

SARS-COV-2 VARIANT TIMEMAPS



RENÉ L WARREN

BC CANCER, GENOME SCIENCES CENTRE

25-JUL-2021

COVID-19: A DIRE SITUATION

- MILLIONS OF LIVES LOST, TOLL ON HUMAN HEALTH

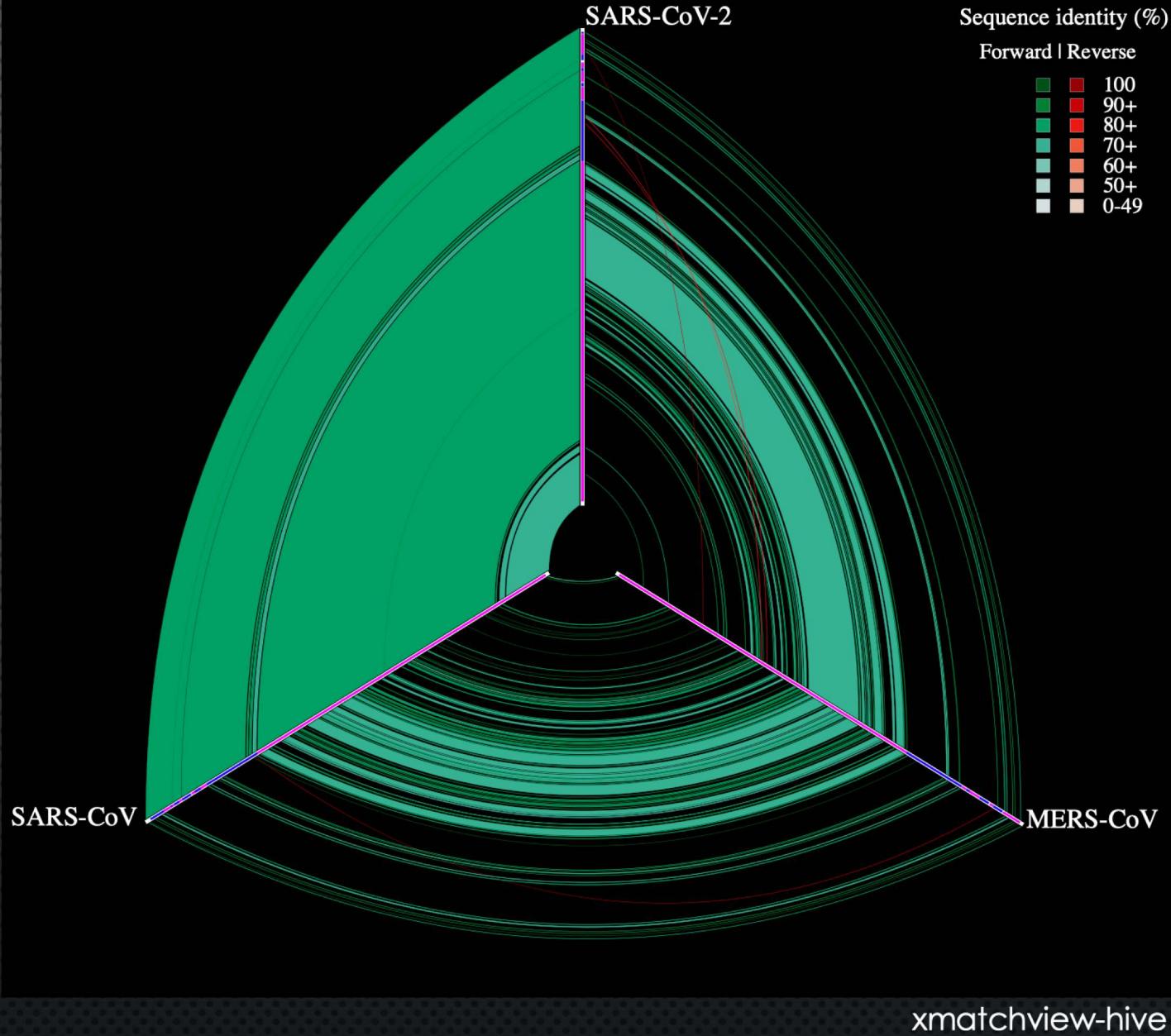


- IMPACTING ALL SECTORS OF HUMAN ACTIVITY
- GRAVE SOCIO-ECONOMIC IMPACT, WILL BE FELT FOR YEARS

July 8th, 2021 <https://coronavirus.jhu.edu/map.html>

SARS-CoV-2

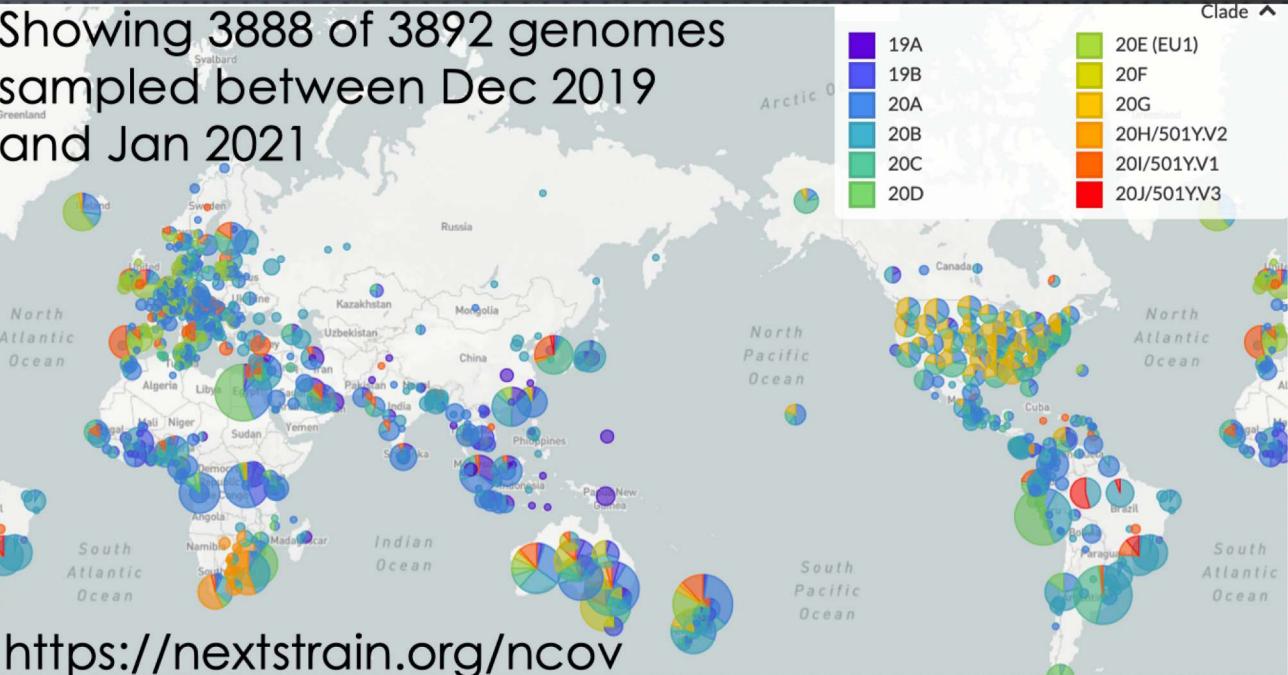
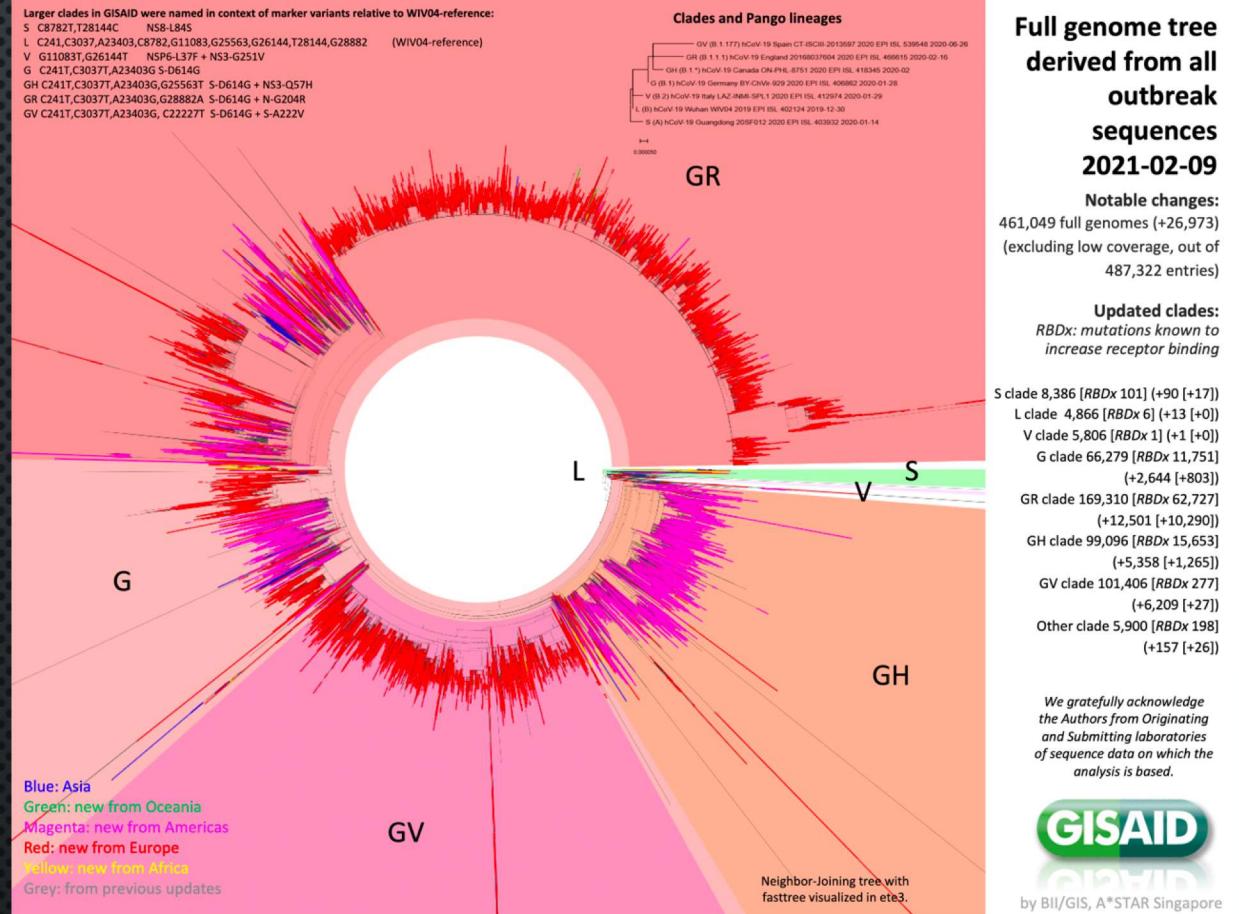
- CORONAVIRUS RELATED TO SARS-CoV
- COVID-19 AGENT
- COMPACT (30KBP) RNA GENOME
 - 10 GENES
 - SURFACE GLYCOPROTEIN (SPIKE)
 - BINDS HOST ACE₂ TO GAIN ENTRY
- AIR BORNE, CONTAGIOUS



GLOBAL EFFORTS

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

TRACKING A CHANGING VIRUS



- GREAT, INTUITIVE UI
- BASED ON PHYLOGENETIC ANALYSES
- LIMITED SAMPLE SIZE

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/SARS2

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

Search

Accession ID: Virus name: complete high coverage
Location: Host: low coverage excl w/Patient status
Collection: 2020-01-01 to Submission: to collection date compl
Clade: all Substitutions: Variants:

<input type="checkbox"/>	Virus name	Passage de	Accession ID	Collection da	Submission D	Length	Host	Location	Originating
<input type="checkbox"/>	hCoV-19/USA/OH-ODH-SC036925/2020	Original	EPI_ISL_2544381	2020-05-07	2021-06-15	29,860	Human	North America / U	Ohio Depar
<input type="checkbox"/>	hCoV-19/USA/OH-ODH-SC036924/2020	Original	EPI_ISL_2544285	2020-05-07	2021-06-15	29,831	Human	North America / U	Ohio Depar
<input type="checkbox"/>	hCoV-19/USA/OH-ODH-SC036908/2020	Original	EPI_ISL_2544284	2020-05-07	2021-06-15	29,859	Human	North America / U	Ohio Depar
<input type="checkbox"/>	hCoV-19/Pakistan/PPHRL-AH62/2021	Original	EPI_ISL_2544200	2021-04-29	2021-06-15	29,812	Human	Asia / Pakistan /	Provincial P
<input type="checkbox"/>	hCoV-19/Pakistan/PPHRL-AH60/2021	Original	EPI_ISL_2544198	2021-04-10	2021-06-15	29,833	Human	Asia / Pakistan /	Provincial P
<input type="checkbox"/>	hCoV-19/Italy/LOM-UniSR10/2021	Vero E6	EPI_ISL_2544194	2021-03-08	2021-06-15	29,813	Human	Europe / Italy / L	Laboratorio
<input type="checkbox"/>	hCoV-19/Pakistan/PPHRL-AH55/2021	Original	EPI_ISL_2544193	2021-03-26	2021-06-15	29,826	Human	Asia / Pakistan /	Provincial P
<input type="checkbox"/>	hCoV-19/USA/OH-ODH-SC036827/2020	Original	EPI_ISL_2543920	2020-05-07	2021-06-15	29,845	Human	North America / U	Ohio Depar

Total: 1,444,967 viruses << < 1 2 3 4 5 > >> Select Analysis

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.

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

- >1,600,000 COMPLETE SARS-CoV-2 GISAID GENOMES
 - 6 CONTINENTS
 - >30,000 DISTINCT NUCLEOTIDE VARIANTS
 - OBSERVED IN 10 OR MORE GENOMES
 - AS HIGH AS 46 PER GENOME
 - 21,593 (72.2%) DISCRETE SITES MUTATED
 - 626 CONVERGENT MUTATIONS
 - S GENE N = 102

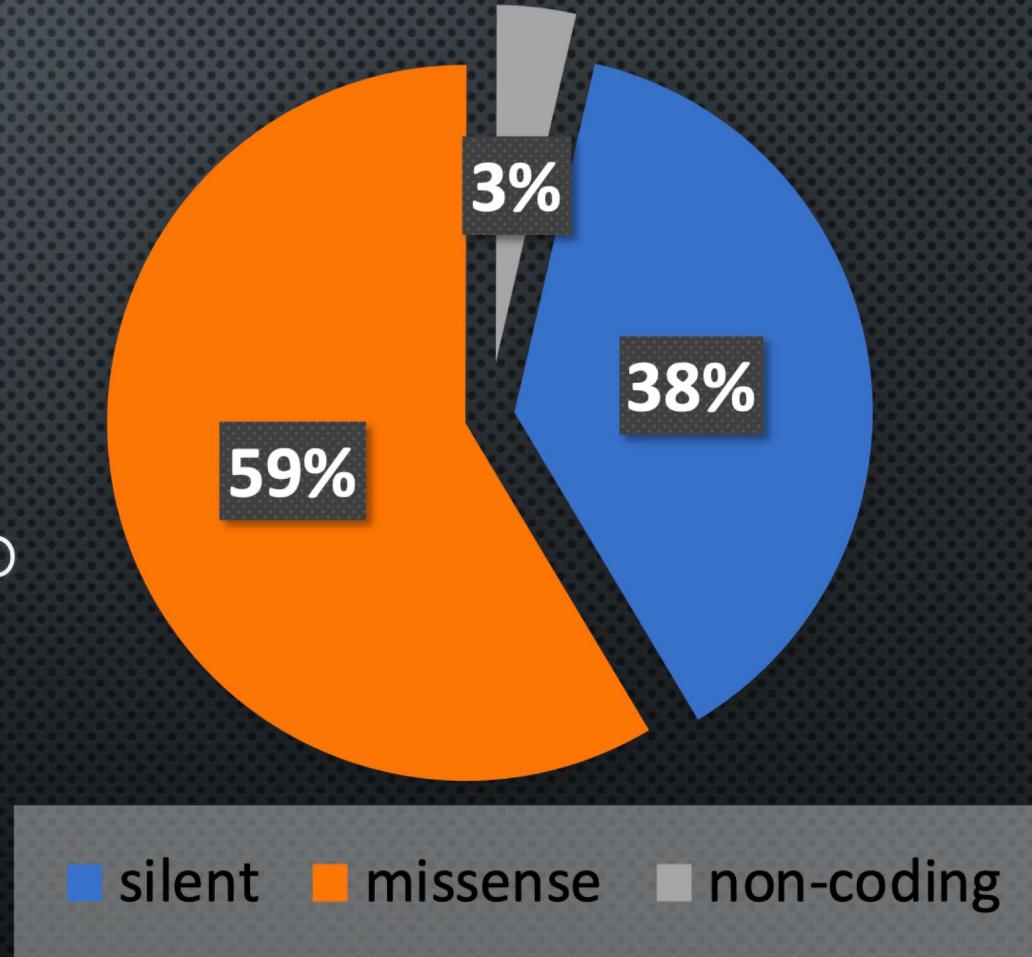
e.g. W152R

T22016A 2020-03-24

T22016C 2020-07-06

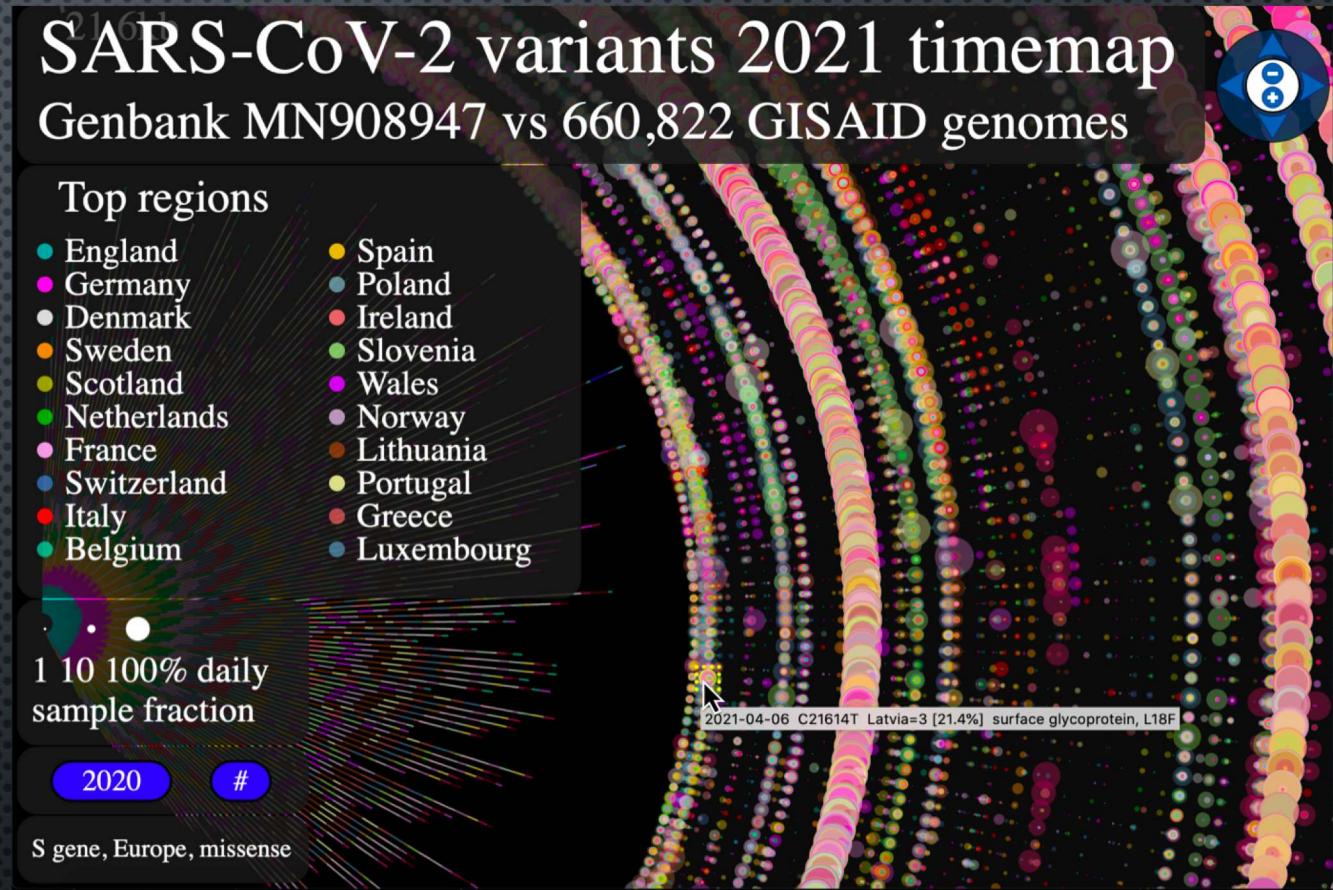
EPI_ISL_457761 USA

EPI_ISL_1424063 Colombia

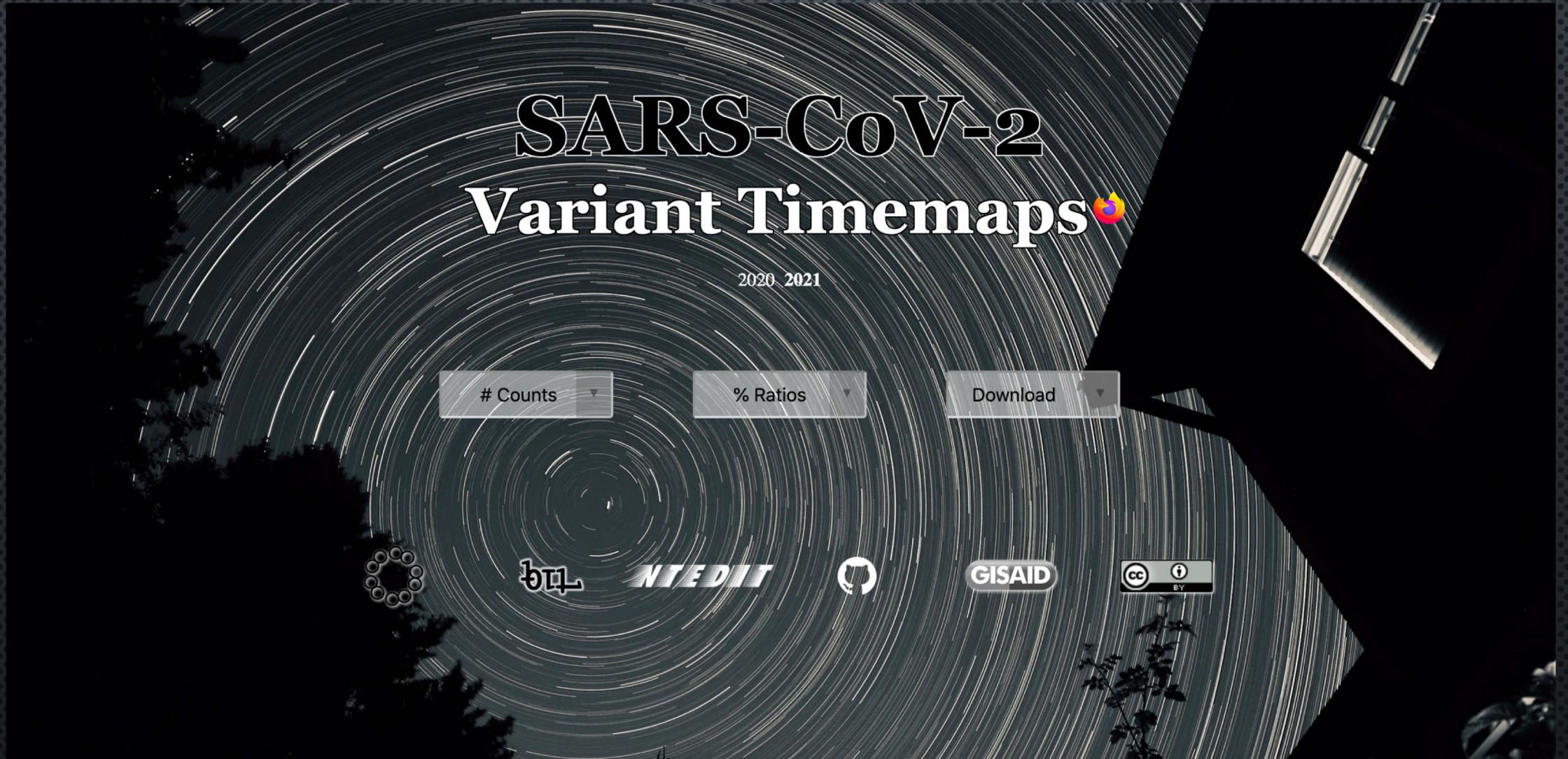


PLOTTING

- COMPREHENSIVE HIGH-RES OUTPUT
- SCALABLE VECTOR GRAPHICS (SVG)
 - VARIANTS ORGANIZED ALONG GENOME
 - YEAR ARRANGED CLOCKWISE
 - 1 RADIUS = 1 DAY
 - COLORED BY REGION
 - SIZED BY SAMPLE COUNT / RATIO
 - INTERACTIVE
 - TOOLTIP, ZOOM, PAN, DRAG, TILT, HIGHLIGHT
 - HOVER
 - REVEALS DATE, VARIANT, REGIONS, COUNT, PROPORTION, GENE, PRODUCT EFFECT



WEB HOSTING



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

A WEALTH OF INFORMATION

SARS-CoV-2 variants 2020 timemap
Genbank MN908947 vs 209,633 GISAID genomes

Top regions

- England
- Denmark
- Wales
- Switzerland
- Scotland
- Netherlands
- Spain
- France
- Germany
- Luxembourg
- Iceland
- Italy
- Belgium
- Sweden
- Norway
- Finland
- Portugal
- Austria
- Russia
- Ireland

1 10 100% daily sample fraction

2021

S gene, Europe, missense

8

Winter

25.4kb

1 radius = 1 day

Fall

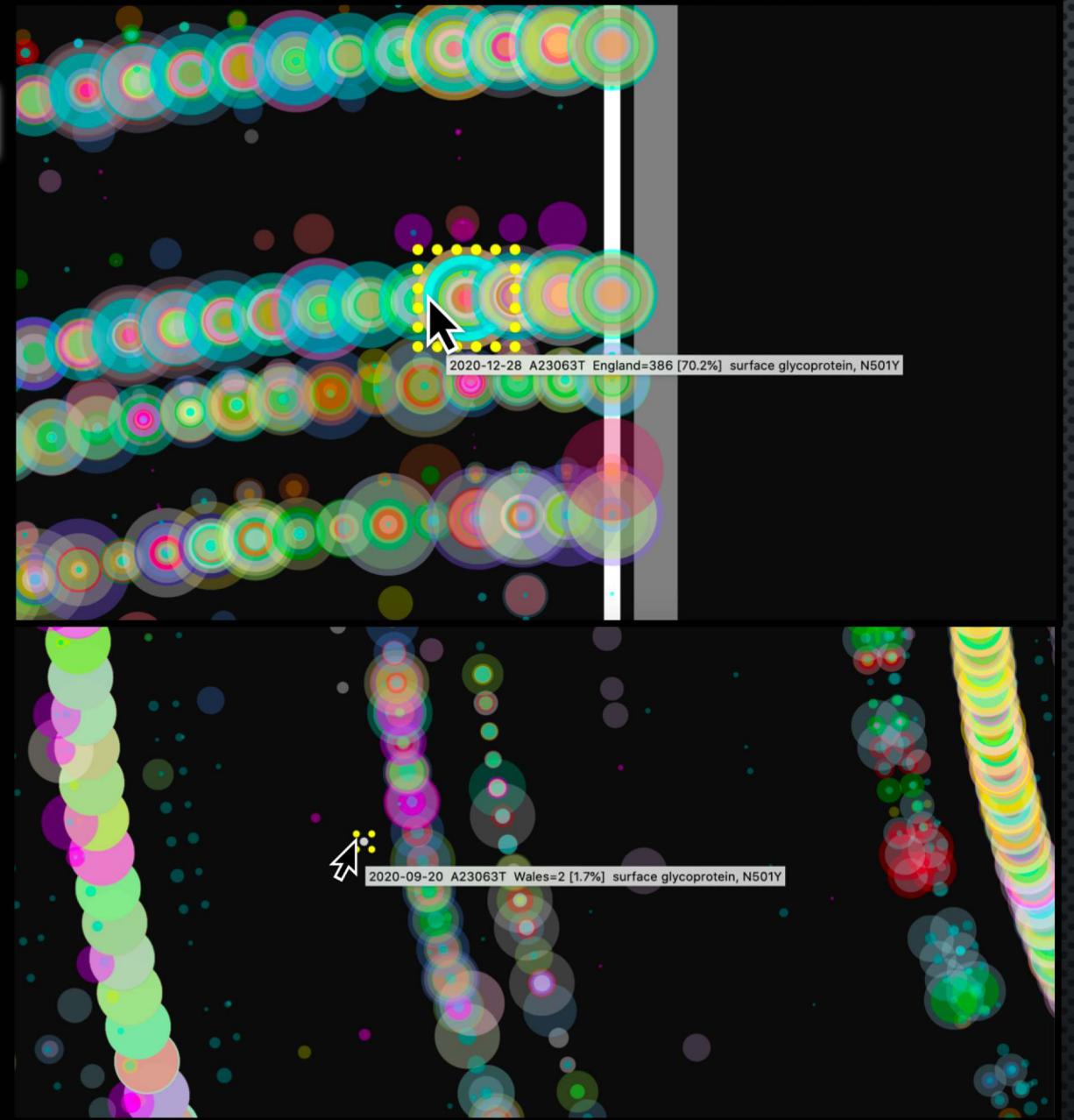
Genome sample counts

Summer

Data from 53 regions in Europe
Total genomes ≥ 10 Genomes/day ≥ 2
902 distinct missense variants mapped
Updated 2021-06-07

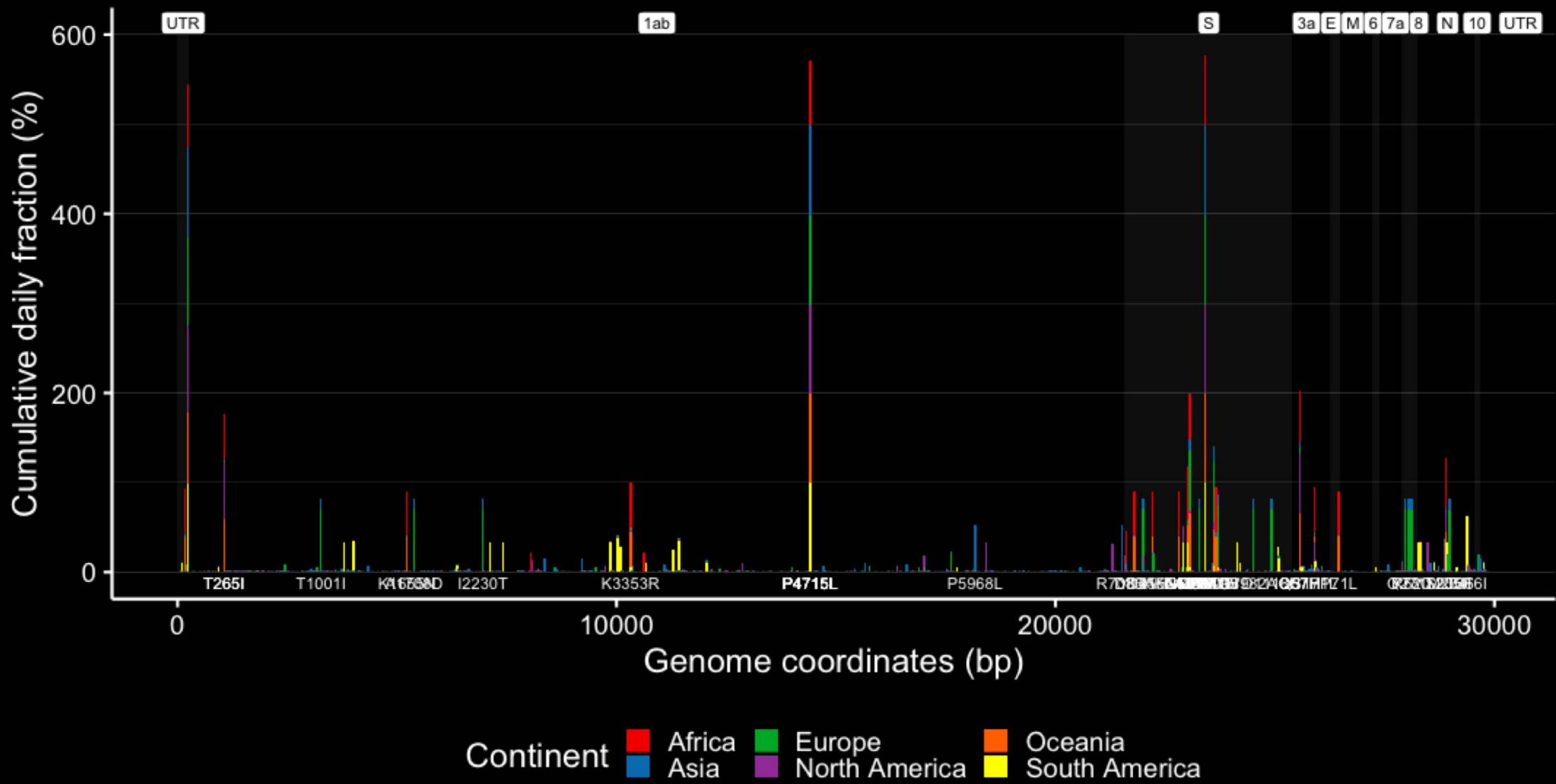
Each spot = mutation
Color-coded by region

Spring



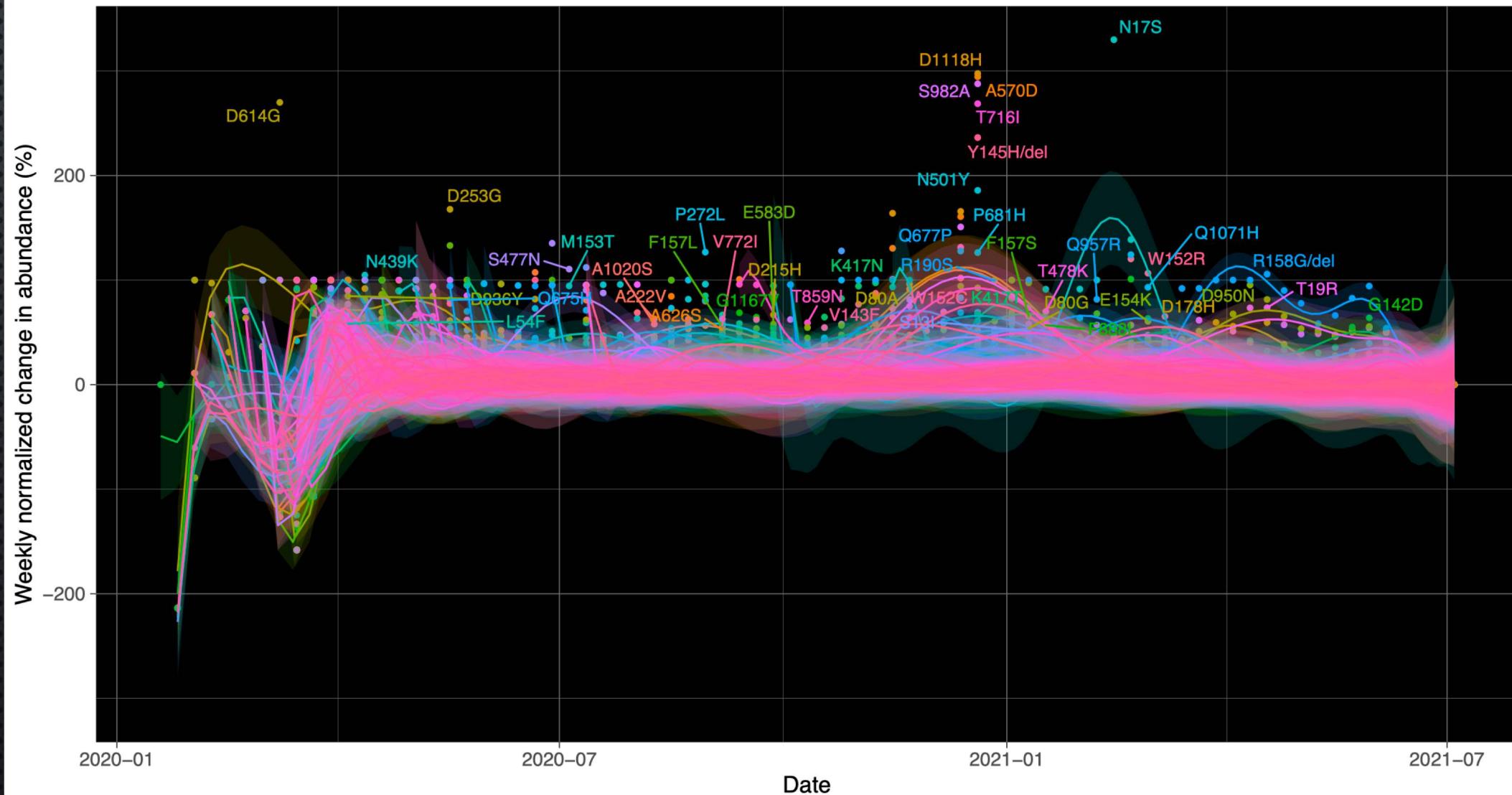
BEYOND THE SNV CALLS

SARS-CoV-2 missense variants. Year 2021, day 17



A MOVING TARGET

Emergence of SARS-CoV-2 spike protein variants



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/)



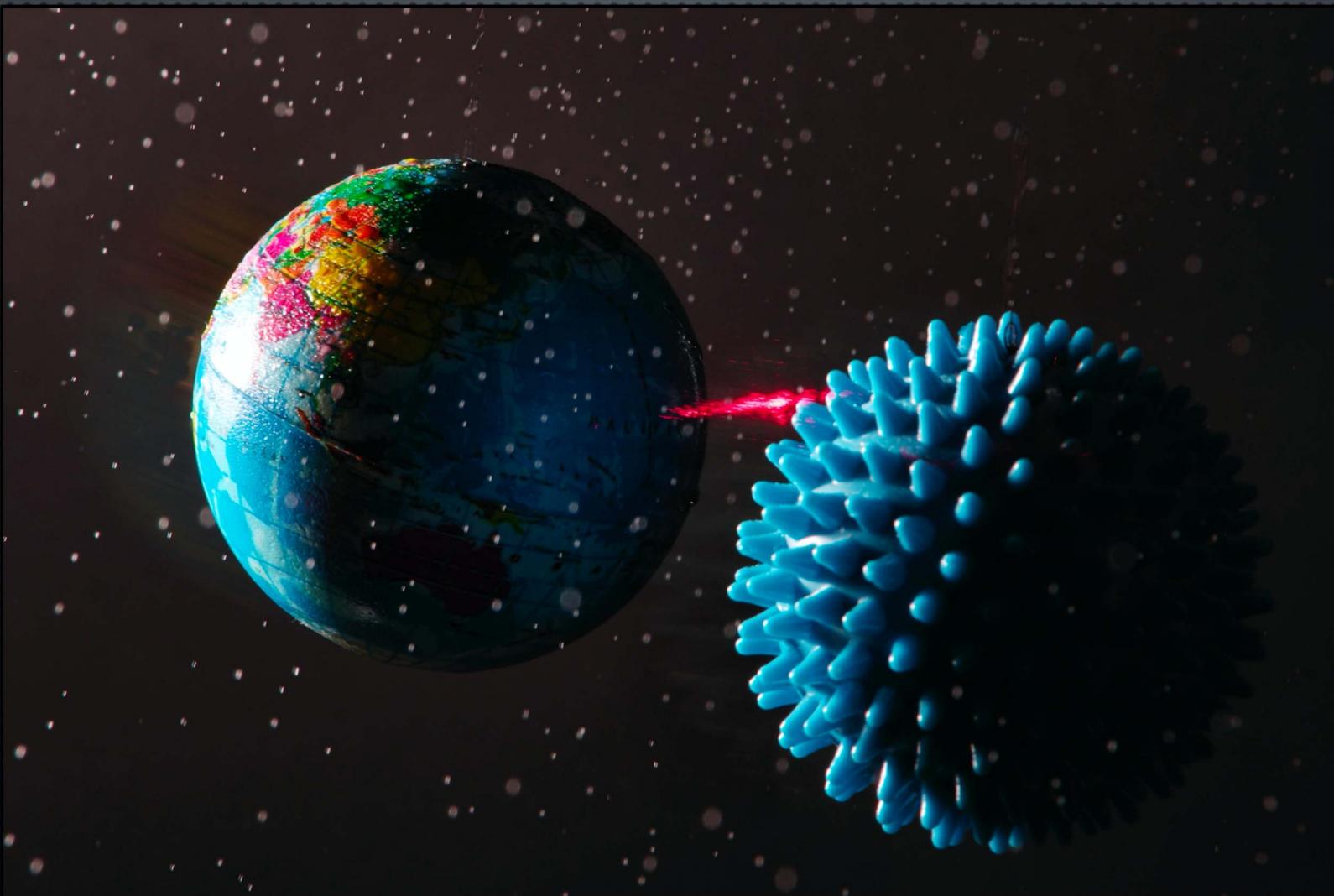
- FUNDING

- NIH (DNATRUER) / GENOME BC, GENOME CANADA (ANNOVIS)
- JOHN JAMBOR EDUCATION FUND



<http://birol-lab.ca>

QUESTIONS?



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

SARS-CoV-2 Variant Timemaps

René L Warren • Inanc Birol

SARS-CoV-2 variants 2020 timemap Genbank MN908947 vs 209,633 GISAID genomes

Top regions

- England
- Denmark
- Wales
- Switzerland
- Scotland
- Spain
- France
- Germany
- Luxembourg
- Iceland
- Italy
- Belgium
- Sweden
- Norway
- Holland
- Portugal
- Austria
- Russia
- Ireland

1 10 100% daily sample fraction

2021 #

S gene, Europe, missense

Fall

SARS-CoV-2, S gene, Europe, missense

2020-09-20 A23083ST Wales+2 (1.7% surface glycosylation, NS39Y)

Winter

1 radius = 1 day

Genome sample counts

Summer

Data from 53 regions in Europe
Total genomes: 10 Genomes/day ≥ 2
902 distinct missense variants mapped
Updated 2021-06-07

Variation "spot"
colored by regions
mouse hover
counts(#)/ratios (%)
region effect

← Arrows indicate emerging missense variations (from top to bottom)
D1118H, S982A, T716I, P681H, A570D and N501Y

- Nucleotide variation in >1.6M GISAID genomes accumulating since ground zero strain**
- Interactive Scalable Vector Graphics (SVG)**
drag, pan, zoom, tilt, highlight, hover / tooltip
- Over 120 SVG timemaps, web-hosted**
world / regional views
genes / genome / variation types
variants of concern, WHO convention, custom
- Panorama of longitudinal strain evolution in human hosts**

SARS-CoV-2 variants 2021 timemap Genbank MN908947 vs 105,279 GISAID genomes

Continents

- Europe
- Asia
- North America
- Africa
- Oceania
- South America

2 10 100 k genome samples

Continent: world, B.1.617.2

Spring

Fall

GENOME DOWNLOAD
GISAID

VARIANT PROFILING
ntEdit

PLOTTING
SVG

WEB HOSTING
bcgsc.github.io

B.1.617.2 [20 A/S]
Delta variant
Genome, world,
missense

Summer

Winter

Data from 6 continents
Total genomes: 10 Genomes/day ≥ 2
3,133 distinct missense variants mapped
Updated 2021-07-09

Weekly emergence of spike protein variants



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

Funding

NIH
Genome BC / Canada
John Jambor Education Fund



<http://warrenlr.github.io>



<http://birol-lab.ca>

