

# Data in the COVID-19 Pandemic

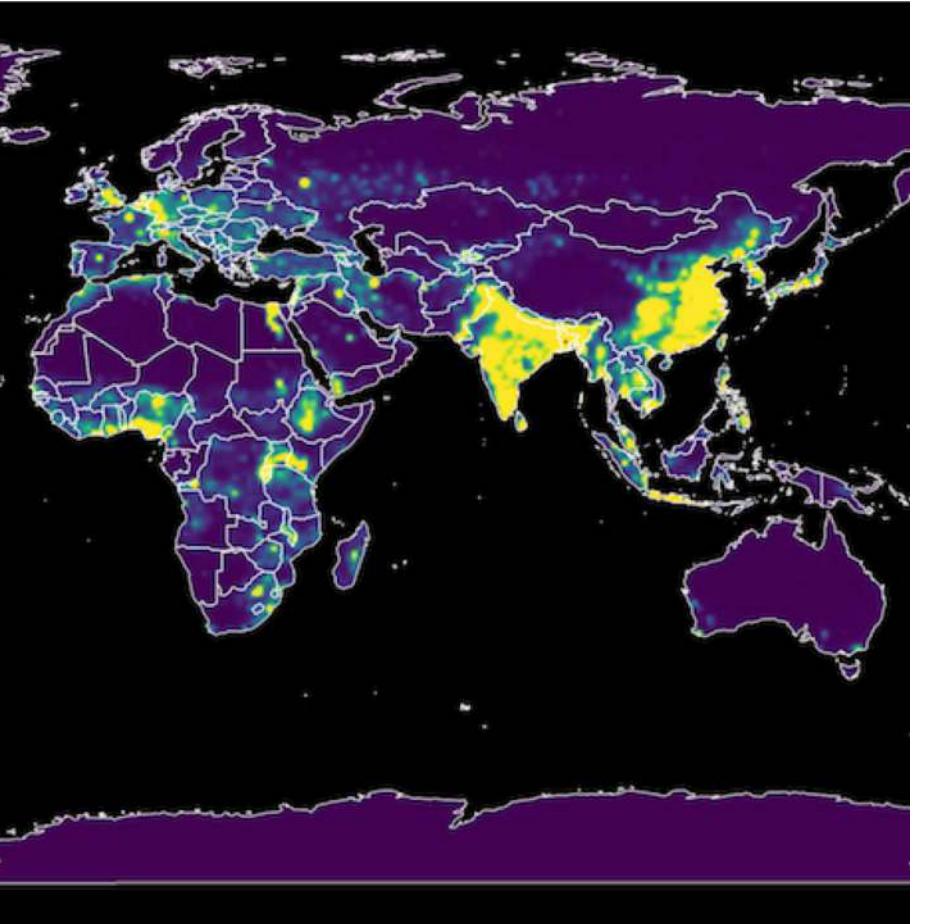
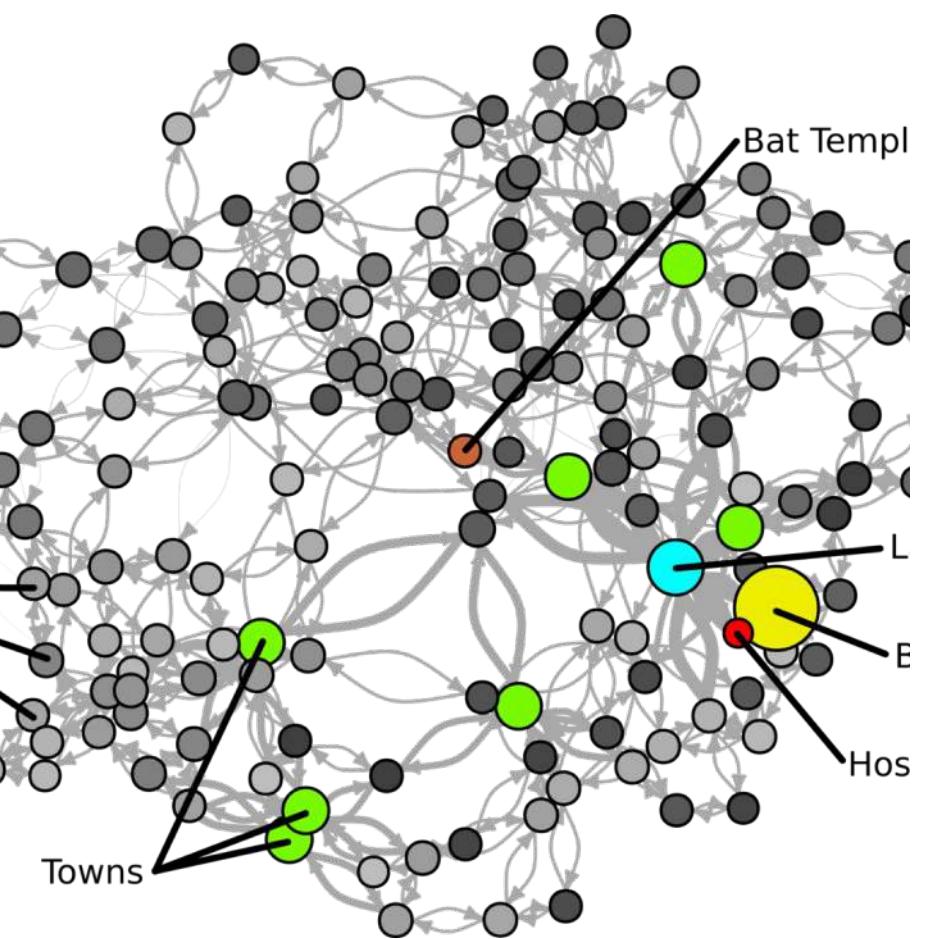
Noam Ross  
Stanford COVID-19 Data Forum  
2020-05-14

noamross @   

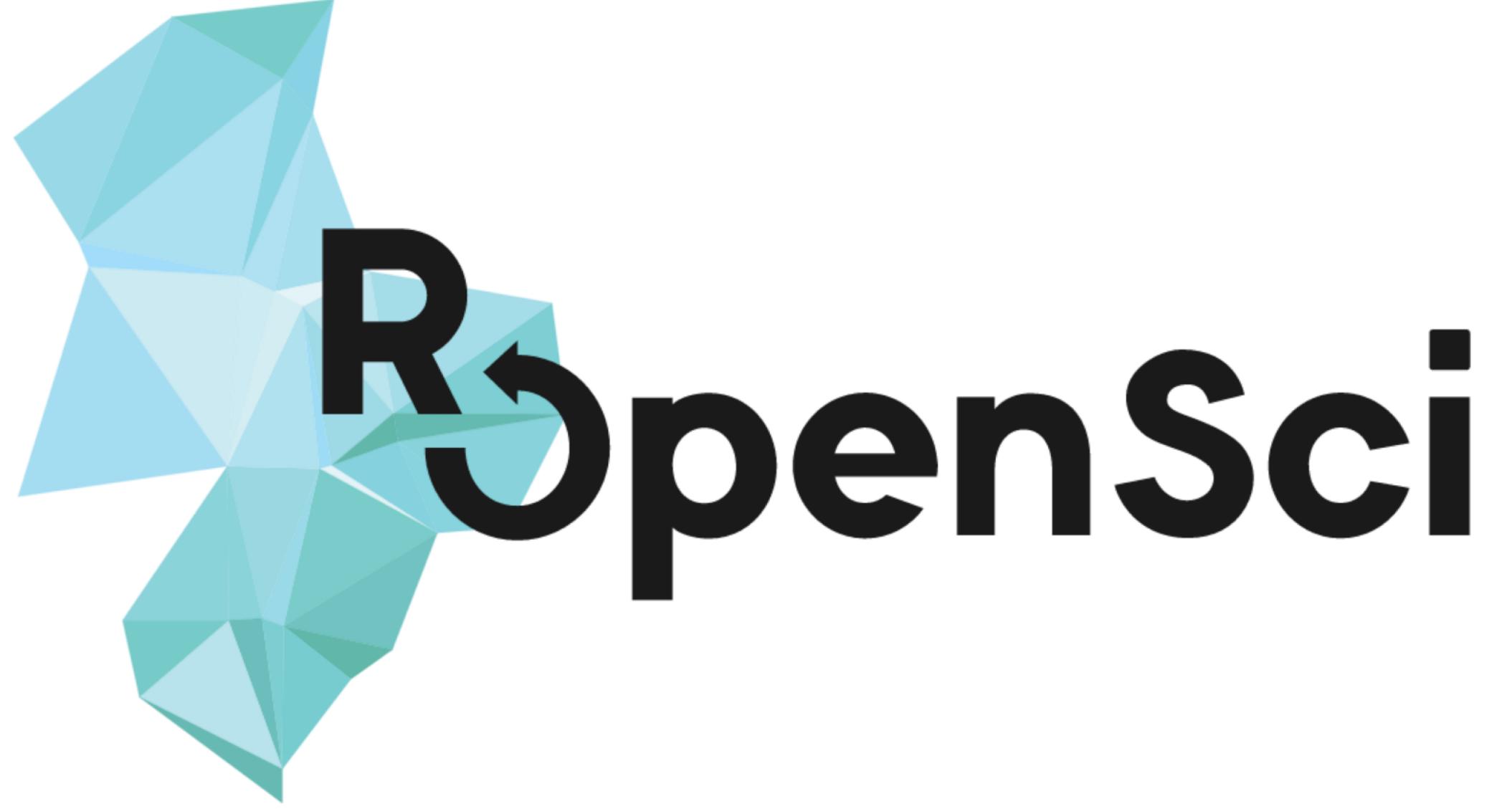


EcoHealth Alliance





@EcoHealthNYC [ecohealthalliance.org](http://ecohealthalliance.org)



Building technical and community  
infrastructure for R to support  
open, reproducible science

@rOpenSci  
[ropensci.org](http://ropensci.org)

# A pneumonia outbreak associated with a new coronavirus of probable bat origin

Peng Zhou, [Xing-Lou Yang](#), [Xian-Guang Wang](#), [Ben Hu](#), [Lei Zhang](#), [Wei Zhang](#), [Hao-Rui Si](#), [Yan Zhu](#), [Bei Li](#), [Chao-Lin Huang](#), [Hui-Dong Chen](#), [Jing Chen](#), [Yun Luo](#), [Hua Guo](#), [Ren-Di Jiang](#), [Mei-Qin Liu](#), [Ying Chen](#), [Xu-Rui Shen](#), [Xi Wang](#), [Xiao-Shuang Zheng](#), [Kai Zhao](#), [Quan-Jiao Chen](#), [Fei Deng](#), [Lin-Lin Liu](#), [Bing Yan](#), [Fa-Xian Zhan](#), [Yan-Yi Wang](#), [Geng-Fu Xiao](#) & [Zheng-Li Shi](#)✉ - Show fewer authors

*Nature* **579**, 270–273(2020) | [Cite this article](#)

**772k** Accesses | **265** Citations | **4061** Altmetric | [Metrics](#)

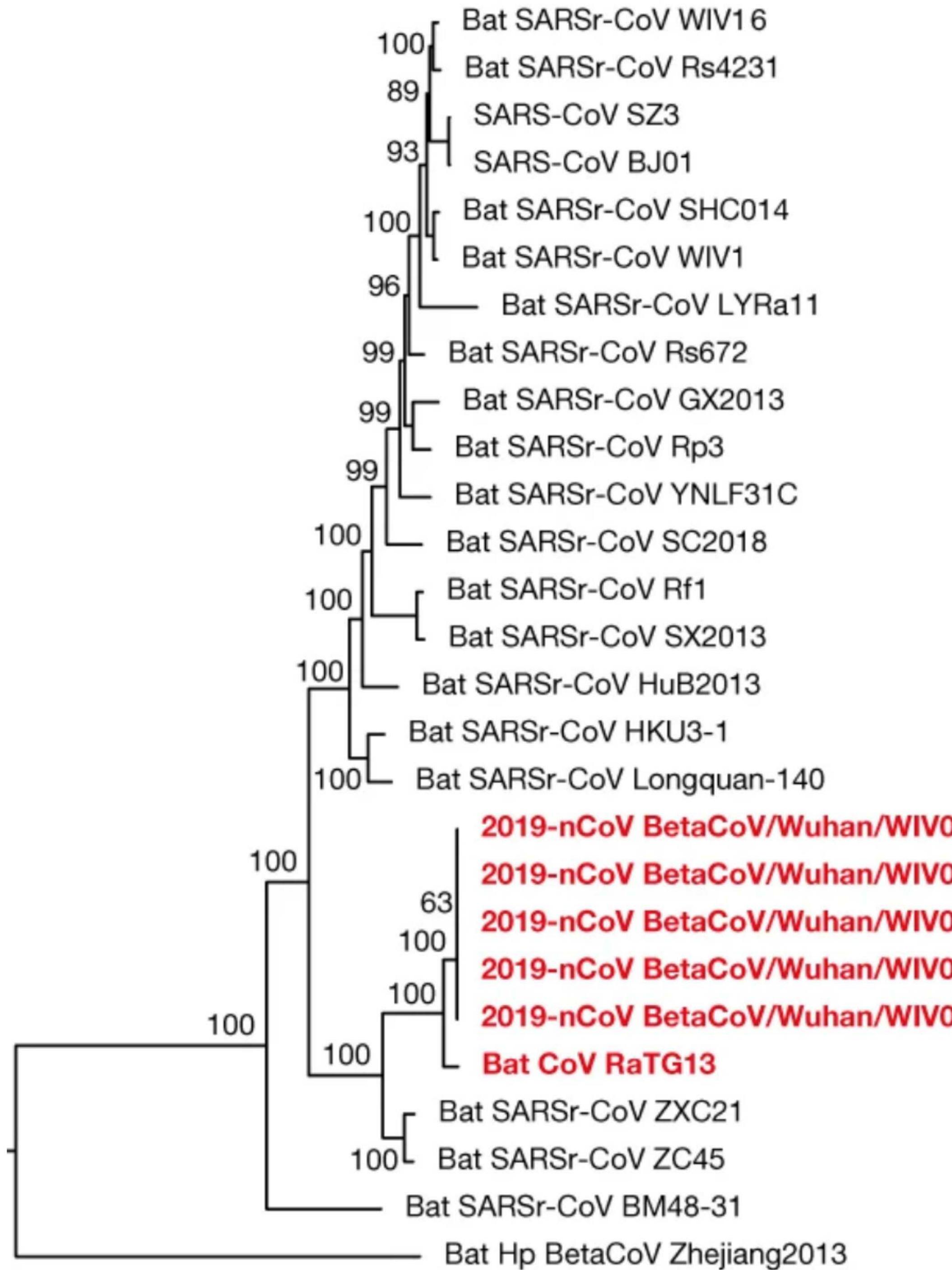
Correspondence | Published: 17 March 2020

# The proximal origin of SARS-CoV-2

Kristian G. Andersen✉, Andrew Rambaut, W. Ian Lipkin, Edward C. Holmes & Robert F. Garry

*Nature Medicine* **26**, 450–452(2020) | [Cite this article](#)

**4.61m** Accesses | **54** Citations | **34235** Altmetric | [Metrics](#)





RESEARCH ARTICLE

## Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus

Ben Hu<sup>1\*</sup>, Lei-Ping Zeng<sup>1\*</sup>, Xing-Lou Yang<sup>1\*</sup>, Xing-Yi Ge<sup>1</sup>, Wei Zhang<sup>1</sup>, Bei Li<sup>1</sup>, Jia-Zheng Xie<sup>1</sup>, Xu-Rui Shen<sup>1</sup>, Yun-Zhi Zhang<sup>2,3</sup>, Ning Wang<sup>1</sup>, Dong-Sheng Luo<sup>1</sup>, Xiao-Shuang Zheng<sup>1</sup>, Mei-Niang Wang<sup>1</sup>, Peter Daszak<sup>4</sup>, Lin-Fa Wang<sup>5</sup>, Jie Cui<sup>1\*</sup>, Zheng-Li Shi<sup>1\*</sup>

# Genetic Data Repositories



GenBank

# A pneumonia outbreak associated with a new coronavirus of probable bat origin

Peng Zhou, [Xing-Lou Yang](#), [Xian-Guang Wang](#), [Ben Hu](#), [Lei Zhang](#), [Wei Zhang](#), [Hao-Rui Si](#), [Yan Zhu](#), [Bei Li](#), [Chao-Lin Huang](#), [Hui-Dong Chen](#), [Jing Chen](#), [Yun Luo](#), [Hua Guo](#), [Ren-Di Jiang](#), [Mei-Qin Liu](#), [Ying Chen](#), [Xu-Rui Shen](#), [Xi Wang](#), [Xiao-Shuang Zheng](#), [Kai Zhao](#), [Quan-Jiao Chen](#), [Fei Deng](#), [Lin-Lin Liu](#), [Bing Yan](#), [Fa-Xian Zhan](#), [Yan-Yi Wang](#), [Geng-Fu Xiao](#) & [Zheng-Li Shi](#)✉ - Show fewer authors

*Nature* **579**, 270–273(2020) | [Cite this article](#)

**772k** Accesses | **265** Citations | **4061** Altmetric | [Metrics](#)

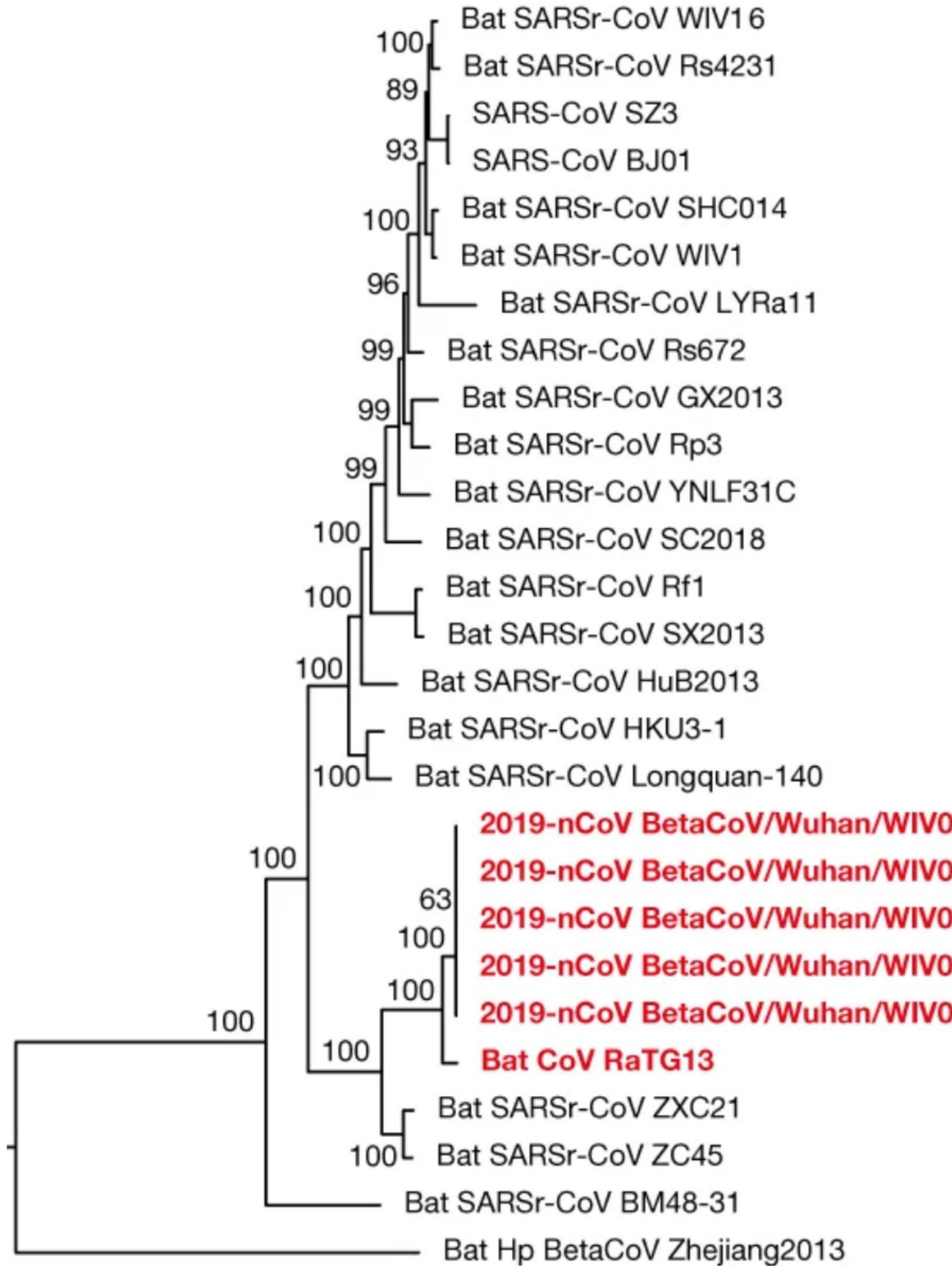
Correspondence | Published: 17 March 2020

# The proximal origin of SARS-CoV-2

Kristian G. Andersen✉, Andrew Rambaut, W. Ian Lipkin, Edward C. Holmes & Robert F. Garry

*Nature Medicine* **26**, 450–452(2020) | [Cite this article](#)

**4.61m** Accesses | **54** Citations | **34235** Altmetric | [Metrics](#)



# Genomic epidemiology of novel coronavirus - Global subsampling

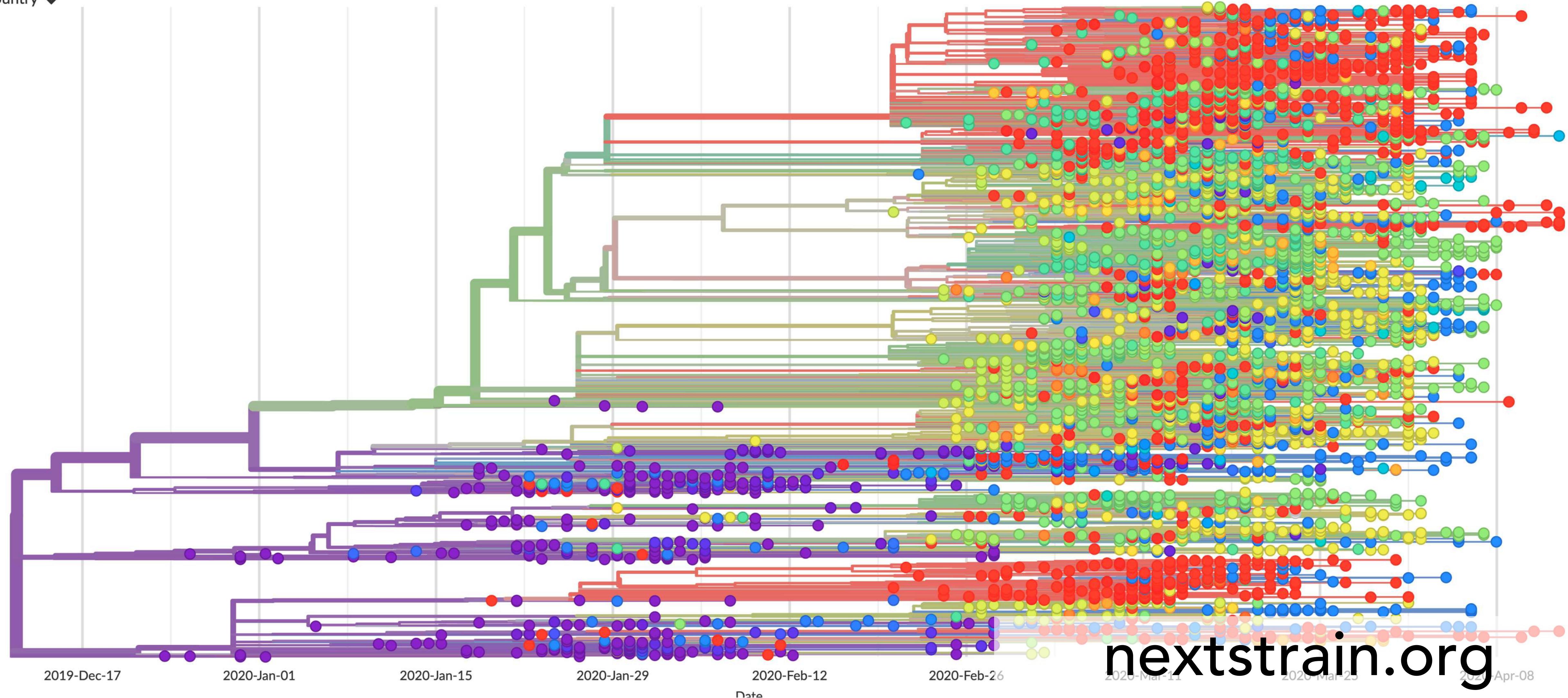
Maintained by the Nextstrain team. Enabled by data from [GISAID](#)

Showing 4537 of 4537 genomes sampled between Dec 2019 and Apr 2020.

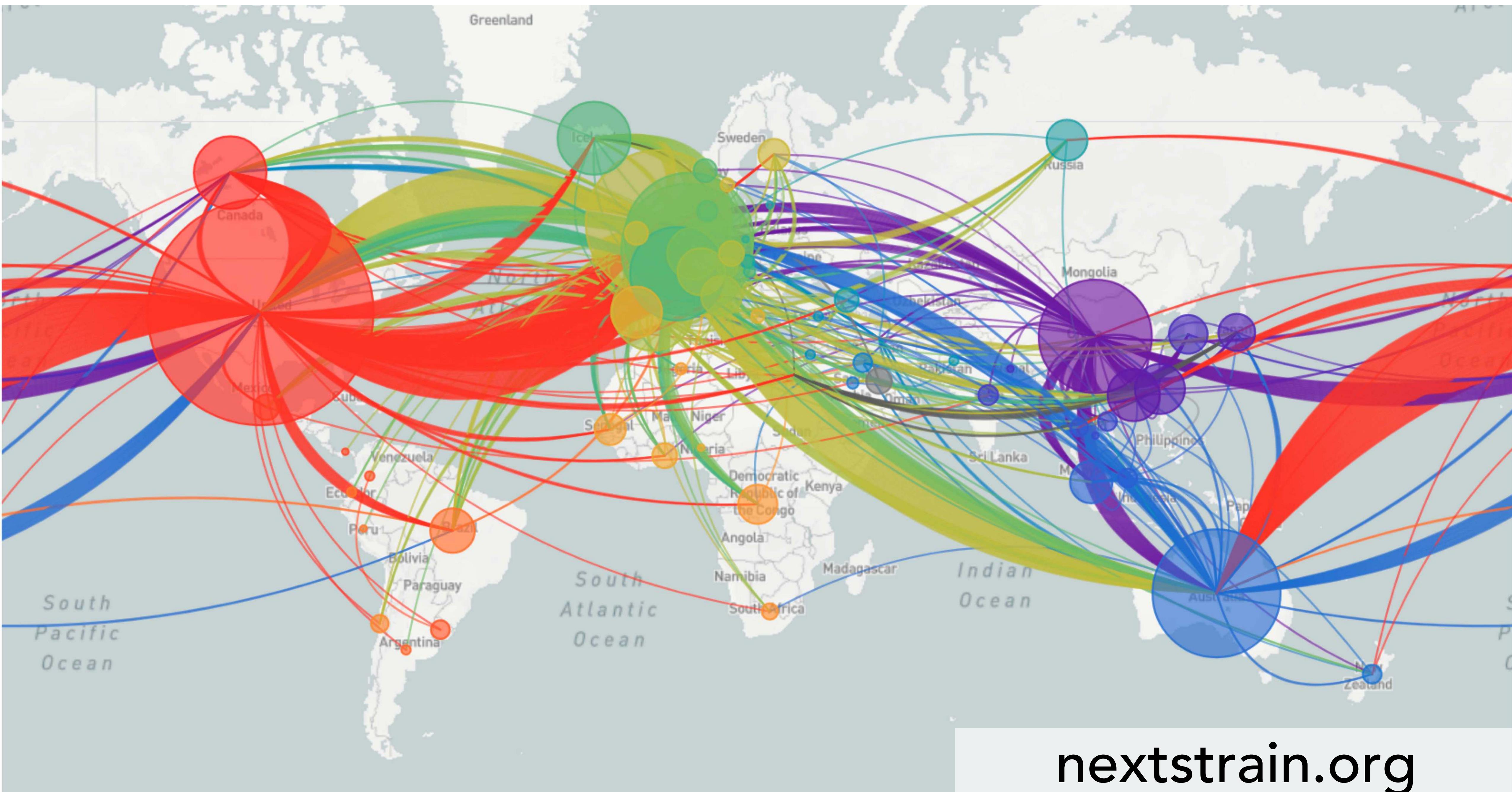
## Phylogeny

Country ▾

RESET LAYOUT



[nextstrain.org](http://nextstrain.org)



# Biodiversity Data

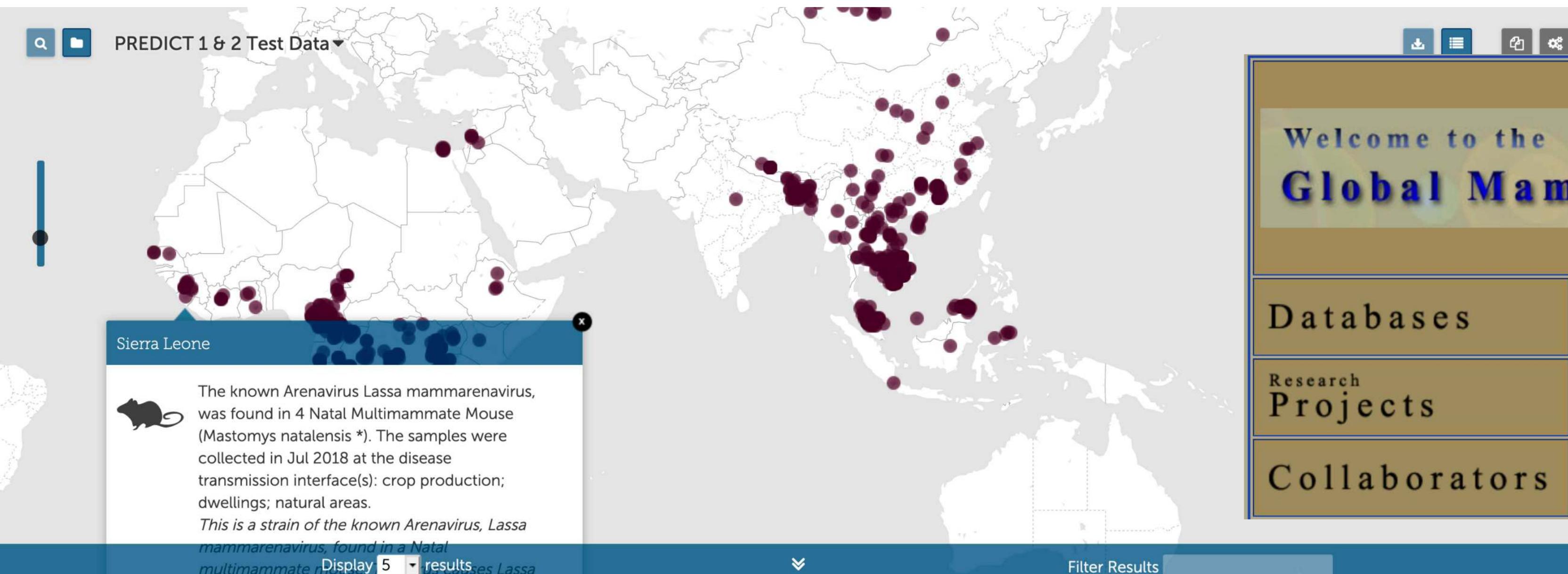


Ver*N*e



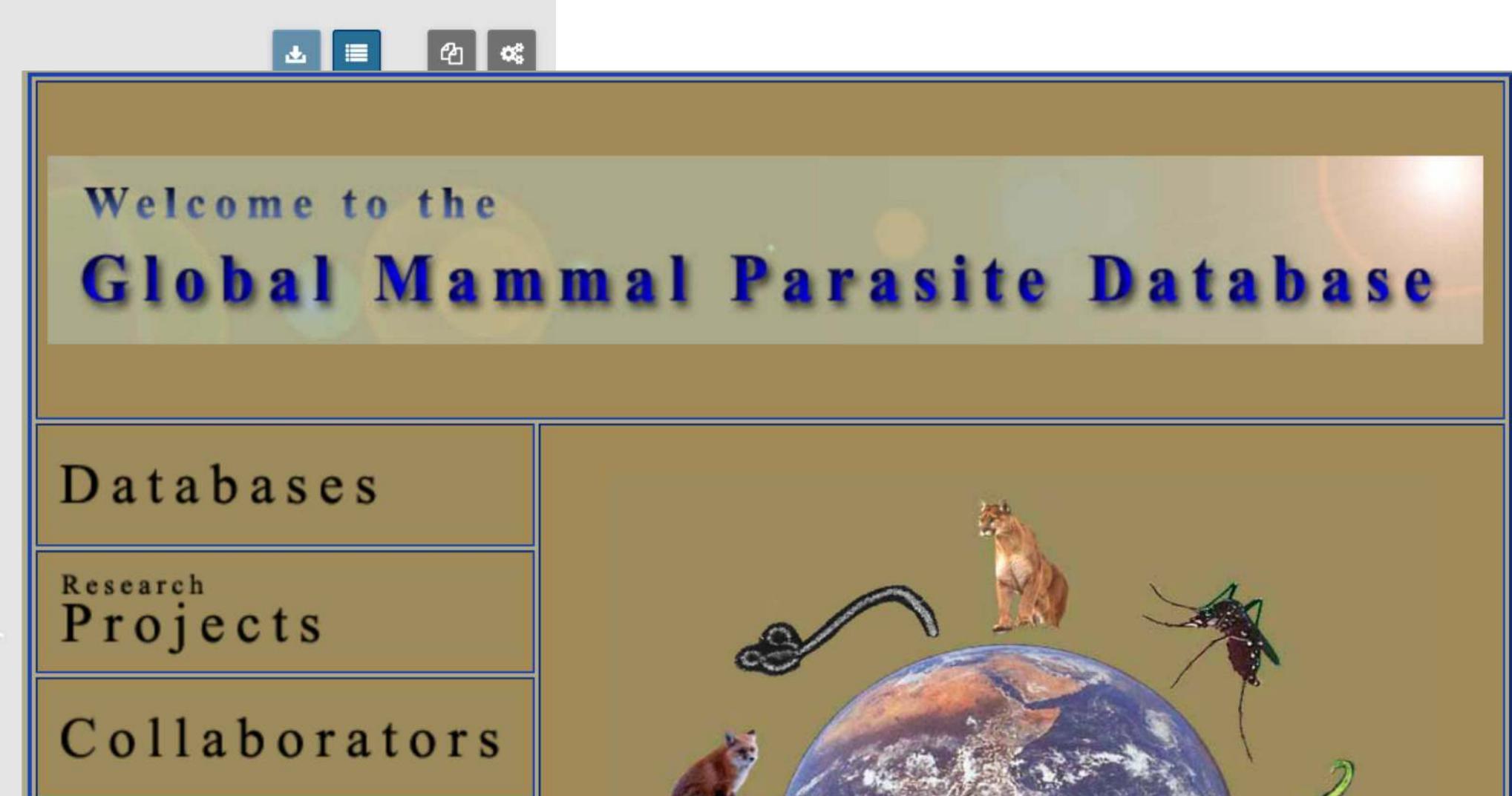
Knowledge Network for  
Biocomplexity (KNB)

# Host-Pathogen Interaction Data



Sample Date	Lat, Long	Country	Disease Transmission Interface	Species Scientific Name	Common Name	Virus	Interpretation	Genbank Entry
Jul 2018	9.91, -11.45	Sierra Leone	dwellings	Chaerephon pumilus *	Little Free-Tailed Bat	Bombali Virus (BOMV)	The is the known ebolavirus, Bombali virus, detected in a little free-tailed bat first found in Sierra Leone as part of the PREDICT project. Further characterization is ongoing to understand the zoonotic potential of this virus.	

<http://www.healthmap.org/predict/>



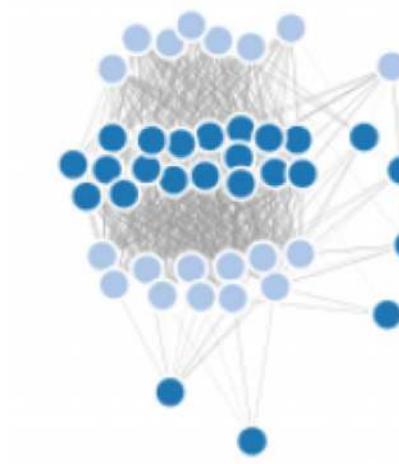
# The Gap for Wildlife Disease Surveillance Data



VerNe

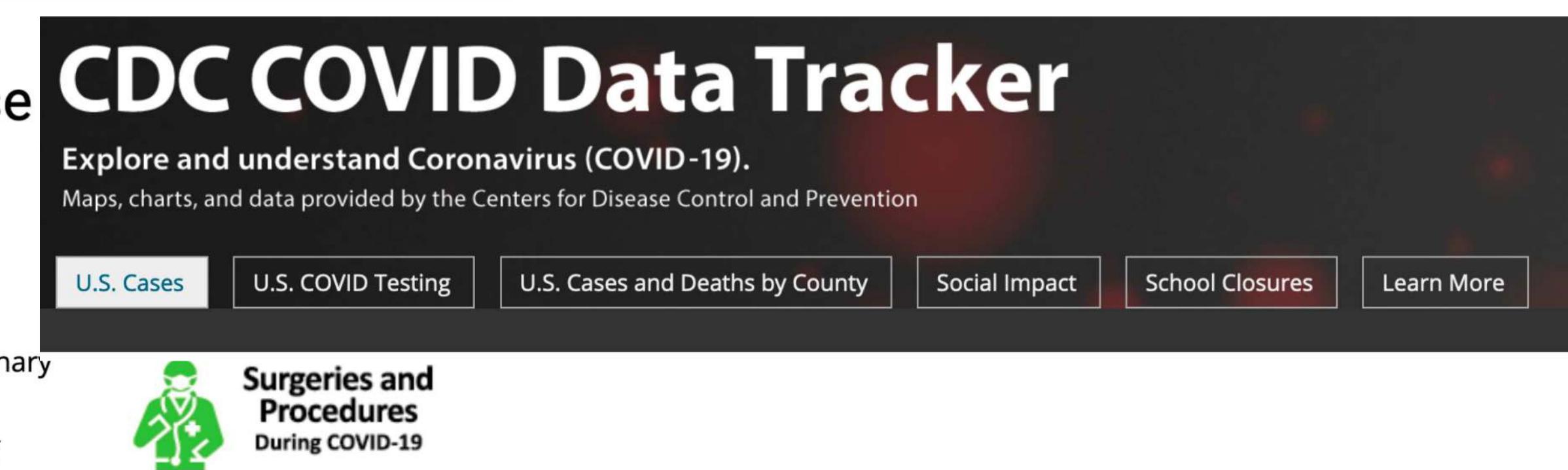
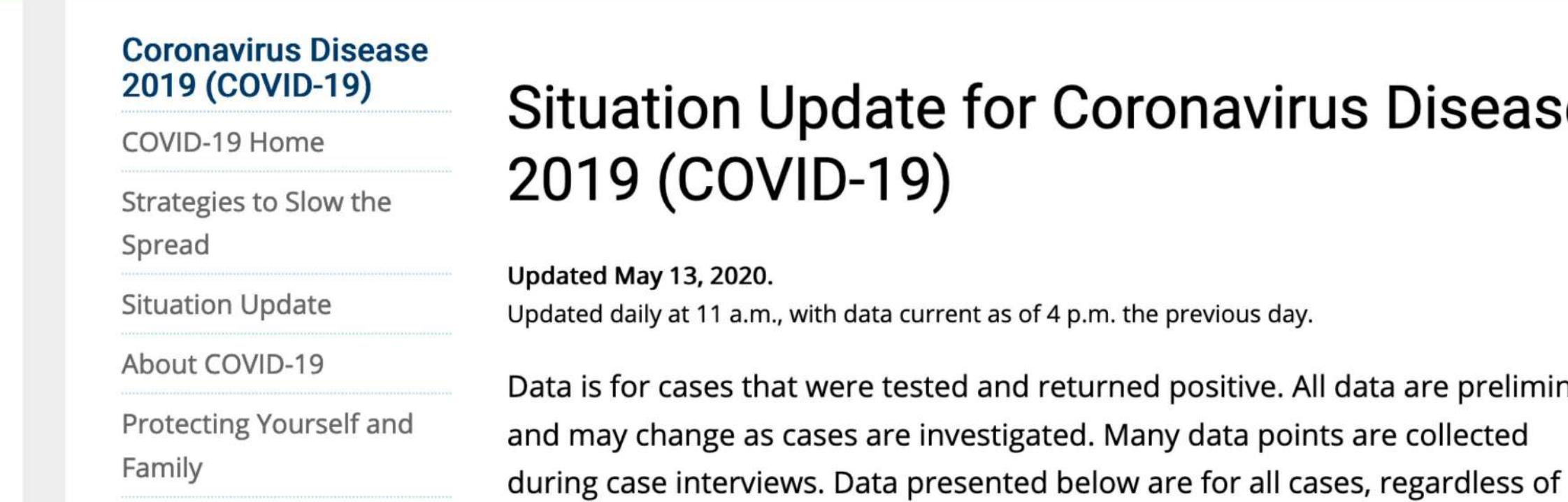
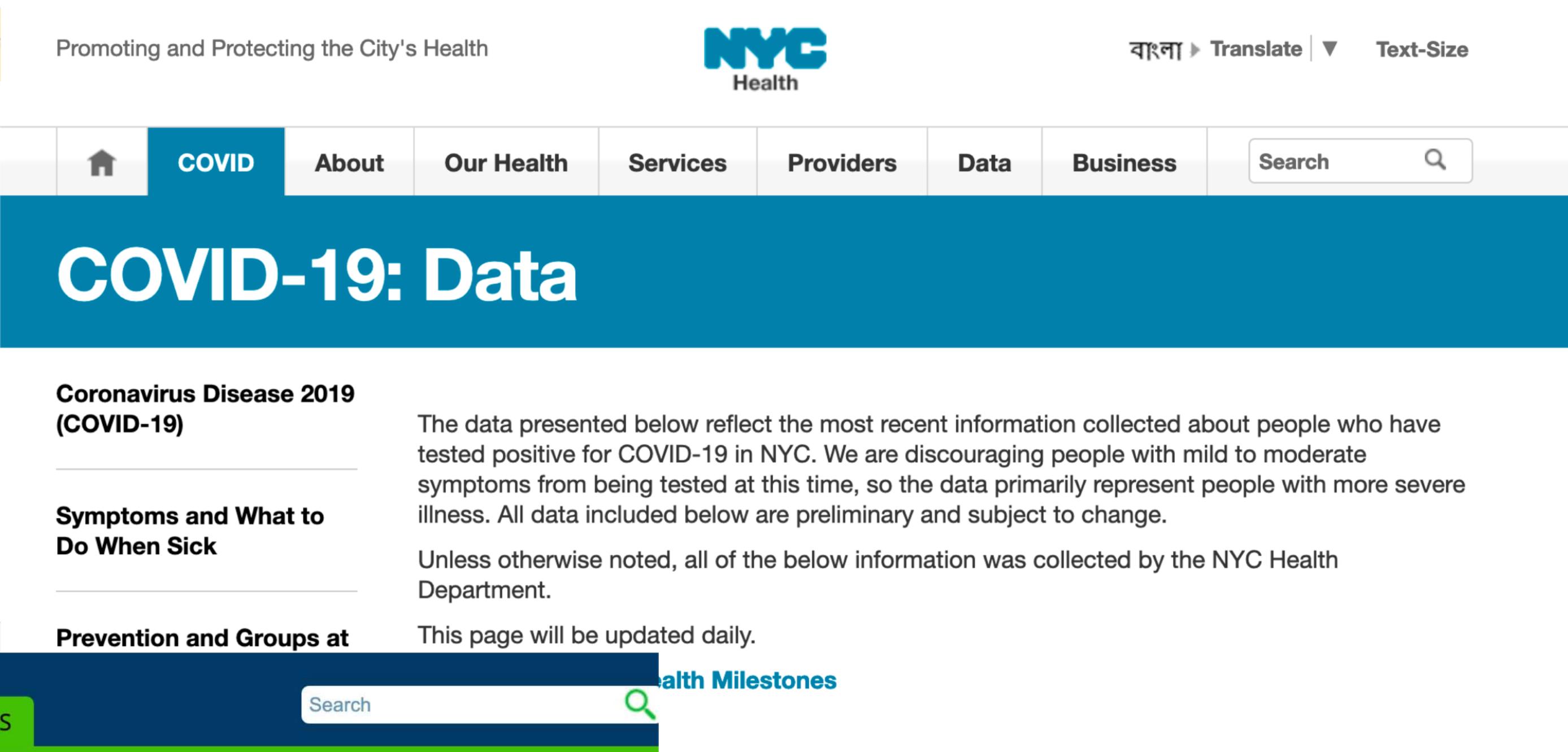
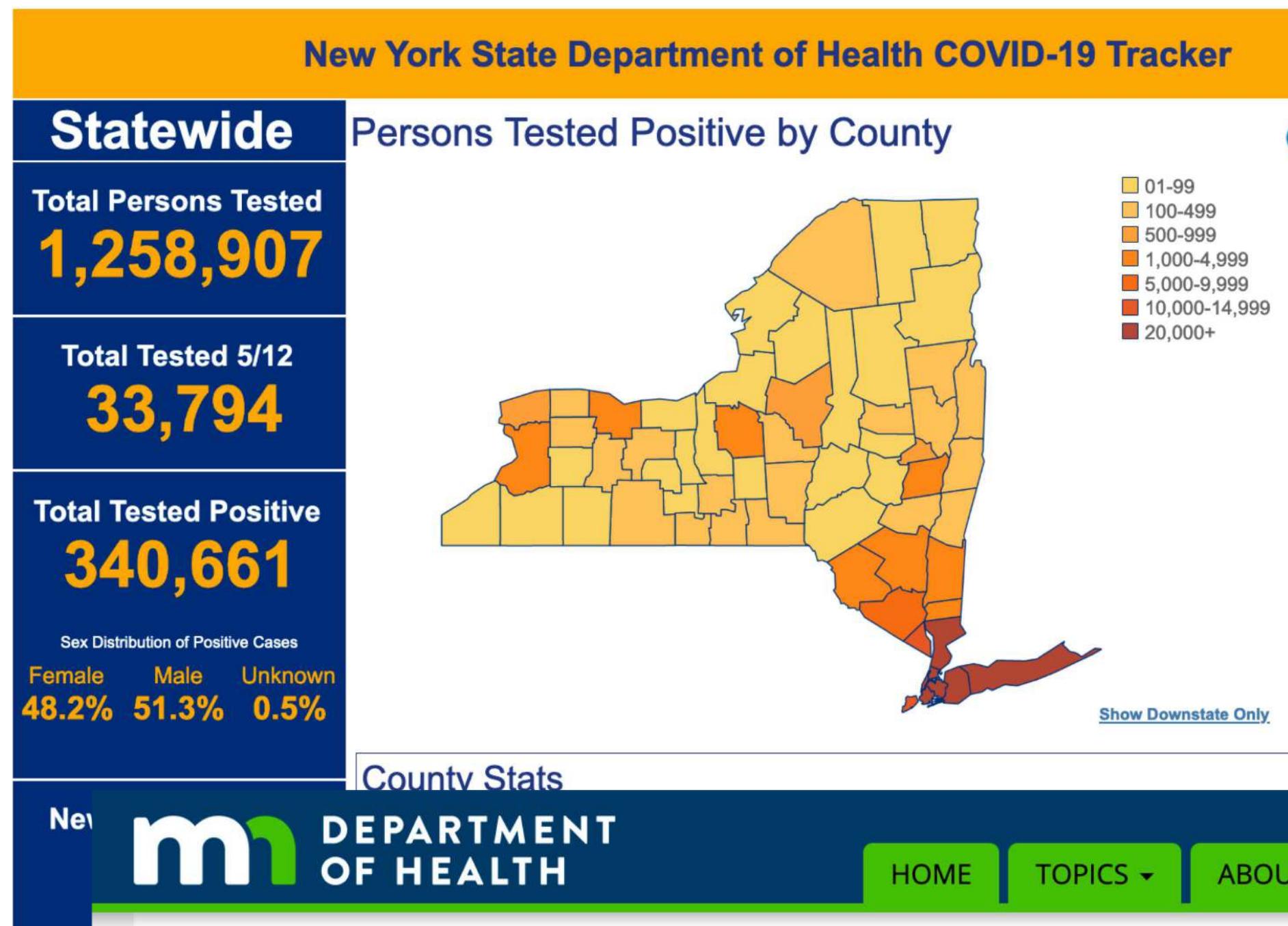


BAT ECO-INTERACTIONS



GenBank

# COVID: Government-Reported Epi Data



# COVID: Government-Reported Epi Data

NYC OpenData

Home Data Ab

 coronavirus

0 Results

No Results

 Open Data Portal

Maryland.gov Council on Open Data Developers Support Local Data ▾ Sign Up

coronavirus

0 Results

No Results

Authority ▾

Official

Community

Categories ▾

Administrative

 CALIFORNIA  
OPEN DATA PORTAL

DATASETS ORGANIZATIONS TOPIC

🏡 / Datasets

Organizations

California Departme... (3)

Topics

Health & Human Serv... (3)

Tags

california departme... (3)

coronavirus (3)

19 (2)

corona (2)

covid (2)

covid-19 (2)

covid19 (2)

adenovirus (1)

coronavirus

3 datasets found for "coronavirus"  
Type search term here

**State Sponsored COVID-19 Test Sites in California**

Location of COVID-19 testing sites sponsored by the State of California. This list does not include private sector sites.

[ZIP](#) [CSV](#)

**Influenza Surveillance**

This dataset contains the following files for California influenza surveillance data: 1)

[CSV](#) [XLSX](#) [ZIP](#)

**California COVID-19 Hospital Data and Case Statistics**

UPDATE: In response to numerous public inquiries asking for more precise definitio

[ZIP](#) [CSV](#) [chart](#) [XLSX](#)

# Aggregation and Standardization

WORLD COUNTRIES ▾ | U.S.A. STATES ▾ N.Y.C.

## Coronavirus in the U.S.: Latest Map and Case Count

By The New York Times Updated May 14, 2020, 8:06 A.M. E.T.

**TOTAL CASES**  
1.3 million+  
**DEATHS**  
84,109  
Includes confirmed and probable cases where available

Map Cases by state Hot spots Clusters Latest news »

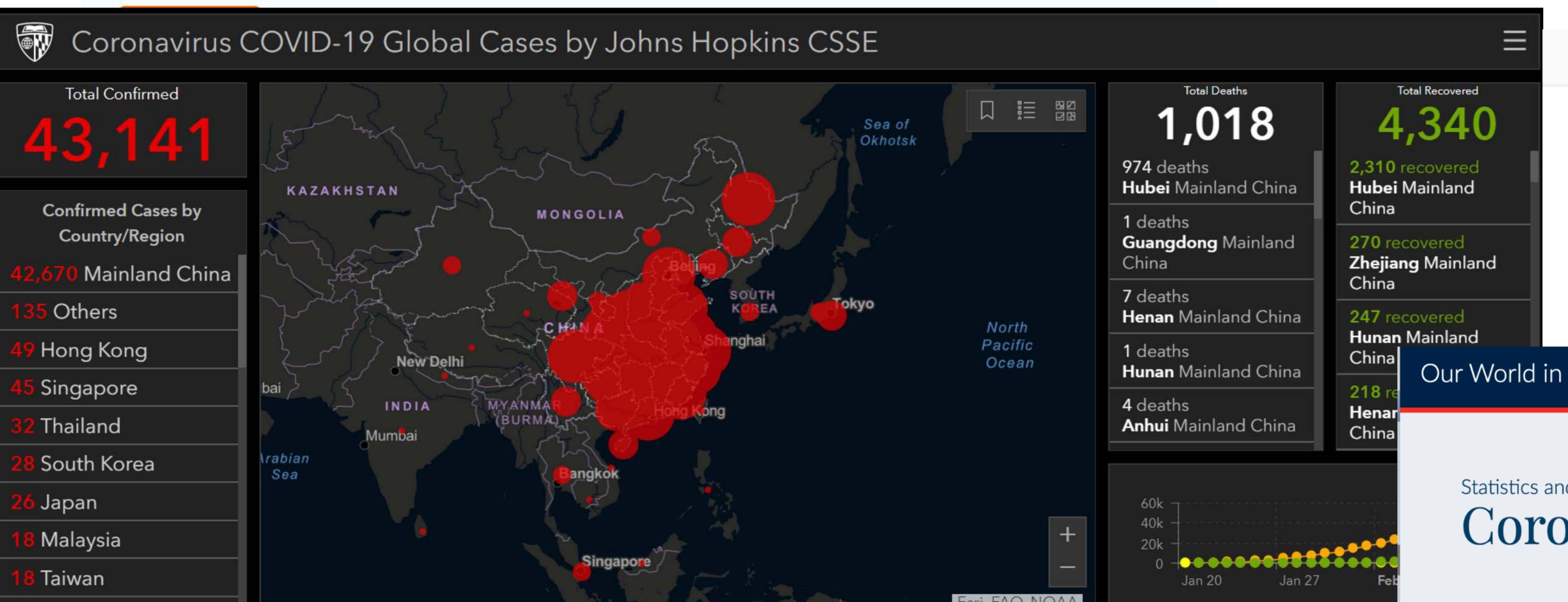
nytimes / covid-19-data

The COVID Tracking Project

FROM The Atlantic

## Most recent data

State by State Totals for the U.S. Data API



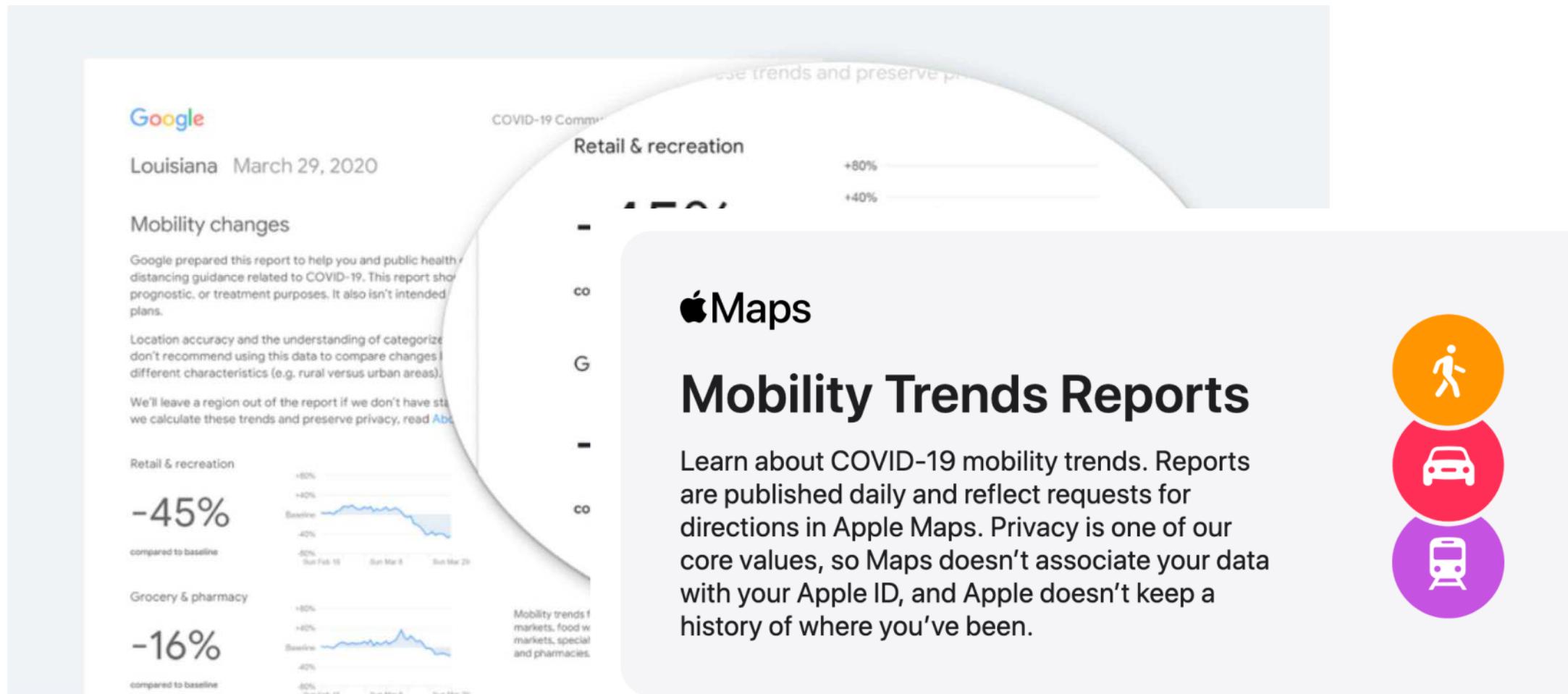
covid19R 0.1.0 Home Reference

## covid19R

The goal of covid19R is to provide a single package that allows users to access all of the tidy covid-19 datasets collected by data packages that implement the covid19R tidy data standard. It provides access to multiple datasets that meet a tidy data standard.

To learn more about the Covid19R project, check our extensive documentation about data standards, how to get your data added to this list, and more.

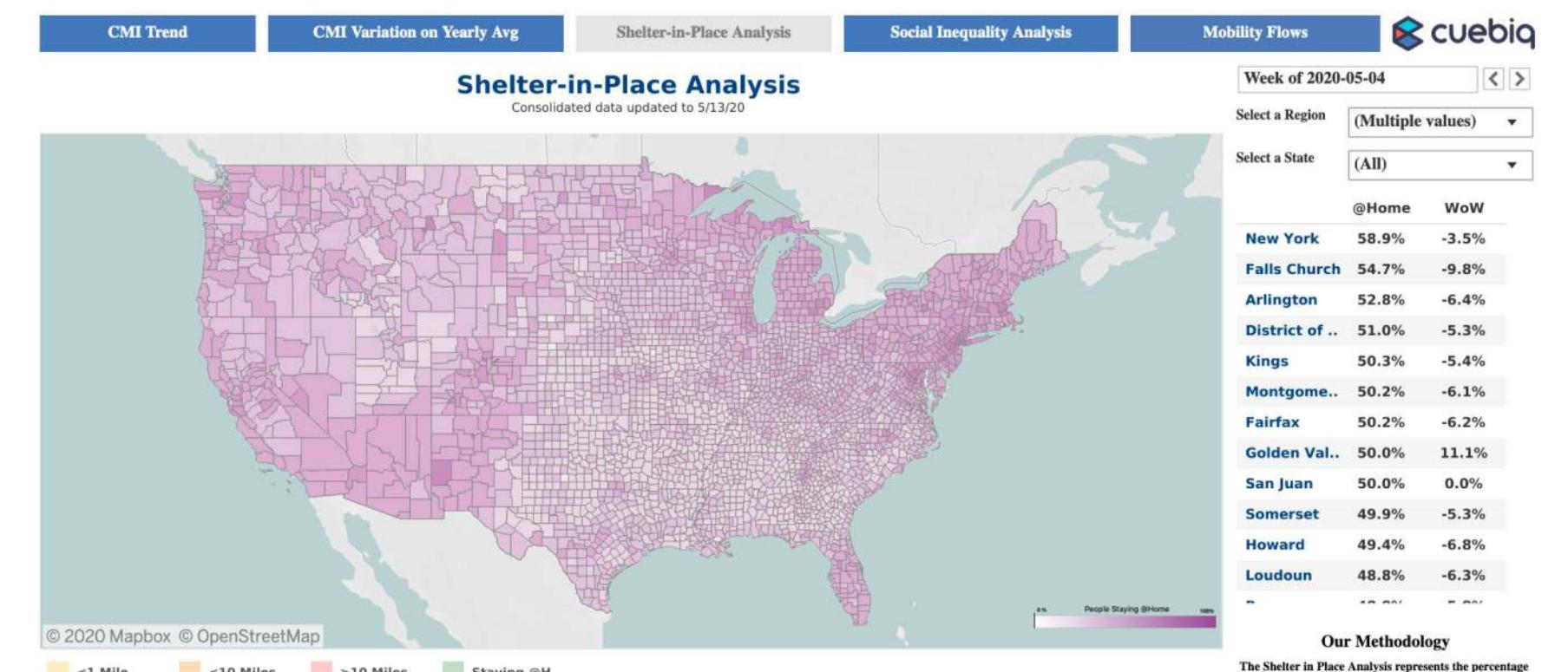
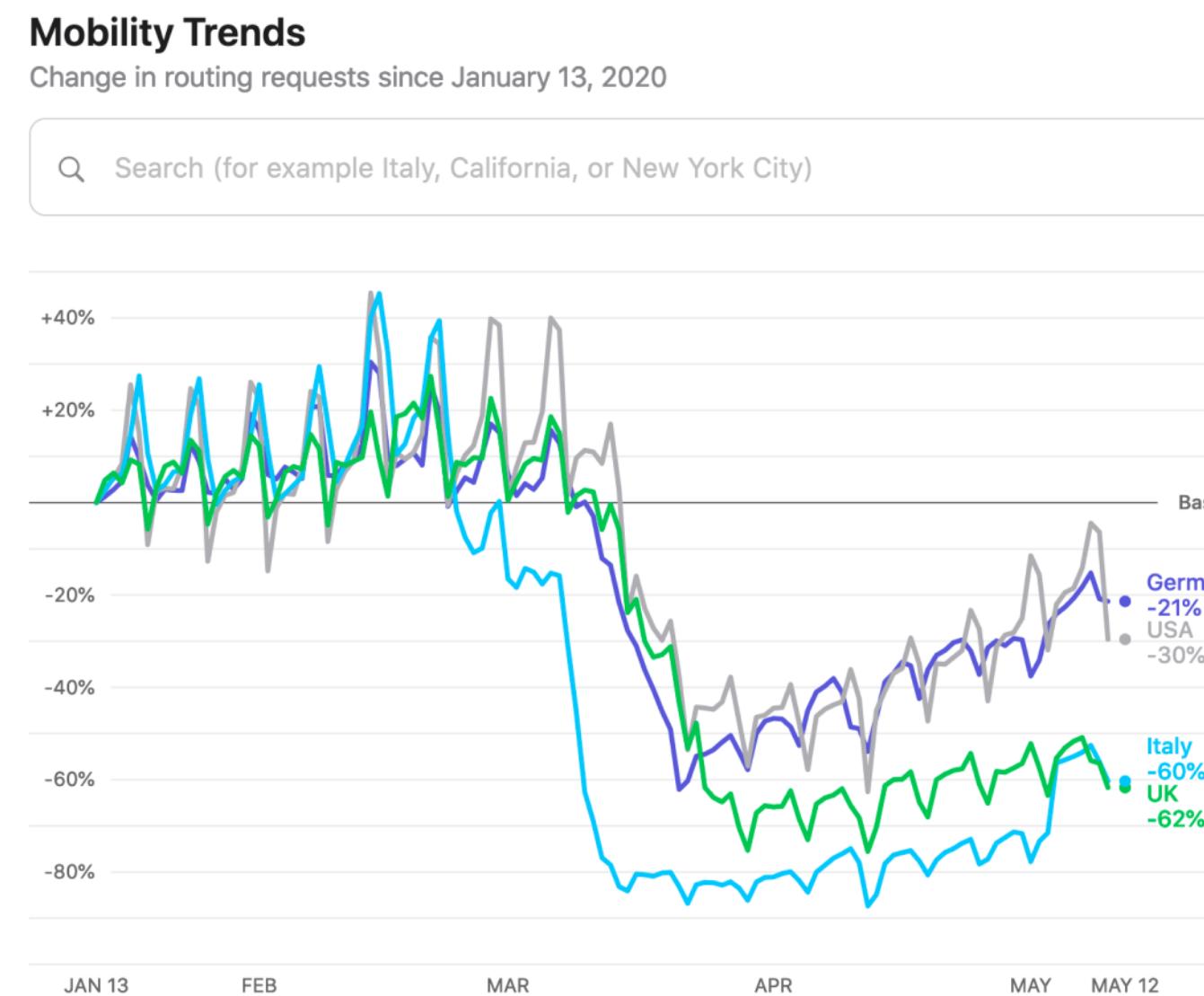
# Mobility Data



## U.S. Geographic Responses to Shelter in Place Orders

### COVID-19 Mobility Insights

We understand this is a tough time for businesses as well as consumers. As part of our commitment to sharing data for the greater good, Cuebiq is providing free access to mobility and store visitation patterns during the COVID-19 crisis to help businesses as they look to adjust their strategies to meet this new and uncertain market.



# Modeling Parameters and Clinical Study Data

## Parameter Estimates - Peer-reviewed

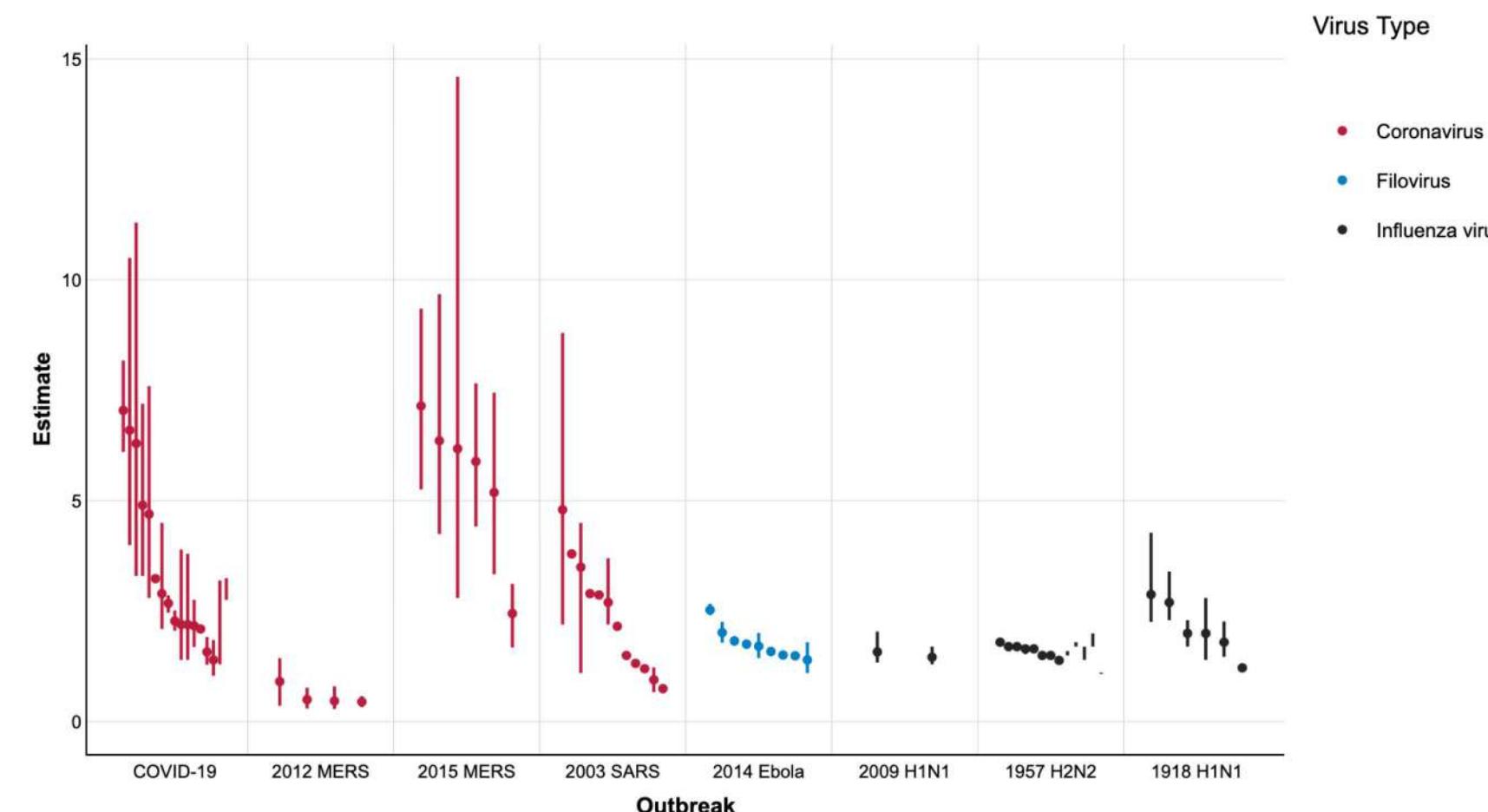
Published estimates of epidemiological characteristics that have been peer-reviewed, encoded by community members and approved by authors. For complete information on each estimate, [view this file on GitHub](#).

Search:									
NAME	VALUE	COUNTRY	LOWER BOUND	UPPER BOUND	AGE RANGE	START DATE	END DATE	PUBLICATION DATE	SOURCE
Time from hospitalization to mechanical ventilation	1.5	United States	0	12	Unspecified	2020-02-20	2020-03-05	2020-03-19	Arentz M, et al.
Time from symptom onset to hospitalization	4.4	China	0	14	Unspecified	2020-12-24	2020-01-27	2020-04-02	Zhang J, et al
Time from symptom onset to hospitalization	3.8								
Time from symptom onset to hospitalization	2.6	China	0	9	Unspecified	2020-01-28	2020-02-17	2020-04-02	Zhang J.
Proportion of icu cases on extracorporeal membrane oxygenation (ecmo)									

<https://midasnetwork.us/covid-19>

## How infectious is COVID-19?

Estimates of the **basic reproduction number ( $R_0$ )**: the average number of new infections that are caused by one infectious person, assuming the whole population is susceptible. This value can differ based on setting, e.g. it is likely higher in more crowded places, or can vary between children and adults.



<http://2019-coronavirus-tracker.com/context.html>



## Literature Review

Using publications from the CORD-19 dataset ([Last updated: 2020-05-12](#))  
Covers 10.1% of the studies published since February 1 (834 of the 8221 papers)

[FEEDBACK](#)

### About This Review

These findings have been extracted from the CORD-19 papers by machine learning algorithms with a human curation overlay (process described in [this thread](#)). The results and quotes on this page should not be relied on without reading and assessing the validity of the underlying research. If you see a conclusion that is misrepresented, please use the [feedback section](#) of this page to report it.

This project is a part of the White House Office of Science and Technology Policy's [call to action](#) for the technology community and addresses research priorities defined by the National Academies and the World Health Organization.

This review can be useful for those wanting a quick overview of what the latest literature is saying on the topics we cover. It might also help those writing local guides, expert opinions or systematic reviews. Click on the topic in the table of contents below to see the results table for that topic.

### Key Scientific Questions about COVID-19

- Persistence of sources
- Persistence on surfaces
- Incubation period
- Physical science
- Diagnostics
- Immune Response
- Asymptomatic shedding
- Movement Control
- Adaptation of Virus
- Information sharing and inter-sectoral collaboration
- Seasonality
- Social and Ethical Considerations
- Vaccine and Therapeutics
- Natural history of the virus
- PPE

### COVID-19 Risk factors

- Hypertension
- Diabetes
- Male gender
- Heart Disease

<https://www.kaggle.com/covid-19-contributions>

# Some Observations

Aggregation happening fast  
but loses useful detail

Linking and metadata key  
but one of the biggest  
challenges

Relevant policy questions  
changing rapidly

Answering the right questions for policymakers on COVID-19

Ellie Graeden • Colin Carlson • Rebecca Katz 

Open Access • Published: April 20, 2020 • DOI: [https://doi.org/10.1016/S2214-109X\(20\)30191-1](https://doi.org/10.1016/S2214-109X(20)30191-1)

**Key questions that officials and experts need to be able to address**

- 1. Clinical presentation and testing**
- 2. Treatment: supplies, hospital beds, workforce**
- 3. Non-pharmaceutical interventions: adherence and mobility**
- 4. Public health response: ability to contact trace and identify exposures**
- 5. Compound hazards and concurrent hazard planning**

# Thanks!



our  
future.

Complete the census today at:  
**2020CENSUS.GOV**  
Paid for by U.S. Census Bureau

Shape  
your future  
START HERE >

We all get  
to shape  
our future.

Complete the census today at:  
**2020CENSUS.GOV**  
Paid for by U.S. Census Bureau

Shape  
your future  
START HERE >