



COVID-19

What is Genomic Surveillance?

Updated Jan. 24, 2022

Key Definitions

- **Mutation:** A mutation refers to a single change in a virus's genome (genetic code). Mutations happen frequently but only sometimes change the characteristics of the virus.
- **Lineage:** A lineage is a group of closely related viruses with a common ancestor. SARS-CoV-2 has many lineages; all cause COVID-19.
- **Variant:** A variant is a viral genome (genetic code) that may contain one or more mutations. In some cases, a group of variants with similar genetic changes, such as a lineage or group of lineages, may be designated by public health organizations as a [variant of concern \(VOC\)](#) or a [variant of interest \(VOI\)](#) due to shared attributes and characteristics that may require public health action.
- **Genomic Sequencing:** Scientists use a process called genomic sequencing to decipher the genetic material found in an organism or virus. Sequences from specimens can be compared to help scientists track the spread of a virus, how it is changing, and how those changes may affect public health.
- **Genomic Surveillance:** Viruses can be tracked using genomic sequence data collected by CDC and its partners. Effective surveillance does not require the sequencing of a specimen from every COVID-19 case. Instead, scientists rely on collecting enough sequence data from representative populations to detect new variants and monitor trends in circulating variants.

All viruses change (or mutate) as they replicate and spread in a population. Viruses that have RNA as genetic material, such as [SARS-CoV-2](#) (the virus that causes COVID-19) and influenza, mutate much faster than viruses with DNA. Every time SARS-CoV-2 replicates, there is an opportunity for the virus to change. Many mutations do not affect the virus's ability to spread or cause disease because they do not alter the major proteins involved in infection and transmission.


When one of these changes does affect the virus's ability to spread or cause disease, there may be a competitive advantage over the other lineages of SARS-CoV-2. Over time, certain lineages with these advantages become more prevalent and circulate in a population. When a lineage or group of lineages have characteristics that impact public health, CDC may classify them as a ["variant of interest"](#), or ["variant of concern"](#).

For more information about how viruses change over time, visit [Understanding Variants](#).

How do scientists track SARS-CoV-2 variants?

Scientists use genomic sequencing to identify which variant of SARS-CoV-2 is in a specimen. Scientists are consistently accumulating sequences and analyzing similarities and differences among these sequences in a process called genomic surveillance. Through genomic surveillance, scientists track the spread of variants, monitor changes to the genetic code of SARS-CoV-2 variants. Collectively, this information is used to better understand how variants might impact public health.

As CDC and public health partners sequence more SARS-CoV-2 genomes, we improve our understanding of which [variants are circulating in the United States](#), how quickly variants emerge, and which variants are most important to characterize and monitor.

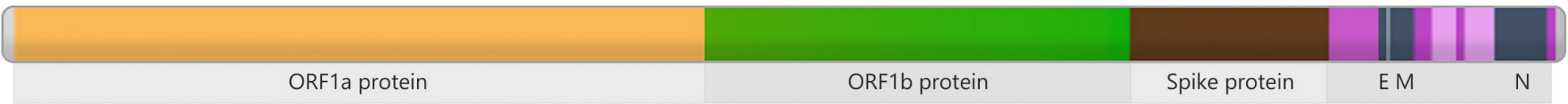


COVID-19 Data and Variant Proportions

Variants in the US

Global Variants

The SARS-CoV-2 Genome



Genes in the SARS-CoV-2 genome contain instructions to build parts of the virus. For example, the brown section in the picture has genetic instructions to build the spike protein, which attaches the virus to human cells during infection. Changes to this section could affect how easily the virus spreads and how effective vaccines are against it. Therefore scientists monitor this section closely to protect the public's health.

Where does the data for genomic surveillance come from?

Scientists use a process called [genomic sequencing](#) to identify [SARS-CoV-2 variants](#). Scientists sequence viral specimens collected from a portion of people who test positive for COVID-19. After a test is determined to be positive, additional time is required to prepare and sequence the viral specimen, and then analyze the sequence data. This can take anywhere from a few days to a few weeks.

Why is genomic sequencing important?

These sequences have valuable information that scientists use to:

1. Characterize the virus
2. Estimate a particular variant's prevalence in a population
3. Evaluate how effective [medical treatments](#), such as monoclonal antibodies, are against variants
4. Investigate the spread of a virus in outbreaks

How does genomic sequencing work?

Genomic sequencing goes beyond testing for SARS-CoV-2 and allows scientists to classify a virus as a particular variant and determine its lineage. Genomic surveillance has been a key component of public health efforts throughout the COVID-19 pandemic. Read more about the [genomic sequencing process](#).

STEP 1

Extraction

First, strands of DNA or RNA are extracted from the bacteria, virus, or other pathogens.



STEP 2

Library Prep

The DNA or RNA to be sequenced must be specially prepared before it can be put into the sequencing machine. The steps in the process may differ depending on the type of sample and specific equipment being used.



STEP 3

Sequencing

The library is loaded into a sequencer, which will identify the nucleotide bases in the DNA fragments.



Some of the steps
include: ✓

- Converting RNA or single stranded DNA into double-stranded DNA if needed
- Chopping strands into shorter pieces to get a desired length
- Modifying the ends of the fragments so they can be recognized by the sequencer. At this point the sample is called a “library” and is ready for sequencing.

STEP 4

Analysis

The sequencer produces data—millions of long strings of letters—which are then assembled together or aligned with a high-quality, reference sequence. Analytical programs compare the new sequence data to the reference sequence and identify variations in the sample that allow scientists to infer an ancestral relationship, much like a family tree.

Reference: CACGTGGAGTCGAGCGTTGTGAACCGTGTAAAGCGTGGTCCTG
CACGTTGAGTCCAGGGTTGT
GAGTCCAGGGTTGTGAAC
GGGTTGTGAACGGTGTAAA

Variants of the Virus

To find more information about mutations and variants CDC is monitoring in the US and globally, go to the [SARS-CoV-2 Variants](#) page.

Related Resources

[Emerging SARS-CoV-2 Variants](#)

[New Variants of the Virus that Causes COVID-19](#)

[CDC's Role in Tracking Variants](#)