3931 |

The filter can be reset to the default values by clicking on the Reset Filter button.

A1226 Sequence Features: Any Country: Any Region: Any Month: Any

Total Isolates >= 0 Lineage Count >= 0 Prevalence >= 0 Growth Rate >= 0 Filter Reset Filter

Genome Browser

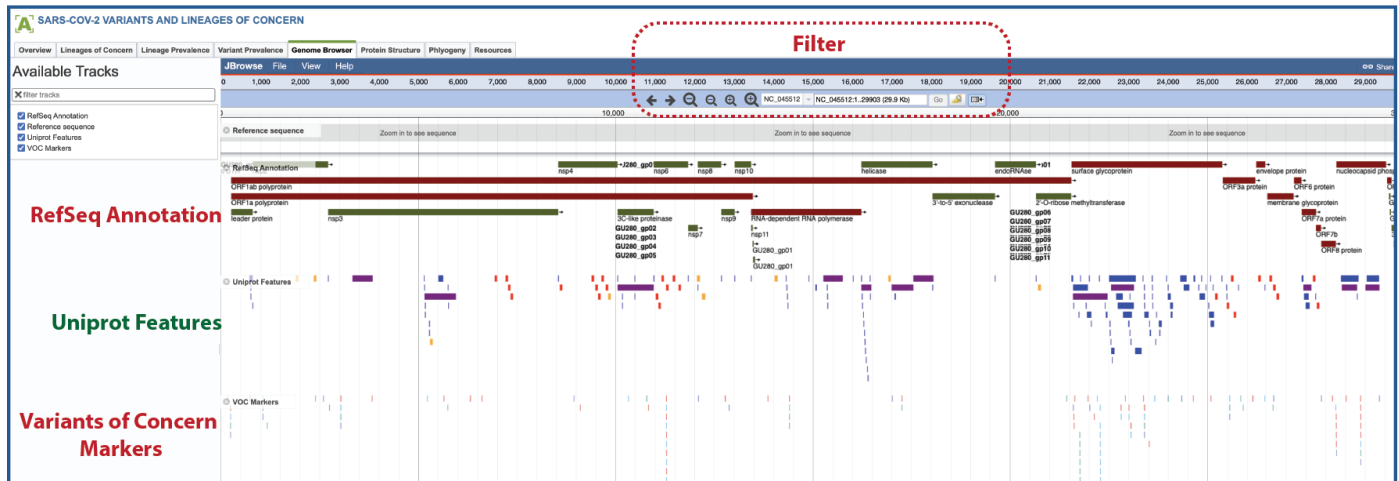
As with other members of the *Coronaviridae* family, SARS-CoV-2 exhibits a relatively high mutation rate. As a result, any given SARS-CoV-2 sequence typically carries multiple nucleotide substitution in comparison with the initial outbreak strain; for this purpose the Wuhan-hu-1 reference strain (NC_045512) can be considered a close representative of the initial outbreak strain. The JBrowse genome browser¹ displays the sequence for Wuhan-hu-1, its RefSeq annotation, corresponding Uniprot sequence features and IEDB B cell/Ab epitopes. In addition, substitutions that result in increased transmissibility or immune escape are variants of concern (VoCs) from a public health perspective, and are also displayed on the browser in their relation to features of the Wuhan-hu-1 reference strain.

To locate the genome browser, click on the tab named Genome Browser

SARS-COV-2 VARIANTS AND LINEAGES OF CONCERN

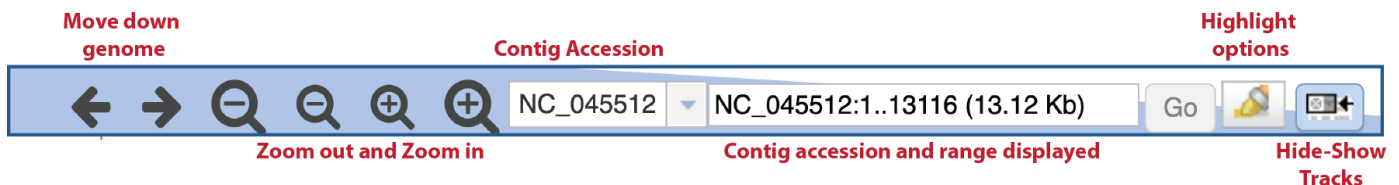
Overview Lineages of Concern Lineage Prevalence Variant Prevalence **Genome Browser** Protein Structure Phylogeny Resources

This will open the genome browser page. The page shows the annotations and the VoCs. The available tracks are shown to the right of the visualization.

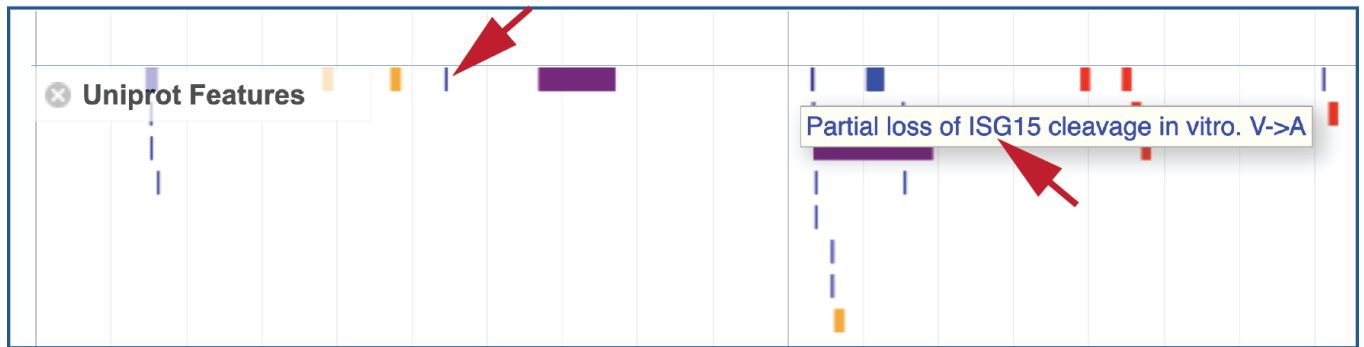


The browser view includes a dynamic filter. Researchers can use the arrow icons to move rapidly in a 5' or 3' direction along the reference genome. Magnifying icons with a minus (zoom out) or a plus (zoom in) sign within them provide an additional way to focus in on different aspects of the view.

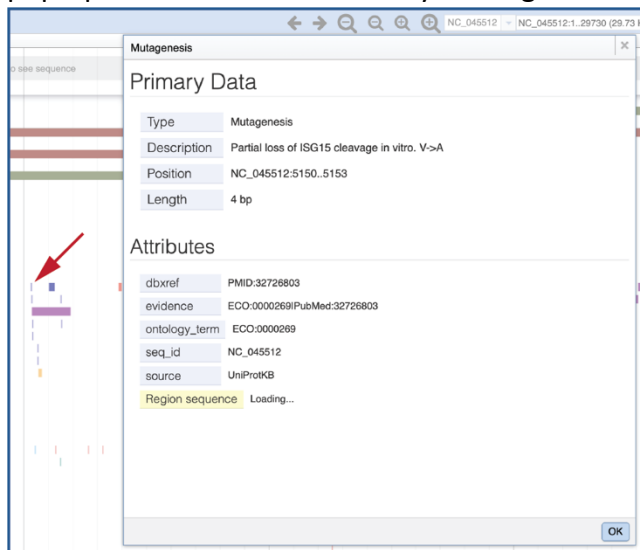
The text box following contig accession shows the contig and the start of the sequence, end of the sequence, and size of the sequence displayed. The display can be changed by adjusting the numbers following the contig (delineated by a ":"). Take care to maintain the format (1..13116) with the "." between the start and end of the sequence. Researchers can use the highlight option to draw emphasis to a specific area of the genome, and also hide or show tracks using that particular icon.



Many features are displayed on the browser. Zooming in can reveal the names of particular features, but this can also be done by mousing over a feature of interest. This will open a pop-up window that provides the name of that particular feature. The pop-up window usually appears to the right of the feature, and, depending on the particular view, could be somewhat downstream from item.

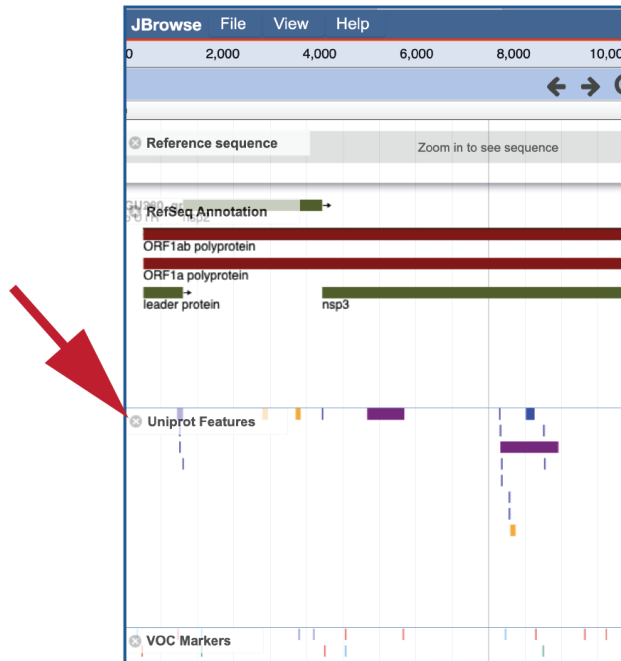


More information about a specific feature can be seen by actively clicking on that feature from the browser view. This will open a pop-up window that shows the data on the specific annotation (Primary Data) as well as additional data linked to that feature (Attributes). The pop-up window can be closed by clicking on the OK button at the bottom right of the window.



Clicking on one of the RefSeq annotations will load more information about that annotation.

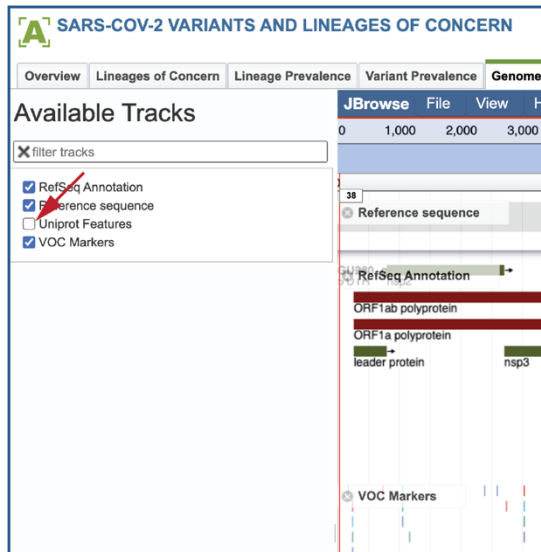
Removing individual tracks from a view is easily implemented by clicking on the "X" in front of the name of the specific track.



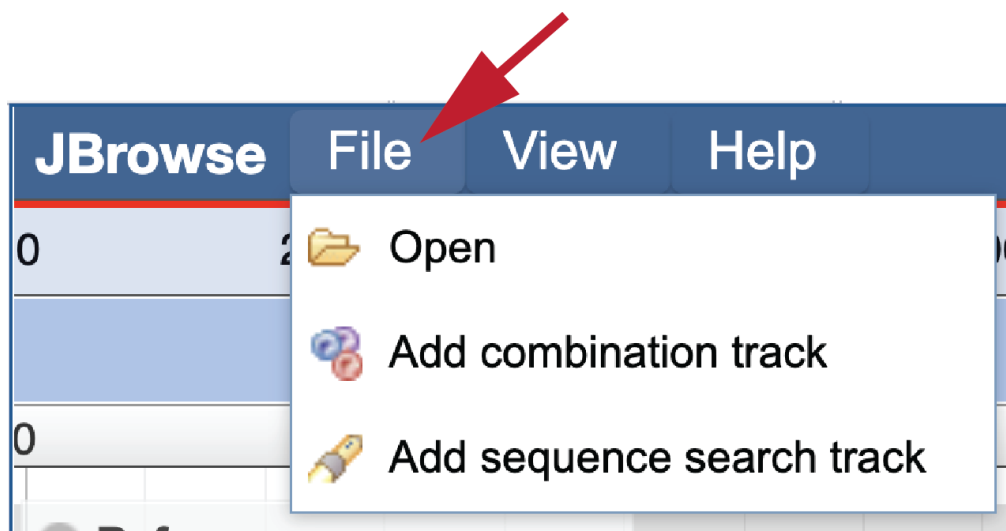
The track will be removed once this has happened.



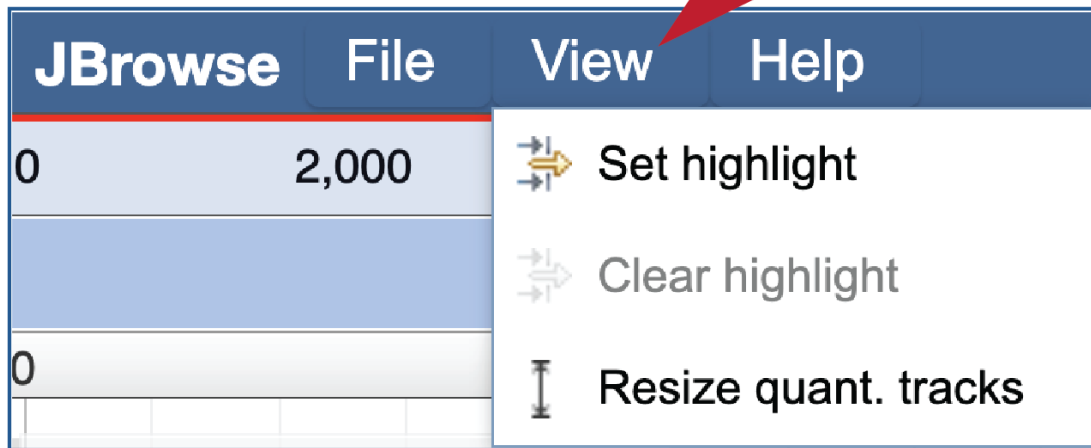
The removed track can be restored by clicking on the check box in front of it in the left side of the browser.



Researchers can also add their own data as an additional track, but it must match the reference genome (in this case the Wuhan-hu-1 genome). Clicking on the File tab and then selecting one of the options will enable this.

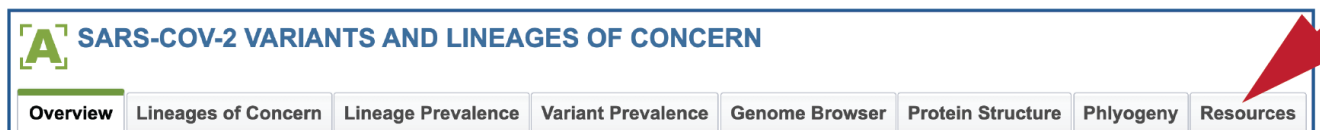


The overall view can also be adjusted by clicking on the View icon.



Additional Resources

Additional resources that provide information on the SARS-CoV-2 strains are provided under the Resources tab. This can be reached by clicking on that tab.



This will open the Resources page, where the name and brief description of the resource and the data types provided by the resource are included. The name of each resource is a hyperlink that will open a new tab to that page.

| SARS-CoV-2 Resources | | |
|---|--|---|
| Resource | Description | Data Types |
| BEI SARS-CoV-2 Resources | Information about SARS-CoV-2 strains and reagents. | biomaterial |
| Bloom Lab: ACE-2 Binding affinities | Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding (Reference) | Experimental data on point mutations affecting receptor binding |
| Broad Terra cloud commons for pathogen surveillance | The Broad Terra cloud workspace with COVID-19 genomics data and workflows for genome assembly, quality control, metagenomic classification, and aggregate statistics. | genomics |
| CDC SARS-CoV-2 Sequencing Resources | crowd-sourced collection of information, documentation, protocols and other resources for public health laboratories intending to sequence SARS-CoV-2 samples. | protocols |
| China National Center for Bioinformation's 2019 Novel Coronavirus Resource (2019nCoV) | A comprehensive resource on COVID-19, combining up-to-date information on all published sequences, mutation analyses, literatures and others. | genomics, variant tracking, literature |
| Coronavirus 3D | Web-based viewer for 3D visualization and analysis of the SARS-CoV-2 protein structures with respect to the CoV-2 mutational patterns | protein structures |
| CoV-GLUE | Amino acid variation database of amino acid replacements, insertions and deletions | variant tracking |
| CoVariants | SARS-CoV-2 variant tracking dashboard largely based on NEXTSTRAIN | Multiple: phylogenomics, variant tracking |
| COVID-19 CoV Genetics Browser | Track transmission, evolution, emergence, immune interactions, diagnostics, therapeutics & vaccines | variant tracking |
| COVID-19 Genome Sequence Dataset on Registry of Open Data on AWS | A centralized sequence repository for all strains of novel coronavirus (SARS-CoV-2) submitted to NCBI. Included are both the original sequences submitted by the principal investigator as well as SRA-processed sequences that require the SRA Toolkit for analysis. | genomics |
| CoViz | Near real-time visualization of nCoV-19 genomic variation | genomics |
| GISAID | International database of nCoV-19 genome sequences and related clinical and epidemiological data | genomics |
| GISAID variant tracking page | GISAID's variant tracking dashboard | Strains, map, graph, metadata, |
| Immune Epitope Database (IEDB) | Experimental data on SARS-CoV-2 antibodies and T cell epitopes | immunology |
| LANL COVID-19 Viral Genome Analysis | Analyses and tools for exploring accruing mutations in SARS-CoV-2, geographically and over time, with an emphasis on the Spike protein. | variant tracking |
| NCBI SARS-CoV-2 Resources | SARS-CoV-2 related data and resources at NCBI, such as nucleotide and protein sequences from GenBank and RefSeq, genomic and metagenomic read sets in SRA, BLAST, PubMed, Clinical Trials. | genomics, literature |
| NextCade | Tool to perform clade assignment, mutation calling, and sequence quality checks | phylogenomics |
| Nextstrain COVID-19 genetic epidemiology | Open-source SARS-CoV-2 genome data and analytic and visualization tools | phylogenomics, variant tracking |
| Pango Lineages | A dynamic nomenclature for SARS-CoV-2 lineages. Pangolin, a tool for global lineage assignment. | phylogenomics, variant tracking |
| Reactome | Open-source curated and peer-reviewed pathway database, including human coronavirus infection pathways. Tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education. | genomics, visualization |
| UCSC SARS-CoV-2 Genome Browser | The UCSC SARS-CoV-2 genome browser and COVID-19 lung gene expression datasets. | genome browser |
| UniProtKB SARS-CoV-2 | SARS-CoV-2 proteins and annotations | protein annotation |

References

- 1 Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J. & Holmes, I. H. JBrowse: a next-generation genome browser. *Genome research* **19**, 1630-1638 (2009).