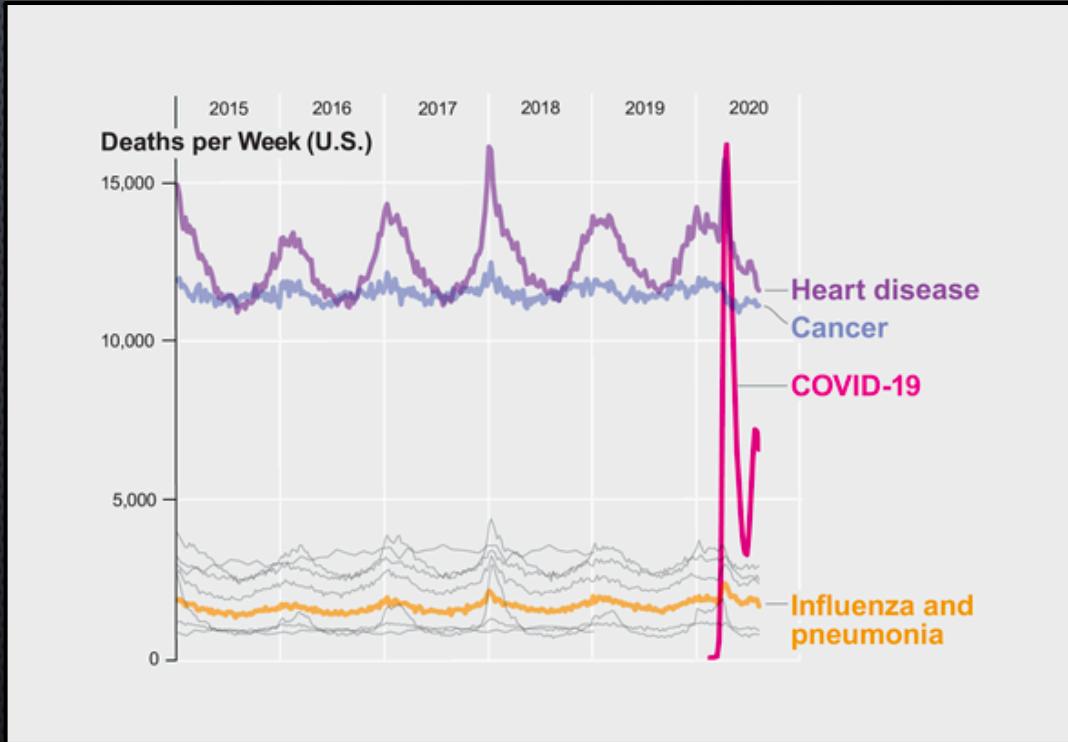


SARS-CoV-2 VARIANT TIMEMAPS

RENÉ L WARREN
BC GENOME SCIENCES CENTRE
01-MAR-2021

COVID-19 PANDEMIC

- INCALCULABLE LOSS OF LIVES, TOLL ON HUMAN HEALTH



Credit: Youyou Zhou
Scientific American

- IMPACTING ALL SECTORS OF HUMAN ACTIVITY
- GRAVE SOCIO-ECONOMIC IMPACT, WILL BE FELT FOR YEARS

A DIRE SITUATION



COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins Universit...



Global Cases

114,223,289

Cases by
Country/Region/Sovereignty

28,606,340 US

11,112,241 India

10,551,259 Brazil

4,209,850 Russia

4,188,827 United
Kingdom

3,815,639 France

3,188,553 Spain

Admin0 ▶

Last Updated at (M/D/YYYY)
3/1/2021, 6:23 AM

Cumulative Cases

Active Cases

Incidence Rate

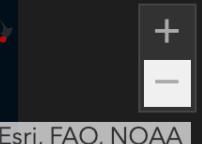
Case-Fatality Ratio

Testing Rate

192
countries/regions

Lancet Inf Dis Article: [Here](#). Mobile Version: [Here](#). Data sources: [Full list](#). Downloadable database: [GitHub](#), [Feature Layer](#).

March 1st, 2021 <https://coronavirus.jhu.edu/map.html>



Esri, FAO, NOAA

Global Deaths

2,533,129

513,092 deaths
US

254,942 deaths
Brazil

185,715 deaths
Mexico

157,157 deaths
India

Global Deaths

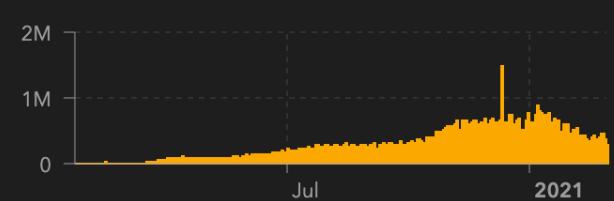
US State Level
Deaths, Recovered

52,158 deaths,
recovered
California US

47,615 deaths,
recovered
New York US

43,697
deaths, **2,422,369**

US Deaths, R...

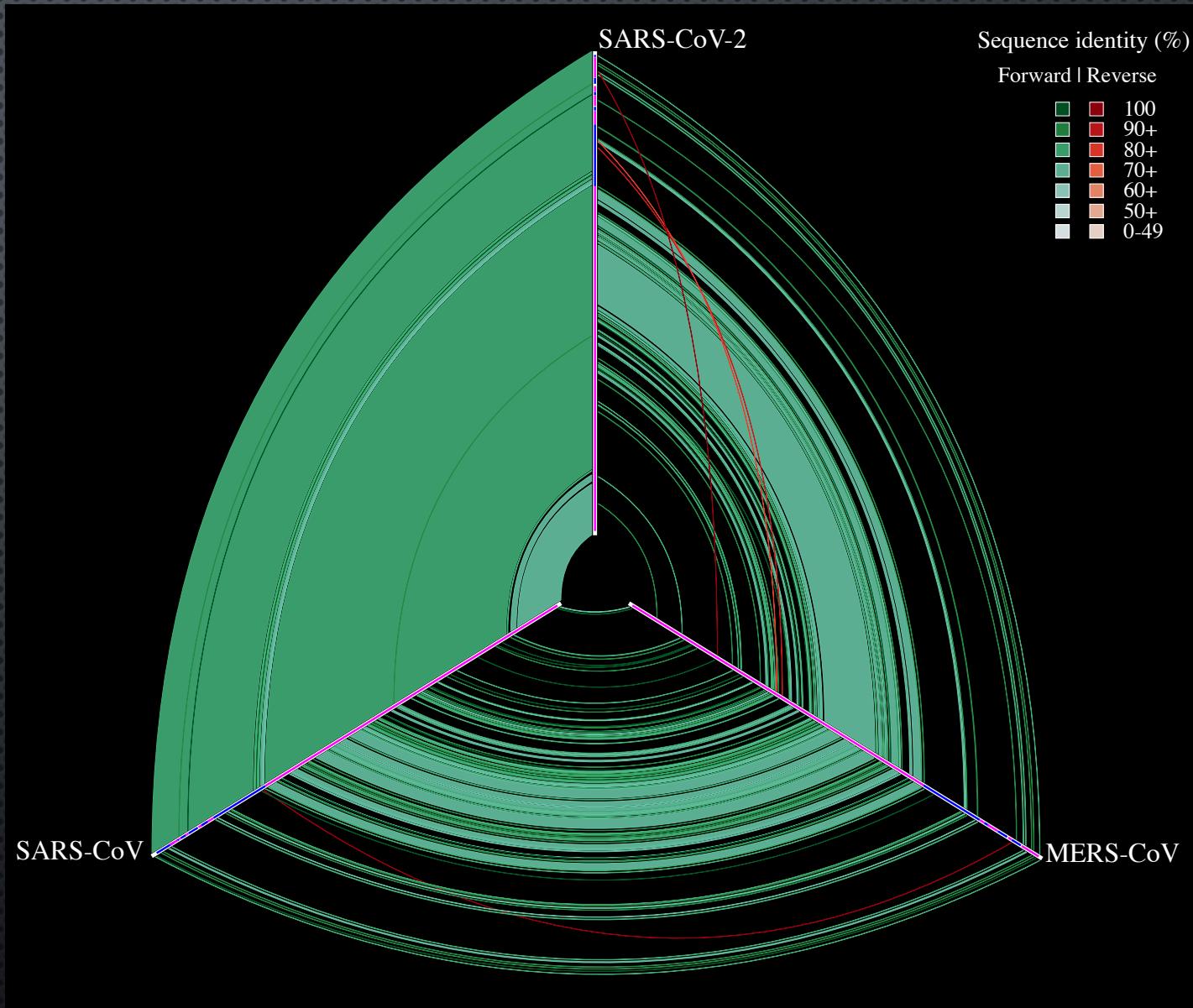


Daily Cases



SARS-CoV-2

- CORONAVIRUS RELATED TO SARS-CoV
- COVID-19 AGENT
- COMPACT (30kbp) RNA GENOME
 - 10 GENES
 - SURFACE GLYCOPROTEIN (SPIKE)
 - BINDS HOST ACE2 TO GAIN ENTRY
 - VARIANTS OF CONCERNS



GLOBAL EFFORTS

- **GISAID** (VIRAL GENOME REPOSITORY)
- SRA (SEQUENCE READ ARCHIVE)
- CANADIAN COVID GENOMICS NETWORK (CANCOGEN)
- MEDICAL, CLINICAL, SEQUENCING CENTRES AND PUBLIC HEALTH LABS

TRACKING INFECTION ORIGINS

TRACKING A CHANGING VIRUS

All viruses, including SARS-CoV-2, mutate slightly as they spread. Since late 2019, when this virus was first reported in humans, more than 70,000 samples from around the globe have been genetically sequenced, allowing researchers to see how it has traveled

HOW TO READ THE CHART

Like a family tree, this visual shows how virus samples are related. For clarity, it only includes about 3,800 genomes from the sample pool. Each dot is a unique infection. Dot colors show where the virus was sampled. Line colors show from where it may have traveled

HOW DO MUTATIONS HAPPEN?

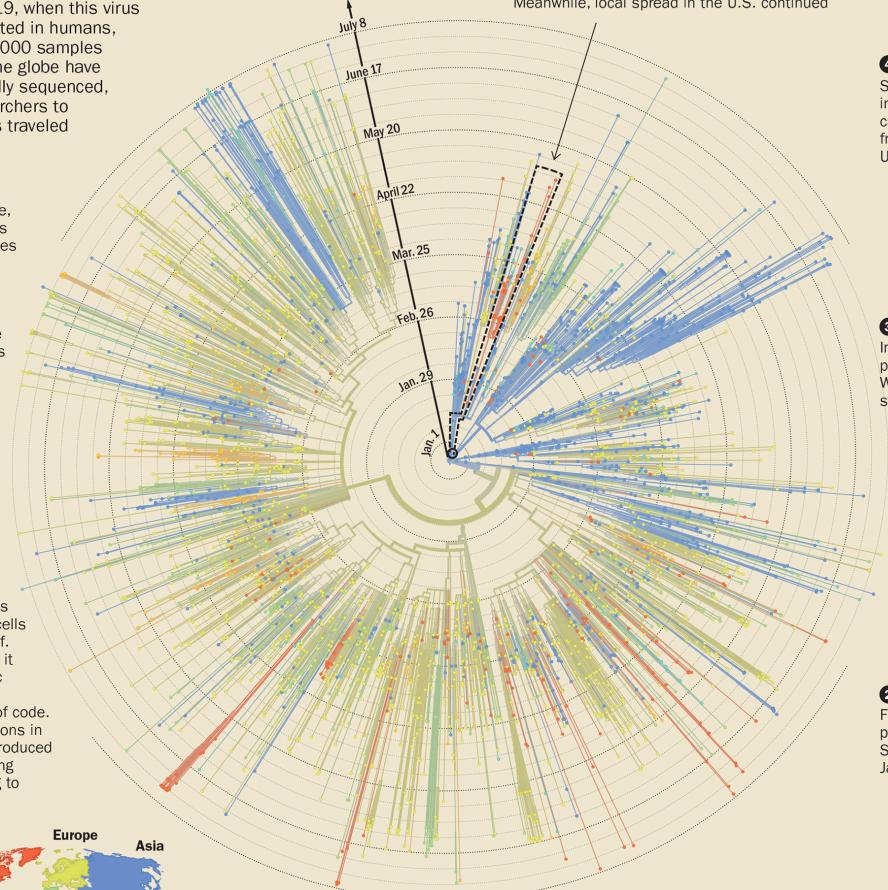
SARS-CoV-2 uses healthy human cells to replicate itself. Once in the cell, it releases genetic material—about 30,000 letters of code. But small variations in the code are introduced during the copying process, leading to slight mutations



NOTE: NUMBER OF SAMPLES IN A REGION IS NOT INDICATIVE OF NUMBER OF CASES IN THAT REGION
SOURCE: NEXTSTRAIN; DATA PULLED JULY 20, 2020

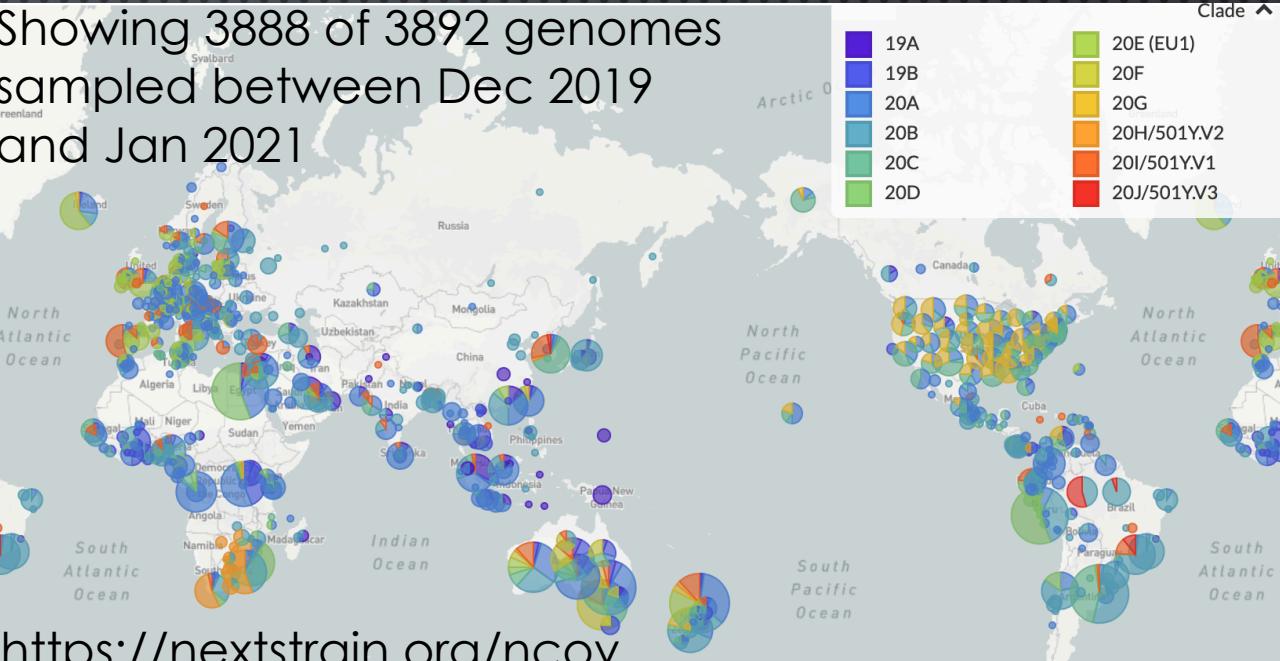
HOW THE CHART CAN BE USED

This analysis genetically traces the earliest known U.S. case, in the Seattle area, to China. That strain spread through Washington state—and from there, on to Europe, South America, Oceania, and even back to Asia. Meanwhile, local spread in the U.S. continued



- 1 Early sequence from Wuhan, China
- 2 First U.S. patient in Seattle, Jan. 20
- 3 Infections proliferate in Washington state
- 4 Some international cases come from the U.S.
- 5 Local spread continues

Showing 3888 of 3892 genomes sampled between Dec 2019 and Jan 2021



<https://nextstrain.org/ncov>

- GREAT, INTUITIVE UI
- BASED ON PHYLOGENETIC ANALYSES
- LIMITED SAMPLE SIZE ↑
- ← NO OR ↓ LIMITED SNV/JURISDICTIONS (HTTPS://COVARIANTS.ORG)

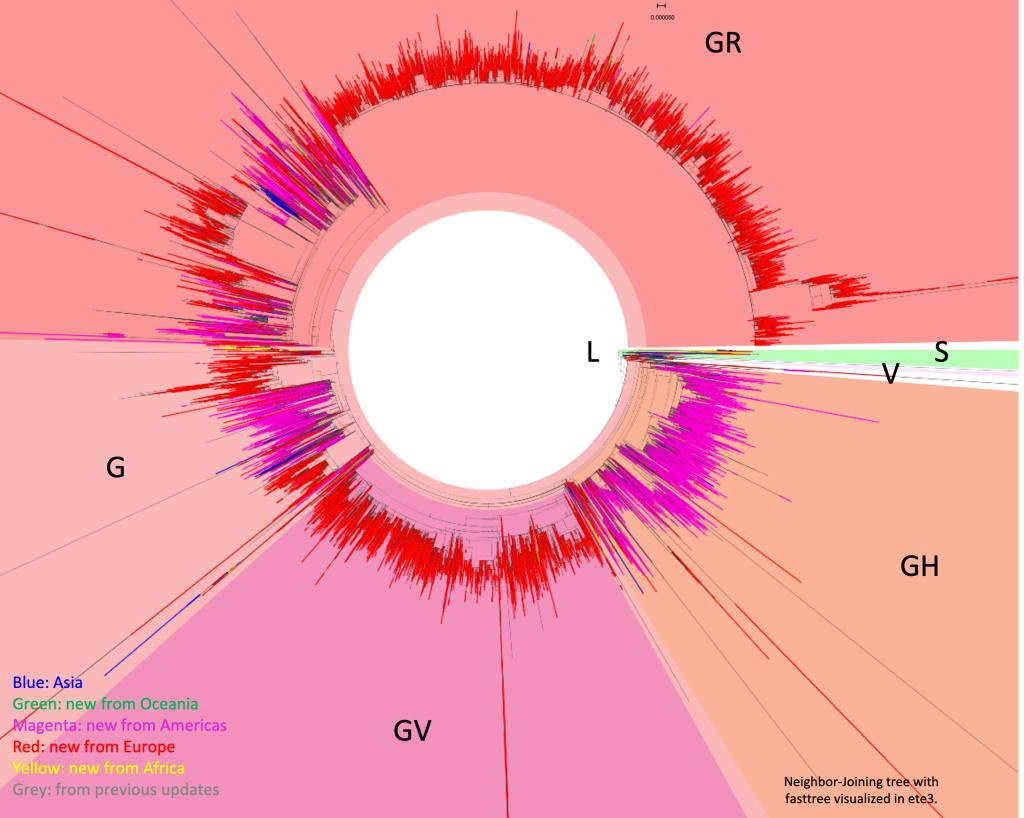
<https://time.com/5870481/coronavirus-origins/>

VIRAL POPULATIONS

Larger clades in GISAID were named in context of marker variants relative to WIV04-reference:

S C8782T,T2814C NS8-L84S
 L C241,C3037,A23403,C8782,G11083,G25563,G26144,T28144,G28882 (WIV04-reference)
 V G11083T,G26144T NSP6-L37F + NS3-G251V
 G C241T,C3037T,A23403G S-D614G
 GH C241T,C3037T,A23403G,G25563T S-D614G + NS3-Q57H
 GR C241T,C3037T,A23403G,G28882A S-D614G + N-G204R
 GV C241T,C3037T,A23403G, C28882T S-D614G + S-A222V

GV C241T,C3037T,A23403G, C28882T S-D614G + S-A222V



Full genome tree derived from all outbreak sequences 2021-02-09

Notable changes:
 461,049 full genomes (+26,973)
 (excluding low coverage, out of 487,322 entries)

Updated clades:
RBDx: mutations known to increase receptor binding

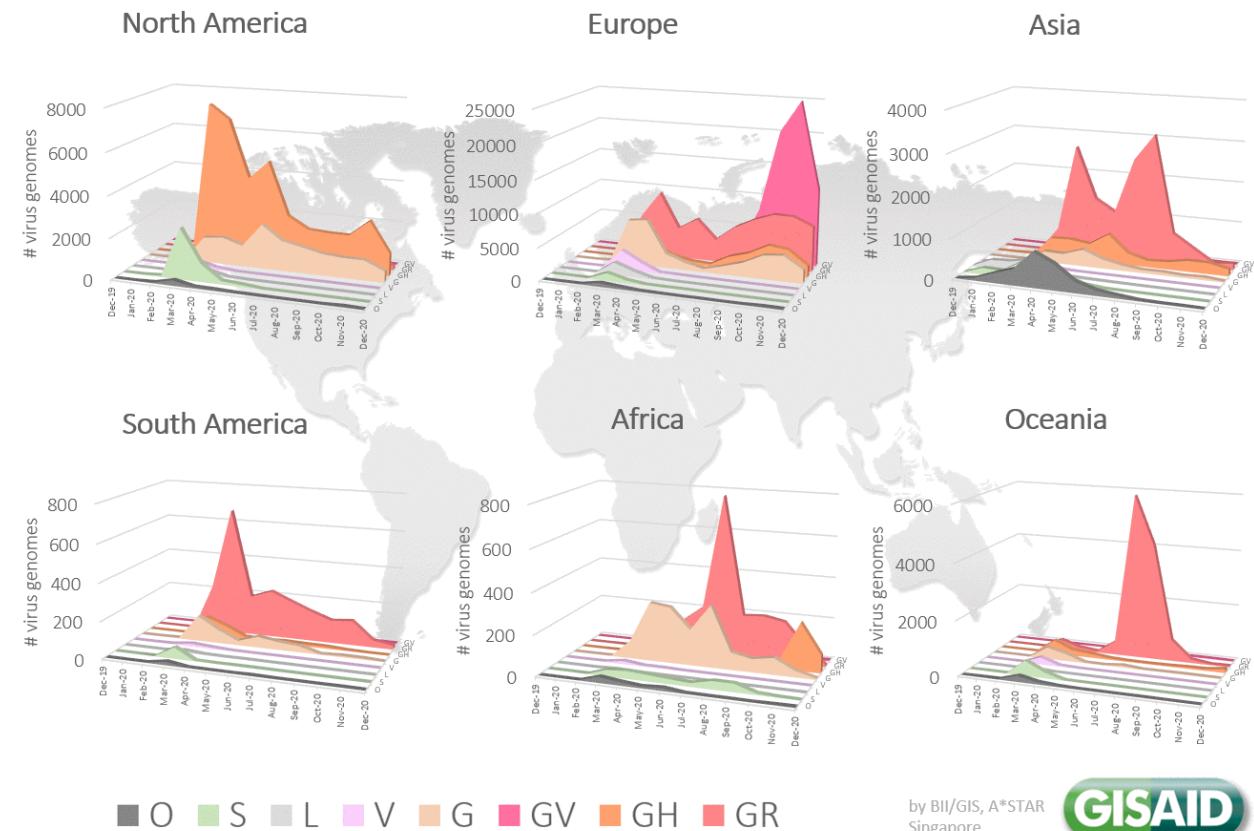
S clade 8,386 [RBDx 101] (+90 [+17])
L clade 4,866 [RBDx 6] (+13 [+0])
V clade 5,806 [RBDx 1] (+1 [+0])
G clade 66,279 [RBDx 11,751] (+2,644 [+803])
GR clade 169,310 [RBDx 62,727] (+12,501 [+10,290])
GH clade 99,096 [RBDx 15,653] (+5,358 [+1,265])
GV clade 101,406 [RBDx 277] (+6,209 [+27])
Other clade 5,900 [RBDx 198] (+157 [+26])

We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.



by BII/GIS, A*STAR Singapore

Temporal and regional distribution of clades in the first year



by BII/GIS, A*STAR Singapore



EVOLUTION IN HUMAN HOSTS

GENOME REPLICATION ERRORS YIELDING FITNESS ADVANTAGE

- WHAT (DISTINCT SNVs) ?
- HOW MANY?
- HOW FAST THEY ACCUMULATE?
- WHEN DID THEY EMERGE?
- WHERE?

VARIANTS **BIRD'S EYE VIEW**

ANALYSIS / DISSEMINATION SARS-COV-2 VARIANTS

GENOME DOWNLOAD

GISAID

VARIANT PROFILING

ntEdit

PLOTTING

SVG

WEB HOSTING

bcgsc.github.io

GENOME DOWNLOAD

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



<https://www.epicov.org/>

You are logged in as **Rene Warren** - [logout](#)

Registered Users

EpiFlu™

EpiCoV™

My profile



EpiCoV™



Search



Downloads



Upload



My Unreleased

Search

Accession ID	<input type="text"/>	Virus name	<input type="text"/>	<input checked="" type="checkbox"/> complete <small>?</small>	<input checked="" type="checkbox"/> high coverage <small>?</small>
Location	<input type="text"/>	Host	<input type="text"/>	<input checked="" type="checkbox"/> low coverage excl <small>?</small>	<input type="checkbox"/> w/Patient status <small>?</small>
Collection	2020-01-01 <small>⌚</small> to <input type="text"/>	Submission	2021-02-04 <small>⌚</small> to <input type="text"/>	<input checked="" type="checkbox"/> collection date compl <small>?</small>	
Clade	<input type="text"/> all <small>▼</small>	Lineage	<input type="text"/>	Variants / Substitutions <small>?</small>	<input type="button" value="Reset"/> <input type="button" value="Fulltext ▲"/>

<input checked="" type="checkbox"/>	Virus name	Passage de	Accession ID	Collection da	Submission d	<small>i</small>	Length	Host	Location	Originating lab
<input checked="" type="checkbox"/>	hCoV-19/Belgium/rega-2110/2020	Original	EPI_ISL_935206	2020-01-10	2021-02-04		29,782	Human	Europe / Belgium	KU Leuven
<input checked="" type="checkbox"/>	hCoV-19/France/IDF_HB_112001900001/2	Original	EPI_ISL_940542	2020-01-29	2021-02-05		29,852	Human	Europe / France	Hôpital Bic
<input checked="" type="checkbox"/>	hCoV-19/France/IDF_HB_112001106076/2	Original	EPI_ISL_940552	2020-01-29	2021-02-05		29,865	Human	Europe / France	Hôpital Bic
<input checked="" type="checkbox"/>	hCoV-19/France/IDF_HB_112002007147/2	Original	EPI_ISL_940551	2020-02-03	2021-02-05		29,867	Human	Europe / France	Hôpital Bic
<input checked="" type="checkbox"/>	hCoV-19/France/IDF_HB_112002096647/2	Original	EPI_ISL_940533	2020-02-26	2021-02-05		29,870	Human	Europe / France	Hôpital Bic
<input checked="" type="checkbox"/>	hCoV-19/France/IDF_HB_112002103234/2	Original	EPI_ISL_940534	2020-02-27	2021-02-05		29,864	Human	Europe / France	Hôpital Bic
<input checked="" type="checkbox"/>	hCoV-19/France/IDF_HB_112002107124/2	Original	EPI_ISL_940535	2020-02-28	2021-02-05		29,862	Human	Europe / France	Hôpital Bic

10

Total: 7,130 viruses

<< < 1 2 3 4 5 > >>

Select

Analysis

VARIANT PROFILING

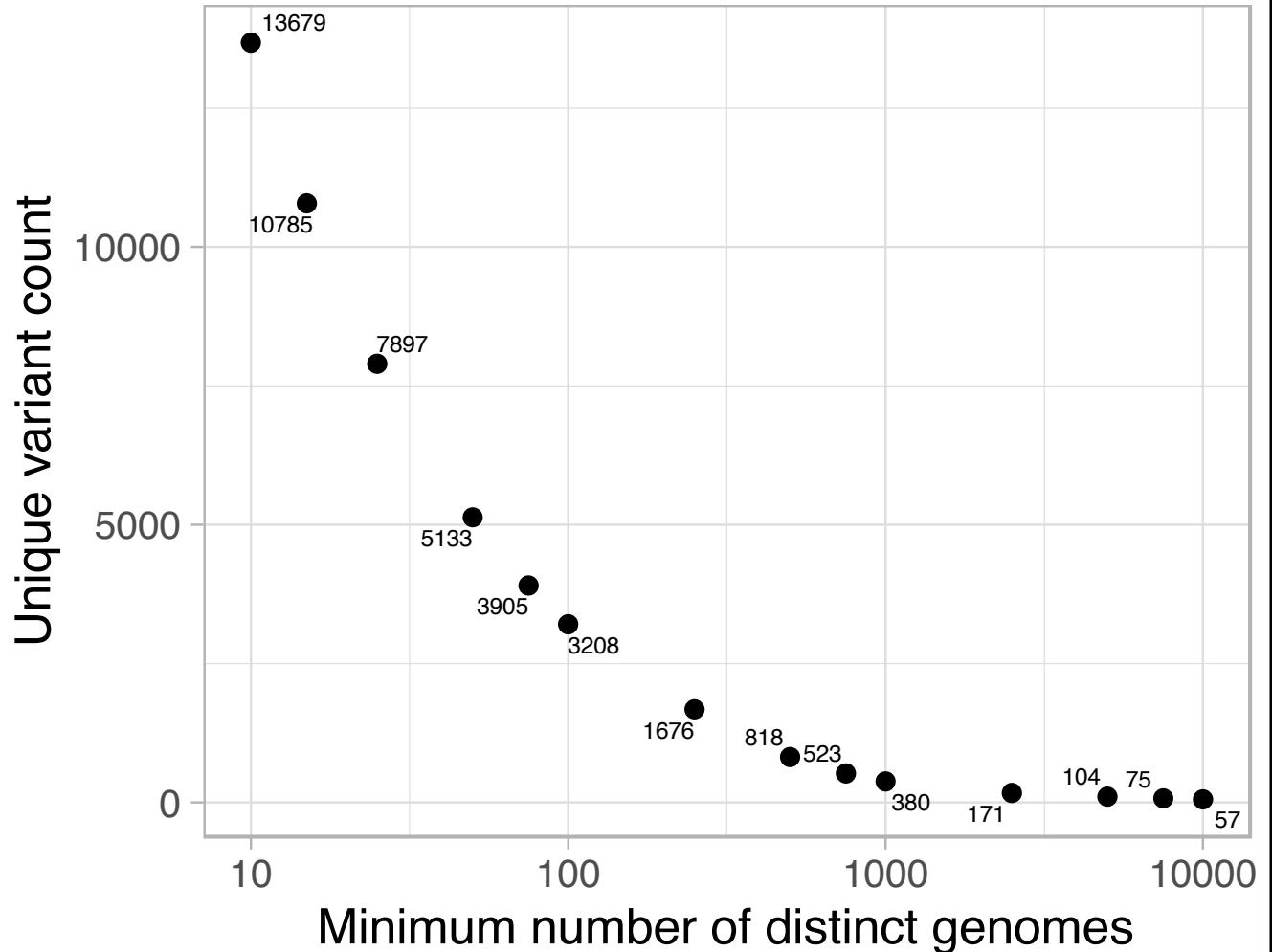
- AUTOMATED PROCESS
 - ITERATE EACH **GISAID** GENOME > **NTHITS** BLOOM FILTER > **NTEDIT** (POLISHING)
 - REFERENCE: WH-HUMAN 1 CORONAVIRUS (WU, F. ET AL. 2020)
 - PARSE VCF FILE AND SUMMARIZE EACH GISAID ID

hCoV-19/Denmark/DCGC-36731/2020|EPI ISL 927872|2020-11-23;C241T,C19032T,C21846T,A23403G,...

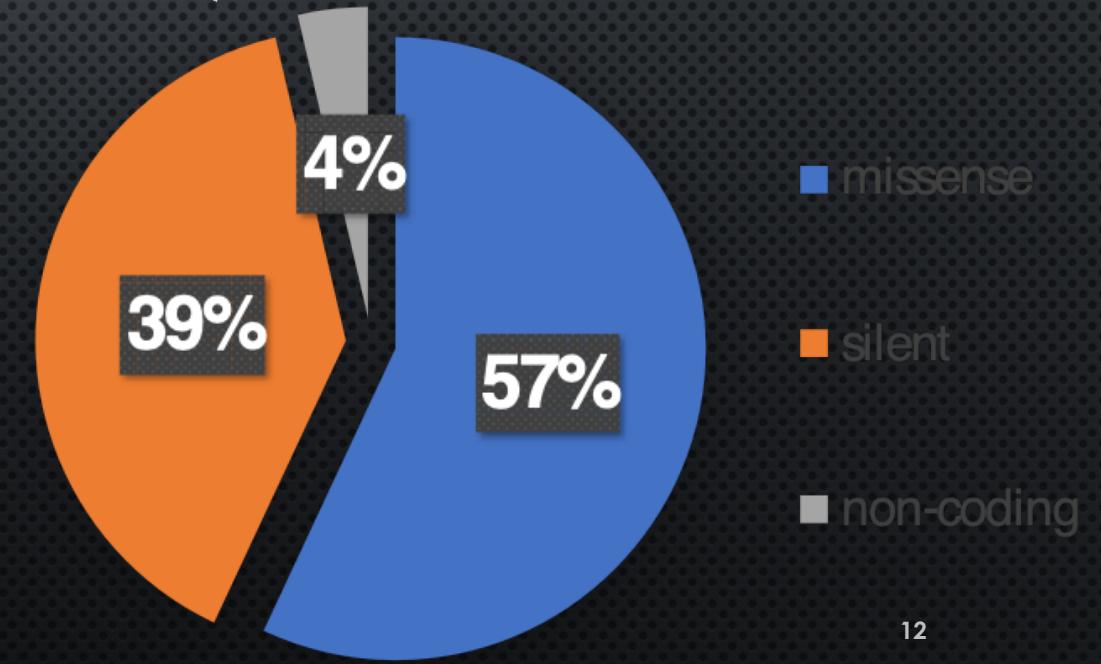
- ADDITIONAL SCRIPTS MAP EFFECT OF NUCLEOTIDE VARIATION

VARIANT PROFILING

Variant [genome] depth distribution

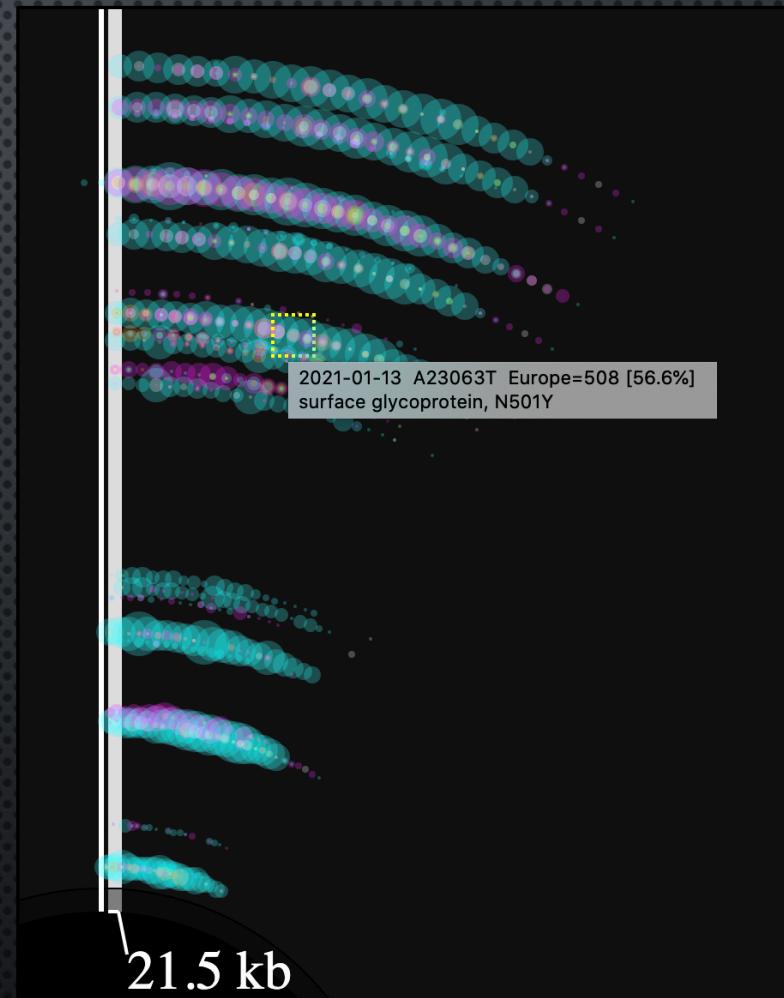


- >300,000 COMPLETE SARS-CoV-2 GENOMES
- 6 CONTINENTS
- >10,000 NUCLEOTIDE VARIANTS



PLOTTING

- COMPREHENSIVE HIGH-RESOLUTION OUTPUT
- SCALABLE VECTOR GRAPHICS (SVG)
 - **VARIANTS ORGANIZED** ALONG SARS GENOME AXIS
 - YEAR DATA **ARRANGED CLOCKWISE** FROM MIDNIGHT
 - **1 DAY = 1 RADIUS**
 - **COLORED** BY REGION
 - **SIZED** BY SAMPLE COUNT / PROPORTION
 - **HOVER** REVEALS DATE, VARIANT, REGIONS
COUNT, PROPORTION, GENE PRODUCT, EFFECT



WEB HOSTING

SARS-CoV-2

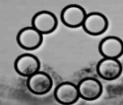
Variant Timemaps

2020 2021

Counts

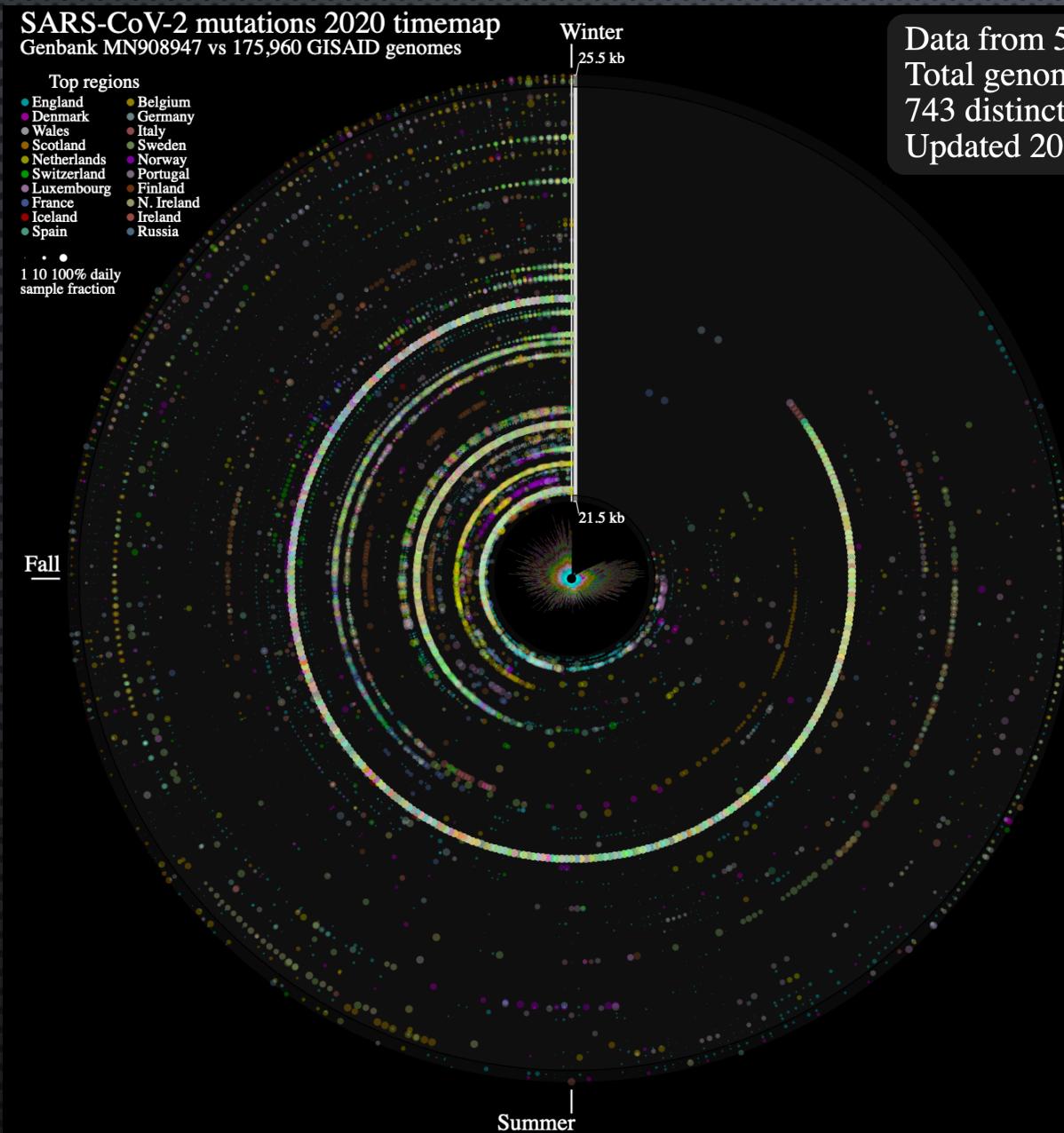
Ratios

Download

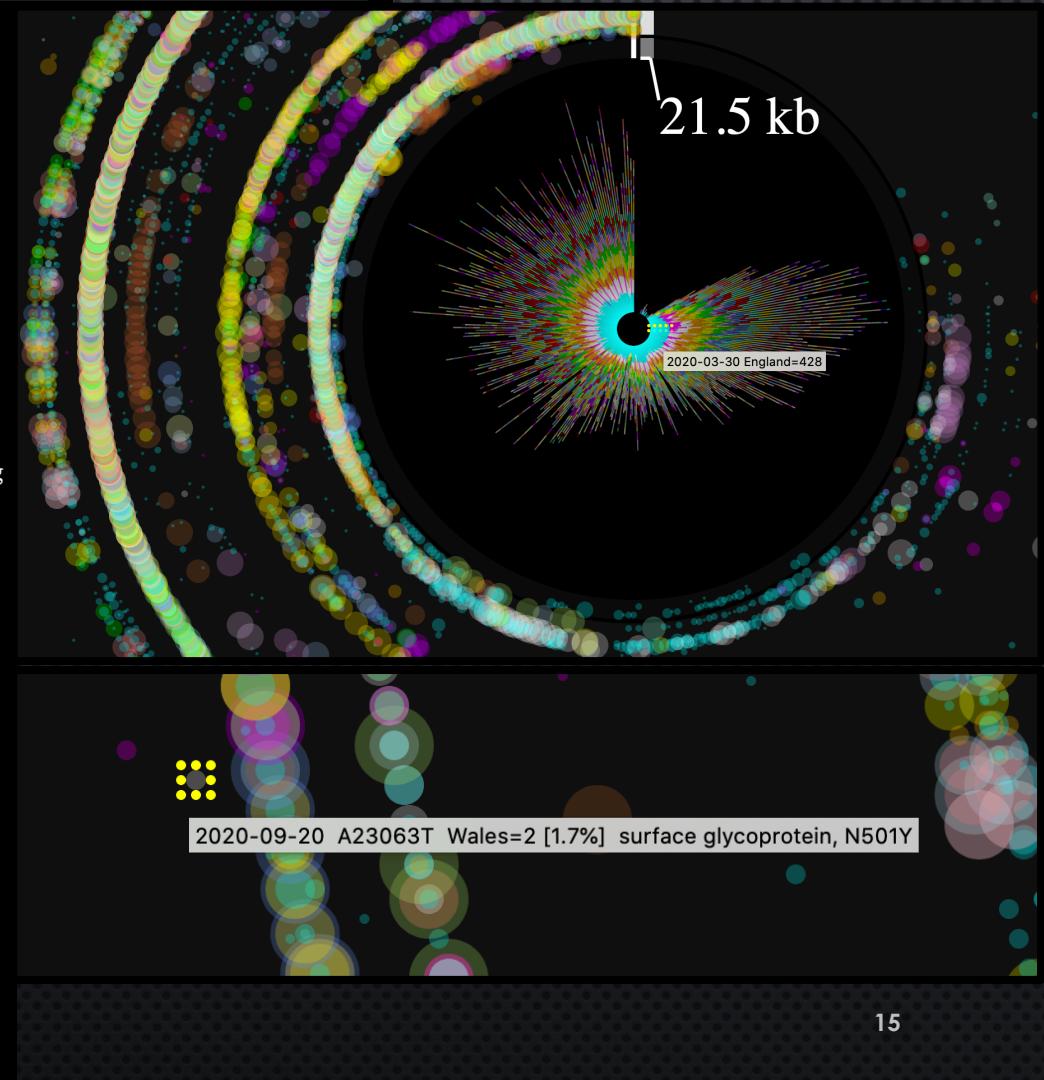


<https://bcgsc.github.io/SARS2>

A WEALTH OF INFORMATION

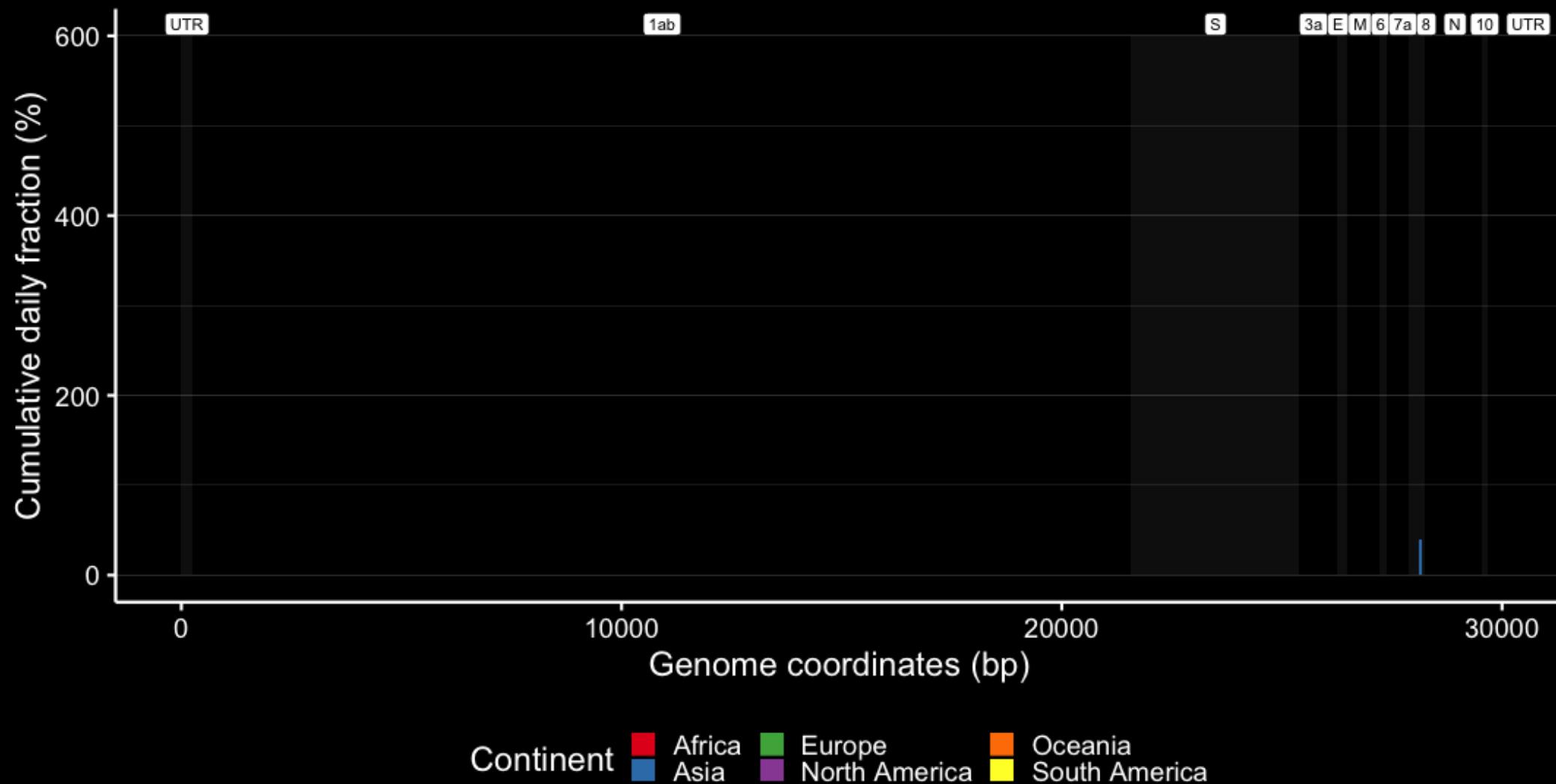


Data from 50 regions in Europe
Total genomes \geq 10 Genomes/day \geq 2
743 distinct missense variants mapped
Updated 2021-02-05



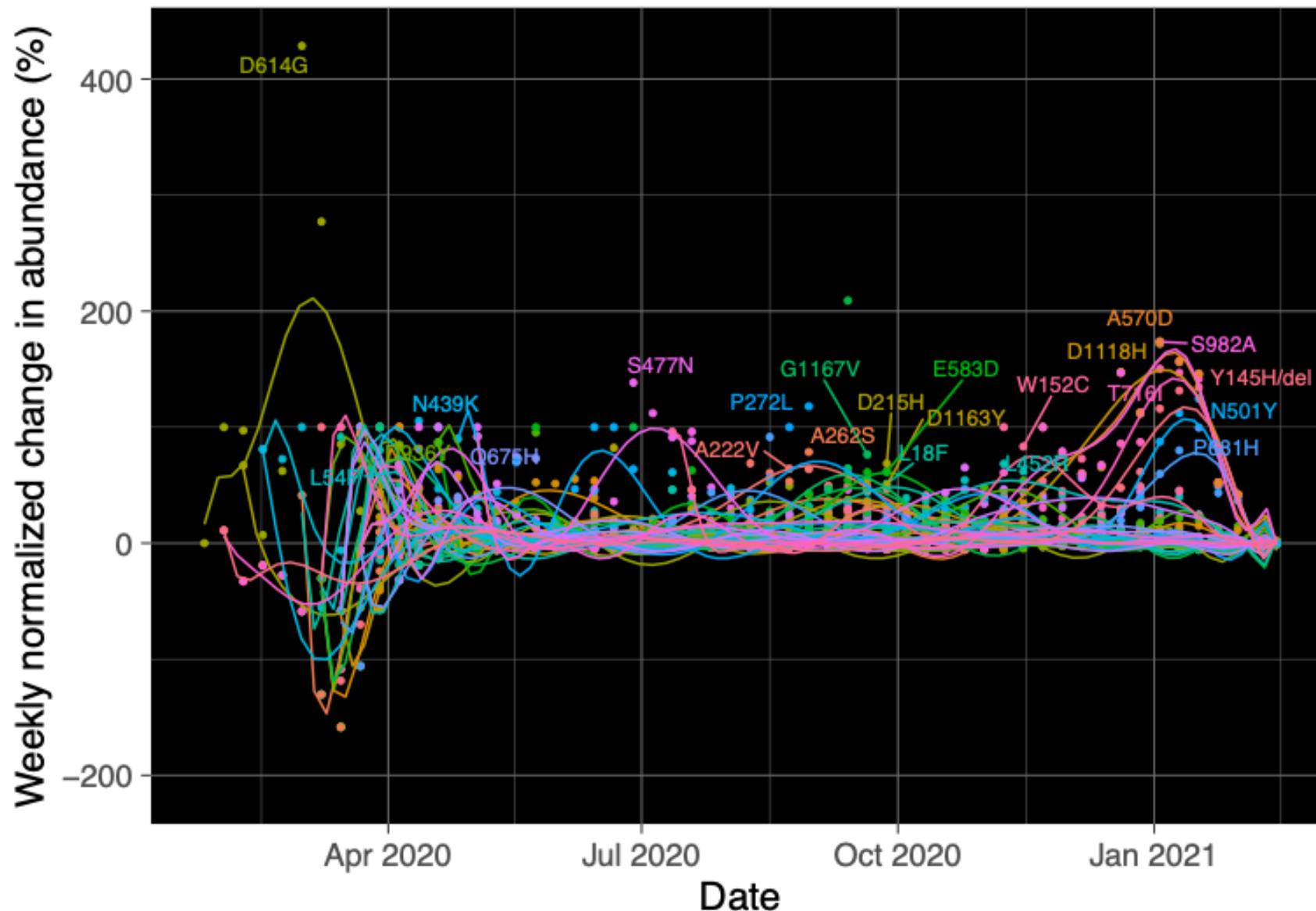
BEYOND THE SNV CALLS

SARS-CoV-2 missense variants. Year 2020, day 13



A MOVING TARGET

Emergence of SARS-CoV-2 spike protein variants



VARIANCE, EMERGENCE, VIRULENCE

SARS-CoV-2 variants 2020 timemap
Genbank MN908947 vs 292,559 GISAID genomes

Continents
 ● Europe
 ● North America
 ● Asia
 ● Oceania
 ● Africa
 ● South America
 ● ● ● 1 10 100 1k genome samples

Winter
25.5 kb

Data from 6 continents
Total genomes>10 Genomes/day≥1
11 distinct custom variants mapped
Updated 2021-02-19

2020-12-31 ATTA21993A Europe=193 [29.9%]
surface glycoprotein, Y145H/del

Fall

Spring

Summer

21.5 kb

SARS-CoV-2 variants 2021 timemap
Genbank MN908947 vs 70,883 GISAID genomes

Continents
 ● Europe
 ● North America
 ● Asia
 ● Africa
 ● South America
 ● Oceania
 ● ● ● 1 10 100 1k genome samples

Winter
25.5 kb

Data from 6 continents
Total genomes>10 Genomes/day≥1
11 distinct custom variants mapped
Updated 2021-02-19

2021-01-31 ATTA21993A Europe=682 [84.7%] surface glycoprotein, Y145H/del

Fall

Spring

Summer

21.5 kb

ACKNOWLEDGEMENTS

- PUBLICATION

- WARREN RL AND BIROL I. INTERACTIVE SARS-CoV-2 MUTATION TIMEMAPS F1000RESEARCH 2021, 10:68
- [HTTPS://BCGSC.GITHUB.IO/SARS2/](https://bcgsc.github.io/SARS2/)

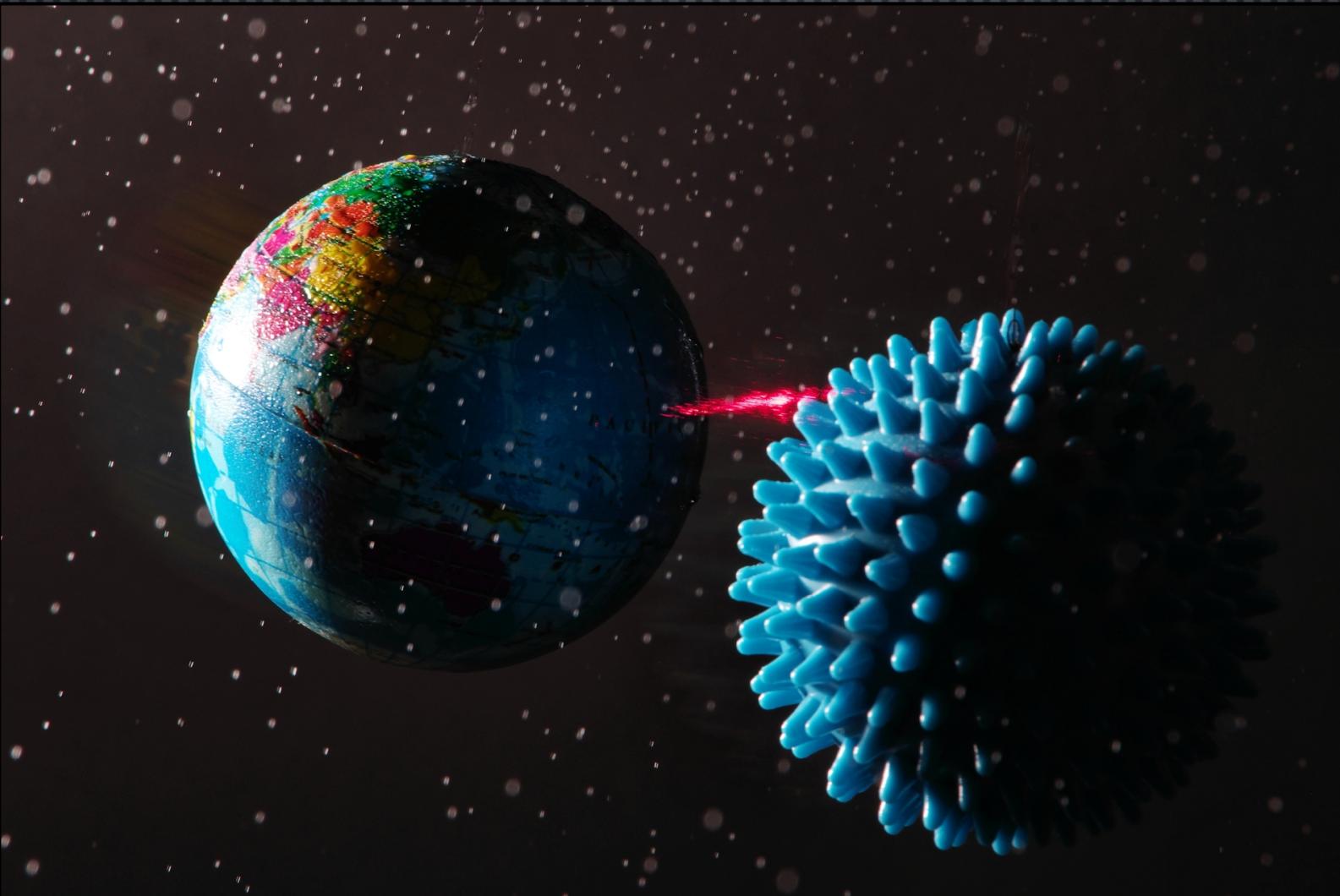
- FUNDING

- NIH (DNATRUER) / GENOME BC, GENOME CANADA (ANNOVIS)

- BTL GROUP

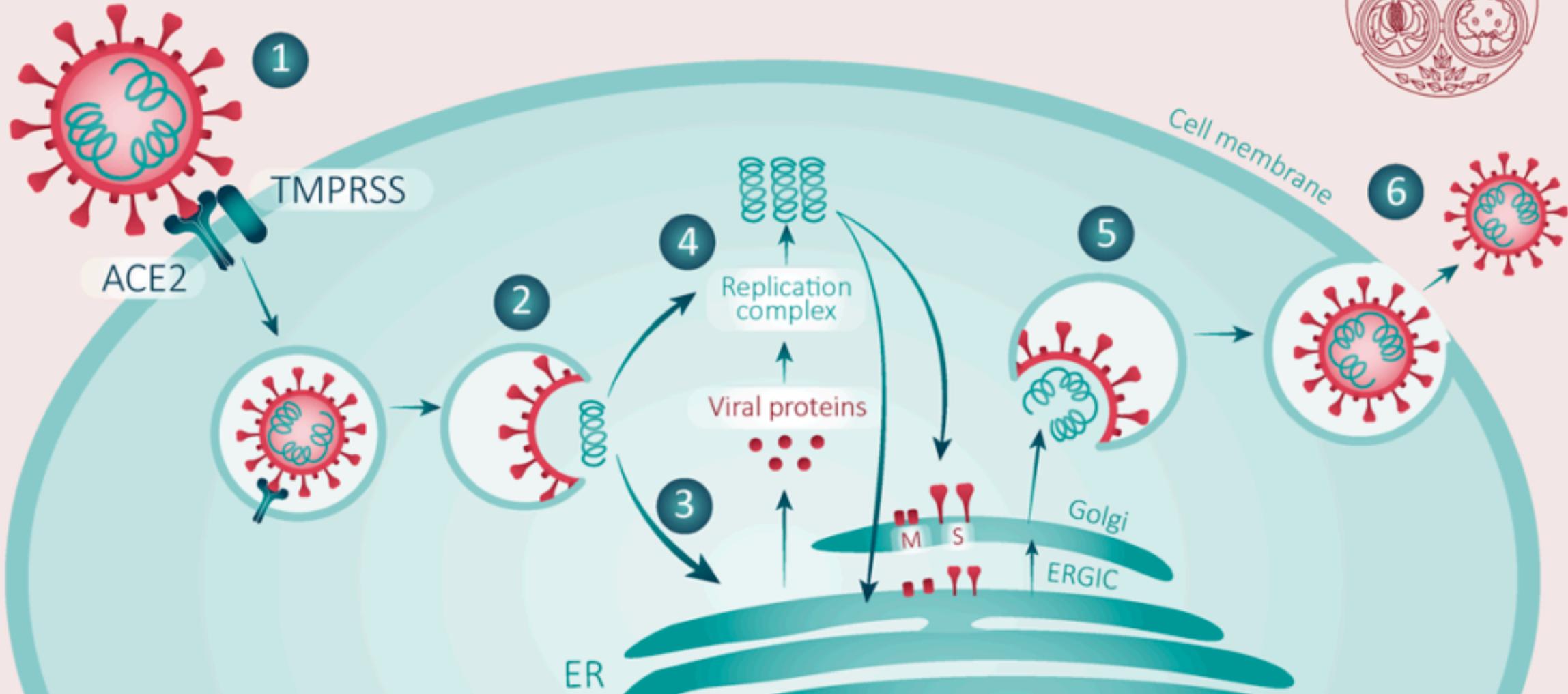
- INANC BIROL
- CECILIA YANG
- DARCY SUTHERLAND
- JOHNATHAN WONG
- HAMID MOHAMADI
- JESSICA ZHANG

QUESTIONS?



“Aucune chance pour la Terre” (no hope for Earth)
photo cr. Robert Warren

SARS-CoV-2



- [1] Spike protein on the virion binds to ACE2, a cell-surface protein. TMPRSS2, an enzyme, helps the virion enter
- [2] The virion releases its RNA, [3] is translated into proteins by the cell's machinery and [4] form a replication complex to make more RNA
- [5] Proteins and RNA are assembled into a new virion in the Golgi and [6] released