

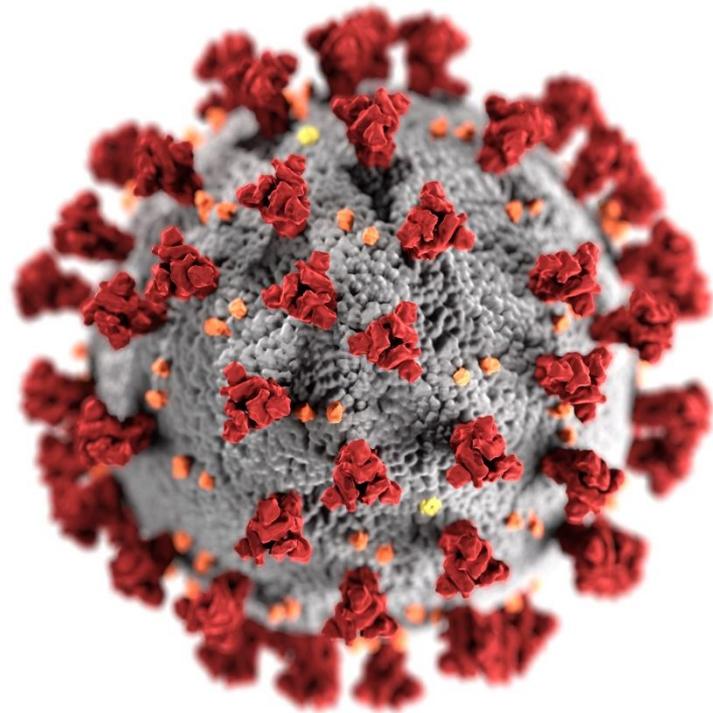
SARS-CoV-2の公開ゲノム

COVID-19ゲノム疫学ツールキット：バスケット3.5

Michael Weigand、PhD

バイオインフォマティクス

疾病管理予防センター



cdc.gov/coronavirus

ウィジェットマップ

パート1：はじめに

- 1.1ゲノム疫学とは何ですか？
- 1.2SARS-CoV-2ゲノム
- 1.3系統樹の解釈方
- 1.4SARS-CoV-2の○な変種

パート2：ケーススタディ

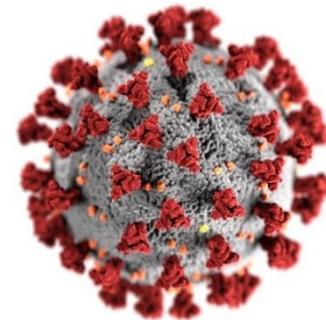
- 2.1アリゾナでのSARS-CoV-2シーケンシング
- 2.2ハウスクエアカの死
- 2.3コミュニティの
- 2.4スーパースプレッダー

パート3：実装

- 3.1Nextstrainの使用する
- 3.2MicrobeTraceの使用をする
- 3.3USHERビジネス
- 3.4Nextstrain3を歩く

3.3。 35.5。 NSパブリックゲノム

SARS-CoV-2を配列決定する理由



- 州/地方
 - ↳ サプリメント管理の
 - ↳ 疫学を理解する
- 国家
 - ↳ 田前
 - ↳ 診断、ワクチン、薬治療法の開発します
- **公開データベースデータベースデータベース**
 - 立物は、SARS-CoV-2ワークフローにあり度があります
 - 疫的、疫的メタデータデータ

SARS-CoV-2ゲノムデータ公開

1.1. GISAID (すべてのインフルエンザの共有共有イニシアチブ)

www.gisaid.org

- インフルエンザ (EpiFlu) ダウンロードSARS-CoV-2 (EpiCoV) からの共有のダウンロードな共有を刺激します

2.2. NCBI / NLM (国立バイオテクノロジーセンター)

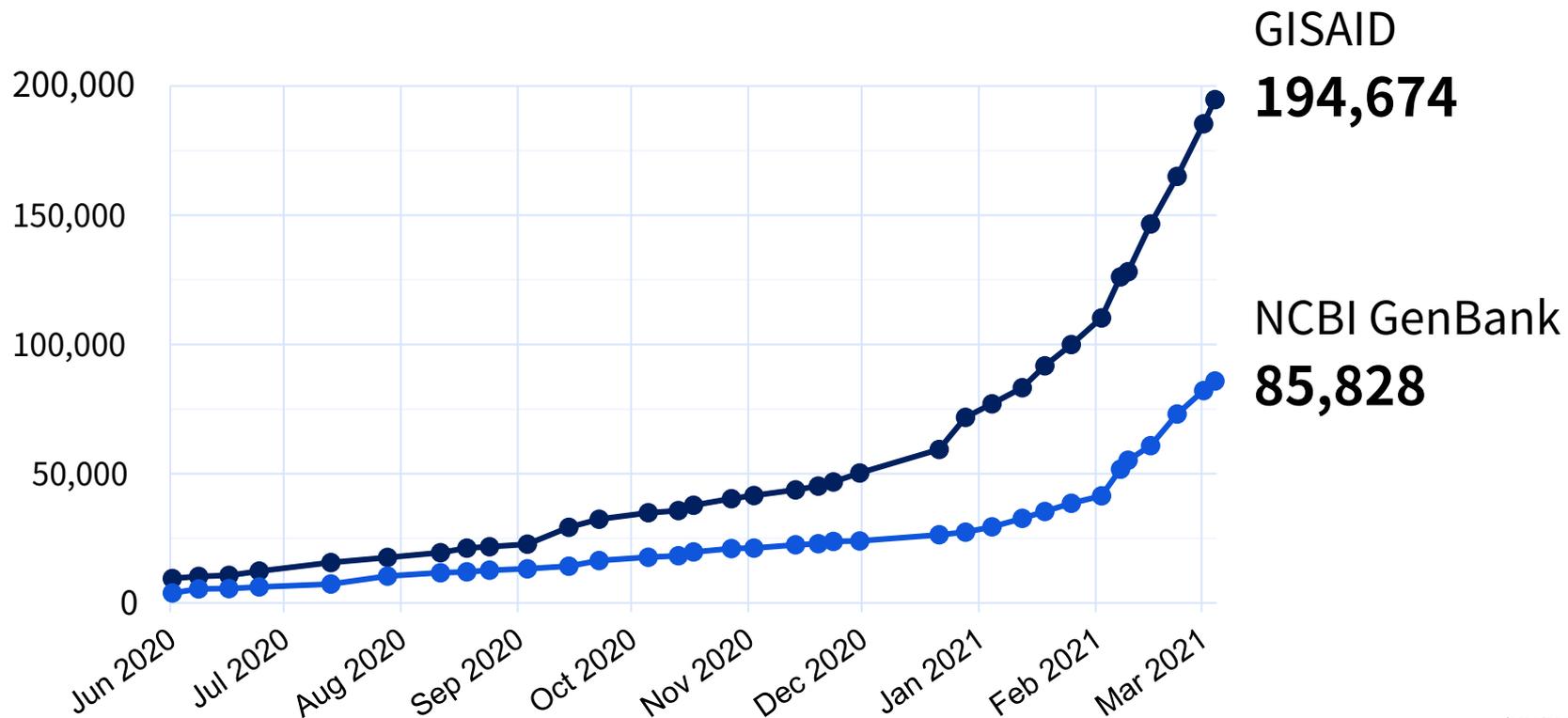
www.ncbi.nlm.nih.gov

- 国立衛生研究所 (NIH)
- GenBank®、Sequence Read Archive (SRA)、PubMed、BLAST®説明。
- SARS-CoV-2配列のパブリックアクセスとドリフトな共有を刺激します
- SARS-CoV-2データを参照 (PubMed / PMC) DoBLAST®ソフトウェアと統合します

- 組織な平基準、組織、主な設定
- 業者が提供します 読みますゲノムデータメタデータのコレクション。

SARS-CoV-2ガイド (GISAID + NCBI)

研究所



GISAID

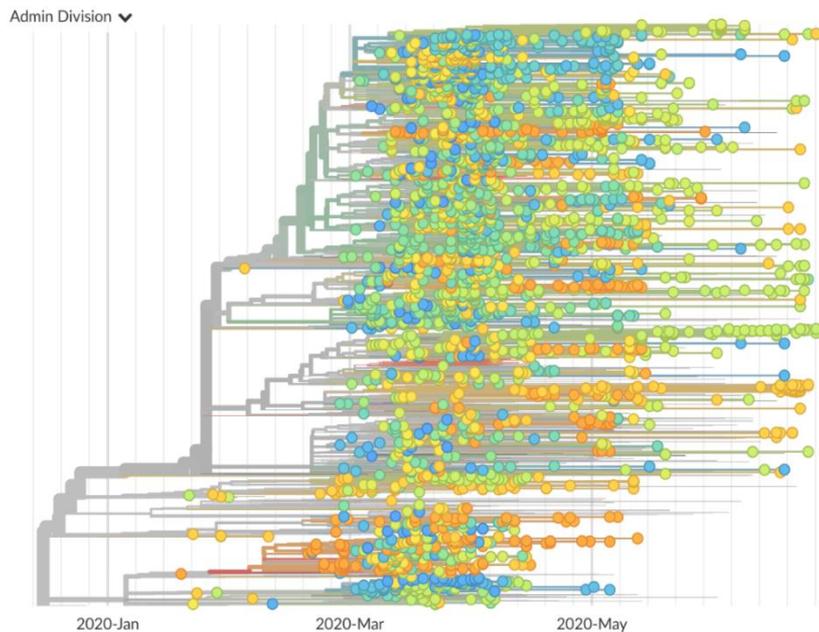
194,674

NCBI GenBank

85,828

鳥インフルエンザ情報共有の種SARS-CoV-2ツールができるになります

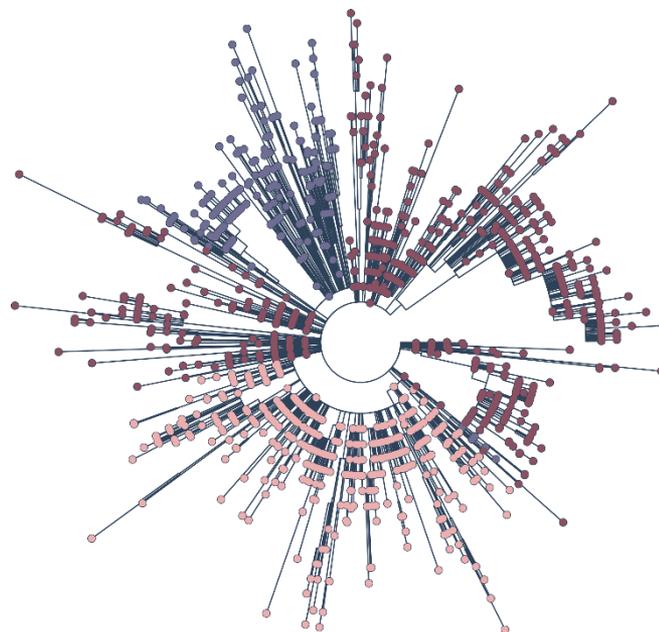
Nextstrain



nextstrain.org

パンゴの血縁

NS化発生 NSの NSamed NSローブ NS Outbreak



cov-lineages.org

処理検索

分析

- 化発生
- 分岐群のささ
- スパイク変異

GISAID

© 2008 - 2021 | Terms of Use | Privacy Notice | Contact

You are logged in as **Michael Weigand** - [logout](#)

Registered Users EpiFlu™ **EpiCoV™** My profile

EpiCoV™ Search Downloads Upload My Unreleased

Pan-ethnic coronavirus causing COVID-19

A previously unknown human coronavirus (hCoV-19) was first detected in late 2019 in patients in the City of Wuhan, who suffered from respiratory illnesses including atypical pneumonia, an illness that has become known as coronavirus disease (COVID-19). The coronavirus originated from an animal host and is closely related to the virus responsible for the Severe Acute Respiratory Syndrome coronavirus (SARS).

On 10. January 2020, the first virus genomes and associated data were publicly shared via GISAID. The World Health Organization announced on 11. March 2020 the first coronavirus pandemic. As the pandemic progresses, scientists from around the globe are tracking the virus and its genome sequences to ensure optimal virus diagnostic tests, to track and trace the ongoing outbreak and to identify potential intervention options. Several analyses to assist with these efforts are offered here, including sequence alignments, diagnostic primer and probe coordinates, 3D protein models, drug targets, phylogenetic trees and many more.

[Search](#)

Analysis Update (2021-02-23)

| | | | | | |
|---|---|--|---|---|---|
| <p>Full genome tree derived from all outbreak sequences</p> | <p>Timecourse of clade distribution in collected sequences</p> | <p>Regional clade distribution of new sequences</p> | <p>Common primer check for high quality genomes</p> | <p>Receptor binding surveillance for complete genomes</p> | <p>Full genome tree of hCoV-19-related precursors</p> |
| <p>Clade evolution in the first year</p> | <p>Temporal and regional distribution of clades in the first year</p> | <p>Spike comparison between pangolin, bat, human</p> | <p>Spike comparison to SARS and bat precursor</p> | <p>Highly conserved drug targets between hCoV-19 and SARS</p> | <p>analysis update.pdf</p> |
| <p>Official GISAID reference sequence</p> | <p>Audacity</p> | <p>Spike glycoprotein mutation surveillance</p> | <p>BLAST</p> | | |

Important note: In the [GISAID EpiFlu™ Database Access Agreement](#), you have accepted certain terms and conditions for viewing and using data regarding influenza viruses. To the extent the Database contains data relating to non-influenza viruses, the viewing and use of these data is subject to the same terms and conditions, and by viewing or using such data you agree to be bound by the terms of the [GISAID EpiFlu™ Database Access Agreement](#) in respect of such data in the same manner as if they were data relating to influenza viruses.

位置

(北米/ EGO /ジョージア)

目日

(2021-01-01から2021-01-31)

代

(Spike_E484K)

追い

(GH / 501Y.V2 / B.1.351)

CoVsurver**ダウンロード**

Search

Accession ID Virus name complete high coverage

Location Host low coverage excl w/Patient status

Collection to Submission to collection date compl

Mode Lineage Substitutions Variants

| <input type="checkbox"/> | Virus name | Passage | Accession ID | Collection date | Submission | <input type="button" value="i"/> | Length | Host | Location | Originating |
|--------------------------|-----------------------------------|----------|-----------------|-----------------|------------|----------------------------------|--------|-------|-------------------|-------------|
| <input type="checkbox"/> | hCoV-19/USA/FL-CDC-LC0021227/2021 | Original | EPI_ISL_1298429 | 2021-03-02 | 2021-03-02 | <input type="button" value="i"/> | 29,695 | Human | North America / U | Laborat |
| <input type="checkbox"/> | hCoV-19/USA/AL-CDC-LC0021169/2021 | Original | EPI_ISL_1298428 | 2021-03-01 | 2021-03-01 | <input type="button" value="i"/> | 29,694 | Human | North America / U | Laborat |
| <input type="checkbox"/> | hCoV-19/USA/FL-CDC-LC0021166/2021 | Original | EPI_ISL_1298427 | 2021-03-01 | 2021-03-01 | <input type="button" value="i"/> | 29,694 | Human | North America / U | Laborat |
| <input type="checkbox"/> | hCoV-19/USA/FL-CDC-LC0021161/2021 | Original | EPI_ISL_1298426 | 2021-02-27 | 2021-03-19 | <input type="button" value="i"/> | 29,694 | Human | North America / U | Laborat |
| <input type="checkbox"/> | hCoV-19/USA/TN-CDC-LC0021173/2021 | Original | EPI_ISL_1298425 | 2021-03-01 | 2021-03-19 | <input type="button" value="i"/> | 29,694 | Human | North America / U | Laborat |
| <input type="checkbox"/> | hCoV-19/USA/GA-CDC-LC0021169/2021 | Original | EPI_ISL_1298424 | 2021-03-01 | 2021-03-19 | <input type="button" value="i"/> | 29,694 | Human | North America / U | Laborat |
| <input type="checkbox"/> | hCoV-19/USA/TN-CDC-LC0021166/2021 | Original | EPI_ISL_1298423 | 2021-03-01 | 2021-03-19 | <input type="button" value="i"/> | 29,694 | Human | North America / U | Laborat |
| <input type="checkbox"/> | hCoV-19/USA/FL-CDC-LC0021161/2021 | Original | EPI_ISL_1298422 | 2021-02-27 | 2021-03-19 | <input type="button" value="i"/> | 29,694 | Human | North America / U | Laborat |
| <input type="checkbox"/> | hCoV-19/USA/GA-CDC-LC0021144/2021 | Original | EPI_ISL_1298421 | 2021-02-28 | 2021-03-19 | <input type="button" value="i"/> | 29,694 | Human | North America / U | Laborat |
| <input type="checkbox"/> | hCoV-19/USA/GA-CDC-LC0021138/2021 | Original | EPI_ISL_1298420 | 2021-02-27 | 2021-03-19 | <input type="button" value="i"/> | 29,694 | Human | North America / U | Laborat |
| <input type="checkbox"/> | hCoV-19/USA/TN-CDC-LC0021137/2021 | Original | EPI_ISL_1298419 | 2021-02-27 | 2021-03-19 | <input type="button" value="i"/> | 29,694 | Human | North America / U | Laborat |
| <input type="checkbox"/> | hCoV-19/USA/GA-CDC-LC0021113/2021 | Original | EPI_ISL_1298418 | 2021-02-27 | 2021-03-19 | <input type="button" value="i"/> | 29,694 | Human | North America / U | Laborat |

Total: 45,927 viruses

<< < 1 2 3 4 5 > >>

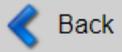
Select

GISAID : ダウンロード

Download

Format

- Sequences (FASTA)
- Patient status metadata
- Sequencing technology metadata
- Acknowledgement (Supplemental table)

 Back  Download

FASTA

```
> EX1  
AAAUGUUAAUUCAUGCU  
> EX2  
AAAUAUUACUCAUGCU  
> EX3  
AAAUAUUACUCAUGCC
```

Metadata.tsv

| アクセス | 日にち | 位置 |
|------|-----------|------------|
| Seq1 | 2021年2月5日 | アメリカ/ジョージア |
| Seq2 | 2021年2月5日 | アメリカ/ジョージア |
| Seq3 | 2021年2月5日 | アメリカ/ジョージア |

鳥インフルエンザ情報共有

www.gisaid.org/registration/terms-of-use/

鳥インフルエンザ情報共有の詩：レコードを検索する

Registered Users EpiFlu™ EpiCoV™ My profile

EpiCoV™ Search Downloads Upload

Search

Accession ID Virus name

Location Host

Collection to Submission to

Clade Lineage Substitutions

| <input type="checkbox"/> | Virus name | Passage de | Accession ID | Collection |
|--------------------------|-----------------------------------|------------|-----------------|------------|
| <input type="checkbox"/> | hCoV-19/USA/GA-CDC-LC0021169/2021 | Original | EPI_ISL_1298424 | 2021-03- |

アクセッションID

Virus detail

Virus name:
Accession ID:
Type:
Clade
Pango Lineage
AA Substitutions

Variant

Passage details/history:

Sample information

Collection date:
Location:
Host:
Additional location information:
Gender:
Patient age:
Patient status:
Specimen source:
Additional host information:
Outbreak:
Last vaccinated:
Treatment:
Sequencing technology:
Assembly method:
Coverage:
Comment:

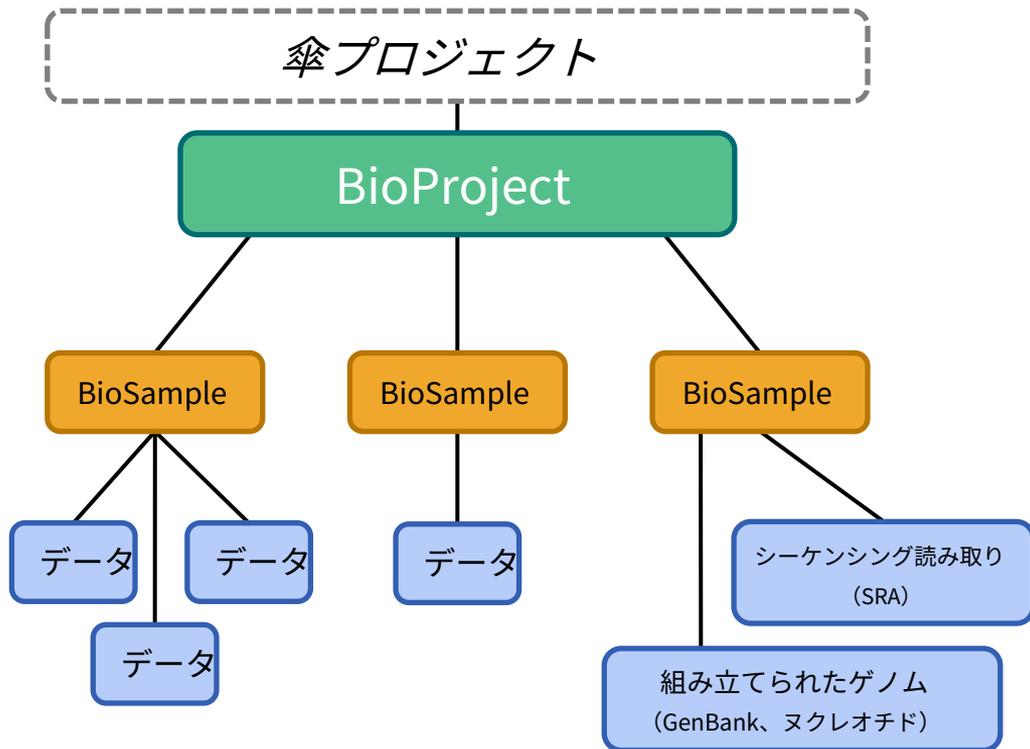
Institute information

Originating lab:
Address:
Sample ID given by the
originating laboratory:
Submitting lab:

鳥インフルエンザ情報共有

www.gisaid.org/registration/terms-of-use/

NCBIデータベース組織



- **BioProject** – 単一のイニシアチブのための生物学的データのコレクション
 - 傘プロジェクト – BioProjectsのコレクション
- **BioSample** – 物理的な標本に関する情報（メタデータ）
- **データ** – ゲノムデータセット; 例えば
 - 生のシーケンス読み取り
 - 組み立てられたゲノム



NCBI SARS-CoV-2 Resources

Quick Navigation Guide

[Sequence Submission](#)

[Literature](#)

[Sequence-Related Resources](#)

[Clinical Resources](#)

[Other Websites](#)

SARS-CoV-2 Data

254,626

[SRA runs](#)

117,986

[Nucleotide records](#)

5,113

[ClinicalTrials.gov](#)

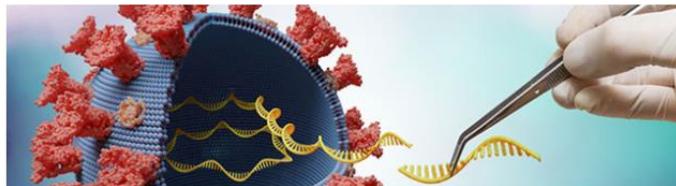
115,033

[PubMed](#)

125,241

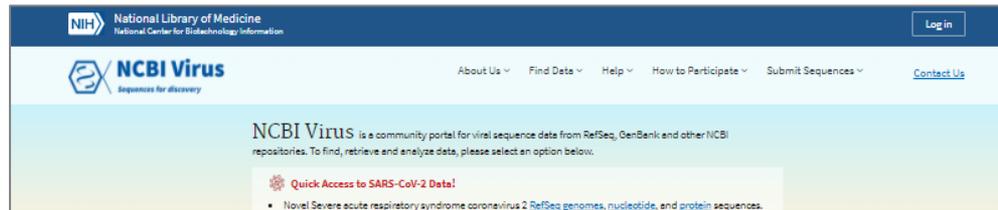
[PMC](#)

Submit SARS-CoV-2 Sequences



Add assembled & raw read data to the growing public archive

[Submit Now](#)



SARS-CoV-2 Data Dashboard

Filters (0)

View results, Analyze, or Download

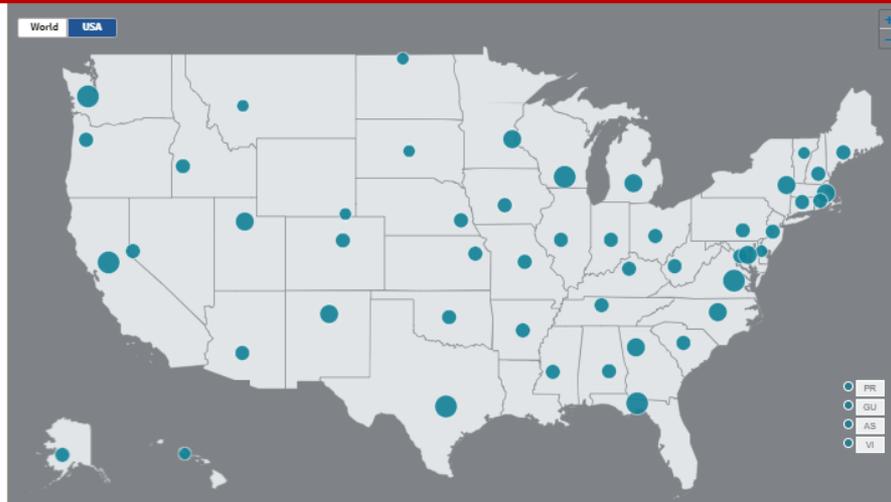
1
RefSeq Genomes

1,317,831
All Proteins

117,986
All Nucleotides

38
RefSeq Proteins

66,259
Complete Nucleotides



SARS-CoV-2 Data Dashboard

Filters (2)

View results, Analyze, or Download

選択表を見る

0
RefSeq Genomes

4,690
All Proteins

403
All Nucleotides

0
RefSeq Proteins

69
Complete Nucleotides



収集日

Collection Date Weekly

Reset

1/20/2021 - 1/26/2021
2/24/2021 - 3/2/2021

Refine Results Reset

Virus +

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049 x

Accession +

Sequence Length +

Sequence Type +

RefSeq Genome Completeness +

Nucleotide Completeness +

Proteins +

Provirus +

Geographic Region +

USA: GA x

Host +

Isolation Source +

Collection Date +

From Jan 20, 2021 To Mar 2, 2021 x

位置

収集日

Selected Results: 0

PubMed

Download

Align

Build Phylogenetic Tree

Expand Table

| | Nucleotide (403) | Protein (4,690) | RefSeq Genome (0) | | | | | | | |
|--------------------------|--------------------------|-------------------|-------------------|-------------------------------|---------------|--------|--------------|----|--|--|
| | Accession | Submitters | Release Date | Species | Molecule type | Length | Geolocation | US | | |
| <input type="checkbox"/> | MW642833 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642834 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642850 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642851 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642852 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642853 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642854 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642855 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642856 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29696 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642925 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642934 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642935 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642936 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642937 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642938 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |
| <input type="checkbox"/> | MW642948 | Cook,P.W., et al. | 2021-02-21 | Severe acute respiratory s... | ssRNA(+) | 29713 | USA: Georgia | | | |

Get Columns

ダウンロード

分析する

NCBIダウンロード

Download Results ×

Step 1 of 3: Select Data Type

| | | |
|---|----------------------------------|----------------------------------|
| Sequence data (FASTA Format) | Accession List | Current table view result |
| <input checked="" type="radio"/> Nucleotide | <input type="radio"/> Nucleotide | <input type="radio"/> CSV format |
| <input type="radio"/> Coding Region | <input type="radio"/> Protein | <input type="radio"/> XML format |
| <input type="radio"/> Protein | <input type="radio"/> Assembly | |

Next

FASTA

```
>シーケンス1  
AAAUGUUAUUCAUGCU  
>シーケンス2  
AAAUAAUACUCAUGCU  
>シーケンス3  
AAAUAAUACUCAUGCC
```

List.txt

```
シーケンス1  
シーケンス2  
シーケンス3
```

Metadata.csv

| 受入日 | | 位置 |
|------|-----------|----------|
| Seq1 | 2021年2月5日 | USA / GA |
| Seq2 | 2021年2月5日 | USA / GA |
| Seq3 | 2021年2月5日 | USA / GA |

GISAID



NCBI



ローカル分析を補足する

ローカル
データ

FASTA

```
>シーケンス1  
AAAUGUUAUUCAUGCU  
>シーケンス2  
AAAUAUUACUCAUGCU  
>シーケンス3  
AAAUAUUACUCAUGCC  
>シーケンス4  
AAACGUUACUAAUGCU
```

Nextstrain

(モジュール3.1)



MicrobeTrace

(モジュール3.2)



アッシャー

(モジュール3.3)



概要

- 公開リポジトリは、組織化されたオープンデータ共有を促進します
- SARS-CoV-2ゲノム配列データの有用性を最大化する
 - パブリックNextstrainビルドなどのサポートリソース
 - ローカル分析または調査を補足するためのクエリ、ダウンロードシーケンス

GISAID



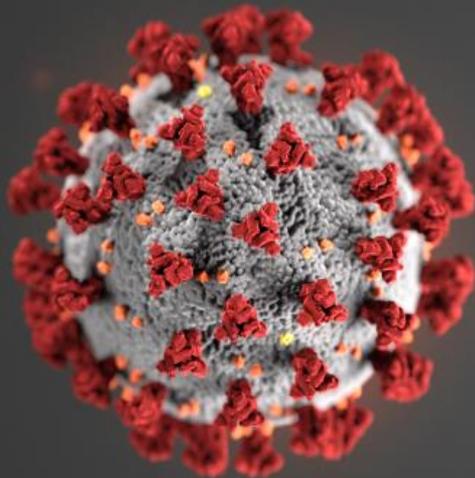
National Library of Medicine
National Center for Biotechnology Information

- あらかじめ支払う
 - にデータを送信する どちらも リポジトリは他の人を助けることができます！
 - 臨床的、疫学的メタデータ（PHA4GE）でより良い

もっと詳しく知る

- その他のモジュール
 - Nextstrain入門-モジュール3.1
 - MicrobeTrace入門-モジュール3.2
 - USHERを使用したリアルタイムの系統発生-モジュール3.3
- COVID-19ゲノム疫学ツールキット
 - さらに読むを見つける
 - フィードバック調査を完了する
 - 新しいモジュールがリリースされたら、それらの更新を受け取るためにサブスクライブします
 - go.usa.gov/xAbMw





詳細については、CDC 1-800-CDC-INFO
(232-4636) にお問い合わせください。
TTY : 1-888-232-6348 www.cdc.gov

このレポートの調査結果と結論は著者のものであり、必ずしも米国疾病予防管理センターの公式の立場を表すものではありません。

